frontiers | Research Topics

Ocean observation based on underwater acoustic technology, volume II

Edited by
Xuebo Zhang, Haixin Sun and Arata Kaneko

Coordinated by
Suleman Mazhar

Published in
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-5751-8
DOI 10.3389/978-2-8325-5751-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Ocean observation based on underwater acoustic technology, volume II

Topic editors
Xuebo Zhang — Northwest Normal University, China
Haixin Sun — Xiamen University, China
Arata Kaneko — Hiroshima University, Japan

Topic coordinator
Suleman Mazhar — Harbin Engineering University, China

Citation
Zhang, X., Sun, H., Kaneko, A., Mazhar, S., eds. (2024). *Ocean observation based on underwater acoustic technology, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5751-8

Table of contents

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 09 August 2023
DOI 10.3389/fmars.2023.1211234

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Keyu Chen,
Xiamen University, China
Rongbin Lin,
Shenzhen Research Institute of Xiamen University, China
Pan Huang,
Weifang University, China

*CORRESPONDENCE
Hao Zhang
zhanghao@ouc.edu.cn

RECEIVED 24 April 2023
ACCEPTED 12 July 2023
PUBLISHED 09 August 2023

CITATION
Chen X, Zhang H, Gao Y and Wang Z (2023) DOA estimation of underwater acoustic co-frequency sources for the coprime vector sensor array. *Front. Mar. Sci.* 10:1211234. doi: 10.3389/fmars.2023.1211234

DOA estimation of underwater acoustic co-frequency sources for the coprime vector sensor array

Xiao Chen[1], Hao Zhang[1,2]*, Yong Gao[1] and Zhen Wang[1]

[1]Department of Electronic Engineering, Ocean University of China, Qingdao, China, [2]Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada

A coprime array with fewer sensors can achieve the same resolution as a uniform linear array. However, when detecting co-frequency targets, there can be prominent false alarms due to overlaps between the main and grating lobes of subarrays. This study proposes a direction-of-arrival (DOA) estimation method to obtain the co-frequency target directions from high grating lobes. The method utilizes joint processing of sound pressure and vibration velocity data from vector hydrophones of a coprime vector hydrophone array and designs joint-cross terms (JCTs) using channel combinations. Based on JCTs, we establish a characteristic data point identification algorithm. The method in this paper can stably and accurately acquire co-frequency target directions from high grating lobes without decoherence operation. Simulation results demonstrate that the proposed algorithm achieves accurate DOA estimation even with reduced signal-to-noise ratio (SNR) and fewer data points. Additionally, a sea experiment confirms the rationality and efficiency of the proposed algorithm, providing new ideas for co-frequency source detection using coprime vector sensor arrays.

KEYWORDS

direction-of-arrival (DOA) estimation, co-frequency sources, coprime vector sensor array, sound pressure and vibration velocity joint processing, vector hydrophone

1 Introduction

Direction-of-arrival (DOA) estimation is an essential aspect of array signal processing that holds immense significance in multiple fields, including acoustics, radar, and wireless communications (Zhang et al., 2022; Xie et al., 2023; Zhang et al., 2023). Classic techniques for DOA estimation involve subspace theory and typically utilize methods such as multiple signal classification (MUSIC) (Schmidt, 1986) and estimating signal parameters via rotational invariance techniques (ESPRIT) (Roy and Kailath, 1989). In these algorithms, uncorrelated incident signals are assumed, and coherent signals will fail due to the covariance matrix's rank deficit. To handle coherent signal situations, several techniques

have been suggested, such as spatial smoothing (SS) (Pillai and Kwon, 1989) and forward/backward SS (FBSS) (Shan et al., 1985). The SS method achieves DOA estimation of coherent signals but at the cost of decreasing array aperture. The FBSS method can enhance estimation accuracy but does not fully address signal decoherence. Furthermore, these techniques typically consider ULAs, with sparse linear arrays being less commonly employed.

Coprime line arrays (CLAs) offer a systematic array setup beyond Nyquist sampling while minimizing mutual coupling between array elements (Vaidyanathan and Pal, 2010; Vaidyanathan and Pal, 2011; Adhikari et al., 2013; Zhang et al., 2013; Tan et al., 2014; Adhikari and Buck, 2015; Di Martino and Iodice, 2017; Qin et al., 2017; Zhou et al., 2017; Alawsh and Muqaibel, 2018; Mei et al., 2018; Adhikari, 2019; Moghadam and Shirazi, 2019; Alawsh and Muqaibel, 2020; Alawsh and Muqaibel, 2021; Moghadam and Shirazi, 2022). The research on DOA estimation of the coprime array is mainly carried out from two aspects. On the one hand, it is implemented from the physical array domain. On the other hand, it is achieved in the virtual array domain. For processing the physical array elements, a DOA estimation method utilizes a decomposed CLA and solves a joint covariance matrix optimization problem. The method enables the reconstruction of the interference-plus-noise covariance matrix and weight vector computation for the minimum variance distortionless response (MVDR) beamformer that minimizes variance distortion (Zhou et al., 2016; Zhou et al., 2017). The DOA estimation in the physical array domain suffers from high grating lobes caused by the intersensor spacing, which is greater than $\frac{\lambda}{2}$ (where λ is the wavelength of the signal), and many methods have been investigated in order to reduce the effect of grating lobes. For scalar CLAs, Product and Min algorithms were proposed to calculate the signal spatial power spectral density (PSD) and resolve the grating lobe problems (Adhikari and Buck, 2017). The array factors can be established based on a uniform linear array for single-target direction estimation. The array factors should satisfy that the beam response of the uniform linear array has the opposite amplitude with one subarray of the CLA (Liu and Buck, 2015). Extending the coprime array is also a way to settle the grating lobe matters. Some methods have been suggested to extend CLAs by changing the positions of grating lobes and sidelobes of the beam output for two coprime subarrays (Adhikari et al., 2013; Adhikari et al., 2014; Chen et al., 2023). The methods above for suppressing grating lobes are developed when the signals are incoherent. In the virtual array domain, the investigation of the coprime vector sensor array has received more attention. Nowadays, DOA estimation for coprime vector sensor arrays has already been developed in the radar field. A six-sensor coprime electromagnetic vector-sensor (EMVS) array (Fu et al., 2021) was used in a new method. The method involved a nuclear norm minimization (NNM) problem to create an extended covariance matrix for DOA information. Then, the issue of DOA estimating in a two-dimensional space was examined for a multiple-input multiple-output (MIMO) radar with coprime EMVS arrays operating in a bistatic configuration (Yang et al., 2021). However, these methods are used to process uncorrelated signals for the coprime vector sensor array, while coherent signals are less considered in the application. Moreover, the main advantage of these algorithms in the radar field is the high degree of freedom for DOA estimation. For underwater array target detection, such a high degree of freedom is not an urgent need to be achieved.

In underwater signal detection, vector hydrophone linear arrays are often used. Each vector hydrophone can be composed of hydrophone and velocity sensors. Owing to the frequency-independent dipole directivity of the vector hydrophone, a vector hydrophone linear array has similar performance but with a smaller array aperture compared with a sound pressure array. Moreover, vector hydrophone arrays have attracted wide attention for their left–right discrimination, which acoustic pressure arrays cannot provide (Hawkes and Nehorai, 1998). Furthermore, when the signals of vibration velocity and sound pressure are combined, the combination holds strong anti-isotropic noise ability (Santos et al., 2011; Felisberto et al., 2016; Felisberto et al., 2018). As for practical applications, fulfilling coprime array configuration in underwater vector sensor arrays is a recently new attempt (Chen et al., 2023). Moreover, the issue about DOA estimation of co-frequency signals for underwater coprime vector sensor array is still expected to be addressed.

When the target is incoherent, the array will output high grating sidelobes but not exceed the magnitude of the output in the direction where the targets are located, and this issue has been studied. However, when the targets are co-frequency, overlapping high grating sidelobes can cause higher array output than the magnitudes of the target directions. As a result, the actual targets may be obscured, and the DOA estimation performance will deteriorate. In this paper, we propose an algorithm that utilizes a coprime vector hydrophone array to achieve DOA estimation of two co-frequency signals. We aim to address the issue of concealed targets due to high grating lobes. Thus, the target directions can be identified accurately from the high grating lobes, thereby avoiding false alarms. To enhance robustness, we employ the conventional beamformer (CBF) based on the entire coprime array as the preprocessing method. Joint-cross terms (JCTs) are constructed based on the vector hydrophone subarrays, and the channel combinations of vector hydrophones are utilized in the algorithm. Additionally, we design a characteristic data point identification method based on JCTs. Unlike existing techniques, the proposed method does not perform spatial smoothing, but it is highly effective in processing coherent signals with the same frequency. Simulation results and experimental data analysis validate the effectiveness of the proposed algorithm. The paper's contributions can be summarized as follows.

1. Firstly, the paper analyzes the cause of the high grating lobes in coprime vector sensor arrays when two co-frequency signals are present. JCTs are constructed using the characteristics of coprime arrays and vector hydrophones, which imply the DOA information.
2. Secondly, a DOA estimation method based on characteristic data point identification algorithm using JCTs is designed, which achieves stable extraction of co-frequency targets' directions.

This paper is organized as follows. In *Section 2*, we establish the mathematical signal model of the coprime vector sensor array and attain the array beam output. Next, we advance the situation in which strong grating lobes appear and present an example. In *Section 3*, we present the DOA estimation method. The DOA

estimation method based on characteristic data point identification is introduced based on JCTs to achieve the direction extraction. We validate the method through simulation in *Section 4* and experimental data processing in *Section 5*. Finally, we summarize the article in *Section 6*.

Notations: The uppercase bold characters denote matrices, while their lowercase counterparts denote vectors. $(\cdot)^{\star}$, $(\cdot)^{H}$, and $(\cdot)^{T}$ represent the complex conjugate, conjugate transpose, and transpose, respectively. $bm\boldsymbol{I}$ stands for the unit matrix. $\otimes$ represents the Kronecker product.

2 Co-frequency signals model

An underwater acoustic vector sensor linear array consists of two sparse uniform vector sensor linear subarrays with M and N physical sensors, respectively. The values of M and N are coprime. The first subarray containing M sensors is spaced apart by Nd, whereas Md spaces apart the second subarray containing N sensors. Here, $d = \lambda/2$ represents the intersensor unit spacing, where λ indicates the wavelength of the narrowband signal received by the array. With two subarrays sharing the first sensor, the other sensors of each subarray are arranged according to the original structure, and the array configuration is represented in Figure 1 and Equation (1).

$$\mathbb{S} = \{Mnd, 0 \le n \le N-1\} \cup \{Nmd, 0 \le m \le M-1\} \quad (1)$$

Assuming that the far-field narrowband co-frequency coherent signal impinges on the coprime vector hydrophone from the direction θ_0, the received signal can be modeled as:

$$\begin{aligned} \boldsymbol{X}(t) &= [\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \cdots \boldsymbol{x}_{3(M+N-1)}(t)]T \\ &= \boldsymbol{a}(\theta_0) \otimes \boldsymbol{u}(\theta_0)\boldsymbol{z}(t) + \boldsymbol{N}(t) \end{aligned} \quad (2)$$

where. denotes the signal waveform vector and $N(t) = [n_1^T(t), n_2^T(t), \cdots n_{3(M+N-1)}^T(t)]^T \sim \mathcal{CN}(0, \sigma_n^2 I)$ denotes statistically independent Gaussian noise component with σ_n^2, where σ_n^2 is the noise power. Here, $n_i(t) = [n_p(t), n_{vx}(t), n_{vy}(t)]^T, i = 1, 2, \cdots M+N-1$ and $n_p(t)$, $n_{vx}(t)$, and $n_{vy}(t)$ denote the pressure component and the horizontal velocity x and y direction components of the noise vector at the ith element, and they are mutually independent. $a(\theta)$ is the steering vector connected with DOA θ_l given by:

$$\boldsymbol{a}(\theta_0) = [1, e^{-j\frac{2\pi}{\lambda}d_2 sin(\theta_0)}, \ldots, e^{-j\frac{2\pi}{\lambda}d_{M+N-1}\sin(\theta_0)}]^T \quad (3)$$

where $[d_1, d_2, \ldots, d_{M+N-1}] \in \mathbb{S}$. Here, $d_1 = 0$ by taking the first array element as a reference, which can be shown in Figure 1. The velocity components of three-dimensional vector hydrophones are displayed in Figure 2. For two-dimensional vector hydrophones in practical application, the 3×1 steering vector can be obtained as:

$$\boldsymbol{u}(\theta_0) = [1, \cos(\theta_0), \sin(\theta_0)]^T \quad (4)$$

Without regard to the noise component, the CBF is given by the following equation (Yang and Ye, 2019):

$$\boldsymbol{B}_{vCLA} = \frac{1}{(M+N-1)^2}\left|\boldsymbol{w}_{vCLA}^H z_{vCLA}\right|^2 \quad (5)$$

$$\boldsymbol{w}_{vCLA} = \boldsymbol{w}_{CLA} \otimes \boldsymbol{u}(\theta) \quad (6)$$

$$\boldsymbol{w}_{CLA} = e^{-j\frac{2\pi}{\lambda}d_{CLA}sin(\theta)} \quad (7)$$

where w_{CLA} denotes the weight of the array beamformer and $d_{CLA} \in \mathbb{S}$. When there is only one source, the beam output of the array can achieve the maximum output in the source direction. However, when two sources have the same frequency, the large cross-term appears in Eq. (5). The large cross-term will result in large beam outputs in other non-target directions, ultimately leading to false alarms or incorrect bearing estimation results. Figure 3 also explains the situation. The positions indicated by the arrows in the figure represent the grating lobe locations for a coprime vector sensor array. When two subarrays' grating lobes caused by co-frequency signals overlap, a high output will be generated for the whole array. The directions with grating lobes relation can be explained by the following formula (Adhikari et al., 2014):

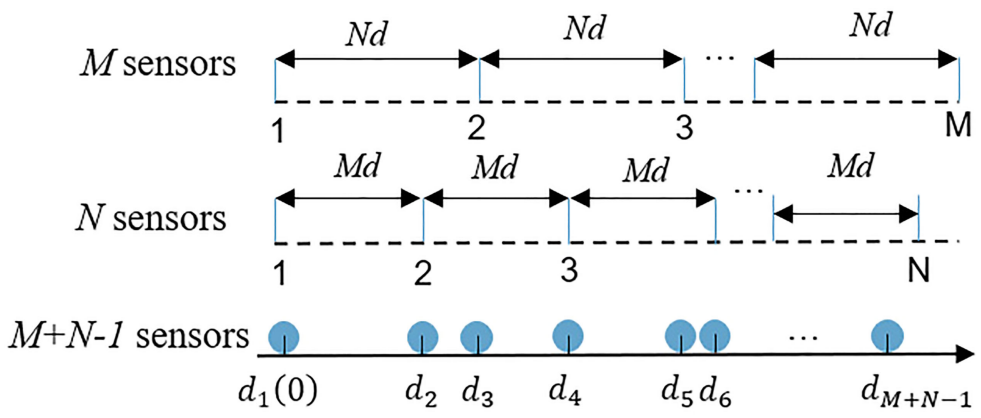


FIGURE 1
Coprime array configuration.

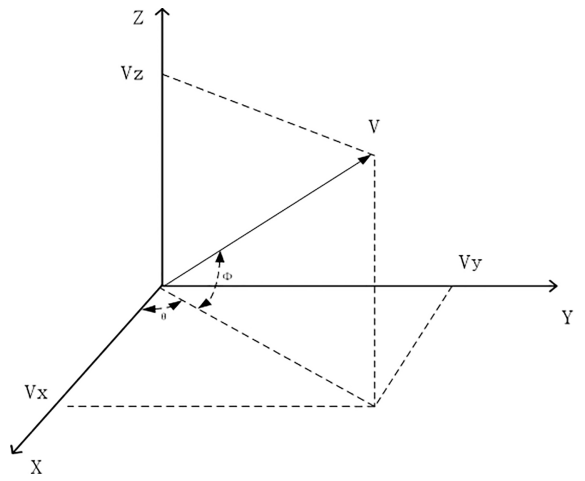


FIGURE 2
The view of a vector hydrophone geometry.

$$\cos\alpha \pm \frac{2}{M}k_1 = \cos\beta \pm \frac{2}{N}k_2 \tag{8}$$

where $k_1 = 0, 1, 2, \ldots$, $k_2 = 0, 1, 2, \ldots$, and α and β are two angles satisfying the overlapping relationship of grating lobes. For instance, as shown in Figure 4, the directions of $\theta_1 = 57^\circ$ and $\theta_1 = 87^\circ$ marked by the black dotted lines are the true co-frequency sources' directions, whereas the directions of $\theta_1 = 32^\circ$ and $\theta_1 = 105^\circ$ marked by the red dotted boxes are the false-alarm directions. In Figure 4A, the main lobe of one subarray and the grating lobe of the other subarray coincide, or the grating lobe of one subarray and the grating lobe of the other subarray coincide (as shown in red dotted boxes). Consequently, ambiguity emerges in comparable amplitude beam output to true sources, as shown in Figure 4B.

3 DOA estimation for two co-frequency sources

3.1 Constructing joint-cross terms for coprime vector hydrophone array

The correlation coefficient between sound pressure and vibration velocity in the isotropic noise field is 0, which means that the joint processing of sound pressure and vibration velocity for the acoustic vector signal suppresses the noise. Therefore, without regard to the noise component, the data channel of the acoustic vector hydrophone is transformed by rotation and combination, and Eq. (9) is obtained

$$\begin{aligned} \boldsymbol{v}_c(t) &= \boldsymbol{v}_x(t)\cos(\varphi) + \boldsymbol{v}_y(t)\sin(\varphi) \\ &= \boldsymbol{s}(t)cos(\theta - \varphi) \end{aligned} \tag{9}$$

$$\begin{aligned} \boldsymbol{v}_s(t) &= -\boldsymbol{v}_x(t)sin(\varphi) + \boldsymbol{v}_y(t)cos(\varphi) \\ &= \boldsymbol{s}(t)sin(\theta - \varphi) \end{aligned} \tag{10}$$

where $v_x(t)$ and $v_y(t)$ represent the velocity components of a vector hydrophone, and they are mutually orthogonal. $v_c(t)$ and

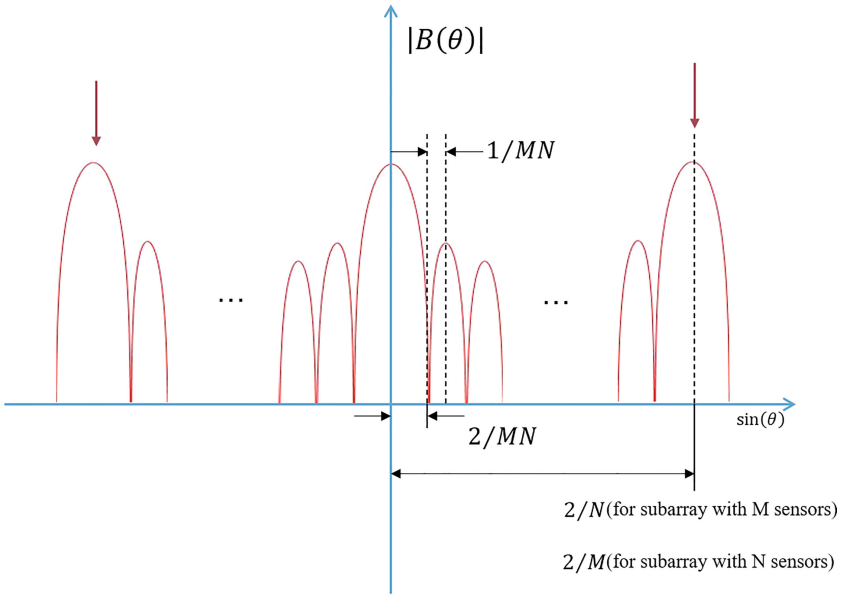


FIGURE 3
The distribution characteristics of zero point, grating lobes, and sidelobes of a coprime vector sensor array output.

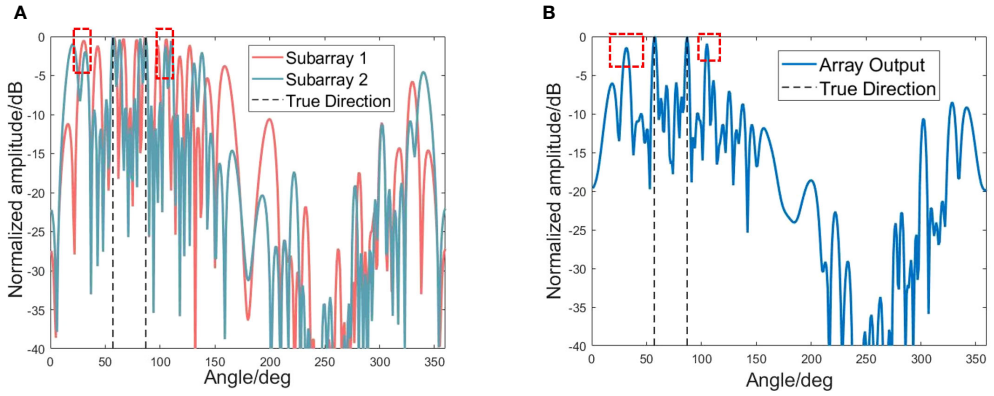


FIGURE 4
The situation of the blurred orientation output caused by the overlapping of subarray grating lobes. **(A)** Beam output for two subarrays of the coprime vector sensor array. **(B)** Beam output of the whole coprime vector sensor array.

$v_s(t)$ stand for the combined transformation of the velocity channel of vector hydrophone, where φ is the electron rotation angle and $s(t)$ is the sound pressure signal received by the hydrophone. In this paper, the combination of sound pressure and vibration velocity is used as

$$[\boldsymbol{p}(t) + \boldsymbol{v}_c(t)]\boldsymbol{v}_s(t) = s^2(t)\boldsymbol{B}_s(\theta) \tag{11}$$

where

$$\boldsymbol{B}_s(\theta) = (1 + \cos(\theta - \varphi))\sin(\theta - \varphi) \tag{12}$$

$p(t)$ indicates the sound pressure of a vector hydrophone. We can see that $B_s(\theta) = 0$ when $\theta = \varphi$ and the noise reduction process is carried out by using the correlation characteristics between signal and noise. Therefore, by rotating the acoustic vector hydrophone data and selecting an appropriate rotation angle φ, the noise can be reduced, thereby reducing the SNR threshold and making it possible to explore weak targets.

For a single uniform sparse vector hydrophone array, when there is a target from a certain orientation (take the target with an orientation of 40° as an example), the spatial spectrum output of the sparse vector array using conventional beamforming obtained based on Eq. (12) can be shown as Figure 5. It can be seen that the spatial spectrum output based on vector hydrophone combined channels shows a concave point at target orientation. However, there is an unreliability in using concave points to determine target orientations when the noise is considered. Moreover, for a single sparse array, this unreliability will become more acute as the spacing of array elements increases.

The CBF for a coprime vector hydrophone array produces two spatial spectra but contains ambiguous orientation concave points due to the spatial undersampling of the subarrays. Inspired by the Product theorem (Adhikari et al., 2014; Adhikari and Buck, 2017), which resolves the spatial frequency ambiguities by performing complex conjugate multiplication between two coprime subarrays (Vaidyanathan and Pal, 2010), we proposed a DOA estimation method based on JCTs for coprime vector hydrophone array. Let $p_M(t)$, $v_{x_M}(t)$, and $v_{y_M}(t)$ be the acoustic pressure and the x-axis and y-axis velocity data of acoustic particles received by the vector hydrophones from the subarray with M sensors, respectively.

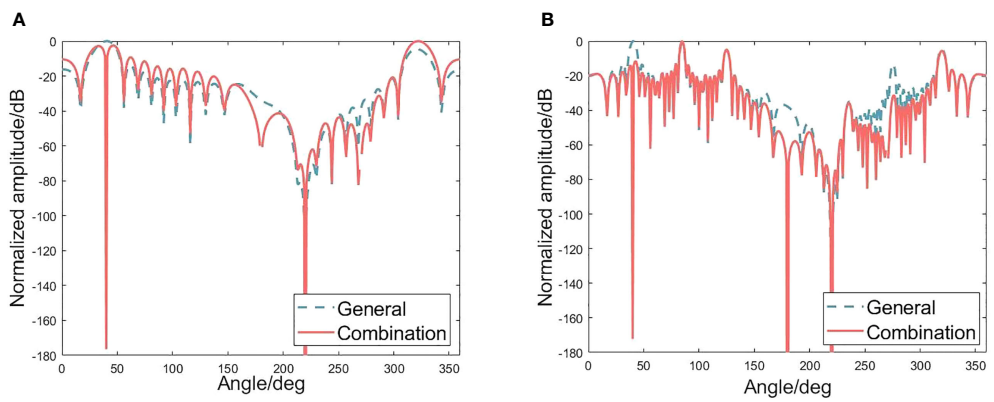


FIGURE 5
Comparison of treating each channel as normal and combining channels for a vector hydrophone. **(A)** Spatial spectrum output of a vector uniform line array without considering noise. **(B)** Spatial spectrum output of a vector sparse line array without considering noise.

Meanwhile, let $p_N(t)$, $v_{x_N}(t)$, and $v_{y_N}(t)$ be the acoustic pressure and the x-axis and y-axis velocity data of acoustic particles received by the vector hydrophones from the subarray with N sensors. The JCTs for coprime vector hydrophone array can be constructed as

$$\begin{cases} \boldsymbol{J}_1(t) = (\boldsymbol{p}_M(t) + \boldsymbol{v}_{cM}(t))\boldsymbol{v}_{sN}(t) \\ \boldsymbol{J}_2(t) = (\boldsymbol{p}_N(t) + \boldsymbol{v}_{cN}(t))\boldsymbol{v}_{sM}(t) \end{cases} \tag{13}$$

where for the subarray with M sensors

$$\begin{cases} \boldsymbol{v}_{cM}(t) = \boldsymbol{v}_{xM}(t)\cos(\varphi) + \boldsymbol{v}_{yM}(y)\sin(\varphi) \\ \boldsymbol{v}_{sM}(t) = -\boldsymbol{v}_{xM}(t)\sin(\varphi) + \boldsymbol{v}_{yM}(t)\cos(\varphi) \end{cases} \tag{14}$$

and for the subarray with N sensors

$$\begin{cases} \boldsymbol{v}_{cN}(t) = \boldsymbol{v}_{xN}(t)\cos(\varphi) + \boldsymbol{v}_{yN}(y)\sin(\varphi) \\ \boldsymbol{v}_{sN}(t) = -\boldsymbol{v}_{xN}(t)\sin(\varphi) + \boldsymbol{v}_{yN}(t)\cos(\varphi) \end{cases} \tag{15}$$

3.2 Estimating DOA based on JCTs for coprime vector hydrophone array

Based on CBF, we define the spatial spectrum output concave point discriminant algorithm

$$\boldsymbol{F}(\theta) = \boldsymbol{B}*_{NM}(\theta) \cdot \boldsymbol{B}^{\star}_{MN}(\theta) \tag{16}$$

where $B_{NM}(\theta)$ and $B_{MN}(\theta)$ are the subarray beam output obtained by beamforming after vector coprime array channel combination based on Eqs. (4) to (7), and Eqs. (13) to (15). Compared with one single sparse array, the relation between two sparse subarrays of the coprime vector hydrophone array is established, thus improving the reliability of the concave points judgment. Let Θ be the search step and θ_s be the suspected target's orientation. The discriminating process can be expressed as

$$\begin{aligned} D_p^1 &= IF(\boldsymbol{F}(\theta_s) - \boldsymbol{F}(\theta_s - \Theta) < 0) \\ &\quad \cdot IF(\boldsymbol{F}(\theta_s) - \boldsymbol{F}(\theta_s + \Theta) < 0), \\ D_p^2 &= IF(\boldsymbol{F}(\theta_s + \Theta) - \boldsymbol{F}(\theta_s) < 0) \\ &\quad \cdot IF(\boldsymbol{F}(\theta_s + \Theta) - \boldsymbol{F}(\theta_s + 2\Theta) < 0), \\ D_p^3 &= IF(\boldsymbol{F}(\theta_s - \Theta) - \boldsymbol{F}(\theta_s - 2\Theta) < 0) \\ &\quad \cdot IF(\boldsymbol{F}(\theta_s - \Theta) - \boldsymbol{F}(\theta_s) < 0). \end{aligned} \tag{17}$$

where "$IF()$" indicates if conditional operation.

3.3 Major steps and practical application

The algorithm steps mainly focus on the data preprocessing, the constructions of JCTs and characteristic data point identification algorithm, and the source directions determination. The preprocessing is conducted based on Eq. (5), and the result can be robust because of the CBF, which can be validated in *Section 4*. The suspected targets' orientations are predetermined with the beam output of the whole coprime vector sensor array. The JCTs are established by taking advantage of the channel data combination of the vector sensor array on Eq. (13). Based on JCTs, the discriminant algorithm for identifying the concave points can be achieved by Eq. (16). In either case, one single target or two detected with a specified detection threshold, the source direction can be determined. Since there is no possibility of false-alarm lobes of array output in either case, only the true output is presented. Furthermore, for more suspected directions, whether there are false targets will be determined according to Eq. (8), and coherent sources can be identified efficiently based on Eq. (17). The pseudo-code of the proposed method is exhibited in Algorithm 1.

Require:
Input data: Array beam data $\boldsymbol{S}_t$
Initialize parameters: Signal integral length T_s, Angle search range Θs, Detection threshold D_T, the flag for grating lobes exist or not $Flag = 0$, Concave point set Θc, Target direction set Θf.
Ensure:
while $Length(\boldsymbol{S}_t) = T_s$ **do**
  for $j = 1 : \Theta_s$ **do**
    Beamforming $\boldsymbol{B}_j$ with Eq. (5)
  end for
  Output $\boldsymbol{B}_\Theta$
  Update Θ_f with D_T
  Update $Flag$ with Eq. (8)
  if $Flag = 1$ **then**
    Update Θ_c with Eq (13) to Eq (17)
    if $((\Theta_f \cap \Theta_c) \neq \varnothing)$ **then**
      Update Θ_f
    else
      D_T adjustment
    end if
  end if
end while
Output Θ_f

ALGORITHM 1
Pseudo code of the major steps for the overall algorithm.

4 Simulation analysis

4.1 Accuracy performance

Numerical simulations are conducted to assess the performance of the proposed method. Furthermore, MUSIC based on the SS (Pillai and Kwon, 1989) and FBSS methods (Shan et al., 1985) are used as comparison methods. A coprime vector hydrophone array with 10 sensors (M = 5 and N = 6) is adopted in all examples.

The first part of the simulations investigates the situation in Figure 4. Two coherent sources with the same frequency, 500 Hz, come from the directions $\theta_1 = 57.8^\circ$ and $\theta_2 = 86.2^\circ$, respectively,

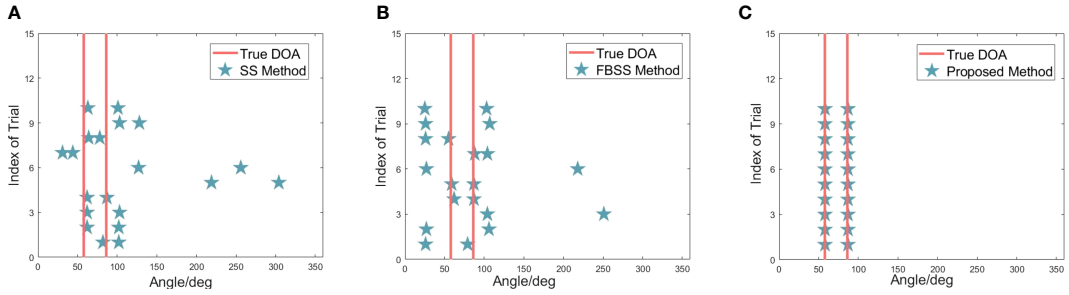


FIGURE 6
Estimation reliability comparison for two coherent sources with the same frequency when SNR = 10 dB and snapshot is 1000. **(A)** Results of the Pillai and Kwon (1989) method. **(B)** Results of the Shan et al. (1985) method. **(C)** Result of the proposed method.

which are shown in red circles in Figure 4B. The grating lobe caused by the same frequency and coherence of the signal leads to wrong target directions of $\theta_3 = 32^\circ$ and $\theta_4 = 105^\circ$, which are presented in blue circles. The DOA estimation performance of different algorithms is evaluated using the root-mean-square error (RMSE), which is described as

$$RMSE = \sqrt{\frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K}(\hat{\theta}_k(i) - \theta_k)^2} \tag{18}$$

where $\hat{\theta}_k(i)$ denotes the estimated DOA of θ_k for the i th independent trial and I and K, respectively, denote the number of Monte Carlo simulation experiments and the number of sources. The suggested method, as shown in Figure 6, consistently produces

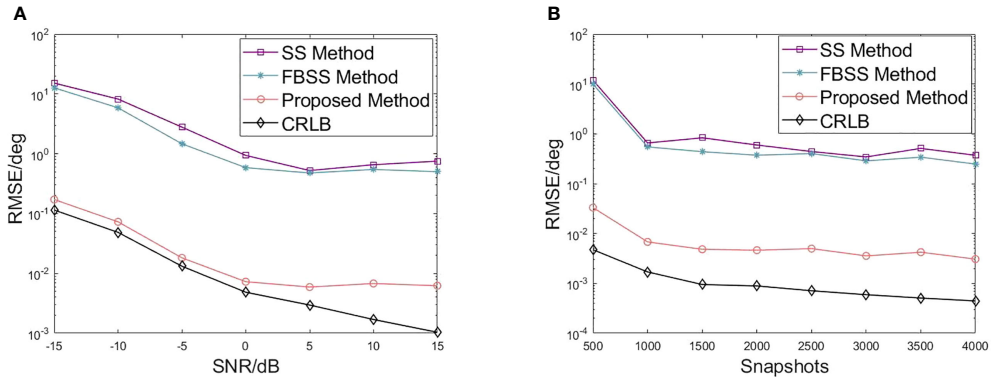


FIGURE 7
RMSE of DOA estimation. Each simulated point is averaged based on 500 trials. **(A)** RMSE versus SNR for two coherent sources with the same frequency. The snapshot is 1,000. **(B)** RMSE versus snapshot for two coherent sources with the same frequency. The SNR is 10 dB.

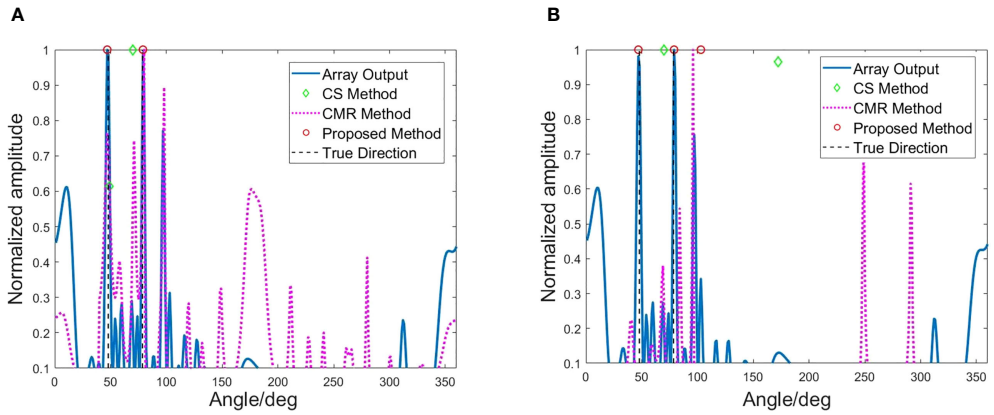


FIGURE 8
Method performance comparison. **(A)** Method performance comparison when the SNR = 10 dB and the signal snapshot is 1,000. **(B)** Method performance comparison when the SNR = −10 dB and the signal snapshot is 1,000.

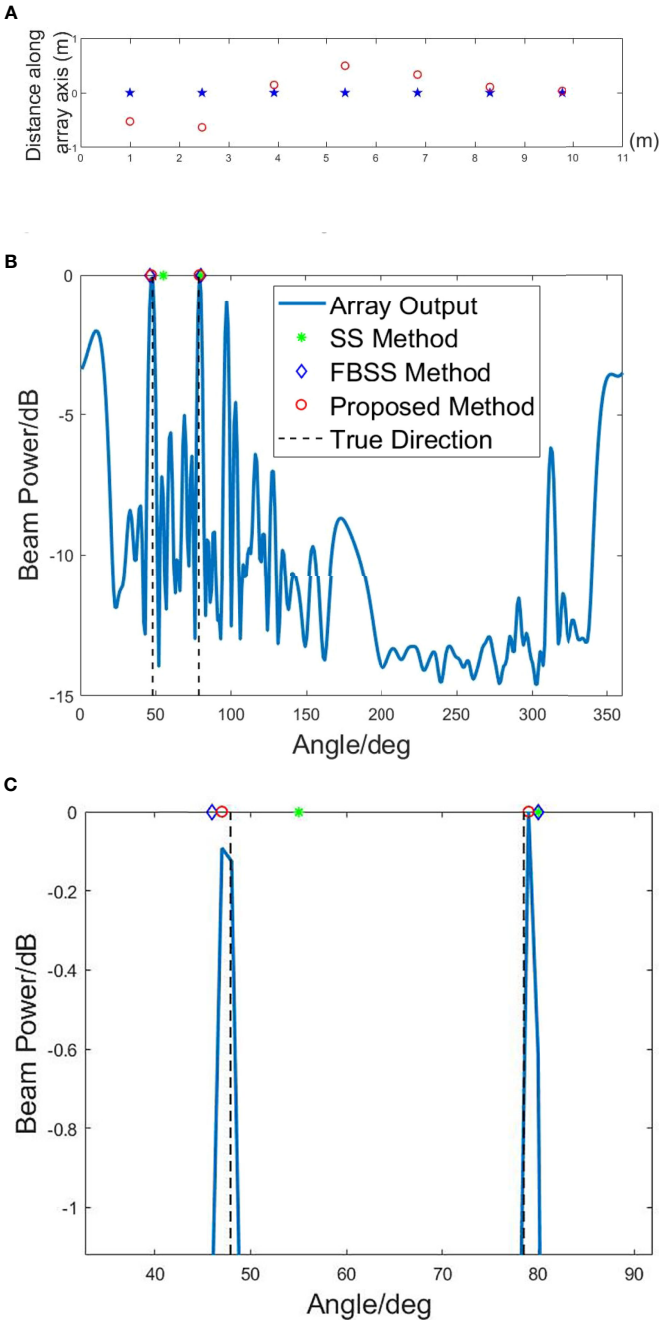


FIGURE 9
Sensitivity of the algorithm to array element position errors. **(A)** Array setup with and without element position errors. **(B)** DOA results based on different methods with element position errors, SNR = 0 dB, $\theta 1$ = 47.9°C and $\theta 2$ = 78.5°C. **(C)** The magnified details for DOA results.

reliable estimates of true DOAs. When the SNR varies from −15 dB to 15 dB, the quantity of snapshots is set to a constant value of 1,000. As shown in Figure 7, the RMSE is reduced with the increase of SNR. Furthermore, when the SNR is fixed at 10 dB, it can be observed that three estimation results become more stable, and the proposed method demonstrates enhanced accuracy as the number of snapshots increases.

The technique of spatial smoothing processing is a widely employed method for decorrelation in practical applications, serving as a foundation for numerous studies. Next, we implement and compare the approaches based on signal covariance matrix recovery (CMR) (Pan et al., 2022) and sparse signal reconstruction using compressive sensing (CS) (Das et al., 2016) with our method. The directions of co-frequency signals are 47.9° and 78.5°, with an

SNR of 10 dB and a signal snapshot of 1000. From Figure 8A, it can be observed that the proposed and the CS methods can obtain the target direction information. The CMR method also achieves high-amplitude output in target directions but suffers from ambiguity. In Figure 8B, when the SNR decreases, both the method in this paper and the CMR method show ambiguous orientation, and the CS method shows direction misjudgment.

4.2 Sensitivity to position errors

In array signal processing, the signal mismatch is a critical issue. In this part, the element position errors are considered to assess the impact of signal mismatch on the proposed method (Yang, 2017). The spacings between array elements were assumed to hold the random error with a mean of 15% of the unit spacing between two sensors, which can be shown as Figure 9A. Receiver positions are represented by symbols, with the desired locations denoted by "☆" and the actual locations denoted by "∘". In Figure 9B, one can find that in the presence of array element errors, the SS method has a DOA estimation bias. In comparison, the FBSS method performs a better DOA estimation accuracy. Compared with the two methods, the method of this work can obtain more accurate DOA estimation results. In addition, it can be seen from Figure 9C that the DOA estimation results of the method depend on the beam output obtained by the conventional beamforming (shown in the legend of "Array output" in the figure). Therefore, the DOA estimation error of the method will be affected by the array beamforming output. However, the algorithm still inherits the robustness of CBFs.

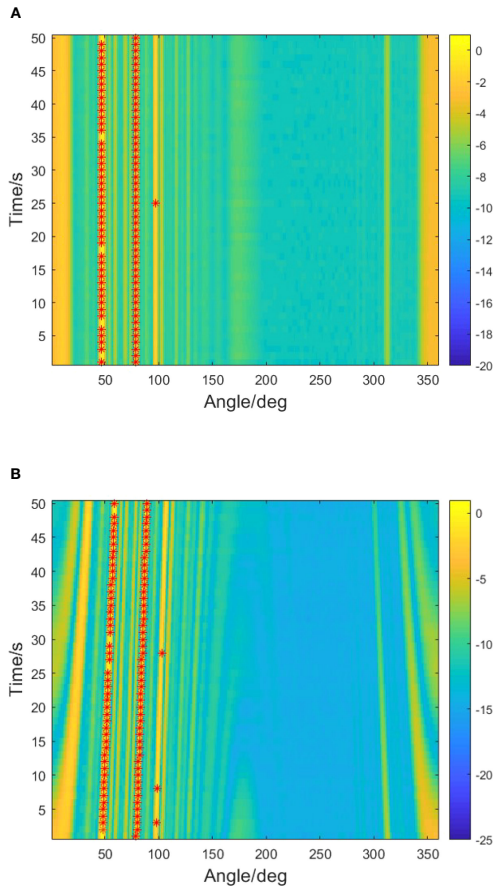


FIGURE 10
BTR for bearing-changing targets. **(A)** SNR = 0 dB. **(B)** SNR = −5 dB.

4.3 Bearing time record performance

This part simulates the bearing time record (BTR) under low SNR. As shown in Figure 10, the red "*" represents the detection result of the algorithm. Figure 10A conducts the simulation for two targets with directions changing. Furthermore, the simulation design ensures high grating lobe interference in the direction change interval (SNR = 0 dB). It can be seen from the figure that many high grating lobe interferences have a severe impact on the target detection results. However, this paper's method can detect targets' actual orientations more stably and accurately. Figure 10B depicts the scenario where two co-frequency targets generate the high grating lobes, with the SNR of both targets being −5 dB, while the targets move in a constant azimuth. It can be seen that there is substantial interference in the direction of the end fire of the array, which will seriously deteriorate the performance of DOA detection. The red "*" shows that the proposed algorithm in this paper demonstrates a stable estimation of true DOAs.

4.4 Attempts in the case of multiple targets

Multiple co-frequency target detection can be divided into three main cases: (1) All targets fall into the relation of overlapping grating lobes. (2) None of the targets fall into the relationship of overlapping grating lobes. (3) Some of the targets hold overlapping grating lobe relation. We set $M = 5$, $N = 6$, SNR = −7 dB, and $f = 500$ Hz. For the first case, there are targets from the directions of 47.9°, 57.8°, 78.5°, and 86.2° respectively, and all of them satisfy the grating lobe overlapping relation. In Figure 11A, one can find that due to the grating lobes, the real directions are submerged within the false alarms, affecting the accurate detection of targets. For the second case, the source directions are set as 30°, 38°, 50°, and 63°. These directions are not in the relationship of grating lobes overlapping. In Figure 11B, without the grating lobes overlapping, array grating lobes will not mask the true direction, and the algorithm can directly obtain the correct target directions' information. However, when some of the targets fall into the relation of overlapping grating lobes, they are from the directions of 20°, 30°, 47.9°, and 78.5°, respectively. As shown in Figure 11C, the method proposed in this article cannot accurately determine the target's true direction from the overlapped lobes of partial targets. Because of the complex grating lobe relationships caused by multiple targets, the feature relationships of the JCTs are affected. Future research will focus on studying and attempting array interference suppression techniques to address this issue.

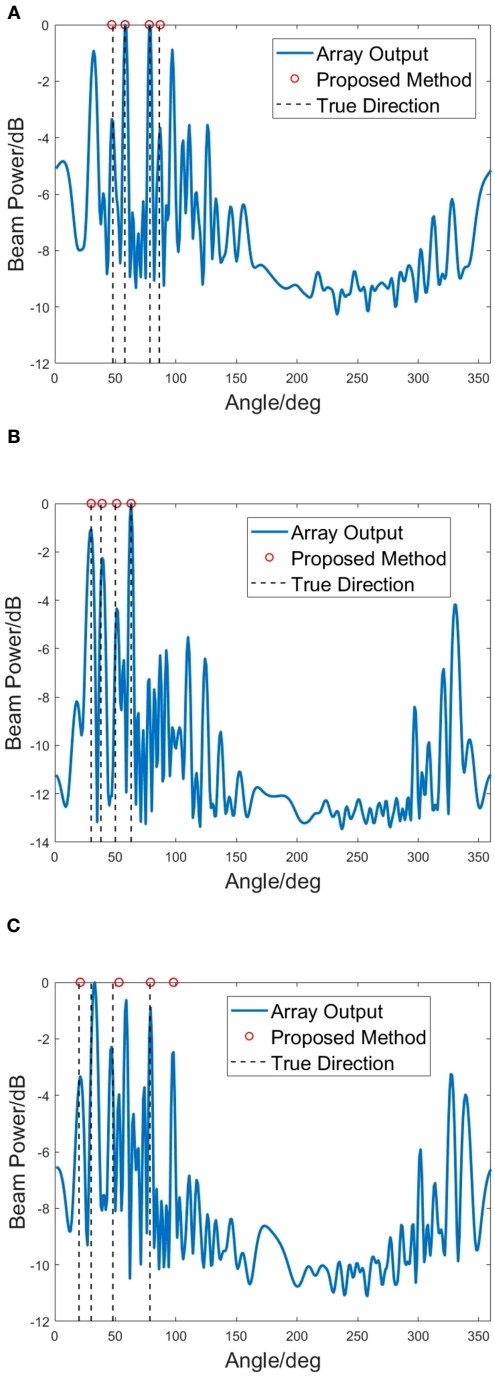


FIGURE 11
Simulation results of algorithm performance in multiple targets case. **(A)** All targets satisfy the grating lobes overlapping relation. **(B)** None of the targets satisfy the grating lobes overlapping relation. **(C)** Some of the targets satisfy the grating lobes overlapping relation.

5 Experiment data analysis

The experiment data analysis has been presented in this part. As shown in Figure 12, the experiment data were collected on an extended coprime vector hydrophone array with 12 sensors on the ocean bottom at a (water) depth of 35 m. The shipborne sound source emits signals to simulate the sound source. The vector hydrophone picks up the underwater sound signal, then transmits the data to the base station through the hydrophone array's data acquisition and transmission system. The base station performs signal processing and realizes the display and reporting of the target detection results. $M = 2$ and $N = 3$ for the extended coprime vector hydrophone and the array expansion factor $e = 3$ (Chen et al., 2023). A moving sound source transmitting at the frequency of 375 Hz moves in a straight line along the direction of 57°. In order to validate the proposed algorithm, the signal data from the direction of 86° have been added to the received signals of the coprime array. The added signal will produce grating lobes in the array output that

FIGURE 12
The experimental layout.

overlap with the source array output. In Figure 13A, one can find that many strong grating lobe interferences show up after the signal was added, deteriorating the performance of source detection and DOA estimation. The proposed method can obtain the DOA information of targets more accurately. Because the signal data from the direction of 86° are artificially added, the DOA estimation method shows more stable results, which are shown in Figure 13B. In the second experiment, the sound source transmits the signal at the frequency of 315 Hz and moves in the direction of 43°, as shown in Figure 14A. In order to increase the grating lobe interferences, the signal from 67° has been included in the original received array signal. Owing to the additional signal, the grating lobes from two sources coincide, resulting in many grating lobe interferences in the array output. The real targets have been buried in strong grating lobe interferences and wide array beams. It can be observed that the proposed method achieves the extraction of real targets from the strong grating lobe interferences and then realizes the targets' DOA estimation, as shown in Figure 14B.

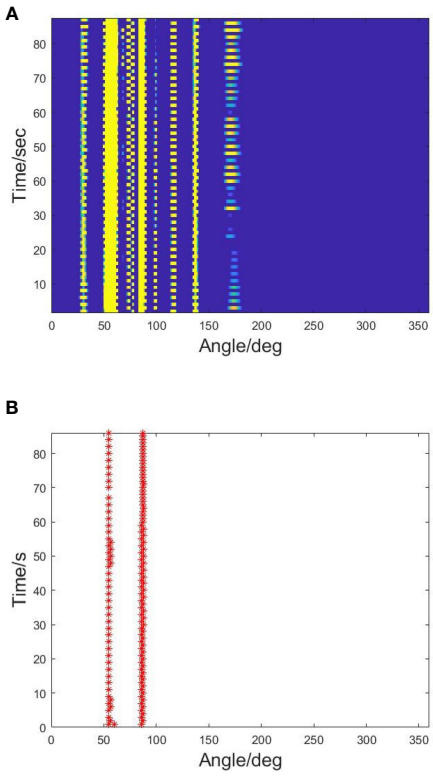


FIGURE 13
BTR in experiment. **(A)** $\theta 1$ = 57°C and $\theta 2$ = 86°C. **(B)** Results of the proposed method.

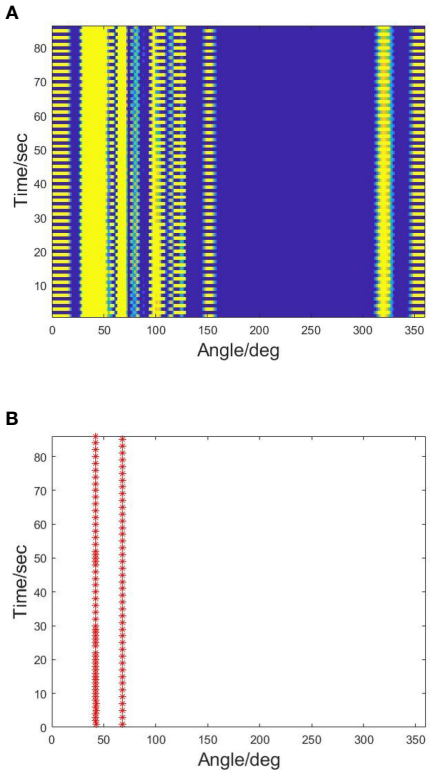


FIGURE 14
BTR in experiment. **(A)** $\theta 1$ = 43°C and $\theta 2$ = 67°C. **(B)** Results of the proposed method.

6 Conclusion

This paper investigates the problem of false alarms that can deteriorate the performance of DOA estimation for two co-frequency sources in a coprime vector hydrophone array. These false alarms are caused by the overlap of main lobes and grating lobes from subarrays. To address this issue, we propose a DOA estimation method that involves JCTs connected with subarrays from a coprime vector hydrophone array. Based on JCTs, we design a method to identify characteristic data points. The proposed method eliminates false-alarm directions without smoothing and detects true DOAs without ambiguity. Simulation and BTR results from the sea experiment data demonstrate that the algorithm performs well and provides a new approach for DOA estimation of coprime vector sensor arrays. Applying large aperture arrays will be a major trend in ocean observation and maritime combat, like marine life detection, UUV (unmanned underwater vehicle), and USV (unmanned surface vehicle) operations. Coprime arrays and their related signal-processing methods will play an important role in the marine domain. The method proposed in this article can also be applied to combined active and passive sonar detection and multi-base sonar cooperative detection. Furthermore, with the application of deep learning in ocean observation, combining deep learning concepts with the method presented in this article may achieve more efficient results in ocean observation, such as target recognition and tracking.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

XC: Conceptualization, methodology, and writing—original draft. HZ: Validation, supervision, funding acquisition, and project administration. YG and ZW: Graphical abstract preparation. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by Marine S and T fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2018SDKJ0210).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adhikari, K. (2019). Beamforming with semi-coprime arrays. *J. Acoustical Soc. America* 145, 2841–2850. doi: 10.1121/1.5100281

Adhikari, K., and Buck, J. R. (2015). "Gaussian signal detection by coprime sensor arrays," in South Brisbane, QLD: *2015 IEEE International Conference on acoustics, speech and signal processing (ICASSP)* (South Brisbane, QLD: IEEE). 2379–2383.

Adhikari, K., and Buck, J. R. (2017). Spatial spectral estimation with product processing of a pair of colinear arrays. *IEEE Trans. Signal Process.* 65, 2389–2401. doi: 10.1109/TSP.2017.2659642

Adhikari, K., Buck, J. R., and Wage, K. E. (2013). "Beamforming with extended co-prime sensor arrays," in *2013 IEEE international conference on acoustics, speech and signal processing* (Vancouver, BC: IEEE), 4183–4186.

Adhikari, K., Buck, J. R., and Wage, K. E. (2014). Extending coprime sensor arrays to achieve the peak side lobe height of a full uniform linear array. *EURASIP J. Adv. Signal Process.* 2014, 1–17. doi: 10.1186/1687-6180-2014-148

Alawsh, S. A., and Muqaibel, A. H. (2018). Multi-level prime array for sparse sampling. *IET Signal Process.* 12, 688–699. doi: 10.1049/iet-spr.2017.0252

Alawsh, S. A., and Muqaibel, A. H. (2020). Achievable degree-of-freedom for three-level prime arrays. *Signal Process.* 171, 107523. doi: 10.1016/j.sigpro.2020.107523

Alawsh, S. A., and Muqaibel, A. H. (2021). "Optimized multi-level prime array configurations," in *2021 International Symposium on Networks, Computers and Communications (ISNCC)* (Dubai, United Arab Emirates: IEEE), 1–4.

Chen, X., Zhang, H., and Lv, Y. (2023). Improving the beamforming performance of a vector sensor line array with a coprime array configuration. *Appl. Acoustics* 207, 109329. doi: 10.1016/j.apacoust.2023.109329

Das, A., Hodgkiss, W. S., and Gerstoft, P. (2016). Coherent multipath direction-of-arrival resolution using compressed sensing. *IEEE J. Oceanic Eng.* 42, 494–505. doi: 10.1109/JOE.2016.2576198

Di Martino, G., and Iodice, A. (2017). Passive beamforming with coprime arrays. *IET Radar Sonar Navigation* 11, 964–971. doi: 10.1049/iet-rsn.2016.0517

Felisberto, P., Santos, P., and Jesus, S. M. (2018). Acoustic pressure and particle velocity for spatial filtering of bottom arrivals. *IEEE J. Oceanic Eng.* 44, 179–192. doi: 10.1109/JOE.2018.2807898

Felisberto, P., Santos, P., Maslov, D., and Jesus, S. (2016). "Combining pressure and particle velocity sensors for seismic processing," in *OCEANS 2016 MTS/IEEE Monterey* (Monterey, CA: IEEE), 1–6.

Fu, M., Zheng, Z., Wang, W.-Q., and So, H. C. (2021). Coarray interpolation for doa estimation using coprime emvs array. *IEEE Signal Process. Lett.* 28, 548–552. doi: 10.1109/LSP.2021.3061272

Hawkes, M., and Nehorai, A. (1998). Acoustic vector-sensor beamforming and capon direction estimation. *IEEE Trans. Signal Process.* 46, 2291–2304. doi: 10.1109/78.709509

Liu, Y., and Buck, J. R. (2015). "Detecting gaussian signals in the presence of interferers using the coprime sensor arrays with the min processor," in *2015 49th Asilomar conference on signals, systems and computers* (Pacific Grove, CA: IEEE), 370–374.

Mei, J., Sun, D., Zhou, Y., and Shi, J. (2018). Study on the problem of coherent source resolution with same frequency for the coprime array. *J. Acoustical Soc. America* 144, 1971–1971. doi: 10.1121/1.5068615

Moghadam, G. S., and Shirazi, A. A. B. (2019). "Doa estimation with extended optimum co-prime sensor array (eocsa)," in *2019 Sixth Iranian Conference on Radar and Surveillance Systems* (Isfahan, Iran: IEEE), 1–6.

Moghadam, G. S., and Shirazi, A. B. (2022). Direction of arrival (doa) estimation with extended optimum co-prime sensor array (eocsa). *Multidimensional Syst. Signal Process.* 33, 17–37. doi: 10.1109/ICRSS48293.2019.9026546

Pan, H., Pan, J., and Zhang, X. (2022). "Time-delay estimation of coherent gpr signal by using sparse frequency sampling and imusic method," in *2022 International Symposium on Wireless Communication Systems (ISWCS)*, (Hangzhou, China: IEEE). 1–6.

Pillai, S. U., and Kwon, B. H. (1989). Forward/backward spatial smoothing techniques for coherent signal identification. *IEEE Trans. Acoustics Speech Signal Process.* 37, 8–15. doi: 10.1109/29.17496

Qin, S., Zhang, Y. D., Amin, M. G., and Himed, B. (2017). Doa estimation exploiting a uniform linear array with multiple co-prime frequencies. *Signal Process.* 130, 37–46. doi: 10.1016/j.sigpro.2016.06.008

Roy, R., and Kailath, T. (1989). Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. acoustics speech Signal Process.* 37, 984–995. doi: 10.1109/29.32276

Santos, P., João, J., Rodríguez, O. C., Felisberto, P., and Jesus, S. M. (2011). "Geometric and seabed parameter estimation using a vector sensor arrayexperimental results from makai experiment 2005," in *OCEANS 2011 IEEE-Spain* (IEEE), 1–10.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Trans. antennas propagation* 34, 276–280. doi: 10.1109/TAP.1986.1143830

Shan, T.-J., Wax, M., and Kailath, T. (1985). On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. Acoustics Speech Signal Process.* 33, 806–811. doi: 10.1109/TASSP.1985.1164649

Tan, Z., Eldar, Y. C., and Nehorai, A. (2014). Direction of arrival estimation using co-prime arrays: A super resolution viewpoint. *IEEE Trans. Signal Process.* 62, 5565–5576. doi: 10.1109/TSP.2014.2354316

Vaidyanathan, P. P., and Pal, P. (2010). Sparse sensing with co-prime samplers and arrays. *IEEE Trans. Signal Process.* 59, 573–586. doi: 10.1109/TSP.2010.2089682

Vaidyanathan, P., and Pal, P. (2011). Theory of sparse coprime sensing in multiple dimensions. *IEEE Trans. Signal Process.* 59, 3592–3608. doi: 10.1109/TSP.2011.2135348

Xie, Z., Lin, R., Wang, L., Zhang, A., Lin, J., and Tang, X. (2023). Data augmentation and deep neural network classification based on ship radiated noise. *Front. Mar. Sci.* 10, 1113224. doi: 10.3389/fmars.2023.1113224

Yang, T. (2017). Deconvolved conventional beamforming for a horizontal line array. *IEEE J. Oceanic Eng.* 43, 160–172. doi: 10.1109/JOE.2017.2680818

Yang, Y., Ruan, N., Huang, G., Shi, J., and Wen, F. (2021). A propagator method for bistatic coprime emvs-mimo radar. *Math. Problems Eng.* 2021, 1–11. doi: 10.1155/2021/9954573

Yang, T., and Ye, Z. (2019). Array gain of coprime arrays. *J. Acoustical Soc. America* 146, EL306–EL309. doi: 10.1121/1.5126924

Zhang, Y. D., Amin, M. G., and Himed, B. (2013). "Sparsity-based doa estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (Vancouver, BC: IEEE), 3967–3971.

Zhang, G., Liu, K., Fu, J., and Sun, S. (2022). Covariance matrix reconstruction method based on amplitude and phase constraints with application to extend array aperture. *J. Acoustical Soc. America* 151, 3164–3176. doi: 10.1121/10.0010383

Zhang, X., Yang, P., and Zhou, M. (2023). "Multireceiver sas imagery with generalized pca," in *IEEE Geoscience and Remote Sensing Letters* (USA: IEEE).

Zhou, C., Gu, Y., He, S., and Shi, Z. (2017). A robust and efficient algorithm for coprime array adaptive beamforming. *IEEE Trans. Vehicular Technol.* 67, 1099–1112. doi: 10.1109/TVT.2017.2704610

Zhou, C., Gu, Y., Song, W.-Z., Xie, Y., and Shi, Z. (2016). "Robust adaptive beamforming based on doa support using decomposed coprime subarrays," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE)* (Shanghai, China). 2986–2990.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 23 August 2023
DOI 10.3389/fmars.2023.1213002

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Xiao Feng,
Xiamen University, China
Zeyad Qasem,
Peking University, China

*CORRESPONDENCE
Cheng Chen
chen.cheng@nwpu.edu.cn

RECEIVED 27 April 2023
ACCEPTED 08 August 2023
PUBLISHED 23 August 2023

CITATION
Liu Y, Chen C and Feng X (2023) Investigating the reliable acoustic path properties in a global scale. *Front. Mar. Sci.* 10:1213002. doi: 10.3389/fmars.2023.1213002

Investigating the reliable acoustic path properties in a global scale

Ying Liu, Cheng Chen* and Xiao Feng

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi, China

Leveraging the benefits of low transmission loss and high signal-to-noise ratio, the reliable acoustic path (RAP) has been extensively employed in various underwater applications. In this study, we investigate RAP properties on a global scale. Acoustic simulations were conducted using global grids with a 0.25° × 0.25° spatial resolution, revealing that RAP range is positively correlated with ocean depth. Contrary to the prevailing belief that RAP properties are relatively unaffected by sound speed variations, our findings indicate that sound speed profiles (SSPs) play a crucial role in determining RAP properties by altering the RAP from 15 km to 50 km at a constant ocean depth of 4000 m. Additionally, the receiver angle can vary by nearly 5 km at the same source location due to SSP variations. Consequently, utilizing highly accurate SSPs can enhance the performance of underwater localization or communication systems that rely on RAP.

KEYWORDS

ocean acoustic propagation, properties of reliable acoustic path, sound speed profile variation, bathymetry variation, transmission loss

Introduction

Ocean acoustic propagation has long been recognized as sensitive to various ocean environmental parameters (Heitsenrether and Mohsen, 2004; Dosso et al., 2007a; Dosso et al., 2007b; Lermusiaux et al., 2010; Pecknold and Osler, 2012; Ngodock et al., 2022; Zhang et al., 2023a; Zhang et al., 2023b), such as bathymetry, geo-acoustic parameters, sound speed fields, and sea surface roughness. This sensitivity renders underwater acoustic activities, including localization and communication, highly dependent on the ocean environment. In deep ocean acoustic propagation studies, sound energy is typically categorized into distinct propagation patterns, including surface duct propagation, convergence zone propagation, bottom bounce propagation, and reliable acoustic path propagation. These propagation patterns exhibit varying levels of sensitivity to changes in the ocean environment. For instance, surface duct energy propagation is sensitive to the

thickness of the surface duct (Chen et al., 2016; Chen et al., 2019; Duan et al., 2016), convergence zone propagation is sensitive to the thermocline gradient (Worcester et al., 2013; Chen et al., 2017; Khan et al., 2021), and bottom bounce propagation is sensitive to geo-acoustic parameters (Choi and Peter, 2004; Heaney et al., 2013; Yang et al., 2017).

The Reliable Acoustic Path (RAP), which represents the direct path between a source and receiver, is generally considered to be less sensitive to ocean environment fluctuations. Moreover, RAP boasts low transmission loss (TL) compared to boundary-reflected paths (Duan et al., 2012) and lower ambient noise levels, with noise below the critical depth being nearly 20 dB less than above it (Gaul et al., 2007). Due to these advantages, such as low TL and high signal-to-noise ratio (SNR), RAP has been widely employed in underwater acoustic localization and communication. Duan et al. utilized RAP multipath time de-lays recorded by a single hydrophone for localizing a moving source (Duan et al., 2014), while the interference structure of RAP was leveraged to estimate source depth with robust performance (Duan et al., 2019). Recently, Qu et al. conducted a comprehensive spatial gain analysis for vertical line arrays in RAP regions (Qiu et al., 2018), and Tompson incorporated RAP as a crucial sound propagation factor in deep ocean acoustic networks, achieving impressive performance in high SNR scenarios (Thompson, 2009).

Besides localization and communication, RAP sound energy has also been applied in geo-acoustic inversion and ocean tomography. Geo-acoustic parameters in the Philippine Sea were obtained using RAP sound energy (Xu et al., 2019), and Varamo et al. examined the feasibility of RAP tomography by employing a mobile ship with an acoustic source transmitting to a fixed bottom hydrophone at the ALOHA Cabled Observatory (Varamo and Howe, 2016).

Despite the growing interest in RAP over the past half-century, the effects of sound speed profiles (SSPs) on RAP properties remain underexplored. Xiao et al. demonstrated that RAP TL is relatively insensitive to seasonal SSP variations (Xiao et al., 2016); however, in some cases, seasonal SSP variations may be minimal compared to global spatial variations.

The properties of the Reliable Acoustic Path (RAP) are highly significant, and prior to our study, there had been no global research conducted on this topic. The prevailing impression was that RAP was insensitive to sound speed profiles; however, our findings have proven otherwise. Through extensive analysis and research, we have demonstrated the sensitivity of RAP to sound speed profiles.

In our study, we conducted a global investigation of Reliable Acoustic Path (RAP) properties and their variations, identifying ocean depth and sound speed profile (SSP) structure as the primary influencing factors. The findings of this paper offer valuable insights for leveraging RAP in underwater acoustic localization and communication. The impact of SSP structure on RAP is found to be significant, with RAP range varying up to 35 kilometers for a depth of 4000 meters. Dynamically complex regions, like the Kuroshio Extension, exhibit substantial variations in SSP even at the same location, resulting in notable changes in RAP characteristics.

Deep-sea positioning methods heavily rely on underwater acoustic propagation characteristics and rely on angle information from measurements for precise underwater localization. Inaccuracies in the sound speed profile can lead to significant reductions in positioning accuracy. For effective underwater acoustic communication, acquiring sufficient environmental information is crucial to obtain more precise underwater acoustic channel characteristics.

In conclusion, this comprehensive study enhances our understanding of RAP properties across different regions worldwide, providing essential guidance for utilizing RAP in underwater applications such as localization and communication.

Properties of reliable acoustic path

Figure 1 presents a typical case of 2D transmission loss of the reliable acoustic propagation. The SSP was the Munk profile as shown in Figure 1A. The source frequency was 200 Hz, the ocean depth was 5000 m, and the source depth was 4800 m. Figure 1B shows that when the source was deployed near the ocean bottom, the coverage of the RAP was bowl shaped. The radiance of the RAP could reach nearly 40 km near the surface, suggesting that the receiver deployed in this case could detect targets within this range with a high SNR ratio.

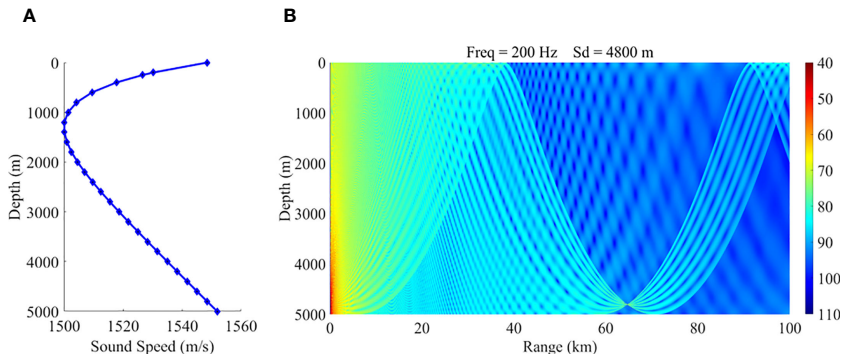


FIGURE 1
2D transmission loss of the RAP. **(A)** Sound speed profile **(B)** 2D transmission loss.

In order to validate the model and simulation method employed in this study, we utilized acoustic experiment data collected from a specific region in the Western Pacific Ocean in 2021 and compared it to the simulated results. For the comparison with the experimental data, we used a ray model and the acoustic reciprocity method, setting the source depth at 5470m and the receiving depth at 300m. In the Figure 2, Figure 2A displays the two-dimensional propagation loss acquired through simulation using real topography, sediment characteristics, and sound speed profiles, while Figure 2B presents a comparison between the propagation loss at 200Hz and the experimental data at a receiving depth of 300m. As evident in the figure, both the simulated results and the experimental data exhibit a rapid decline trend at the edge of the reliable acoustic paths, with similar magnitudes, implying that the model and modeling methods used in this study are relatively trustworthy. Moreover, the ray model has mature applications in acoustic field simulation. The results of the model are reliable in deep-sea areas, and are effective for the analysis of reliable acoustic paths in this study.

Here, we focused on two RAP properties, namely, the RAP radiance near the sea surface and the receiver direction that varies with the source range. In the previous studies of Duan et al (Heaney et al., 2013), they showed that the receiver direction is sensitive to the source locations. Figure 3 presents the receiver directions and the source locations. The source/receiver set is the same as that in Figure 1. At a water depth of 5000 meters and a source depth of 4800 meters, the sound speed profile follows the Munk curve as shown in Figure 1A. The seafloor has a minor impact on the arrival structure and is set with a sound speed of 1600 m/s, density of 1.8 g/cm^3, and an attenuation coefficient of 0.8 dB/wavelength. In Figure 3, the receiver directions of the direct path (also named RAP), the bottom reflected path with no surface reflection, the surface reflected path with no bottom reflection and the surface and bottom reflected path are denoted as B0T0, B1T0, B0T1 and B1T1, respectively. We notice that the receiver directions of the RAP arrival and the other paths are sensitive to the source ranges. Thus, determining the source range with the RAP arrivals has practical significance. However, when we use the physics-based RAP arrival to estimate the source location, we should be aware that the RAP properties could be affected by the ocean acoustic environment. For the RAP case, the SSP structure and the ocean depth should be considered. Detailed analyses were carried out on a global scale to study the variation pattern.

Characterization of the reliable acoustic path worldwide

The Etopo1 database was used as the bathymetry data (Hirt and Rexer, 2015). The WOA18 annual database (Locarnini et al., 2018; Zweng et al., 2018) was used in this paper to provide the temperature and salinity profiles for the computation of the SSPs. The empirical formula (Lovett, 1978) is shown in Equation (1), where C (m/s) refers to the sound speed, T (°C) refers to the temperature, and S (‰) refers to the salinity. P (kg/cm^2) is the static pressure of the immediate ocean water column.

$$C = 1449.22 + \Delta C_T + \Delta C_S + \Delta C_P + \Delta C_{STP} \quad (1)$$

Here,

$$\Delta C_T = 4.6233T - 5.4585(10)^{-2}T^2 + 2.822(10)^{-4}T^3 - 5.07(10)^{-7}T^4$$

$$\Delta C_P = 1.60518(10)^{-1}P + 1.0279(10)^{-5}P^2 + 3.451(10)^{-9}P^3 - 3.503(10)^{-12}P^4$$

$$\Delta C_S = 1.391(S-35) - 7.8(10)^{-2}(S-35)^2$$

$$\Delta C_{STP} = (S-35)[-1.197(10)^{-3}T + 2.61(10)^{-4}P - 1.96(10)^{-1}P^2 - 2.09(10)^{-6}PT] + P[-2.796(10)^{-4}T + 1.3302(10)^{-5}T^2 - 6.644(10)^{-8}T^3] + P^2[-2.391(10)^{-1}T + 9.286(10)^{-10}T^2] - 1.745(10)^{-10}P^3T$$

To study the effects of the ocean depth and the SSPs on the RAP properties, we conducted acoustic model simulation on each grid of the WOA18 database. The SSPs were obtained from the WOA18 data of the same grids, and the ocean depths were obtained by

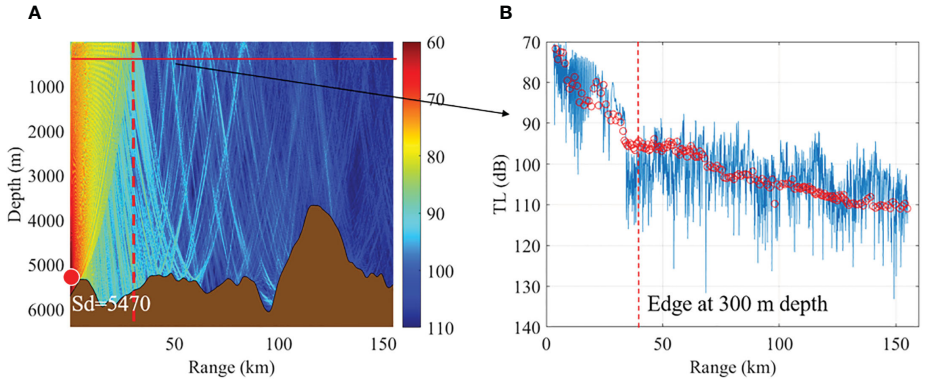


FIGURE 2
The measured data compared to the simulated result. **(A)** Two-dimensional Propagation loss **(B)** Propagation loss comparison.

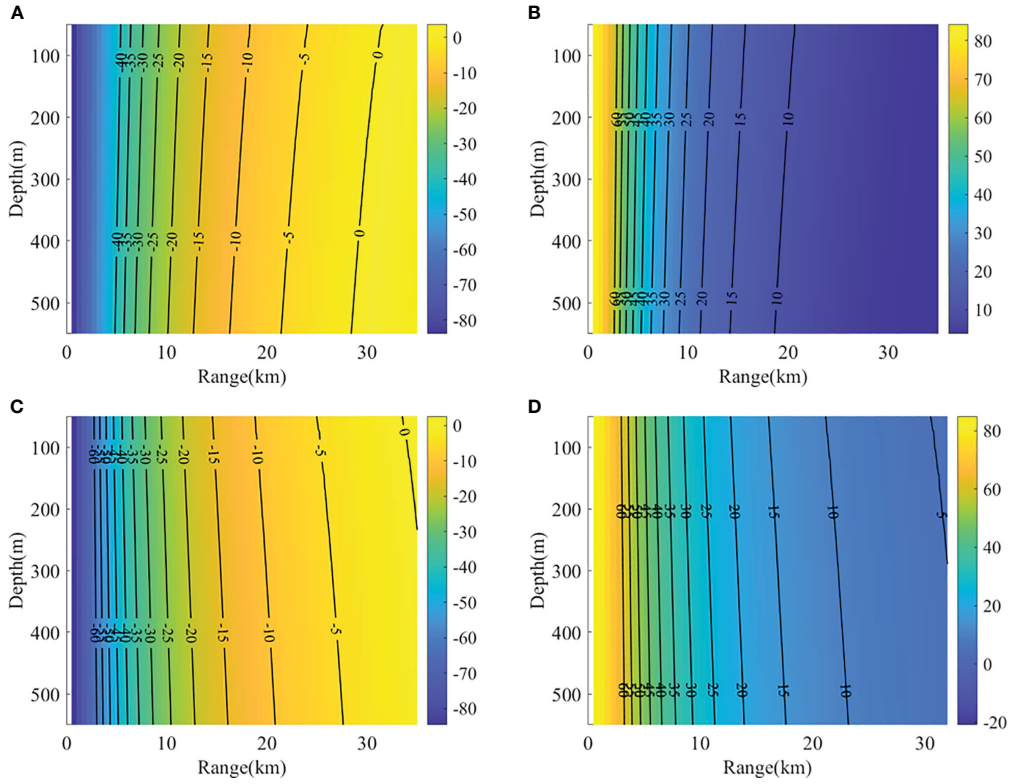


FIGURE 3
Receiver directions and source locations for different arrivals: **(A)** B0T0 **(B)** B1T0 **(C)** B0T1 **(D)** B1T1.

interpolating the ETOPO1 bathymetry over the grids. On each grid, we set up a reasonable range-independent case model because the RAP does not interact with the sea bottom. Thus, the combined effects of the ocean depth and SSP variations on the RAP properties could be obtained on the global scale.

Figure 4A presents the bathymetry from the ETOPO1 database. In Figure 4B, the maximum range of the RAP at the surface has a 0.25° × 0.25° spatial resolution. On each grid point, we set up an acoustic model to calculate the maximum range of the RAP near the surface. The source 200 m above the ocean bottom, and the source frequency was 200 Hz. The SSP was obtained from the WOA18 annual database. The grids with an ocean depth smaller than 2500 m were excluded. The reason for choosing a grid of less than 2500m lies in the focus of our study, which is primarily concerned with deep-sea areas. Given that the sound speed profile inflection points in most sea areas are around 1000m, we wanted to take into account more common situations, enabling us to observe and analyze a more comprehensive structure of the sound speed profile. As a compromise, we chose a depth of 2500m as a boundary point.

Figure 4B indicates that the maximum range of the RAP varies from 20 km to 60 km. The comparison of Figures 4A, B shows that the maximum range of the RAP is weakly related to the ocean depth. Figure 4C presents the relation between the ocean depth and the maximum range of the RAP with a scatter plot. The maximum range of the RAP increases with the ocean depth. For example, at the ocean depth of 4000 m, the maximum RAP range is 15–50 km. Given that the two factors affecting the RAP range are the ocean depth and the SSP structure, the effect of the SSP structure on the RAP range should be examined.

Figure 5A presents the latitudinal Sound Speed Profiles (SSPs) at a longitude of 160.125°E, revealing significant variations in the SSP structure. In Figure 5B, the bathymetry along the same line is displayed, excluding depths smaller than 2500 meters. Figure 5C shows the relationship between the maximum range of the Reliable Acoustic Path (RAP) and ocean depth, indicating a linear trend between the range and the ocean depth.

The structurally consistent sound speed profile was utilized to examine the impact of ocean bathymetry on the reliable acoustic path range under the same sound speed profile conditions. Figure 6 illustrates the variation in the RAP range. The left subplot depicts the SSP used, while the right subplot exhibits the RAP variation as we vary the water depth from 2600 to 5500 meters.

From analyzing Figures 5, 6, it is evident that there is a linear trend between the RAP range and ocean depth. However, the range can vary significantly even at the same ocean depth. This variation can be attributed to the considerable differences in the SSP structure for the same ocean depth, resulting in variations in the RAP range.

To determine the effects of the SSP structure on the RAP properties on the global scale, we clustered the global SSPs obtained from the WOA18 data into 12 groups. There are numerous approaches for classifying transonic airfoil shapes, each resulting in different categorizations. The 12 categories used in this study were chosen purely to facilitate analysis, and the results could be presented even more concisely in a 4 x 3 matrix. The profiles with

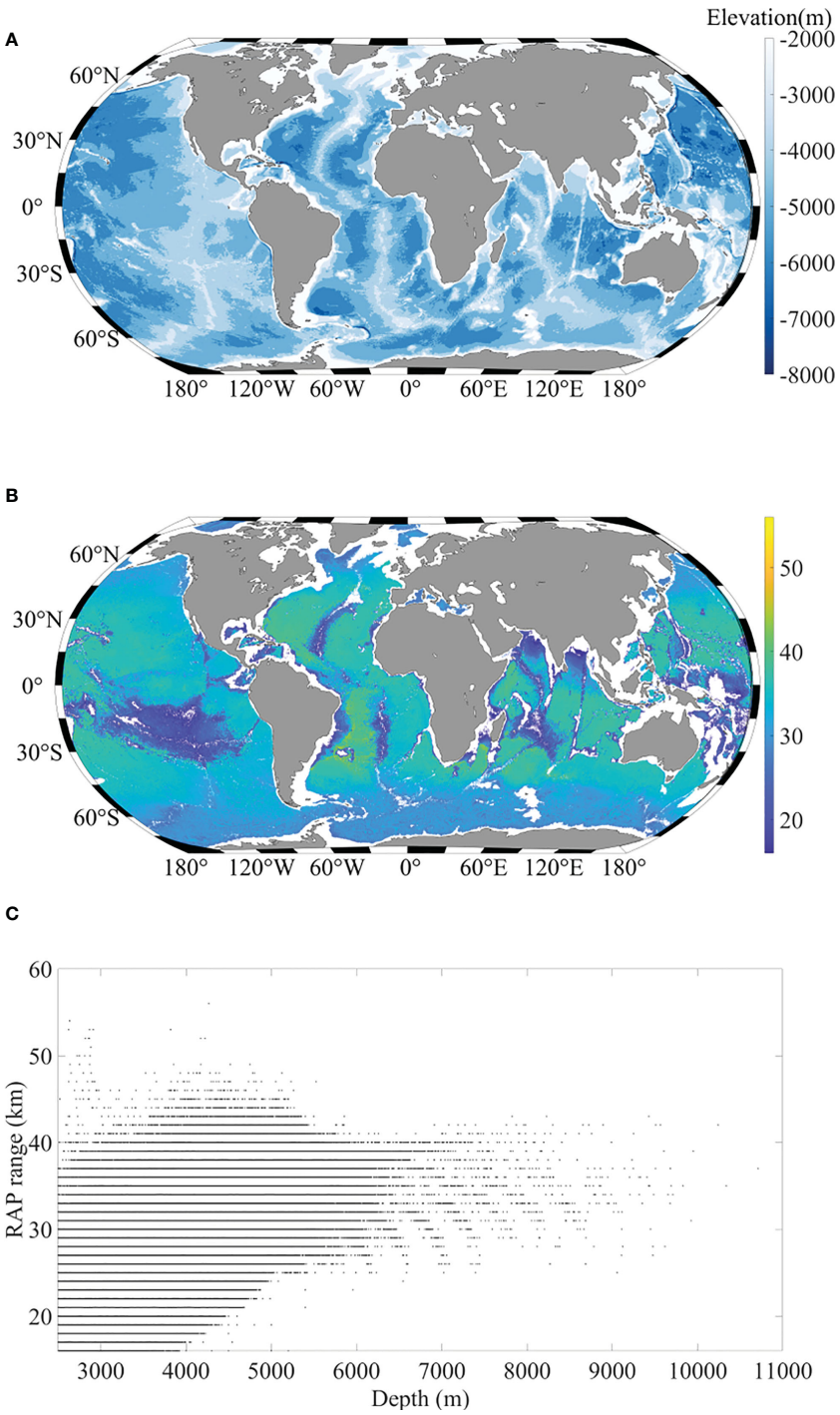


FIGURE 4
(A) Bathymetry of the global ocean, **(B)** the maximum range of the RAP in the global ocean scale, and **(C)** the relation between the ocean depth and the maximum range of the RAP.

a maximum depth exceeding 4500 were used, and a total of 242934 profiles were available for use. The temperature and salinity profiles were used to obtain the SSPs on each grid. Figure 7A shows the 12 types of SSPs, and Figure 7B presents their distributions around the world.

The 12 types of SSPs were then used to study the effects of the SSP structure on the RAP properties. Figure 8 presents the TL and the rays corresponding to the 12 types of SSPs. In Figure 8A, the 2D TLs for the 12 cases with a maximum depth of 4500 m are presented. The source depth was 4300 m, and the source frequency was 200 Hz. Figure 8B presents the 2D TLs for the upper 500 m for the 12 cases. The major difference could be observed near the edge of the RAP, and the SSPs could result in a significant variation pattern of the TL on the upper 500-m deep

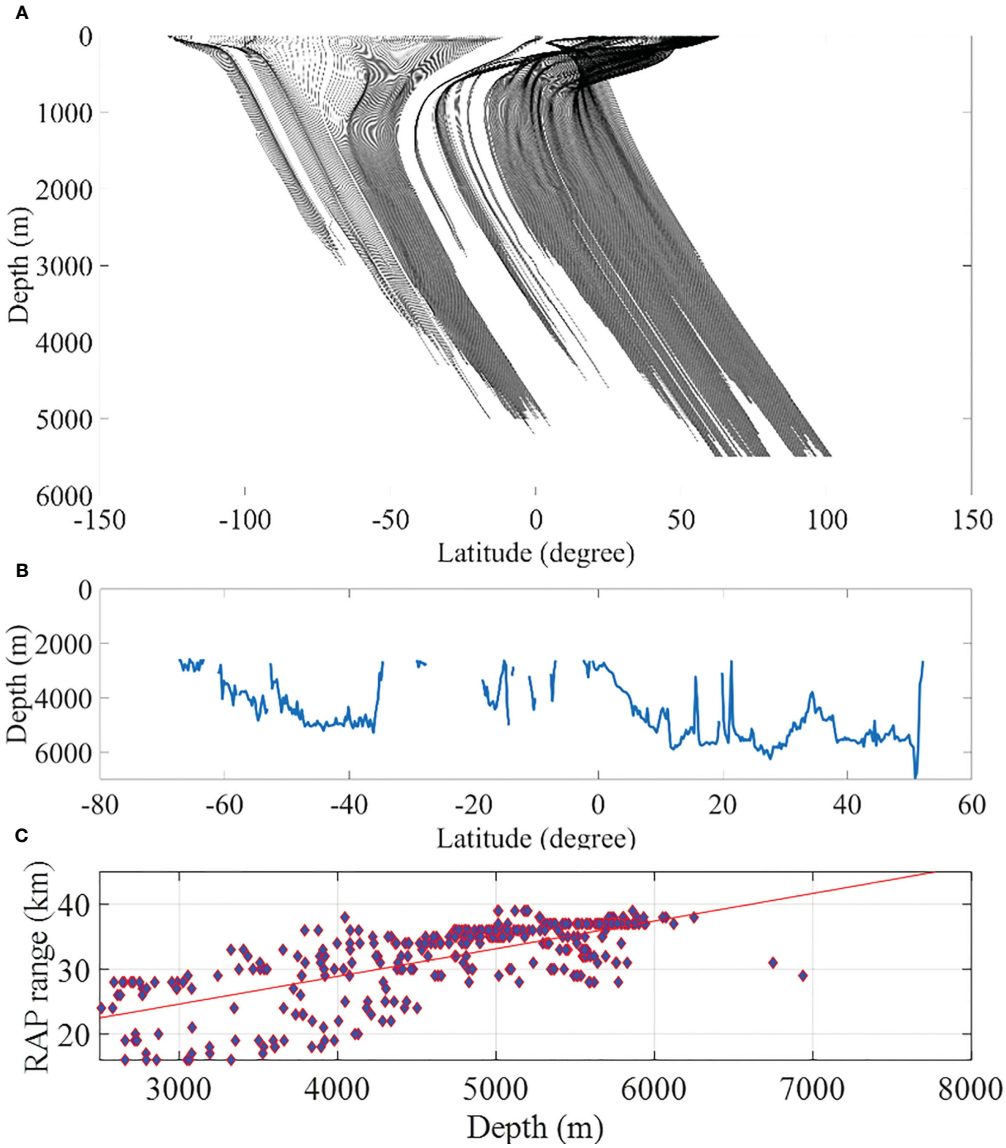


FIGURE 5
(A) SSPs, **(B)** the bathymetry and **(C)** the relation between maximum range of the RAP and the ocean depth along the latitudinal direction at the longitude 160.125°E.

layers. In Figure 8C, the ray's geometry is illustrated with green lines, indicating the case with surface and non-bottom reflections, and the red lines represent the case with no surface or bottom reflection. Figure 8C clearly indicates that the variation of the SSP structure could result in a significant change in RAP properties.

Table 1 lists the maximum range of the RAP for the 12 types of SSPs with an ocean depth 4500 m and a source depth of 200 m above the ocean bottom. The maximum range varies from 25 km to 42 km, which could be significant in many cases when the RAP is used for underwater localization or communication.

Then, we investigated the possible effect of SSP variation on underwater localization. Figure 9 shows the effects of SSP variation on the receiver angle. Figure 9 reveals that the effects of the SSP variation on the receiver angles increase with the source range. For example, at the receiver angle of -10°, the source range could vary by approximately 1 km when the source range is lower than 15 km. At the receiver angle of -5°, the source range could vary by nearly 5 km when the source range reaches or exceeds 20 km. The variation of the source range at the same receiver angle is large, even in the case of 1-km variation because the 1 km variation could be a significant localization error when we attempt to detect underwater sources with the RAP rays.

Specifically, we selected the Kuroshio Extension region to study the effect of the Kuroshio Extension front on the RAP properties. In Figure 10, the left subplot is of the SSPs on the two sides of the Kuroshio Extension front. The right subplot is of the effects of the two SSPs on the receiver angles of the RAP. Figure 10 suggests that the SSPs could induce great variation in the receiver angle. At the 25-km range, the variation of the source range could vary by approximately 1 km in the upper 500-m deep layers at the same

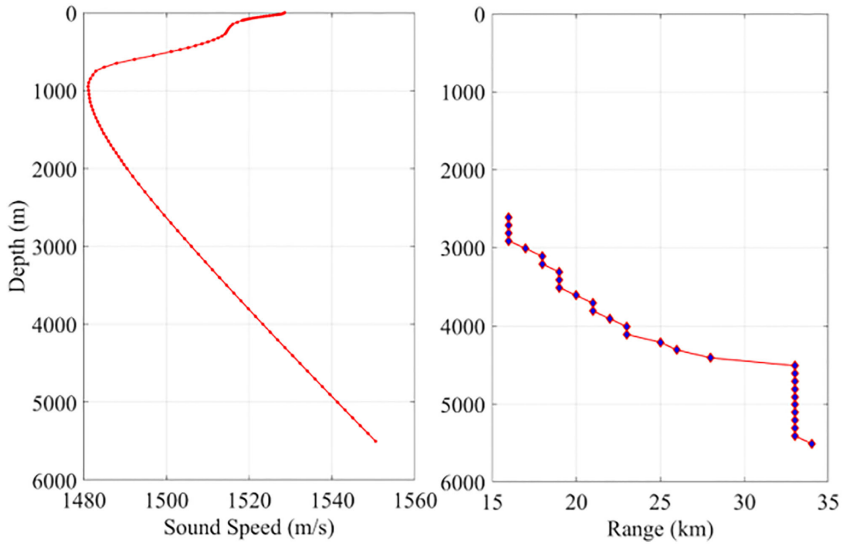


FIGURE 6
The range variation of the Reliable Acoustic Path (RAP) along with the bathymetry variation. The left subplot represents the Sound Speed Profile (SSP) used, while the right subplot shows the RAP variation as we vary the water depth from 2600 to 5500 meters.

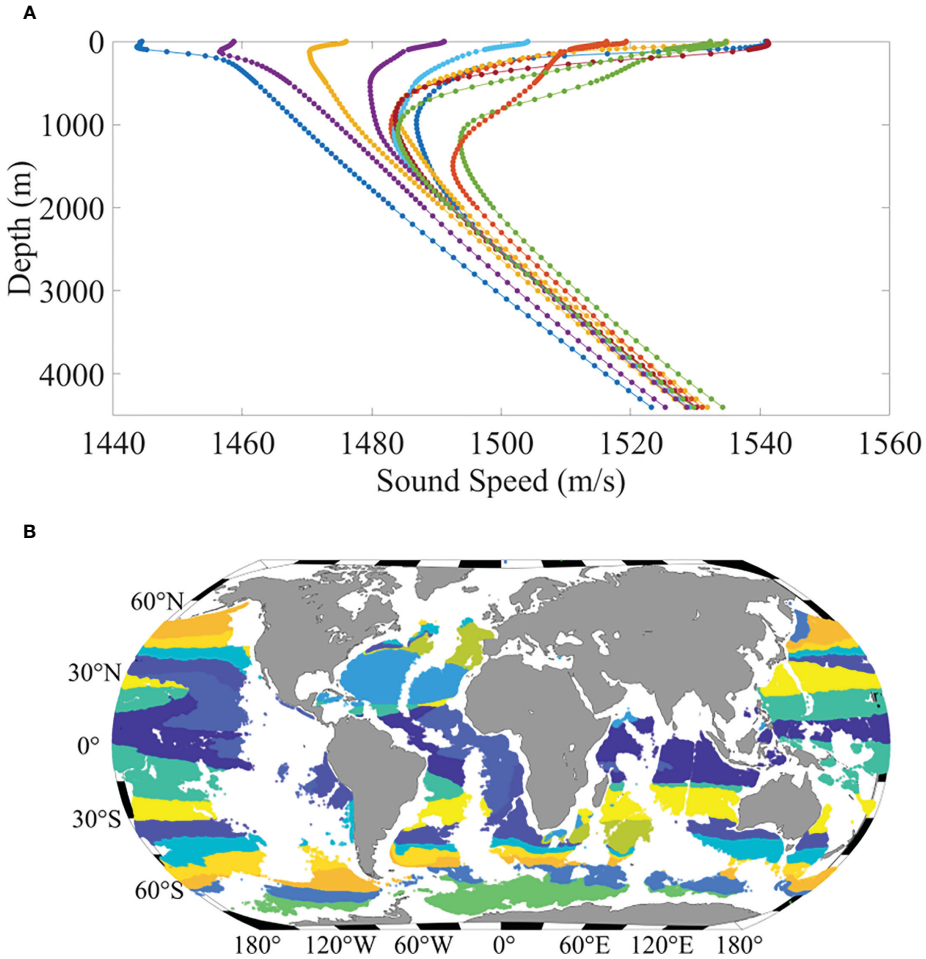


FIGURE 7
(A) 12 types of SSPs around the global ocean, and **(B)** the distribution of the 12 groups of SSPs (each color referring to one group of SSPs).

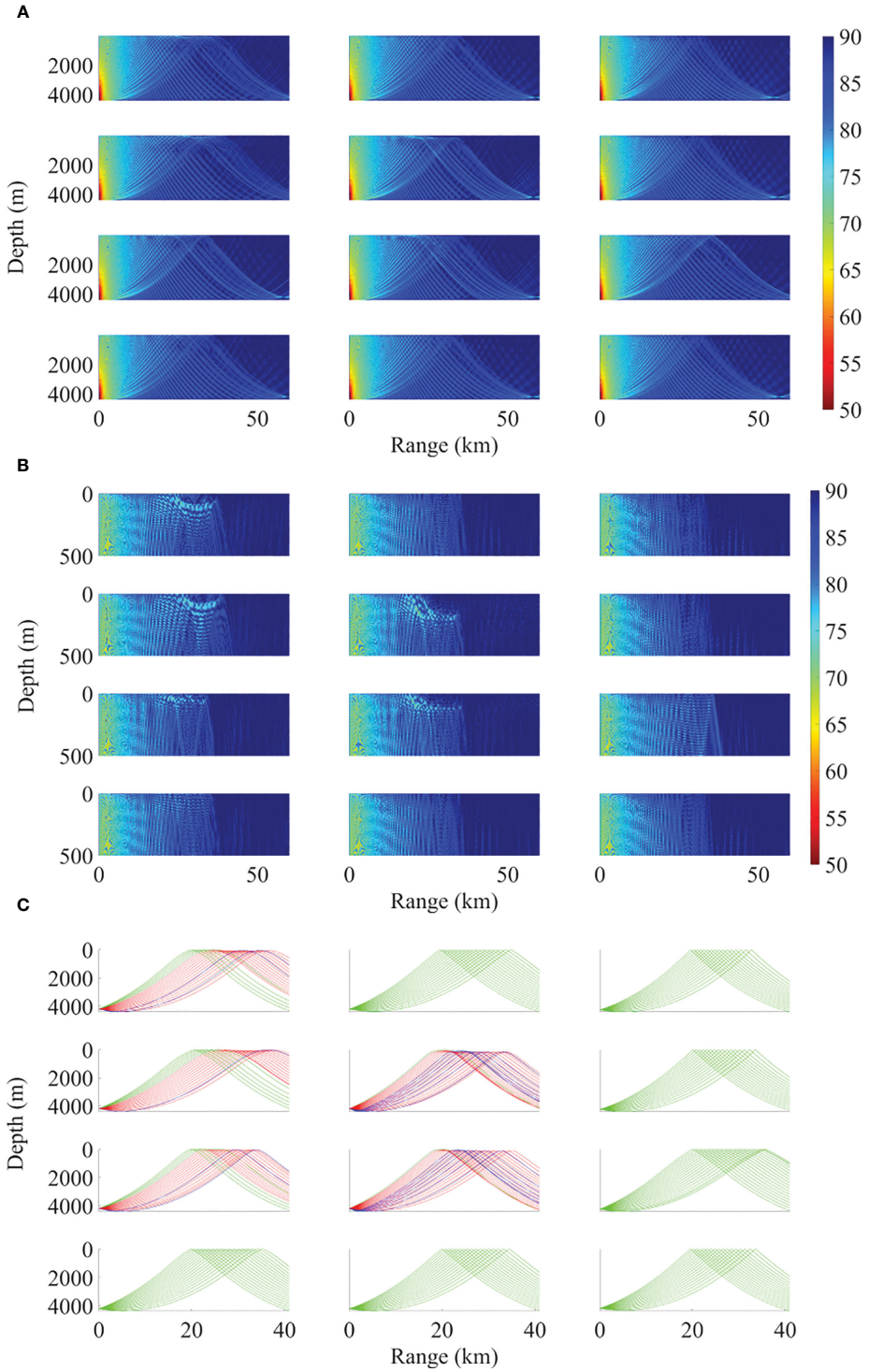


FIGURE 8
TLs and rays corresponding to the 12 types of the SSPs. **(A)** 2D TL with the maximum depth down to 4500 m, **(B)** 2D TL with the maximum depth range down to 500 m, **(C)** the ray figure (the grey lines refer to the eigenrays with one time of surface reflection and non-bottom reflection, the red lines refer the eigenrays with no bottom or surface reflection).

TABLE 1 Maximum range of the RAP for the 12 types of SSPs.

Type	1	2	3	4	5	6	7	8	9	10	11	12
Range	32	31	33	32	34	40	42	37	34	25	34	37

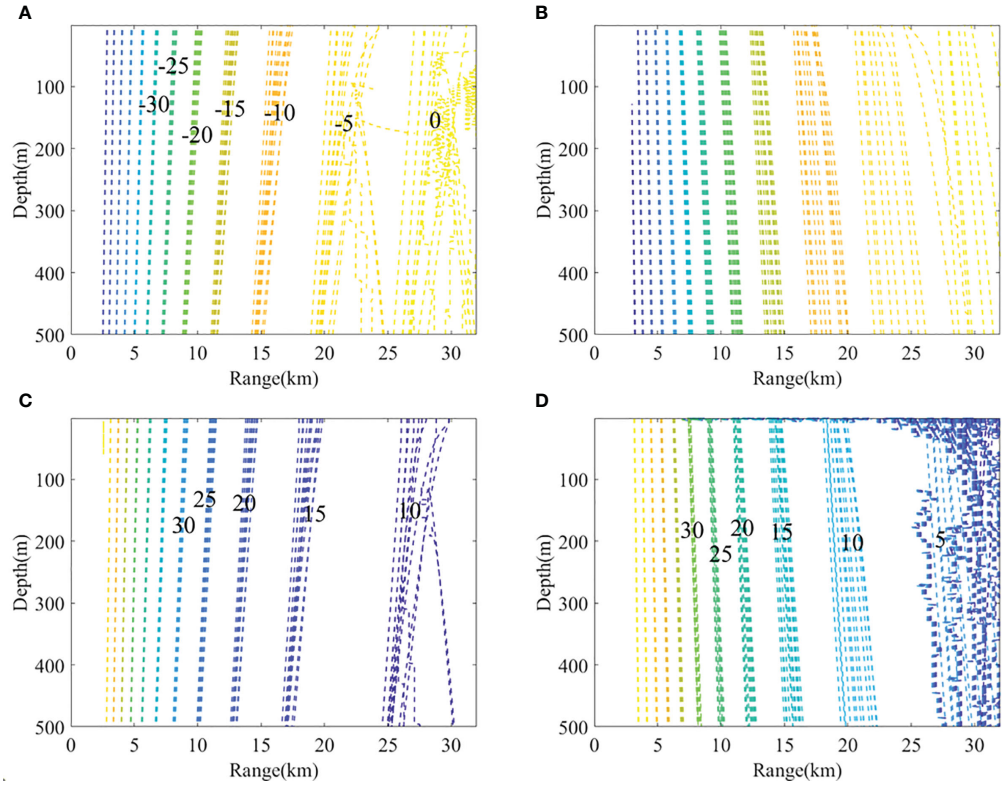


FIGURE 9
Effects of the SSP variation on the receiver angle **(A)** B0T0, **(B)** B0T1, **(C)** B1T0, **(D)** B1T1, (the SSPs used here were the 12 types from the cluster result).

receiver angle. Even in some specific regions, the dynamics of the ocean front could result in a significant change in RAP properties. Thus, if we want to make full use of the RAP for underwater activities, then SSPs with high accuracy should be used to avoid a large error.

Conclusions

In this study, we examined the RAP properties on a global scale using SSPs calculated from temperature and salinity profiles within the WOA18 annual database, and ocean depth data from the ETOPO1 dataset. Acoustic modeling was conducted on each grid of the WOA18 data at a 0.25° × 0.25° spatial resolution, and the maximum RAP range across the global ocean was calculated. Results indicate that the RAP range increases with ocean depth, albeit with a weak correlation. At a consistent ocean depth of 4000 meters, the RAP range can vary between 15 and 50 kilometers, with the SSP structure being the main variable.

We also investigated the effect of SSPs on RAP properties by clustering global SSPs into 12 groups, yielding 12 distinct SSP types. Acoustic simulations revealed that the RAP range could vary between 25 and 42 kilometers when using these 12 SSPs. Additionally, the impact of SSP variation on receiver angles was

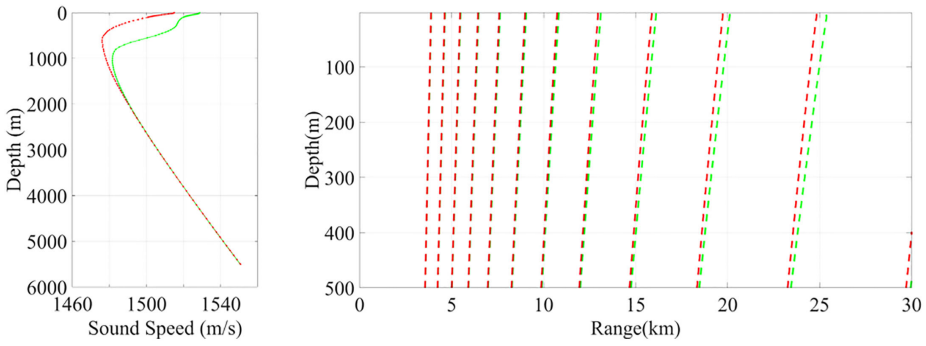


FIGURE 10
SSPs in the Kuroshio Extension region (left) and the effect of SSPs on the receiver angles of the RAP (right).

studied. Results demonstrated that receiver angle variation increases with source range, reaching up to 5 kilometers at a 20-kilometer distance. This variation can introduce significant localization errors when using the RAP for underwater source localization with inappropriate SSPs. Therefore, careful consideration of SSP structure is essential when leveraging the RAP for underwater activities.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, YL, CC; methodology, YL software, CC; validation, YL; formal analysis, YL.; investigation, YL, CC; resources, CC; data curation, CC, XF; writing—original draft preparation, YL; writing—review and editing, CC; visualization, CC; supervision, CC; project administration, CC; funding acquisition, CC. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 52101389); the Natural Science Foundation of Shandong Province, China (Grant No. ZR201910220437); the Open Fund of State Key Laboratory of Acoustics (Grant No. SKLA202103); and Fundamental Research Funds for the Central Universities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chen, C., Yang, K. D., Ma, Y. L., and Duan, R. (2016). Comparison of surface duct energy leakage with bottom-bounce energy of close range propagation. *Chin. Phys. Lett.* 33 (10), 104302. doi: 10.1088/0256-307X/33/10/104302

Chen, C., Yang, K., Duan, R., and Ma, Y. (2017). Acoustic propagation analysis with a sound speed feature model in the front area of Kuroshio Exten-sion. *Appl. Ocean Res.* 68, 1–10. doi: 10.1016/j.apor.2017.08.001

Chen, C., Tao, J., and Zhiquan, Z. (2019). Effect of eddy on acoustic propagation from the surface duct perspective. *Ap-plied Acoustics* 150, 190–197. doi: 10.1016/j.apacoust.2019.02.019

Choi, J. W., and Peter, H. D. (2004). Mid-to-high-frequency bottom loss in the East China Sea. *IEEE J. Oceanic Eng.* 29 (4), 980–987. doi: 10.1109/JOE.2004.834178

Dosso, S. E., et al. (2007a). Linear and nonlinear measures of ocean acoustic environmental sensitivity. *J. Acoust. Soc. Am.* 121 (1), 42–45. doi: 10.1121/1.2382719

Dosso, S. E., Morley, M. G., Giles, P. M., Brooke, G. H., McCammon, D. F., Pecknold, S., et al. (2007b). Spatial field shifts in ocean acoustic environmental sensitivity analysis. *J. Acoust. Soc. Am.* 122 (5), 2560–2570. doi: 10.1121/1.2783123

Duan, R., Yang, K., Ma, Y., and Chapman, N. R. (2016). A simple expression for sound attenuation due to surface duct energy leakage in low-latitude oceans. *J. Acoust. Soc. Am.* 139 (5), EL118–EL123. doi: 10.1121/1.4947012

Duan, R., Yang, K., Li, H., Yang, Q., Wu, F., and Ma, Y. (2019). A performance study of acoustic interference structure applications on source depth estimation in deep water. *J. Acoust. Soc Am.* 145, 903–916. doi: 10.1121/1.5091100

Duan, R., Yang, K. D., Ma, Y. L., and Hui, L. (2014). Moving source localization with a single hydrophone using multipath time delays in the deep ocean. *J. Acoust. Soc Am.* 136, 159–165. doi: 10.1121/1.4890664

Duan, R., Yang, K. D., Ma, Y. L., and Lei, B. (2012). A reliable acoustic path: Physical properties and a source localization method. *Chin. Phys. B* 21, 276–289. doi: 10.1088/1674-1056/21/12/124301

Gaul, R. D., Knobles, D. P., Shooter, J. A., and Wittenborn, A. F. (2007). Ambient noise analysis of deep-ocean measurements in the Northeast Pacific. *IEEE J. Ocean. Eng.* 32, 497–512. doi: 10.1109/JOE.2007.891885

Heaney, K. D., Campbell, R. L., Murray, J. J., Baggeroer, A. B., Scheer, E. K., Stephen, R. A., et al. (2013). Deep water towed array measurements at close range. *J. Acoust. Soc. Am.* 134 (4), 3230–3241. doi: 10.1121/1.4818869

Heitsenrether, R. M., and Mohsen, B. (2004). "Modeling acoustic signal fluctuations induced by sea surface roughness," in *HIGH FREQUENCY OCEAN ACOUSTICS: High frequency ocean acoustics conference*, Vol. 728 La Jolla, California (USA): (American Institute of Physics).

Hirt, C., and Rexer, M. (2015). Earth2014: 1 arc-min shape, topography, bedrock and ice-sheet models – Available as gridded data and degree-10,800 spherical harmonics. *Int. J. Appl. Earth Obs. Geoinf.* 39, 103–112. doi: 10.1016/j.jag.2015.03.001

Khan, S., Song, Y., Huang, J., and Piao, S. (2021). Analysis of underwater acoustic propagation under the influence of mesoscale ocean vortices. *J. Mar. Sci. Eng.* 9 (8), 799. doi: 10.3390/jmse9080799

Lermusiaux, P. F., Xu, J., Chen, C. F., Jan, S., Chiu, L. Y., and Yang, Y. J. (2010). Coupled ocean–acoustic prediction of transmission loss in a continental shelfbreak region: Predictive skill, uncertainty quantification, and dynamical sensitivities. *IEEE J. Oceanic Eng.* 35 (4), 895–916. doi: 10.1109/JOE.2010.2068611

Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., et al. (2018). *World ocean atlas 2018, volume 1: temperature*. Ed. A. Mishonov (NOAA Atlas NESDIS 81), 52.

Lovett, J. R. (1978). Merged seawater sound-speed equations. *J. Acoust. Soc Am.* 63, 1713–1718. doi: 10.1121/1.381909

Ngodock, H. E., Carrier, M. J., Fabre, J., Zingarelli, R., Smith, S., and Souopgui, I. (2022). "Sensitivity analysis in ocean acoustic propagation," in *Data assimilation for atmospheric, oceanic and hydrologic applications*, vol. IV. (Switzerland: Springer, Cham), 427–438.

Pecknold, S., and Osler, J. C. (2012). Sensitivity of acoustic propagation to uncertainties in the marine environment as char-acterized by various rapid environmental assessment methods. *Ocean Dynamics* 62 (2), 265–281. doi: 10.1007/s10236-011-0497-1

Qiu, C., Ma, S., Chen, Y., Meng, Z., and Wang, J. (2018). Reliable Acoustic path and direct-arrival zone spatial gain analysis for a vertical line array. *Sensors.* 18 (10), 3462. doi: 10.3390/s18103462

Thompson, S. R. (2009). *Sound propagation considerations for a deep-ocean acoustic network*. Monterey, California. Naval Postgraduate School.

Varamo, V. J., and Howe, B. (2016). Reliable acoustic path tomography at Aloha Cable Observatory. *J. Acoust. Soc Am.* 140, 3184–3184. doi: 10.1121/1.4970013

Worcester, P. F., Dzieciuch, M. A., Mercer, J. A., Andrew, R. K., Dushaw, B. D., Baggeroer, A. B., et al. (2013). The North Pacific Acoustic Laboratory deep-water acoustic propagation experiments in the Philippine Sea. *J. Acoust. Soc. Am.* 134 (4), 3359–3375. doi: 10.1121/1.4818887

Xiao, P., Yang, Y., Yang, L., and Shi, Y. (2016). "Seasonal effects of sound speed profile on mid-range acoustic propagations modes: reliable acoustic path and bottom bounce," in *Theory, methodology, tools and applications for modeling and simulation of complex systems* (Singapore: Springer).

Xu, L., Yang, K., and Yang, Q. (2019). Experimental study of geoacoustic inversion with reliable acoustic path in the Philippine Sea. *J. Theor. Comput. Acoust.* 27 (04), 1850061. doi: 10.1142/S2591728518500615

Yang, K., Xiao, P., Duan, R., and Ma, Y. (2017). Bayesian inversion for geoacoustic parameters from ocean bottom reflection loss. *J. Compu-tational Acoustics* 25 (03), 1750019. doi: 10.1142/S0218396X17500199

Zhang, X., Yang, P., and Sun, H. (2023a). An omega-k algorithm for multireceiver SAS. *Electron. Lett.* 59 (13), 1–3. doi: 10.1049/ell2.12859

Zhang, X., Yang, P., and Zhou, M. (2023b). Multireceiver SAS imagery with generalized PCA. *IEEE Geosci. Remote Sens. Lett.* 20, 1502205. doi: 10.1109/LGRS.2023.3286180

Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., Grcia, H. E., et al. (2018). *World ocean atlas 2018, volume 2: salinity*. Ed. A. Mishonov (NOAA Atlas NESDIS 82), 50.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 October 2023
DOI 10.3389/fmars.2023.1279693

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Li Dawei,
Naval Aviation University, China
Hanhao Zhu,
Zhejiang Ocean University, China

*CORRESPONDENCE
Sen Zhang
johnson_xh@sina.com

RECEIVED 18 August 2023
ACCEPTED 13 September 2023
PUBLISHED 17 October 2023

CITATION
Zhang S, Wu J and Yin T (2023) Simulation of sonar reverberation signal considering the ocean multipath and Doppler effect. *Front. Mar. Sci.* 10:1279693. doi: 10.3389/fmars.2023.1279693

Simulation of sonar reverberation signal considering the ocean multipath and Doppler effect

Sen Zhang[1]*, Jian Wu[1,2] and Tianqi Yin[1]

[1]Naval University of Engineering, Wuhan, Hubei, China, [2]91497 Unit, Ningbo, Zhejiang, China

In order to achieve accurate modeling and simulation of sonar reverberation signals, four types of multi-path underwater reverberation models are established considering Doppler effect under the condition of separating the sound source and hydrophone. The simulation of underwater reverberation signals under static or uniform linear motion conditions is carried out for single point for the separating the sound source and hydrophone transceiver, as well as horizontal linear array. The non-stop-and-hop model of reverberation signals is presented. And the underwater reverberation signals in the array element domain and beam domain are obtained. From the simulation results of the improved model, it can be seen that the spatiotemporal two-dimensional characteristics and Doppler expansion are consistent with theoretical analysis. The frequency shift of the horizontal linear array reverberation signal is approximately sinusoidal with the directionality angle of the linear array. Comparing the simulation results of the improved model with traditional models, the improved model can more accurately simulate sonar reverberation signals.

KEYWORDS

reverberation, ray acoustics, ocean multipath, bottom scattering, Doppler, accurate modeling, signal simulation, towed linear array

1 Introduction

Ocean reverberation refers to the acoustic signal generated at the receiving point caused by the scattering of a large number of random inhomogeneous bodies in the undulating sea surface, uneven seabed, and seawater medium during the propagation of sound waves (Yangang et al., 2020). Consequently, a sonar reverberation signal will have a negative impact on the precise reception and identification of the target underwater acoustic signal (Bing et al., 2016). In addition, the movement of the signal transceiver will inevitably introduce a frequency shift in the ocean reverberation signal caused by the Doppler effect (Yuliang, 2020). Therefore, it is important to introduce a more accurate model of ocean reverberation signals. The present study establishes four types of multipath sound rays,

which are then modeled and simulated under the consideration of the Doppler effect (Yulu et al., 2017).

2 Models and methods

When it comes to simulating the ocean reverberation signal, reference (Danping, 2020) followed four distinct steps to obtain the simulated reverberation: a) start from the shallow sea environment, b) adopt the normal mode propagation model, c) introduce the probability density function of Rayleigh distribution, and d) accumulate the reverberation generated by each scatterer at different distances. In contrast, Zhou et al. (2020) based their method on the ray-normal mode analogy, using the normal mode to simulate the reverberation field in shallow water (Zhou et al., 2020). In both studies, the reverberation is simulated under the condition that both the sound source and the hydrophone are placed close to each other (Liya, 2018). established the attenuation model of deep seabed reverberation intensity with time and the model of seabed reverberation signal based on the principles of statistical physics. In reference (Yangang et al., 2020), the reverberation sequence signal was obtained by convoluting the equivalent reverberation scattering sequence with the transmitted signal. Lijun et al. (2021) used the small slope approximation and the ray theory sound field algorithm to evaluate the scattering effect of the rough interface in the full grazing angle range, and the multipath factor was then employed to establish the reverberation intensity model of the sea surface and seabed. Based on the ray acoustic model, reference (Teng et al., 2021) used the channel convolution method and the echo signal to derive the echo signal in the ideal environment and the shallow water environment, respectively, with reverberation interference. In reference (Runze et al., 2021), the interface reverberation was described as the incoherent superposition result of different multipath reverberation fading processes, and a reverberation intensity model was established, using the physical parameters of the sea surface and seabed as variables. However, a limiting factor of these studies was that they did not consider the influence of the Doppler effect.

Siwei et al (Kou et al., 2021). proposed that when the sonar platform moves, the reverberation and echo entering the sonar array from different incidence cone angles have different Doppler frequency shifts; however, this study only examined the case of direct incidence of the receiver through the first scattering on the seabed. In addition, the ocean multipath factor was not taken into consideration. In reference (Sibo, 2018), three-dimensional bistatic multipath reverberation signals were modeled and simulated, while at the same time, the authors analyzed the space-time characteristics of bistatic reverberation, including Doppler frequency shift and reverberation directivity. In addition, that study investigated the suppression of reverberation signals using the space-time optimal processing method. However, the influence of the Doppler stretching effect on the pulse width of the reverberation signal was still not regarded.

Therefore, it becomes evident that current research on simulating ocean reverberation signals tends to ignore the Doppler stretching effect on the pulse width of the reverberation signal. Furthermore, several research studies have not considered the influence of the Doppler frequency shift on the reverberation signal, while others have not considered the multipath factor of the ocean. Consequently, to realize the accurate modeling and simulation of sonar reverberation signals, based on the ray acoustics theory and the principle of sound field superposition (Jun et al., 2012; Tao, 2007), the influence of the Doppler stretching effect on the signal pulse width has been analyzed under the condition that the sound source and the hydrophone are separated. As a result, four types of ocean reverberation models considering the Doppler effect have been established. On this basis, the reverberation model of the sonar signal is simulated, the single-point transceiver is extended to the horizontal towed linear array, and the seafloor reverberation signals in the array element space and beam space are obtained (Jincheng, 2019). The space-time two-dimensional characteristics and Doppler spread in the simulation results are consistent with the theoretical analysis. Comparing the simulation results of the improved model and the traditional model, the improved model can simulate the sonar reverberation signal more accurately.

2.1 Ocean multipath model for reverberation signal simulation

In the present study, the marine environment refers to the environment in which the depth of the seawater is much lower than the length of the sound propagation path in the seawater. In this environment, reverberation in the seawater stems largely from the scattering of sound waves on the seafloor, and the intensity of seafloor reverberation is mainly contributed by four types of multipath sound rays (Minghui, 2011). Hence, the simulation of seafloor reverberation signals mainly considers four types of multipath sound rays, as shown in Figure 1 (Sibo and Song, 2016).

In this figure, H represents the depth of the sea. The paths of these four types of sound rays involve the following: a) sound source, scattering at the bottom surface, and hydrophone; b) sound source, scattering at the bottom surface, sea surface reflection, and hydrophone; c) sound source, sea surface reflection, scattering at the bottom surface, and hydrophone; d) sound source, sea surface reflection, scattering at the bottom surface, second sea surface reflection, and hydrophone.

In general, the combined transmitter and receiver can be regarded as a special case of a separated transmitter and receiver. Therefore, considering that the towed linear array sonar to be analyzed is a separated transmitter and receiver, the present study investigated the establishment of a sonar reverberation simulation model under the condition of a separated sound source and hydrophone.

2.2 Reverberation signal model considering the Doppler effect

The model in this paper is based on the following three hypotheses:

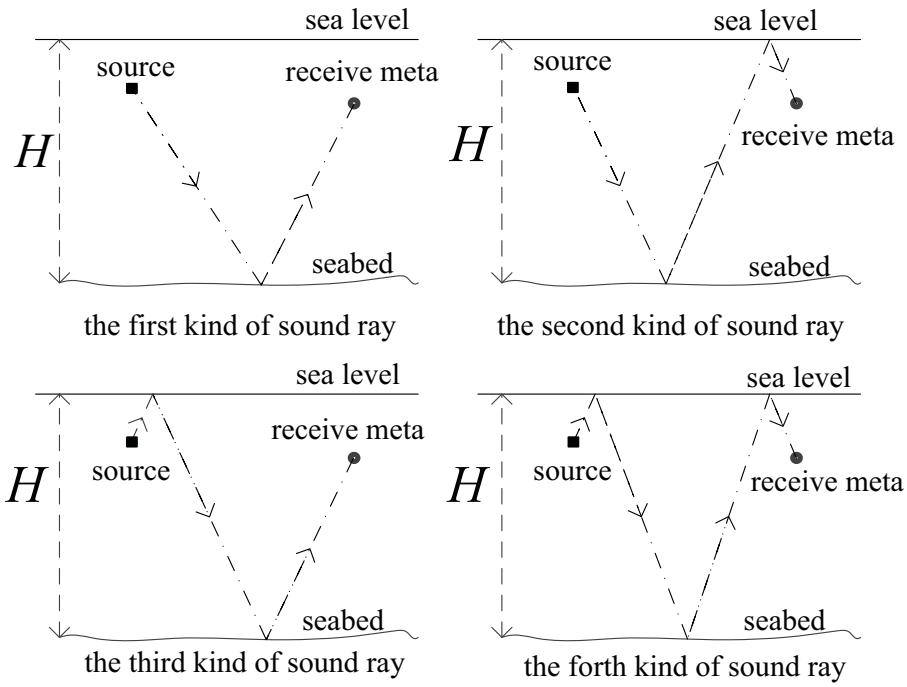


FIGURE 1
The four types of multipath sound rays contributing to the intensity of seafloor reverberation.

Hypothesis 1: Sound waves propagate in the form of spherical waves.

Hypothesis 2: The absorption of sound waves is neglected, and thus scattering is calculated at the sea bottom, and reflection is calculated at the sea surface.

Hypothesis 3: Scattering of the sea bottom is uniform.

The influence of the Doppler effect on sonar reverberation signal mainly affects signal frequency and signal pulse width.

2.2.1 Doppler effect on signal frequency

Let us consider a sinusoidal signal where the signal (Jian, 2019) source moves at a radial rate v relative to the hydrophone. Let the velocity of the sound source close to the hydrophone be positive and the velocity of the sound source far away from the hydrophone negative. If the frequency of the signal is f, the wavelength is λ, and the propagation speed of the signal in the medium is c, the frequency of the signal after the Doppler effect becomes f', and the wavelength becomes λ'. Consequently (Xianwen et al., 2022), the Doppler shift is given by $\Delta f = f'\text{-}f$ or

$$\Delta f = \frac{v}{c - v} \cdot f \tag{1}$$

In the rectangular coordinate system, a sound source is assumed to be moving with a velocity v_t, the hydrophone moves with a velocity v_r, and the bottom scatterer dA is static, whereas all other environmental conditions remain unchanged. The Doppler shift models of four types of multipath sound rays are discussed respectively in the following.

2.2.1.1 Doppler shift model of the first type of sound ray

As shown in Figure 2-1, SS represents the sound source, RE is the receiving element, r_1 is the propagation vector of the first segment of the sound ray, and r_2 is the propagation vector of the second segment of the sound ray. Furthermore, φ is the scattering azimuth angle, θ^I is the grazing angle of the incident sound ray, θ^s is the grazing angle of the scattered sound ray, and n is a scattering element serial number (Sheng and Xucheng, 2010).

First, we investigate the section of the first type of sound ray from the sound source SS to the bottom scatterer dA. Let the frequency shift of the signal received by the seafloor scatterer be Δf_{1t}. In accordance with the physical meaning of the vector dot product, the radial velocity v_t on r_1 can be expressed as

$$v_1 = \frac{V_t \cdot r_1}{|r_1|} \tag{2}$$

If we substitute the relevant parameters of the first type of sound ray into Equation (1), the variation of the signal frequency Δf_{1t} of the first type of sound ray transmitted from the sound source to the scattering element dA can be obtained as follows (Zhongchen et al., 2013):

$$\Delta f_{1t} = \frac{v_1}{c - v_1} \cdot f \tag{3}$$

If we substitute Equation (2) into Equation (3), we can obtain the frequency shift of the sound source as it hits the seafloor scattering element via r_1:

$$\Delta f_{1t} = \frac{V_t \cdot r_1}{c \cdot |r_1| - V_t \cdot r_1} \cdot f \tag{4}$$

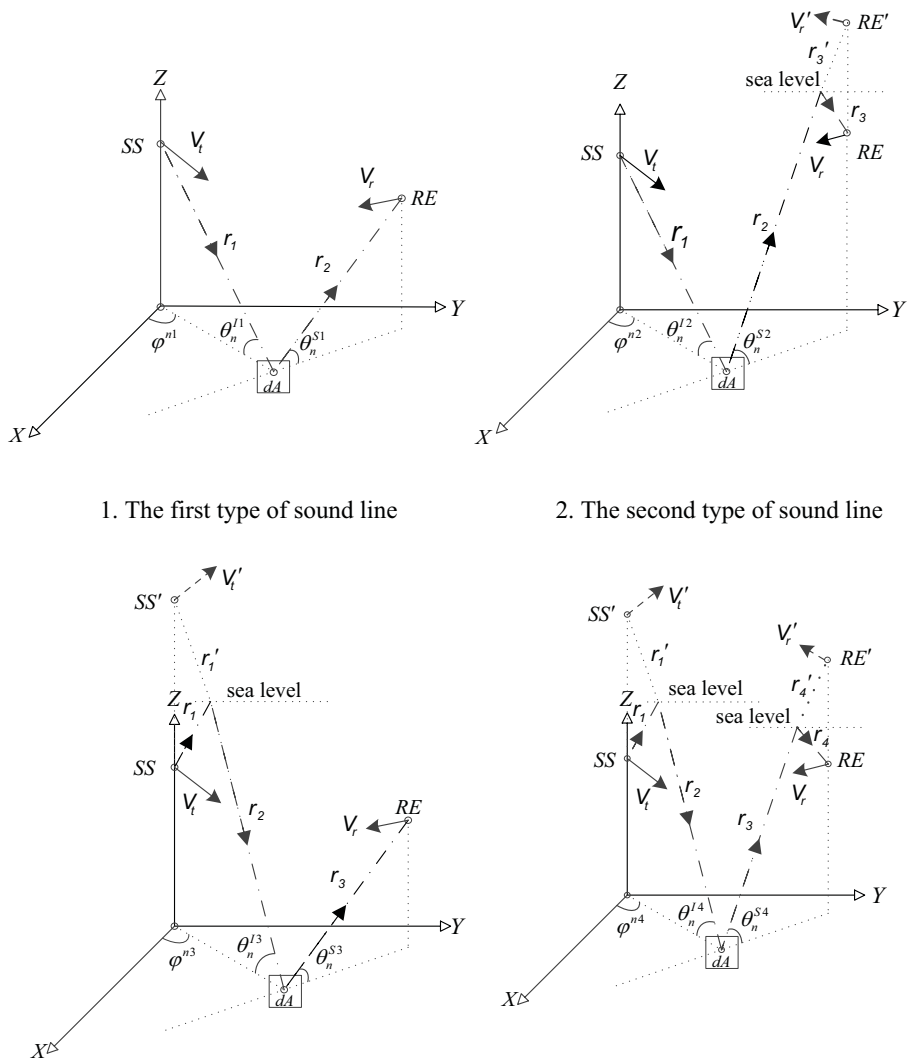


FIGURE 2
Four types of sound line.

Similarly, considering the propagation of the first type of sound ray from the scattering element dA through r_2 to the hydrophone RE, we can assume that the frequency of the signal scattered by the first type of sound ray on the seabed is f_{1b}.

Consequently, the relationship between f_{1b} and the original frequency f of the signal is:

$$f_{1b} = f + \Delta f_{1t} \tag{5}$$

Assuming that the velocity of the hydrophone is v_r, the frequency shift Δf_{1r} of the received signal for the first type of sound ray from the seafloor scatterer dA to the hydrophone is

$$\Delta f_{1r} = \frac{V_r \cdot r_2}{c \cdot |r_2| - V_r \cdot r_2} \cdot f_{1b} \tag{6}$$

and the total Doppler shift of the first type of sound ray is: (Yao, 2013)

$$\Delta f_1 = \Delta f_{1t} + \Delta f_{1r} = \left[\frac{V_t \cdot r_1}{c \cdot |r_1| - V_t \cdot r_1} + \frac{V_r \cdot r_2}{c \cdot |r_2| - V_r \cdot r_2} \cdot \left(1 + \frac{V_t \cdot r_1}{c \cdot |r_1| - V_t \cdot r_1} \right) \right] \cdot f \tag{7}$$

According to formulas (5 – 5) and (5 – 13) in (Minghui, 2011), the first type of sound ray reverberation signal model $p_1(t_{1k})$ can be obtained as (Yali, 2018):

$$p_1(t_{1k}) = \sum_{n=1}^{N_1} \frac{1}{|r_1| \cdot |r_2|} \cdot s(t - t_k) \cdot \sqrt{\sin \theta_n^{I1} \sin \theta_n^{S1}} dA \cdot \mu_p^{'n1} \cdot \exp\left(j\varphi^{n1} \right),$$

where $s(t - t_k)$ is the signal emitted by the sound source, and $s(t)$ passes through t_k.

10.3389/fmars.2023.1279693

Considering the Doppler shift, after replacing $s(t-t_k)$ with a complex signal $\exp\left[j2\pi\left(f+\Delta f_1\right)\left(t-t_{1k}\right)\right]$ (Zhang et al., 2021a; Zhang et al., 2022a), we can get the first type of sound line reverberation signal model $p_1(t_{1k})$ as:

$$p_1(t_{1k}) = \sum_{n=1}^{N_1} \frac{1}{|r_1|\cdot|r_2|}\cdot\exp[j2\pi(f+\Delta f_1)(t-t_{1k})] \cdot\sqrt{\sin\theta_n^{I1}\sin\theta_n^{S1}}dA\cdot\mu_p^{'n1}\cdot\exp\left(j\varphi^{n1}\right) \tag{8}$$

where the first term $(|r_1|\cdot|r_2|)^{-1}$ reflects the signal propagation loss; r_1 is the vector of sound sources to seafloor scattering elements; the second term $\exp\left[j2\pi\left(f+\Delta f_1\right)\left(t-t_{1k}\right)\right]$ is the complex signal arriving at the hydrophone after the frequency shift and time delay; f is the frequency of the transmitted signal; Δf_1 is the total Doppler shift of the first type of sound; $t_{1k}=(|r_1|+|r_2|)/c$ is the signal propagation time delay; and the third term $\sqrt{\sin\theta_n^{I1}\sin\theta_n^{S1}}dA\cdot\mu_p^{'n1}\cdot\exp(j\varphi^{n1})$ is the scattering coefficient of the seabed sound pressure; dA is the scattering element area; $\mu_p^{'n1}$ is a proportional constant, which is subject to the Gaussian distribution (Xiaohui et al., 2017); φ^{n1} is the transient phase, which is subject to the $(0\sim 2\pi)$ uniform distribution; n is the serial number of seabed scattering elements; θ_n^{I1} is the incidence grazing angle; θ_n^{S1} is the scattering grazing angle; N_1 is the total number of scattering elements.

Similarly, the three remaining types of sound ray Doppler shift models and reverberation signal models can be deduced accordingly.

2.2.1.2 The Doppler shift model of the second type of sound ray

As shown in Figure 2-2, the frequency change of the second type sound ray signal is given by the following equations:

$$\Delta f_{2t} = \frac{V_t\cdot r_1}{c\cdot|r_1|-V_t\cdot r_1}\cdot f \tag{9}$$

$$\Delta f_{2r} = \frac{V_r\cdot(r_2+r_3')}{c|r_2+r_3'|-V_r'\cdot(r_2+r_3')}\cdot f_{2b} \tag{10}$$

where V_r' is the velocity of the virtual source RE' of the hydrophone RE which is symmetrical to the sea surface, r_3' is the vector from the intersection of the scattered sound ray and the sea surface to the virtual source RE', and r_3 is the propagation vector of the third segment of the sound ray, $f_{2b}=f+\Delta f_{2t}$.

The total Doppler shift of the second type of sound ray is

$$\Delta f_2=\Delta f_{2t}+\Delta f_{2r} =\left[\frac{V_t\cdot r_1}{c\cdot|r_1|-V_t\cdot r_1}+\frac{V_r'\cdot(r_2+r_3')}{c|r_2+r_3'|-V_r'\cdot(r_2+r_3')}\cdot\left(1+\frac{V_t\cdot r_1}{c\cdot|r_1|-V_t\cdot r_1}\right)\right]\cdot f \tag{11}$$

and the model of the second type of sound ray reverberation signal $p_2(t_{2k})$ is

$$p_2(t_{2k}) = \sum_{n=1}^{N_1}\frac{m}{|r_1|\cdot|r_2+r_3'|}\cdot\exp[j2\pi(f+\Delta f_2)(t-t_{2k})] \cdot\sqrt{\sin\theta_n^{I2}\sin\theta_n^{S2}}dA\cdot\mu_p^{'n2}\cdot\exp(j\varphi^{n2}) \tag{12}$$

Here, the first term $m\cdot\left(|r_1|\cdot|r_2+r_3'|\right)^{-1}$ represents the signal propagation loss, the second term $\exp\left[j2\pi\left(f+\Delta f_2\right)\left(t-t_{2k}\right)\right]$ is the complex signal arriving at the hydrophone after the frequency shift and time delay, and the third term $\sqrt{\sin\theta_n^{I2}\sin\theta_n^{S2}}dA\cdot\mu_p^{'n2}\cdot\exp(j\varphi^{n2})$ is the sea bottom sound pressure scattering coefficient, m is the sea surface reflectivity, $\mu_p^{'n2}$ obeys a Gaussian distribution, φ^{n2} obeys a $(0\sim 2\pi)$ uniform distribution, and $t_{2k}=(|r_1|+|r_2|+|r_3|)/c$ is the signal propagation delay.

2.2.1.3 The Doppler shift model of the third type of sound ray

As shown in Figure 2-3, the amount of change in the frequency of the third type of sound ray signal:

$$\Delta f_{3t} = \frac{V_t'\cdot(r_1'+r_2)}{c|r_1'+r_2|-V_t'\cdot(r_1'+r_2)}\cdot f \tag{13}$$

$$\Delta f_{3r} = \frac{V_r\cdot r_3}{c\cdot|r_3|-V_r\cdot r_3}\cdot f_{3b} \tag{14}$$

where V_t' is the velocity of the virtual source SS', whose sound source SS is symmetrical to the sea surface, and r_3' is the vector from the intersection point of the incident sound ray and the sea surface to SS', $f_{3b}=f+\Delta f_{3t}$.

The total Doppler shift of the third type of sound ray is:

$$\Delta f_3=\Delta f_{3t}+\Delta f_{3r} =\left[\frac{V_t'\cdot\left(r_1'+r_2\right)}{c|r_1'+r_2|-V_t'\cdot\left(r_1'+r_2\right)}+\frac{V_r\cdot r_3}{c\cdot|r_3|-V_r\cdot r_3}\cdot\left(1+\frac{V_t'\cdot\left(r_1'+r_2\right)}{c|r_1'+r_2|-V_t'\cdot\left(r_1'+r_2\right)}\right)\right]\cdot f \tag{15}$$

and the model of the third type of sound ray reverberation signal $p_3(t_{3k})$ is

$$p_3(t_{3k}) = \sum_{n=1}^{N_1}\frac{m}{|r_1'+r_2|\cdot|r_3|}\cdot\exp[j2\pi(f+\Delta f_3)(t-t_{3k})] \cdot\sqrt{\sin\theta_n^{I3}\sin\theta_n^{S3}}dA\cdot\mu_p^{'n3}\cdot\exp\left(j\varphi^{n3}\right) \tag{16}$$

The first term $m\cdot\left(\left|r_1'¢+r_2\right|\cdot|r_3|\right)^{-1}$ represents the signal propagation loss, the second term $\exp[j2\pi(f+\Delta f_3)(t-t_{3k})]$ is the complex signal arriving at the hydrophone after the frequency shift and time delay, and the third term $\sqrt{\sin\theta_n^{I3}\sin\theta_n^{S3}}dA\cdot\mu_p^{'n3}\cdot\exp\left(j\varphi^{n3}\right)$ is the scattering coefficient of the seabed sound pressure, $\mu_p^{'n3}$ obeys a Gaussian distribution, φ^{n3} obeys a $(0\sim 2\pi)$ uniform distribution, and $t_{3k}=(|r_1|+|r_2|+|r_3|)/c$ is the signal propagation delay.

2.2.1.4 Doppler shift model of the fourth type of sound ray

As shown in Figure 2-4, the amount of change in the frequency of the fourth type of sound ray signal is

$$\Delta f_{4t} = \frac{V_t'\cdot(r_1'+r_2)}{c|r_1'+r_2|-V_t'\cdot(r_1'+r_2)}\cdot f \tag{17}$$

$$\Delta f_{4r} = \frac{V_r' \cdot (r_3 + r_4')}{c|r_3 + r_4'| - V_r' \cdot (r_3 + r_4')} \cdot f_{4b} \quad (18),$$

where r_1' is the vector from the intersection point of the incident sound ray and the sea surface to SS', r_4 is the propagation vector of the fourth segment of the sound ray, and r_4' is the vector from the intersection point of the scattered sound ray and the sea surface to the virtual source RE', $f_{4b} = f + \Delta f_{4t}$.

Consequently, the total Doppler frequency shift of the fourth type of sound ray is (Zhang et al., 2021b):

$$\Delta f_4 = \Delta f_{4t} + \Delta f_{4r} = \left[\frac{V_t' \cdot (r_1' + r_2)}{c|r_1' + r_2| - V_t' \cdot (r_1' + r_2)} + \frac{V_r' \cdot (r_3 + r_4')}{c|r_3 + r_4'| - V_r' \cdot (r_3 + r_4')} \cdot \left(1 + \frac{V_t' \cdot (r_1' + r_2)}{c|r_1' + r_2| - V_t' \cdot (r_1' + r_2)}\right)\right] \cdot f \quad (19)$$

and the model of the fourth type of sound ray reverberation signal $p_4(t_{4k})$ is:

$$p_4(t_{4k}) = \sum_{n=1}^{N_1} \frac{m^2}{|r_1' + r_2| \cdot |r_3 + r_4'|} \cdot \exp[j2\pi(f + \Delta f_4)(t - t_{4k})] \cdot \sqrt{\sin\theta_n^{I4} \sin\theta_n^{S4}} dA \cdot \mu_p'^{n4} \cdot \exp(j\varphi^{n4}) \quad (20)$$

The first term of Equation 20 $m^2\left(|r_1' + r_2| \cdot |r_3 + r_4'|\right)^{-1}$ represents the signal propagation loss, the second term $\exp[j2\pi(f + \Delta f_4)(t - t_{4k})]$ is the complex signal arriving at the hydrophone after the frequency shift and time delay, and the third term $\sqrt{\sin\theta_n^{I4} \sin\theta_n^{S4}} dA \cdot \mu_p'^{n4} \cdot \exp(j\varphi^{n4})$ is the scattering coefficient of the seabed sound pressure, $\mu_p'^{n4}$ obeys a Gaussian distribution, φ^{n4} obeys a $(0 \sim 2\pi)$ uniform distribution, and $t_{4k} = (|r_1| + |r_2| + |r_3| + |r_4|)/c$ is the signal propagation delay.

2.2.2 Influence of the Doppler effect on the signal pulse width

As shown in Figure 3, c is the signal speed, λ is the wavelength, τ is the pulse width, and k is the number of cycles, i.e., the signal contains k wavelengths. The signal will be affected by the Doppler stretching effect, and will thus have a new wavelength λ' and a new pulse width τ' (Xiye et al., 2009). According to the relationship between distance, speed, and time, we can easily obtain that:

$$\tau = \frac{k \cdot \lambda}{c} \quad (21)$$

$$\tau' = \frac{k \cdot \lambda'}{c} \quad (22)$$

If we substitute $f \cdot \lambda = c$ and $f' \cdot \lambda' = c$ into Equation (21) and Equation (22), respectively, then we can obtain the relationship between the signal pulse width before and after the Doppler stretching effect:

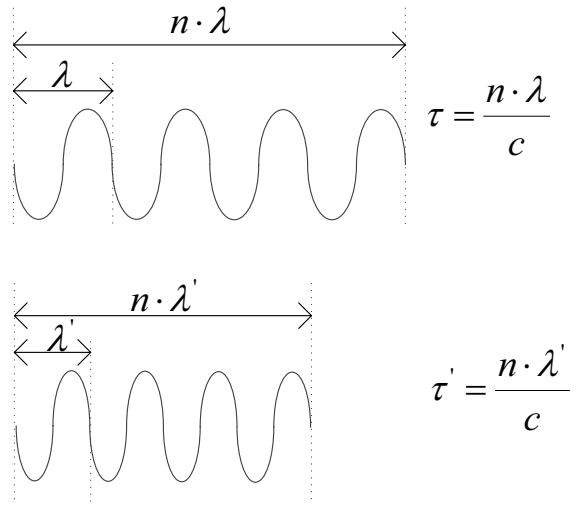


FIGURE 3
Influence of the Doppler stretching effect on the signal pulse width.

$$\tau' = \frac{f}{f'}\tau \quad (23)$$

Simulation and results.

2.3 Underwater reverberation simulation

The geometric model of seabed reverberation simulation is shown in Figure 4 (Zhang and Yang, 2022). The towed linear array is horizontally arranged along the negative direction of the Y axis, where SS is the transmission source of the towed linear array located on the Z axis (Xiaohui et al., 2017). Furthermore, RE is the receiving element of the towed linear array, i is the serial number of the receiving element, N is the total number of receiving elements, d is the distance between the receiving elements, H is the depth of the sea water, and h is the distance of the towed linear array from the sea floor. The distance between the transmitting source SS and RE_1 is $2d$ (Jinhua et al., 2020; Yonghong, 2011). The motion states of the

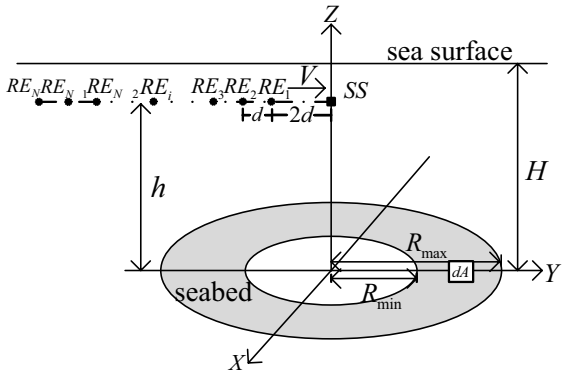


FIGURE 4
Geometric modeling of seabed reverberation simulation.

transmitting source and the receiving element are the same. The sea floor scattering area is an annular area, which assumes the origin as the center of a circle with an inner diameter R_{min} and an outer diameter R_{max}, and dA is a sea floor scattering element (Meina et al., 2017; Zhao et al., 2011). The transmitting source sends out a complex signal described by $s(t) = \exp(j2\pi ft)$ (Zhiguang and Zhiqiang, 2016), with a pulse width of τ, sea surface reflectivity of m, and speed c (Zhang et al., 2022b).

2.3.1 Reverberation signal simulation of a single scattering unit by a single sound source and a single hydrophone

In this section, we simulate the reverberation signal received by RE_1, which is emitted by the sound source SS and is incident to RE_1, via a single scattering element dA (Yuqiang et al., 2018). The simulation parameters are shown in the following Table 1:

In this paper, the reverberation model, represented by Equations (24), (25), (26), and (27), takes into consideration both the ocean multipath and Doppler effect. Depending on whether the transmitting source and the single receiving element RE_1 are stationary or moving, the simulation results are shown in Figures 5, 6, respectively.

Figure 5 demonstrates a time domain diagram of a single bottom scatterer's reverberation signal. A single source emits a sinusoidal pulse signal with a fixed frequency pulse width of 0.5(s). After passing through the single scattering element dA, the single hydrophone RE_1, receives the time domain map of the signal. The blue part of the figure is the time domain plot of the signal received by RE_1 while the trailing linear array is stationary. Conversely, the red part of the figure is the time domain plot of the signal received by RE_1 when the linear array is dragged. Thus, when the transceiver device is stationary, the signal frequency of the four types of voice lines remains unchanged because there is no Doppler shift. Hence, the superimposed signal of the four types of voice lines still basically maintains the shape of sinusoidal pulses (blue part). When the transceiver device moves, the signals of the four types of voice lines undergo different degrees of Doppler shift. The superposition of four types of sound lines with different frequencies forms a signal (red part), in which the envelope amplitude changes following the law of sine and cosine. Moreover, under the same motion state, the Doppler frequency shift of the four types of voice lines is calculated by the Doppler shift model of the four types of voice lines, and the accurate modeling and simulation of sonar reverberation signals are realized. Additionally, owing to the influence of the Doppler effect on the signal pulse width, the signal pulse width of the light-colored part of the figure is shortened, and the analysis graph shows the pulse width change $\Delta\tau = \tau' - \tau = -0.005$ (s).

In addition, Figure 6 exhibits a spectrum diagram of the reverberation signal generated by a single bottom scatterer (Yanzi et al., 2018). It can be seen that the peak center of the spectrum increases following the movement of the transceiver, as opposed to when the transceiver remains stationary, and the analysis graph demonstrates that the frequency shift is $\Delta f = 35$ Hz (Huang and Gao, 2014). When in motion, the spectrum is extended due to the different Doppler frequency shifts of the four types of sound rays moving at the same speed.

Although (Minghui, 2011) provides the models of four types of multipath sound ray reverberation signals, the authors do not investigate the influence of the Doppler effect on the four types of sound ray reverberation signals in detail. To compare the traditional model, which does not consider the multipath and the Doppler effect, with the improved model presented in this paper, we employed the first type of bistatic sound ray model equations (5 – 12) stated in Chapter 5 of (Minghui, 2011) to compare the simulation results with the present model.

In the case of motion, Figure 7 shows a comparison of the simulation time domain of a single seafloor scatterer:

The blue graph in Figure 7 is a time domain plot of the received signal simulated based on a traditional model. Conversely, the red graph is a time domain plot of the received signal simulated using the improved model. Figure 7 reveals that the time domain of the reverberation signals of the improved and traditional models is as follows: 1. The pulse width of the improved model signal is longer than that of the traditional model signal, which is due to the influence of the multipath, and part of the sound ray propagation path is longer. Since the Doppler frequency shift is not considered by the traditional model, the reverberation amplitude will remain constant. Compared with the traditional model, the reverberation envelope of the improved model changes according to the sine and cosine law, which is due to the superposition of reverberation signals formed by the different Doppler shifts of the four types of sound rays. The reverberation time domain amplitude of the traditional model is smaller than the signal amplitude under static conditions in Figure 5 because the traditional first type of sound ray model does not consider multipath superposition. It can be seen that the improved model can reflect the space-time characteristics of reverberation signals more accurately in the case of a single source, single hydrophone, and single scatterer.

TABLE 1 Parameters used to simulate the single seafloor scattering element reverberation.

Title	Symbol	Value
Type of sea		Homogeneous
Type of seabed		Flat seabed
Reflection coefficient of the seafloor		Obeys a (0–1) distribution
Sea depth (m)	H	60.0
Distance between the line array and the seabed (m)	h	45.0
Line array speed (m/s)	V	10.0 or 0
Line array element spacing	d	0.25
Complex signal frequency (Hz)	f	3000
Complex signal pulse width (s)	τ	0.50
Sea surface reflectivity	m	0.80
Velocity (m/s)	c	1500
Seafloor scattering element coordinates	(x_b, y_b, z_b)	(0,101,0)

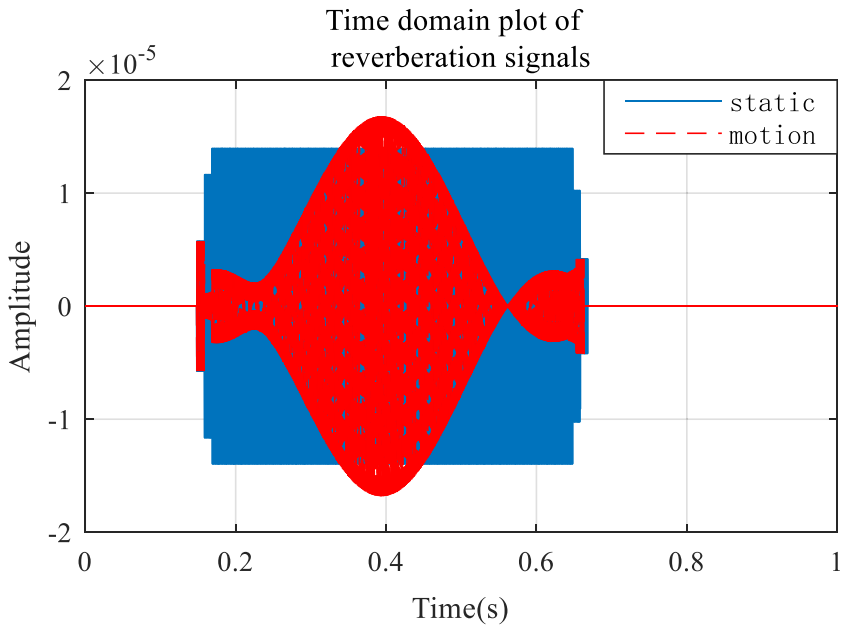


FIGURE 5
Time domain map of a single seafloor scattering element reverberation signal.

The frequency domain comparison between the improved model and the traditional model is shown in Figure 8:

From Figure 8, we can see that the peak frequency of the improved model increases due to the influence of the Doppler shift, and the spectrum of the improved model is extended compared with the traditional model because of the different Doppler frequency shifts of different sound rays caused by the ocean multipath. Therefore, the improved model can reflect the spectrum characteristics of the reverberation signal more accurately. In addition, the improved model presented in this study can simulate the reverberation signal more accurately under the condition of a single source, single hydrophone, and single scattering unit by analyzing the respective space-time and spectrum characteristics.

2.3.2 Reverberation signal simulation of a towed linear array

In this section, we simulate the reverberation signal of a towed linear array using the parameters shown in Table 2 and the simulation results are shown in Figures 9–12.

The "angle" in Figures 9–12 describes the directivity angle of the dragged line array (Zhe et al., 2017; Zelin, 2019; Junchao, 2021).

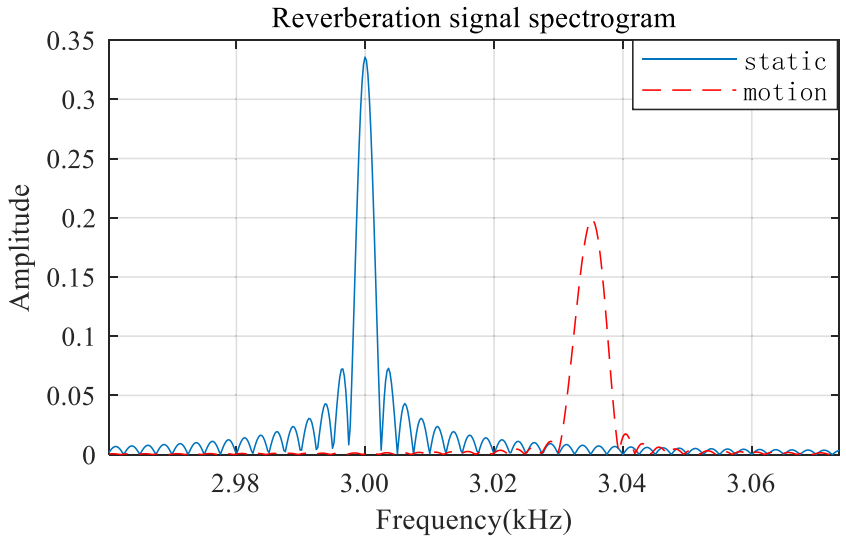


FIGURE 6
Spectrogram of a single seafloor scattering element reverberation signal.

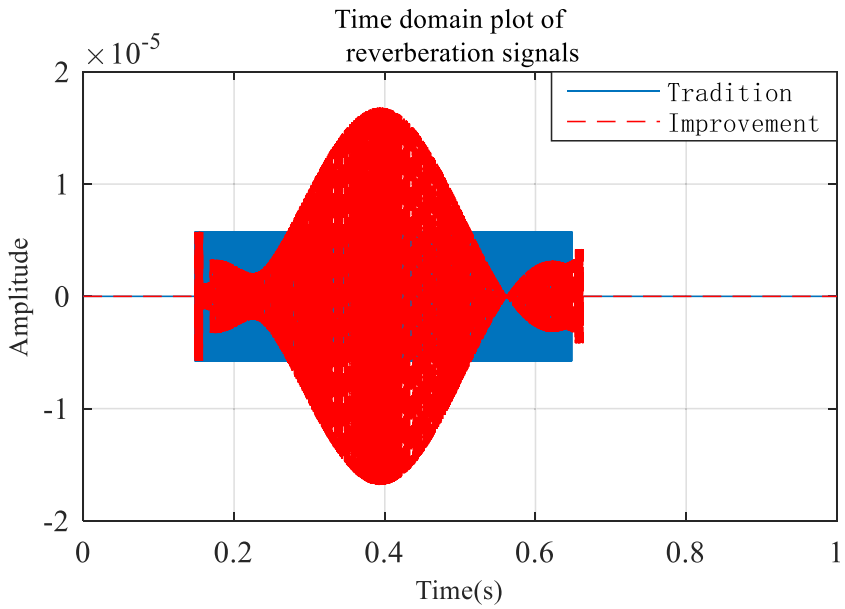


FIGURE 7
Comparison of the time domain diagram of a single seafloor scattering element reverberation signal.

Furthermore, different angles correspond to scattering elements at different positions. A comparison between Figures 11, 12 demonstrates that the reverberation signal frequency decreases when the directivity angle is negative, and increases when the directivity angle is positive. Figure 12 also shows that the frequency shift of the reverberation signal is approximately sinusoidal with the directivity angle of the linear array, a finding which is consistent with the theoretical analysis (Xiaodong et al., 2011; Yanqiu, 2015; Zhu et al., 2023; Zhang et al., 2023a; Zhang et al., 2023b).

Figures 13, 14 show the reverberation signal simulation of the traditional model under the condition of towed linear array motion

A comparison between Figures 13 and 10 reveals that the maximum amplitude of the reverberation signal of the traditional model is 0.12, which is smaller than the maximum amplitude of the reverberation signal of the improved model (0.22). This improvement can be explained by the fact that the improved model considers the superposition of multipath sound rays. Furthermore, a comparison between Figures 14 and 12 demonstrates that potential changes in the

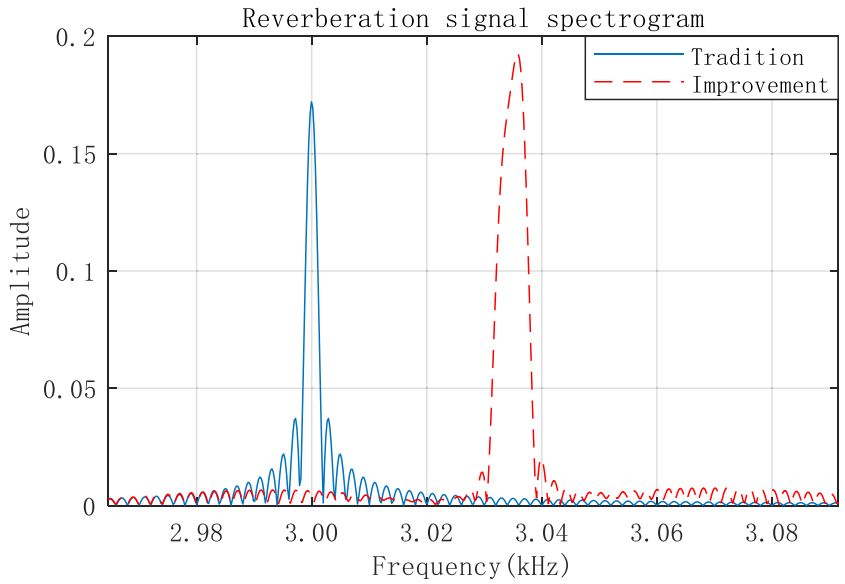


FIGURE 8
Frequency domain comparison diagram of a single seafloor scattering element reverberation signal.

TABLE 2 Dragging line array reverberation simulation parameters.

Title	Symbol	Value
Type of sea		Homogeneous
Type of seabed		Flat seabed
Reflection coefficient of the seafloor		Obeys a (0–1) distribution
Sea depth (m)	H	1200
Distance between the line array and the seabed (m)	h	900
Line array speed (m/s)	V	10 or 0
Number of line array receiving elements	N	64
Element spacing	d	0.25
Complex signal frequency (Hz)	f	3000
Complex signal pulse width (s)	τ	0.50
Sea surface reflectivity	m	0.80
Velocity (m/s)	c	1500
Inner radius (m)	$d_{\min}$	1200
Outer radius (m)	$d_{\max}$	12500

directivity angle of the linear array do not induce any frequency shift in the reverberation signal of the traditional model. This is because the traditional model does not consider the Doppler frequency shift factor.

Overall, our findings clearly show that the improved model presented in this study can simulate the reverberation signal significantly more accurately compared to the traditional models by analyzing space-time and spectral characteristics under the condition of a towed linear array.

3 Conclusion and discussion

Under the condition of a single sound source, single hydrophone, and single scattering unit, our simulation data

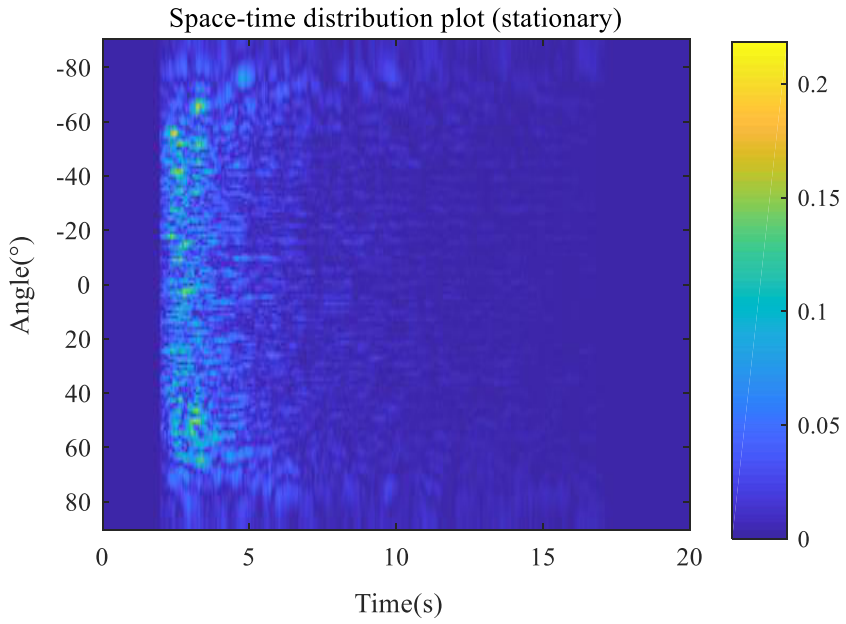


FIGURE 9
Space-time distribution of a stationary dragging line array reverberation signal.

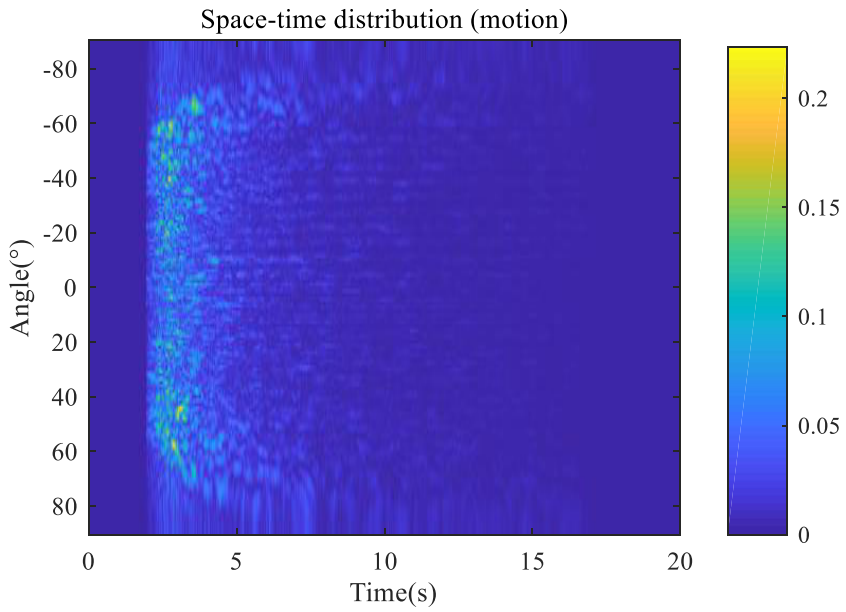


FIGURE 10
Space-time distribution of a panned dragging line array reverberation signal.

showed that the envelope amplitude of the reverberation signal changes according to the sine and cosine law, while changes in the pulse width and spectrum of the signal will occur when the transceiver moves in response to the Doppler effect. A comparison between the simulation results of the improved and the traditional models clearly demonstrates that the reverberation pulse width of the improved model is longer than that of the traditional model due to the ocean multipath. In addition, the reverberation envelope of the improvement model changes according to the sine and cosine law, caused by the superposition

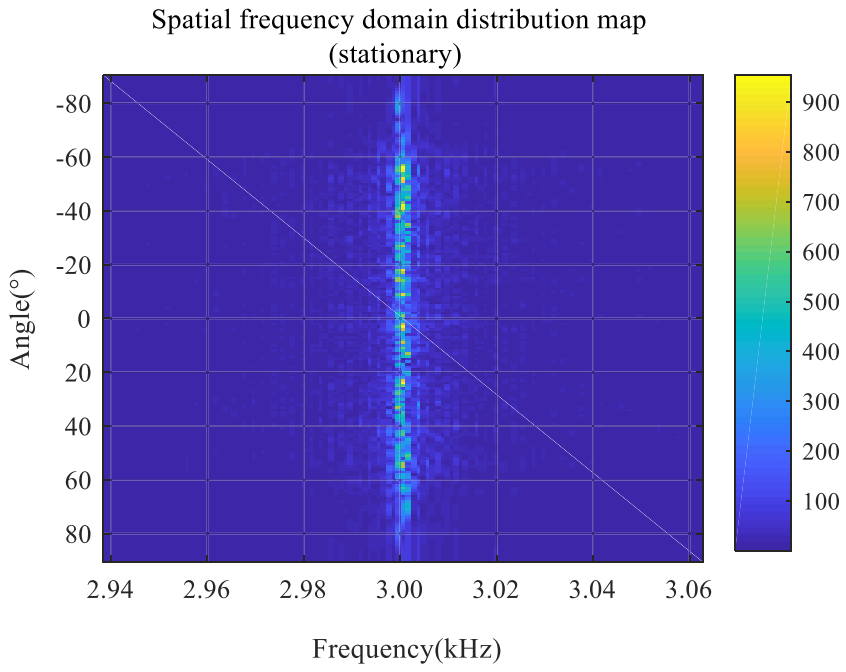


FIGURE 11
Space-frequency distribution of a stationary dragging line array reverberation signal.

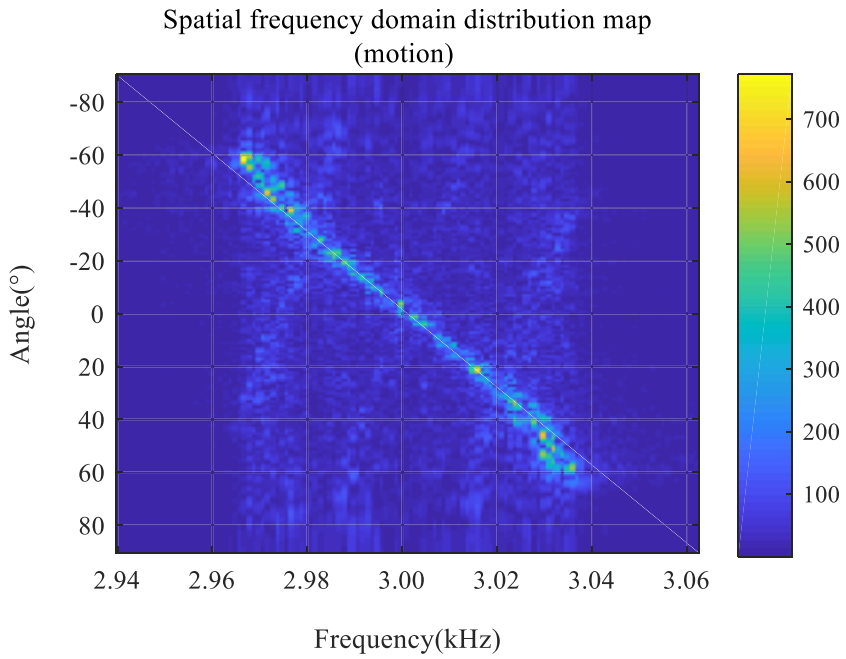


FIGURE 12
Space-frequency distribution of a panned dragging line array reverberation signal.

of reverberation signals formed by the different Doppler frequency shifts of the four types of sound rays considered, as opposed to the reverberation signal amplitude of the traditional model which remains unchanged. When assessing the space-frequency characteristics of the two models, the peak frequency of the improved model was increased and the spectrum width was extended due to the different Doppler frequency shift of the multipath sound ray. It can also be seen that the reverberation

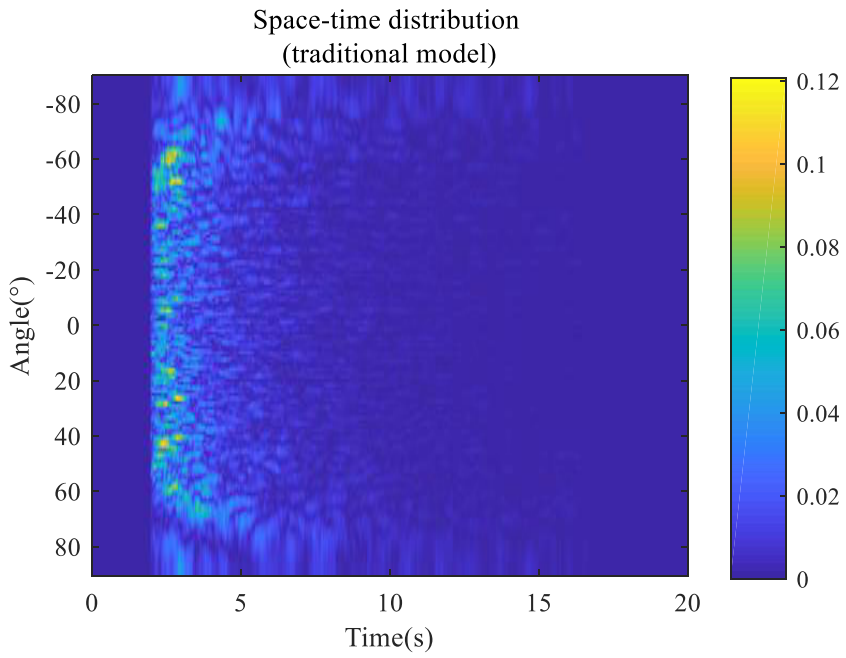


FIGURE 13
Space-time distribution of a panned dragging line array traditional model reverberation signal.

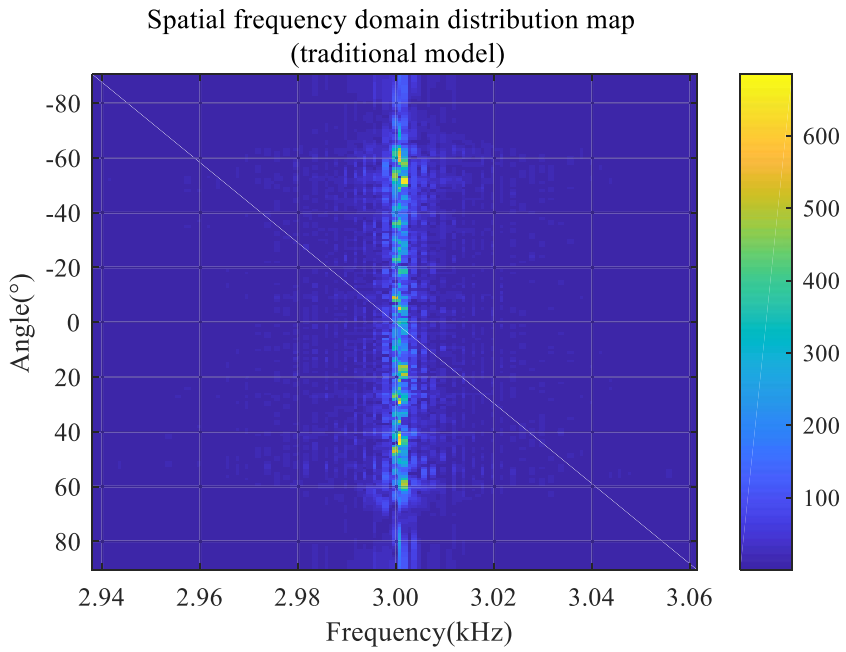


FIGURE 14
Space-frequency distribution of a panned dragging line array traditional model reverberation signal.

signal model that considers both the ocean multipath and the Doppler effect can reflect the variation in the frequency and pulse width of the reverberation signal far more accurately.

Under the condition of a towed linear array, the relationship between the frequency shift of the reverberation signal and the directivity angle of the linear array is approximately sinusoidal when the linear array is in uniform linear motion, and the space-time two-dimensional characteristics and Doppler spread in the simulation results are consistent with the theoretical analysis. Our findings confirm that the model established on the premise of single-point transceiver separation can be well-extended to the case of multi-point separation. Consequently, the proposed model has broad universal applicability and can be used to simulate more diverse combinations of sonar array elements.

When it comes to space-time characteristics, our data showed that the reverberation amplitude of the improved model is larger than that of the traditional model due to the superposition of multipath sound rays in the ocean. Finally, pertaining to the spatial frequency characteristics, the improved model can reflect the frequency domain characteristics of reverberation signals more accurately than the traditional model.

Conclusively, the reverberation signal model considering both the ocean multipath and the Doppler effect can simulate the sonar reverberation signal more accurately than the traditional models presented in current literature.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

SZ: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. JW: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. TY: Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (NO.41976177).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bing, J., Yunfei, C., Yang, Z., and Rui, W. (2016). Analysis of ocean reverberation characteristics based on the vector hydrophones. *Tech. Acoustics* 35 (3), 110–114.

Danping, S. (2020). Modeling and simulation of seabed reverberation in shallow seas. *Aud Eng.* 44, 31–32,36. doi: 10.16311/j.audioe.2020.08.008

Huang, X., and Gao, T. (2014). Reverberation spectra modeling and simulation of arbitrary sea bottom shape. *J. Syst. Simul* 26, 2576–2580.

Jian, L. (2019). *Interception detection and parameter estimation based on time-frequency image of underwater acoustic pulse signal* (China: Southeast University).

Jincheng, Z. (2019). *Research and implement on robust beamforming algorithm based on towed array* (China: Harbin Engineering University).

Jinhua, L., Jinsong, T., and Haoran, W. (2020). A MSR based range Doppler algorithm for the moderate squint multi-aperture synthetic aperture sonar. *Tech. Acoustics* 39(3), 354–359. doi: 10.16300/j.cnki.1000-3630.2020.03.017

Jun, F., Weilin, T., and Linkai, Z. (2012). Planar elements method for forecasting the echo characteristics from sonar targets. *J. Ship Mechanics* 16 (1-2), 171–180.

Junchao, W. (2021). *Research on self-noise suppression method of autonomous underwater vehicle* (China: Zhejiang University).

Kou, S., Feng, X. A., Bi, Y., and Huang, H. (2021). High-resolution angle-Doppler imaging by sparse recovery of underwater acoustic signals. *Chin. J. Acoustics* 46, 519–528. doi: 10.15949/j.cnki.0371-0025.2021.04.004

Lijun, Y., Jinrong, W. U., Shengming, G., Zhang, J., Hou, Q., and Ma, L. (2021). Modeling and analysis of reverberation intensity in incomplete SOFAR channel. *J. Harbin Eng. Univ* 42, 325–330. doi: 10.11990/jheu.201912045

Liya, X. U. (2018). *Researches on the method of geoacoustic inversion with bottom parameters in deep ocean* (China: Northwestern Polytechnical University).

Meina, W., Gang, B., Chao, S., and Wang, W. (2017). Simulation analysis of vertically divided seafloor reverberation signal in ocean bathymetry. *Hydrogr Surv Charting* 37, 75–78. doi: 10.3969/j.issn.1671-3044.2017.01.019

Minghui, Z. (2011). *Research on bottom sound scattering strength measurement method and reverberation properties in irregular sea area* (China: Harbin Engineering University).

Runze, X., Rui, D., Kunde, Y., Ma, Y., and Guo, Y. (2021). Modeling and analysis of monostatic incoherent boundary reverberation intensity in deep water. *Acta Acust* 46, 926–938. doi: 10.15949/j.cnki.0371-0025.2021.06.014

Sheng, L., and Xucheng, C. (2010). Simulation research on reverberation for bistatic sonar system. *Tech Acoust* 29, 355–360. doi: 10.3969/j.issn1000-3630.2010.04.001

Sibo, L., and Song, L. (2016). A modeling and simulation based on K-distribution model. *Ship Sci. Technol.* 38, 158–161. doi: 10.3404/j.issn.1672-7619.2016.S1.029

Sibo, L. I. (2018). *The characteristics of transceiver separation reverberation in shallow water* (China: Harbin Engineering University).

Tao, H. (2007). *The study on visual simulation techniques of underwater target* (China: Harbin Engineering University).

Teng, Z., Chunhua, Z., and Peng, W. (2021). Three-dimensional imaging sonar echo modeling simulation and analysis. *Netw. New Media Technol.* 10, 22–29.

Xianwen, Z., Zhou, C., Zhengwei, W., Zhigang, L., Pengpeng, H., and Mengxin, W. (2022). Echo signal modeling and simulation for moving target in complex marine environment. *Mobile Commun.* 46 (7), 82–87. doi: 10.3969/j.issn.1006-1010.2022.07.015

Xiaodong, C., Jianguo, H., and Qunfei, Z. (2011). Modeling and simulation of spacetime reverberation for active sonar array. *Comput. Simul* 28, 365–368.

Xiaohui, H., Gang, L., Cheng, W., and Chen, W. (2017). Multi-beam bathymetry seabed reverberation simulation study. *Hydrogr Surv Charting* 37, 22–25. doi: 10.3969/j.issn.1671-3044.2017.01.006

Xiye, G., Shaojing, S., and Yueke, W. (2009). Research on modeling and simulating seafloor reverberation with the moving sonar. *J. Natl. Univ Def Technol.* 31, 92–96.

Yali, S. (2018). *Reverberation signal simulation and inhibition technology of reverberation based on multistatic sonar* (China: Harbin Engineering University).

Yangang, H., Zhenhua, Z., and Hanfen, L. I. (2020). Modeling technology for active sonar target echo signal. *Command Inf Syst. Technol.* 11, 70–75. doi: 10.3969/j.issn.1671-3044.2017.01.006

Yanqiu, Z. (2015). *Simulation and research of underwater target acoustic imaging* (China: Southeast University).

Yanzi, G., Guoliang, L., and Fusheng, Z. (2018). Research on an engineering simulation model of ocean reverberation. *Ship Electron Eng.* 38, 123–125. doi: 10.3969/j.issn.1672-9730.2018.10.030

Yao, Z. (2013). *Research on methods of active sonar target detection in shallow water condition* (China: Harbin Engineering University).

Yonghong, Y. (2011). *Research on multi-beam synthetic aperture sonar imaging technology* (China: Harbin Engineering University).

Yuliang, H. (2020). *Research on target echo detection technology of active sonar for small and moving platform* (China: Harbin Engineering University).

Yulu, C., Xiaogang, Y., and Tuo, Z. (2017). Numerical compute about near distance bottom reverberation of typical sound speed profiles in shallow sea. *Ship Sci. Technol.* 39, 94–97. doi: 10.3404/j.issn.1672-7619.2017.06.019

Yuqiang, L., Yan, X., and Shengzeng, Z. (2018). Analysis of characteristics and research of modeling simulation on marine reverberation. *Ship Electron Eng.* 38, 86–88. doi: 10.3969/j.issn.1672-9730.2018.11.022

Zelin, S. (2019). *Research of highlight feature extraction and recognition based on time-frequency domain filtering* (China: Harbin Engineering University).

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021a). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Sel Top. Appl. Earth Obs Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., and Yang, P. (2022). Back projection algorithm for multi-receiver synthetic aperture sonar based on two interpolators. *J. Mar. Sci. Eng.* 10, 718. doi: 10.3390/jmse10060718

Zhang, X., Yang, P., Huang, P., Sun, H., and Ying, W. (2022a). Wide-bandwidth signal-based multireceiver SAS imagery using extended chirp scaling algorithm. *IET Radar Sonar Navig* 16, 531–541. doi: 10.1049/rsn2.12200

Zhang, X., Yang, P., and Sun, H. (2022b). Frequency-domain multireceiver synthetic aperture sonar imagery with Chebyshev polynomials. *Electron Lett.* 58, 995–998. doi: 10.1049/ell2.12513

Zhang, X., Yang, P., and Sun, M. (2023a). An omega-k algorithm for multireceiver synthetic aperture sonar. *Electron Lett.* 59, 1–3. doi: 10.1049/ell2.12859

Zhang, X., Yang, P., and Zhou, M. (2023b). Multireceiver SAS imagery with generalized PCA. *IEEE Geosci Remote Sens Lett.* 20, 1502205. doi: 10.1109/LGRS.2023.3286180

Zhang, X., Ying, W., Liu, Y., et al. (2021b). "Processing multireceiver SAS data based on the PTRS linearization," in *IEEE international geoscience and remote sensing symposium*, 11-16 July 2021, Brussels, Belgium. 167–517 (IEEE: Brussels, Belgium). doi: 10.1109/IGARSS47720.2021.9553688

Zhao, Y., Xian, F., Yufeng, Z., and Wang, Q. (2011). Simulation and space-time model of seafloor reverberation. *Comput. Simul* 12, 398–401.

Zhe, C., Zhizhong, L., Xin, G., and Qijun, L. (2017). Simulation of sonar target tracking and locating based on pressure hydrophone in ocean. *Comput. simulation* 34 (5), 1–6.

Zhiguang, X., and Zhiqiang, L. (2016). Analysis of small target sonar detective performance based on LFM signal. *Audio Engi-neering* 40 (4), 48–50. doi: 10.16311/j.audioe.2016.04.11

Zhongchen, D., Ya-an, L., and Yanfeng, J. (2013). Shallow seafloor reverberation modeling and simulation of torpedo. *Torpedo Technol.* 21, 100–104.

Zhou, Y., Li, Q., Li, P., Gong, J., and Yang, R. (2020). Active sonar signal modeling and simulation. *Electron Opt Control* 27, 14–17. doi: 10.3969/j.issn.1671-637X.2020.02.004

Zhu, J., Song, Y., Jiang, N., Xie, Z., Fan, C., and Huang, X. (2023). Enhanced Doppler resolution and sidelobe suppression performance for Golay complementary waveforms. *Remote Sens* 15, 2452. doi: 10.3390/rs15092452

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 23 October 2023
DOI 10.3389/fmars.2023.1280305

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Xuebo Zhang,
Northwest Normal University, China
Mingzhang Zhou,
Xiamen University, China

*CORRESPONDENCE
Hao Zhang
zhanghao@ouc.edu.cn
Wei Huang
hw@ouc.edu.cn

RECEIVED 20 August 2023
ACCEPTED 26 September 2023
PUBLISHED 23 October 2023

CITATION
Wang Z, Zhang H, Huang W, Chen X, Tang N and An Y (2023) A Wasserstein generative adversarial network with gradient penalty for active sonar signal reverberation suppression.
Front. Mar. Sci. 10:1280305.
doi: 10.3389/fmars.2023.1280305

A Wasserstein generative adversarial network with gradient penalty for active sonar signal reverberation suppression

Zhen Wang[1], Hao Zhang[1,2*], Wei Huang[1*], Xiao Chen[1], Ning Tang[1] and Yuan An[1]

[1]Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China, [2]Open Studio for Marine High Frequency Communications, Pilot National Laboratory for Marine Science and Technology, Qingdao, China

Reverberation is the primary background interference of active sonar systems in shallow water environments, affecting target position detection accuracy. Reverberation suppression is a signal processing technique used to improve the clarity and accuracy of received signals by eliminating the echoes, reverberations, and noise that occur during underwater propagation. Existing reverberation suppression methods include algorithms based on Time-Frequency domain processing, noise reduction, adaptive filtering, and spectral subtraction, but their performance in high-reverberation environments (echo of small targets) still does not meet the requirements of target detection. To address the impact of high reverberation environments, we propose a structural suppression method based on the Wasserstein gradient penalty generative adversarial network (RSWGAN-GP). The reverberation suppression generation network uses a one-dimensional convolutional network structure to process normalized time-domain signals and achieves the reconstruction of the reverberation signal through Encoder-Decoder. The proposed method is verified through accurate and effective data collection during sea trials. Comparative results show that RSWGAN-GP effectively suppresses reverberation in observation signals with multiple bright spots, improving the signal-to-reverberation ratio by approximately 10 dB compared to other excellent algorithms and enhancing the information analysis and feature extraction capabilities of active sonar signals.

KEYWORDS

active sonar signal, reverberation suppression, generative adversarial network (GAN), high reverberation environment, one-dimensional convolution Frontiers

1 Introduction

Sonar can accurately detect fixed targets and determine target distance, it has poor concealment and limited range and is susceptible to reverberation interference. The detection signal emitted by active sonar encounters obstacles or targets in the propagation path of water, and is then affected by reflection, refraction, and scattering, resulting in signal delay and overlap. The signals are subsequently received at the receivers, forming echo signals with reverberation (Huang and Wang, 2019). The influence of reverberation on active sonar like synthetic aperture sonar (Zhang et al., 2023a; Zhang et al., 2023b), communication sonar (Yu et al., 2018), multibeam sonar (Neasham et al., 2007) and other active sonars should be completely reviewed. Underwater interface reverberation is an important factor limiting the detection performance of active sonar in shallow water environments. Unlike noise interference, reverberation has non-smooth statistical characteristics and is usually mixed with the target echo, which is challenging to distinguish (Faure, 1964).

Previous works on reverberation in signal processing mainly focus on the study of detectors under specific reverberation conditions (Bharathi and Mohanty, 2019) and the spatiotemporal distribution characteristics of reverberation; researchers have studied robust detection performance under various reverberation distribution conditions. Some researchers have attempted to reduce the effect of reverberation on target echo by designing a transmission waveform, such as frequency-hopping signals encoded with particular frequencies (Costas, 1984), Q-function sonar signals (Cox and Lai, 1994), and SFM signals (Ward, 2001). However, enough high frequency is needed to achieve a reverberation suppression effect, which leads to a low utilization rate of the low-frequency band and affects the range accuracy.

Some researchers study anti-reverberation processing on signals received by sonar. Marine reverberation has a strong temporal correlation with target echo signals.Target echo signals cannot be effectively found by regularly matched filtering methods as the two spectra overlap in the frequency domain. In order to improve the performance of coherent processing in reverberation, Kay et al. used the AR pre-whitening processing method to filter out reverberation as white noise under certain conditions (Kay and Salisbury, 1990). Higher gain and more effectively detected target echoes can be obtained through matched filter by Wu et al. (Wu et al., 2018), but local stationarity of the reverberation is required as a premise (Widrow et al., 1967), which is widely used in ALE (Adaptive Line Enhancement) algorithm (Ma et al., 2021). However, it has strict requirements for the channel environment. H.M. Ozaktas and L.B. Almeida filtered the signal based on the time-frequency focus difference between the echo and reverberation in the Fourier transform domain to achieve reverberation suppression (Ozaktas et al., 1996; Zhang et al., 2019; Mejjaoli and Omri, 2020). However, The LMS algorithm performs adaptive filtering based on the error between input and output of the channel, which has strict requirements on the channel environment. Freburger et al. used the principal component inversion algorithm to project the received signal into two subspaces based on the power difference between different backgrounds, thereby achieving reverberation separation (Freburger and Tufts, 1997). When the power of the target echo signal is similar to that of the reverberation signal, distinguishing between the two becomes difficult.

With the development of artificial intelligence (AI) technology, deep neural networks have brought new research ideas to solve the shallow sea sonar reverberation problem. As a hot research direction in the field of machine learning, GAN (Ashraf et al., 2021) has become a popular model in the field of deep learning due to its advantages of generating high-quality samples, learning unlabeled data, supporting multi-modal data and innovation (Zhan et al., 2019) (Dong and Yang, 2018). Recently, Gans have evolved from image generation to reverberation data generation Hu et al. (2023). Gans are also used to generate spatial impulse responses, with the aim of enhancing high-quality RIRs with existing real RIRs (Ratnarajah et al., 2023). In the field of underwater acoustic engineering, it is theoretically feasible to use GAN for active sonar reverberation suppression to solve the problem of reverberation suppression under high reverberation environment.

This paper proposes a Wasserstein generative adversarial network model with a gradient penalty (RSWGAN-GP) to solve reverberation suppression of sonar signals. Sonar signals are different from the picture, and a one-dimensional convolutional approach is built in this paper to process the signal data. The generation side of the adversarial network is made according to the U-net network (Ronneberger et al., 2015) to encode the original reverberation signal data, and the decoder generates the anti-reverberant signal data. The discriminator uses the design idea of SkipNet (Abrahamyan et al., 2021) that the discriminator to achieve a balance of speed and accuracy. In order to realize the fast and accurate training of the countermeasure network, the structure design referred to WGAN-GP comprehensively. In the experiment, simulation is used to supplement the data set to solve the difficulty of Marine experiment data sampling and insufficient data set. Our main contributions are concluded as follows:

1) In order to solve the difficulty of feature extraction in a reverberation environment, we propose a Wasserstein generative adversarial network model with a gradient penalty method
2) Underwater active sonar reverberation simulation with echo targets is proposed to obtain many marine reverberation signals, solving the problem of insufficient training sets.

The rest of the article is organized as follows. Section 2 starts with a brief review of some related works. In Section 3, some data preliminary work is presented, which includes RSWGAN-GP reverberation data generation, signal time gain control, and automatic gain control. In Section 4, the reverberation suppression method based on RSWGAN-GP is proposed, and the generation network, discrimination network, and error loss are explained respectively. In Section 5, experiments are given to verify the effectiveness of the method, and in Section 6, conclusions are given. In active sonar, the commonly used detection signals include single continuous wave (CW), LFM, NLFM, BPSK and other signals. In this paper, the research object of active sonar signal feature enhancement is selected as the typical CW signal for research.

2 Related works

2.1 Reverberation suppression methods

Traditional methods for reverberation suppression typically focus on mapping the feature subspace. The differences between reverberation and target echoes have been investigated in the domains of Doppler space, discrete wavelet, and fractional Fourier. Previous research provides essential features for reducing reverberation and aiding target detection. For moving targets, in particular, target tracking can achieve reverberation suppression. However, these methods could be limited in low signal-to-reverberation ratio (SRR) and high reverberation scenes. When the echo of a tiny target is received, the target's echo is weak and covered by reverberation, and its feature subspace is weak and difficult to find.

Low-rank and sparsity theories developed a decade ago have found wide applications in image processing for tasks such as background modeling, camera calibration, and optical character recognition (Chandrasekaran et al., 2011). They have also been introduced to underwater acoustic engineering for reverberation suppression and target detection (Qian and Cao, 2019). In the case of fixed-position active sonar, the received signal from multiple pings exhibits significant stationarity. Considering the echo data from a single ping as a frame, multiple frames can be constructed over time and decomposed into dynamic and steady components. The steady components display similar strength distributions over time and can be viewed as a low-rank matrix. On the other hand, the dynamic components, consisting of reverberation fluctuations and target echoes, can be treated as a sparse matrix.

Consequently, reverberation suppression methods based on low-rank and sparse matrix decomposition have been proposed. These include techniques such as non-negative matrix factorization, principal component analysis, and robust principal component analysis (RPCA) (Chalapathy et al., 2017). For the implementation of processing large matrix factorization, alternative methods have been developed to expedite the process, such as accelerated proximal gradient, augmented Lagrange multiplier, and alternate direction multiplier methods (ADMM) (LiXiukun et al., 2015). Zhu et al. applied low-rank and sparse matrix estimation to decompose received data, enhancing reverberation suppression techniques' robustness (Zhu et al., 2022).

These reverberation suppression methods can achieve the purpose of reverb suppression to a certain extent. However, they still perform limited at a low signal-to-reverberation ratio.

2.2 Application of artificial intelligence in reverberation suppression

With the development and rise of artificial intelligence in recent years, algorithms combining artificial intelligence with anti-reverberation technology continue to surge, such as support vector machines, CNN (Song et al., 2019), RNN (Chen et al., 2022), and GAN In the beginning, it was simply a simple addition to machine learning. For example, Zhu et al. designed a feature kernel function SVM based on the non-Gaussian difference between reverberation and target echo to detect the signal in the reverberation background. This method improves the recognition quality of reverberant background, and its effect is better than the adaptive filtering algorithm (Wu et al., 2008). Jiang Keyu et al. processed the lake test data (Jiang et al., 2007) and used RBF neural network to detect the target echo in reverberation to be better. Xiang et al. proposed a reverberation suppression method for underwater moving target detection based on a robust autoencoder (Zhu and Sun, 2008). Xiao et al. proposed an ABNN focusing on the frequency domain characteristics of the target, which suppresses environmental noise and ship interference and makes the accuracy of target detection and recognition higher (Xiao et al., 2021).

The deep learning technology's continuous development and innovation, many neural network architectures with good performance and robust stability have emerged. For instance, multilayer perceptrons (MLP) and long short-term memory (LSTM) networks have been developed to learn mappings from a window of reverberated frames (or "context" windows) to a source frame, thus learning to deliberate by inverse transformations Han et al. (2015); Wang et al. (2017); Wuth et al. (2020). Additionally, Zhao et al. Zhao et al. (2018) proposed an LSTM-based late reverberation suppression strategy that learned the difference between the source and reverberated signals; therefore, dereverberation is performed by subtracting the late reverberation estimation from the observed reverberated signal.

The application of deep learning provides another effective method for reverberation suppression. Artificial intelligence has relatively excellent performance and effect. It can achieve many effects that cannot be achieved by traditional methods, which makes the development of anti-reverberation technology in recent years mainly biased to- wards the direction of artificial intelligence.

The above studies show that the combination of deep learning has specific feasibility for sonar signal reverberation suppression. However, reverberation suppression still needs to be improved under high reverberation environments and different underwater signal environments. At the same time, the extraction ability of effective information in the signal still cannot meet the needs of the complex environment.

2.3 The relationship between artificial intelligence methods and traditional methods

In terms of underwater reverberation suppression, the initial reference of artificial intelligence (AI) and machine learning methods is to make up for the shortcomings of traditional methods and complement and combine them. In a new study, it was found that artificial intelligence could complete the task better to replace it entirely (Koh et al., 2020).

Traditional methods are mainly based on signal processing and digital filtering techniques, which involve preprocessing, filtering, and noise reduction operations to suppress reverberation in underwater sound signals (Singer et al., 2009). These methods

often rely on domain knowledge and expertise to analyze and model the reverberation characteristics, followed by the design of corresponding algorithms for processing. While traditional methods can reduce the impact of underwater reverberation, their effectiveness is limited when dealing with complex reverberation environments and signals.

AI methods, on the other hand, utilize machine learning and deep learning techniques to learn and recognize reverberation features and perform suppression automatically (Hao et al., 2023). By training models with large amounts of data, AI methods can possess more substantial generalization and adaptability, making them capable of handling more complex underwater reverberation environments and signals. Compared to traditional methods, AI methods exhibit higher levels of automation and intelligence in underwater reverberation suppression.

Traditional methods and AI methods can be combined in the context of underwater reverberation suppression. Traditional methods can provide basic processing techniques and approaches for preprocessing and initial reverberation suppression, which AI methods can further optimize and enhance (Yin et al., 2023). For instance, traditional methods can be used for filtering and noise reduction of underwater sound signals, and the processed signals can be used as training data for training AI models to achieve better reverberation suppression.

In the current research, some scholars have found that traditional methods and artificial intelligence methods are complementary in terms of underwater reverberation suppression and can be combined. Other scholars have used AI alone to replace traditional methods and improve reverberation suppression (Weiss et al., 2023).

3 Preliminary

After the signal is received, the received signal will be processed by the active sonar system. In this part, the hardware implementation of the processing will be reproduced by the following algorithm. At the same time, the signals that generate the training set data will be processed in the same way.

The active sonar device processes the received signal in the following way. After the hardware receives the sonar signal, the sonar signal will go through time-varying gain and automatic signal gain control processing so that the long-distance echo signal power is stronger and more convenient for subsequent processing. After processing, the generated training set is closer to the actual data. Figure 1 shows the signal state of each process.

The processing of the signal in the hardware device after receiving is shown in Figures 1B–D, and the signal processing process will be explained below.

The echo received by the active sonar system, and the generated echo signal with target information in Section 4.1 are shown in Figure 1A, and the circled position is the target echo.Time-Variable Gain (TVG) Innami and Kasai (2012). According to the sonar equation, it can be obtained that the echo margin of the sonar is determined by the difference between the echo signal level and the background interference level.

$$DT = (SL - 2TL + TS) - (NL - DI) \tag{1}$$

In Eq. (1), DT represents the detection threshold and is the strength of the echo signal received by the sonar, TL represents the propagation loss because the active sonar is bidirectional, so $2TL$ defines its complete propagation loss; SL represents the sound source level; NL represents the noise level; DI represents the directivity index; TS represents the target strength. For most sonar systems, including multibeam sonar, the propagation loss (TL) is compensated by the TVG device inside the receiver. The ideal TVG curve should follow the expectation of sonar propagation loss, i.e.

$$TL = 10log_{10}r + ar \tag{2}$$

In Eq. (2), r represents the action distance, and a is the loss factor, a function of frequency. Figure 1B shows that the signal passes through the TVG and that the distant signal is no longer attenuated as the distance increases.

Signal Automatic Gain Control (AGC) (Zhang et al., 2017). In practice, automatic gain control is typically implemented through circuit design. However, AGC needs to be implemented for the experimental simulation of the sonar data set. The signal adjustment must be automatically adjusted based on the input and output data size, which initially requires numerous logarithmic operations. In order to implement these operations are avoided, a simple comparison operation is used with a gain lookup table instead. The algorithm can be described as follows.

$$20lg(G(n+1)) - 20lg(G(n)) = -10lg((\frac{R^2}{R'^2})^{\mu'}) \tag{3}$$

The left side is transformed into the adjusted value between two adjacent gain coefficients $G(n+1)$ and $G(n)$ in dB format for better clarity. Following the estimation of the average power of the output signal, a comparison is made with the reference power to determine the appropriate method for adjusting the gain coefficient based on the outcome of this comparison. In Figure 1C, the AGC equalizes the signal strength and partially suppresses reverberation through processing. There are hardware limitations to consider in practical usage. The hardware restricts the received signal and simulates how the part of the signal that exceeds the limit would appear. Figure 1D illustrates this simulation. Figure 1E displays the desired output result of the network, which will be utilized to calculate the Jensen-Shannon Divergence (JSD), aiming to bring the output result closer to it.

In both experimental and simulation data, some signals exhibit peak clipping phenomena and cannot be restored to complete sine waves, making traditional methods ineffective. After undergoing the above (B), (C), and (D) processing and normalization, the simulated signal has inputs with the same dynamic range.

4 Reverberation suppression in RSWGAN-GP

In order to solve the problem of limited suppression effect in a high reverberation environment, a reverberation suppression

10.3389/fmars.2023.1280305

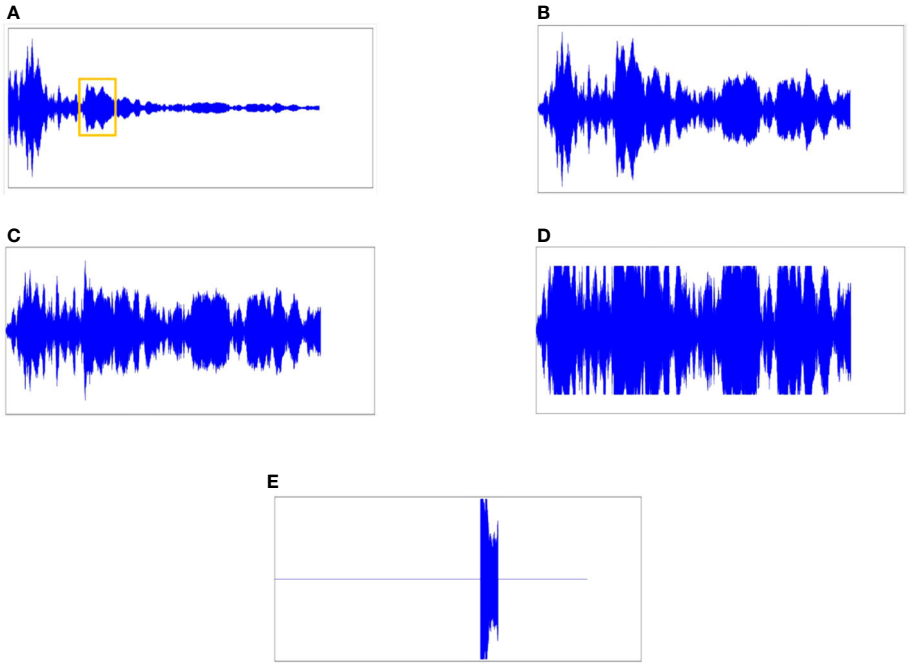


FIGURE 1
Signal processing. **(A)** The echo received by the active sonar system. **(B)** The echo signal is processed by TVG. **(C)** The echo signal is processed by AGC. **(D)** The echo signal received and processed by sonar. **(E)** Target echo signal hidden in sonar signal.

framework based on the Generative adversarial network is proposed. A nonlinear mapping from the sonar detection signal with reverberation to the sonar signal with reverberation suppression is established, where the input data is the signal propagation time series. The output is the corresponding time series after reverberation suppression. Due to the scarcity of actual data with reverberation signals, the training of RSWGAN-GP will face the problem of overfitting, reducing its generalization performance. To this end, the generation of virtual reverberation data is implemented in this section by combining statistical modeling and a multi-highlight model. The training data set is expanded by mixing real and virtual reverberation data while ensuring the consistency of its distribution pattern. Then, the theoretical signal gain control was calculated by signal propagation theory to form a time series signal for artificial intelligence model learning. The implementation of this part is shown in Figure 2.

The data generated in Part 4.1 is mixed with real data for Part 4.3, 4.4, and the gradient penalty part is invoked for training.

4.1 Underwater active sonar reverberation simulation

Underwater reverberation consists of volume reverberation, surface reverberation, and submarine reverberation. Sea surface reverberation and submarine reverberation are collectively referred to as interface reverberation.

To model the network, a significant amount of experimental data is necessary. Therefore, in this paper, we will simulate sonar reverberation data. The reverberation simulation comprises three parts: the generation of reverberation, the simulation of echoes, and environmental noise simulation.

As shown in Figure 3, at Point M, a non-directional signal is emitted to activate the ring energizer.. The reverberation model is illustrated in the figure. Point M represents the transducer, the distance from M to the interface xoy is h, and the signal is emitted without any specific direction. At time t, the ring of scatterers contributing to the reverberation is inside the ring. (Sun et al., 2010).

It is assumed that the number of scatterers generating reverberation on the i-th ring is Q. The emitted signal is denoted as $s(t)$, and its strength is A. The length of the MA is represented by r, and the wave number is k, where $k = 2\pi f/c$. The speed of sound in the ocean is represented as c, and the scattering coefficient of the q-th scatterer is $R_{iq} = a_{iq}e^{j\psi_{iq}}$. Where $r = ct/2$, $MB = c(t - T)/2$, the ring area $S = \pi r(r^2 - MB^2)$. If the unit area is $\triangle S$, so the number of scatterers contributing to the reverberation in the ith ring is N, $N = [S/\triangle S]$, and $[\,]$ denotes the integer command, $1 \le q \le Q$. Therefore, the scattering characteristic function at time t can be expressed as:

$$P_i(t) = \sum_{q=1}^{Q} \frac{A}{r} e^{-jkr} R_{iq} \frac{1}{r} e^{-ikr} \tag{4}$$

$$= \frac{A}{r^2} \sum_{q=1}^{Q} e^{-j2kr} a_{iq} e^{j\psi_{iq}} \tag{5}$$

where a represents the amplitude and ψ represents the phaseain, a and ψ in are both random numbers following the

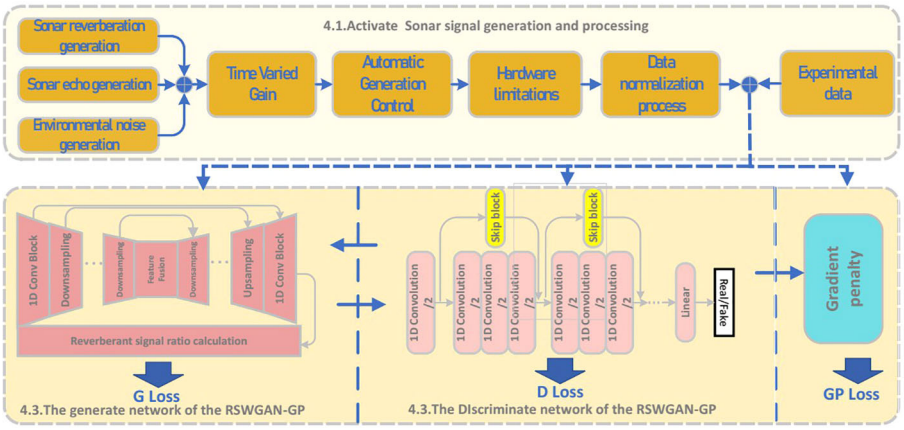


FIGURE 2
Reverberation suppression RSWGAN-GP construction.

Gaussian distribution with the constraint of, $0 \le a_{iq} \le 1, 0 \le \psi_{iq} \le 2\pi$. The process of generating reverberation involves convolving the emission signal with the scattering characteristic function. Therefore, the reverberation at time t can be expressed as the following formula:

$$R(t) = s(t) \otimes P_i(t) \tag{6}$$

Active sonar is utilized for detecting underwater targets, which involves reverberation and capturing the target echo signal. When the sonar signal hits the object, it generates a new echo through the multi-point superposition of the target body, enabling the active sonar to receive it (Hodges, 2011).

In addition to reverberation noise, target echo is an essential component of active sonar signals. The sonar is assumed to have an array of $[1 \cdots m \cdots Z]$ elements for reception. The target echo signal received by the m-th array element can be expressed as:

$$E(t) = \sum_{m=1}^{Z}\sum_{i=1}^{I} b_i S(t - \tau_{mi}) exp[W_{mi}(t) + 2\pi f_d(t) \cdot (t - \tau_{mi}) - \psi_i] \tag{7}$$

In Eq. (7) the coordinate of the m-th array element (x_m, y_m, z_m) is represented by a vector $\vec{r}_m$. I represents the number of highlights of the target, b_i represents the reflection coefficient of the i-th highlight, $S(t)$ represents the envelope of the transmitted signal, τ_{mi} represents the time delay experienced by the sound wave incident on the i-th highlight and then reflected the m-th array element, $W_{mi}(t)$ represents the angular frequency change of the sound wave irradiated to the i-th highlight and then reflected the m-th array element, f_d represents the Doppler shift, and ψ_i represents the random phase shift of the i-th highlight echo, uniformly distributed between $(0 \sim 2\pi)$. Different objects can be simulated by adjusting the number of highlights.

For environmental noise simulation, the spatial and physical characteristics of the Marine environment are complex, and the noise level depends on mixing multiple noise sources. This paper adopts an AR modeling method to simulate and synthesize Marine environmental noise (Chen et al., 2018). Firstly, an uncorrelated Gaussian white noise sequence $v(n)$ is generated, and the Marine environmental noise can be obtained by passing $v(n)$ through an AR filter with a specific temporal correlation. Mark the Marine environmental noise as $W(n)$, then the generation process of Marine environmental noise is shown in Eq. (8):

$$W(n) = -\sum_{k=1}^{p} a_p(k) W(n-k) + \sigma_v b_0 v(n) \tag{8}$$

p represents the order of the AR filter, $v(n)$ Gaussian white noise, and $\sigma_v^2 b_0^2$ represents variance of Marine ambient noise.AR filter coefficients $a_p(k)$ and b_0 can be solved by Levinson-Durbin (Diniz et al., 2010) method.

The data generated by the above three equations are normalized respectively, and the corresponding weight is assigned. The obtained signal is denoted as $S(t)$, and the obtained signal is shown in Figure 1A, where the signal marked in the yellow box represents the echo position of the target.

$$S(t) = R(t) + \alpha E(t) + \lambda W, \qquad \alpha, \lambda \in [0, 1] \tag{9}$$

Here, α represents the trade-off between $E(t)$ and $R(t)$. λ is an adjustable parameter that controls the degree of the strength fluctuation. Through the adjustment of parameters α and λ, enable diversity in the data set, the desired SIR (Signal to interference ratio, $10log(\alpha^2 P(E(t))/P(R(t)))$) and SINR (Signal to interference plus noise ratio, $10log(\alpha^2 P(E(t))/[P(R(t)) + \lambda^2 P(W)]$)) are achieved, where P represents power.

4.2 Implementation mechanism of the generative adversarial network

GAN is an effective data generation network, including Generator (G) and Discriminator (D). The G-analysis process is a minimal game process, and the discriminator and generator finally reach Nash equilibrium.

The adversarial training optimization process for generators and discriminators can be expressed as follows:

$$\min_G \max_D V(G, D) = E_{x \sim P_r}[log(D(x))] + E_{\tilde{x} \sim P_g}[log(1 - D(\tilde{x}))] \tag{10}$$

In Eq. (9), x is the actual data, P_r is the actual data distribution, P_g is the generated data $\tilde{x} = G(z)$ distribution. The objective

function expressed by $\min_G \max_D V(G,D)$ is to minimize the JSD between the expectation data distribution $P_r$and the generated data distribution, provided that the D is optimal.

In that case, the JSD cannot measure the distance between the generated and actual data distribution. Training the GAN by optimizing the JSD will result in not finding the correct optimization target, which is prone to the problem of unstable training gradient and model collapse.

In order to solve the problems mentioned above, the Wasserstein Generative Adversarial (Wasserstein GAN, WGAN) network proposes to use Wasserstein distance as an optimization method for training GANs. To satisfy the Lipschitz continuity, WGAN limits the weights to a specific range to enforce the Lipschitz continuity, but it is leads to poor generation results. WGAN-GP is a gradient penalty-based WGAN. WGAN-GP improves the Lipschitz continuity constraint by using gradient penalty instead of weight clipping in WGAN.

The objective function of WGAN-GP is:

$$\max_D V(G,D) = E_{x\sim P_r}[D(x)] - E_{\tilde{x}\sim P_g}[D(\tilde{x})] - \lambda E_{\hat{x}-P_{\hat{x}}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2] \quad (11)$$

$$\min_G L(G,D) = E_{x\sim P_r}[D(x)] - E_{\tilde{x}\sim P_g}[D(\tilde{x})] \quad (12)$$

In the formula, λ is the gradient penalty term coefficient, $P_{\hat{x}}$ is the sampling distribution of the gradient penalty term, the discriminator maximization $\max_D V(G,D)$, and the generator minimization $\min_G L(G,D)$.WGAN-GP provides a stable training method that requires little parameter tuning to solve training gradient disappearance and gradient explosion.

4.3 Reverberation suppression generator network

In the previous step, the shape of processed data is $[B,N]$, meaning that there are B test data of length N. B is divided into multiple b. Our goal is to separate a mixture signal $S \in [-1,1]^{N_m \times b}$ into K source signals $S'^1, \ldots S'^K$ with $S^k \in [-1,1]^{N_s \times B}$ for all $k \in \{1,\ldots,K\}$, K is set to 1 by default in this paper, B as the batch size at training time and N_m and $N_s$as the respectivenumbers of signal length. For model variants with no extra input context, we have $Nm = Ns$ and make predictions for the echo part of the input. Here we input the data S into the neural network structure and perform feature extraction on the data. It is divided into two parts to introduce the G network. The first part is the realization of data crop and concat, and the second part will introduce the whole generator network.

4.3.1 Data concatenation

It is challenging work to extract sonar signal features using one-dimensional convolution. A well-designed deep network structure is crucial for obtaining more valuable dataset recognition features. As the number of network layers increases, training deep networks becomes labor-intensive due to the common insurmountable problem. To address this issue, optimizing data concatenation when passing network parameters can be more effective. The implementation of this approach in the paper is shown in Figure 4.

In Figure 4, the signal feature x is passed from the previous layer and is processed by the $conv_1 \ldots conv_n$ layers to obtain the data x_n. $conv_y$ processes x through another branch line to obtain the result $conv_y(x)$. Then, the result $x_n \oplus conv_y(x)$ is obtained, and subsequent processing continues, done to prevent the loss of original features after multiple convolutions. The convolution of the branch is used to process the data and obtain the final data. In Figure 4A, the data crop operation is equivalent to ⊕, which is the operation of skip connect in Figure 4B.

The network construction will be built with network blocks and have the u-net network structure. The network has $\ell + n$ layers, and each layer is labelled $[1,\ldots,\ell-1,\ell,\ell+1,\ldots,\ell+n]$, where $\ell - n = 1$.

For a stacked-layer structure (consisting of several stacked layers), the learned feature is recorded as $H(x_{\ell-n})$ when the input is x. When ℓ is 0, the accumulation layer only performs identity mapping, and the network performance will not be degraded. It allows the accumulation layer to learn new features based on the input features, resulting in better performance. A convolutional block of length $\ell + n$ can be expressed as:

$$\begin{cases} x_{\ell+n} = F_d(x_{\ell-n-1}, w_{\ell-n-1}), & n < 1, n \in \mathbb{N} \\ x_{\ell+n} = F_u(x_{\ell+n-1}, w_{\ell+n-1}), & n > 1, n \in \mathbb{N} \\ x_\ell = F_m(x_\ell, w_\ell), & n = 1 \end{cases} . \quad (13)$$

$$x_{\ell+n} = F(x_{\ell+n}, w_{\ell+n}) \oplus H(x_{\ell-n}) \quad (14)$$

The convolution result can be obtained by iteratively convolving the output x from $y_{\ell-n}$ block to $y_{\ell+n-1}$ block, where $\ell + n$ ranges from $\ell - n - 1$ to $\ell + n - 1$. The input and output of the

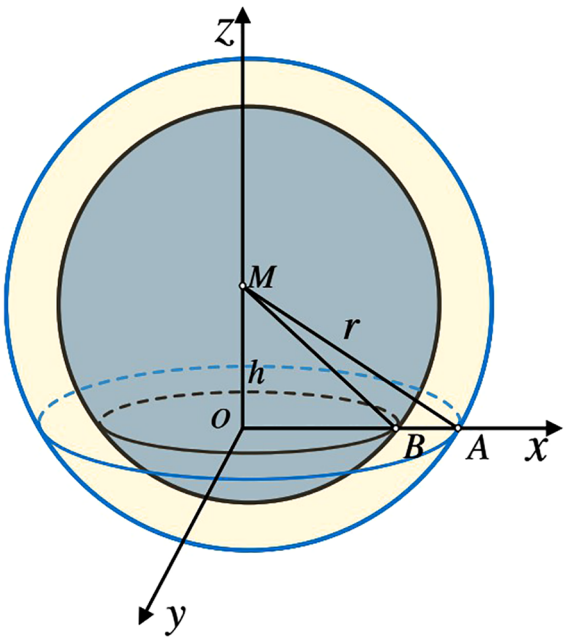


FIGURE 3
Reverberation model.

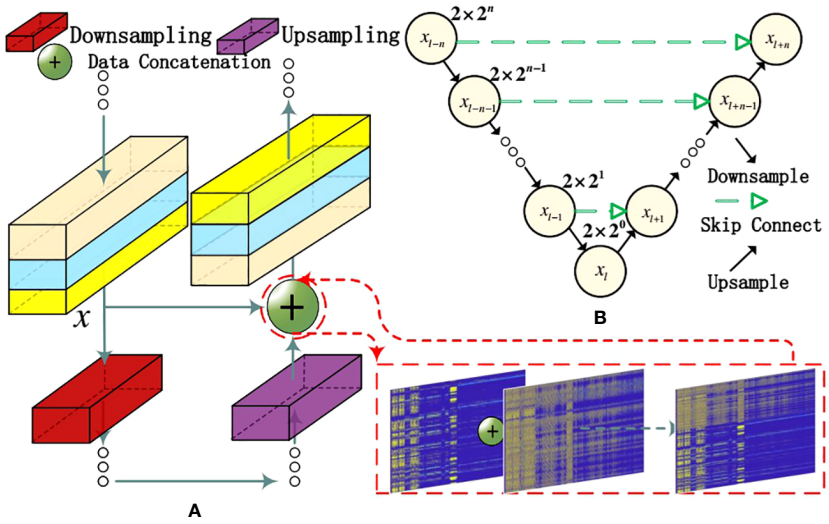


FIGURE 4
Data concatenation. **(A)** There are changes in the signal data processing process. **(B)** Show the process of signal data transmission.

formula are represented by x_ℓ, and each residual unit typically contains a multi-layer structure. F_d is the downsampling block function, F_u is the upsampling block function, representing the learned residual, and $H(x_\ell) = x_l$ represents the identity mapping. The learned features from the shallow layer ℓ to the deep layer L are expressed as:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, w_i) \tag{15}$$

The determination of L depends on the shortest distance $\mathcal{L}$ detected by the sonar, that is, the number of data points N processed by the sonar equipment. The size of L can be solved by the formula $L = log_2(\mathcal{L} \times Fs/N \times c)$, where Fs is the sampling rate of the active sonar brother and c represents the speed of sound propagation in water (m/s).

4.3.2 The generator network structure of RSWGAN-GP

The generator side of RSWGAN-GP is called G, which is constructed by a U-shaped network. It utilizes a one-dimensional convolution network that convolves specifically on signals while adding skip connections based on their original basis to enhance accuracy in signal feature extraction, as shown in Figure 5.

As shown in Figure 5, the signal data S is directly input into the encoder layer X_{En}^1 to start the one-dimensional convolution operation. The network structure's transmission process and main characteristics are shown in the figure, where G isspecially designed for processing sonar acoustic signals. The role of the encoder is to transform the input sequence into a low-dimensional representation that can capture the critical features of the input sequence. The decoder transforms the encoding vector into the target sequence and dynamically generates the content related to the target, as shown at X_{De}^3. The decoder receives the feature map from the same-scale encoder layer X_{En}^3 directly. Its data scale will not change, still 96×25600. The convolution of multiple neural networks may weaken data features with the increase of convolutional layers, so the data crop structure is utilized to reduce the loss of information, as shown in Figure 4.

In the generation part of the RSWGAN-gp network, we formulate the network running result X_{De}^1 as follows: let i indexes the down-sampling layer along the encoder, N refers to the total number of the encoder. The stack of feature maps represented by X_{De}^i, is computed as:

$$X_{De}^i = \begin{cases} \mathcal{H}\left(\underbrace{[\mathcal{C}\left(\mathcal{D}\left(X_{En}^k\right)\right)_{k=1}^{i-1}, \mathcal{C}\left(X_{En}^i\right)}_{\text{Scales: } 1^{th} \sim i^{th}}, \underbrace{\mathcal{C}\left(\mathcal{U}\left(X_{De}^k\right)\right)_{k=i+1}^{L}}_{\text{Scales: } (i+1)^{th} \sim N^{th}} \right)^{X_{En}^i} \\ i = 1, \cdots, L-1 \end{cases}^{i=L} \tag{16}$$

Where function $\mathcal{C}(\cdot)$ denotes a convolution operation, $\mathcal{H}(\cdot)$ realizes the feature aggregation mechanism with a convolution followed by a batch normalization and a Leaky-ReLU activation function. $\mathcal{D}(\cdot)$ and $\mathcal{U}(\cdot)$ indicate up- and down-sampling operation respectively, and $[,]$ represents the concatenation.

The convolution operation $\mathcal{C}$ of the signal is shown in the following formula:

$$X_j^{i+1} = (\sum_{k}^{K} \sum_{c=0}^{F} \omega_c^{i+1} X_{j+c}^i) + bias \tag{17}$$

Where X_l, X_{l+1} and ω are inputs, outputs and weight parameters, respectively; $1 \times F$ is the size of a single kernel; K is the number of kernels. Here, the kernel size of $1 \times F$ is shared for the whole input feature maps, called weight sharing.

It is worth mentioning that our proposed generator network is more efficient with fewer parameters. As for the decoder of the generator part, the depth of the feature map in a generation is symmetric to the encoder, and thus X_{De}^i, also has 12×2^i, channels. The number of parameters in i^{th} decoder stage of $\mathcal{P}_{De}^i$ can be computed as:

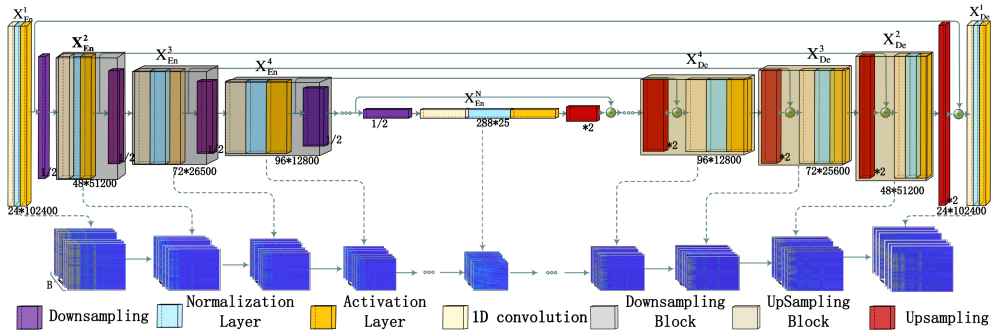


FIGURE 5
Generator network.

$$\mathcal{P}_{De}^{i} = D_F \times D_F \times \left[d\left(X_{De}^{i+1}\right) \times d\left(X_{De}^{i}\right) + d\left(X_{De}^{i}\right)^2 + d\left(X_{En}^{i} + X_{De}^{i}\right) \times d\left(X_{De}^{i}\right) \right] \tag{18}$$

where D_F is the convolution kernel size, $d(\cdot)$ denotes the depth of the nodes.

In G, the convolution with a stride of 1 maintains the output length equal to the input length. A downsampling method $\mathcal{D}(\cdot)$ is employed to increase the receptive field of the original data by $[1/2]$. The signal data $S^{N_m \times B} = [s^1, s^2, s^3, s^4, \cdots, s^{m-3}, s^{m-2}, s^{m-1}, s^m]$ is downsampled to $[s^1, s^3, \cdots, s^{n-2}, s^n]$. After convolving the data to obtain its minimum scale, corresponding upsampling $[\times 2]$ is used along with interpolation to restore the data to its original scale. The signal is transformed from S_n to $S_n^{'}$ after processing, while the signal length remains unchanged.

The network is symmetric, with the first half using downsampling and the second half using upsampling. The network's construction affects the length of data processing and the shortest distance for processing sonar signals. When the downsampling block has L layers, the number of input points is at least 2^L. An 11-layer symmetric network structure is used in the experiments, so the minimum input signal points are $2^{11} = 2048$. However, if only the signal data with a length of 2048 is input, it will output only one value after 11 downsamplings, leading to less feature representation. For X_{En}^{L} to be greater than 1 in the middle of the convolution, the data signal length for training should be at least $J = 2 \times 2048 = 4096$. The shortest detection distance of convolution is $\mathcal{L}$, and its formula is as follows:

$$\mathcal{L} = \frac{J}{2 \times Fs} \times c \tag{19}$$

Where Fs stands for the sampling frequency, based on the example calculation, we can determine that the shortest detection range of the active sonar after processing is 12.3m in the network constructed with an 11-layer downsampling block.

4.4 Reverberation suppression discriminator network

This section describes the discriminator part of building the adversarial network (D). The discriminator plays a crucial role in the GAN. It helps the generator to generate more realistic signals after the downsample (achieving unity in the frequency domain). It improves the GAN's ability to understand the training data, which lays the foundation for generating higher-quality sonar signals. In this will generate the signal $S_n^{'}$ and the required $E(S_n)$ input. In order to improve the accuracy of D, a one-dimensional convolutional discriminator network of SkipNet suitable for underwater acoustic signals is constructed by referring to the DenseNet Gao et al. (2020) structure. The main structure and the overall structure are introduced in the following.

When training the network, the complexity of the GAN network will bring problems such as long training time and difficulty in discovering signal features. The discriminator network uses traditional convolution to process longer underwater acoustic signals, which requires the design of a deeper network, which wastes time and may lead to feature disappearance and network degradation problems. Introducing SkipNet blocks can reduce the construction of network depth and training time. The skip connection makes it a flexible and efficient neural network architecture with good accuracy and resource efficiency performance.

In the more compact convolution, where rich features are less readily available due to the limited number of parameters, the different features that emerge from activation map-pings derived from data points during model inference may indicate the existence of a set of unique descriptors that are necessary to distinguish between different classes of objects. In contrast, data points with low feature diversity may need to provide sufficient unique descriptors to make valid predictions, called random predictions. Random prediction can negatively affect the optimization process and impair the final performance. This paper presents a series of Skipnet block models structured to fuse the previously convolved information using a skip net whenever a portion of the convolution passes to show the importance of diversity.

As shown in Figure 6, After downsampling the signal, the signal features will be put into the next layer and skipnet for multiple convolution. Finally, the results of two parts will be added. The feature signal obtained by addition is subjected to the same processing after downsampling once. This can be expressed using the following formula:

$$x_\ell = h_\ell(x_{\ell-1}) + H_\ell(H_{\ell-1} \cdots H_{\ell-n}(x_{\ell-n})) \tag{20}$$

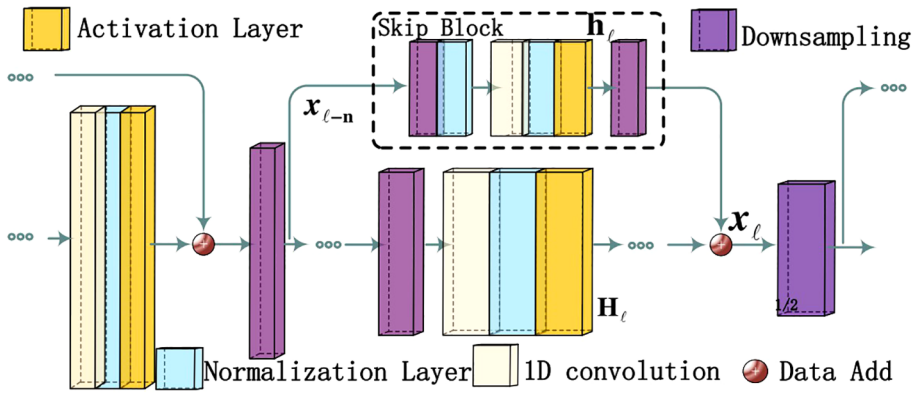


FIGURE 6
Discriminator SkipNet.

In Eq. (20), X represents the feature data after the convolution operation. $h(\cdot)$ stands for SkipBlock, which can be expressed as $h = (Downsampled\,|BN\,|\,Conv\,|\,BN\,|\,Activation|\,Downsampled)$, and H stands for the main convolution process. x_ℓ adds the results above the main line and the sideline.

The RSWGAN-GP network uses convolution with a skip network with skipBlock to implement the discriminative signal network. Due to the difference between signal and image processing, the feature extraction is carried out in the form of one-dimensional convolution, and the subsequent dimension reduction processing is transformed from pooling to down-sampling processing, which is more suitable for the processing of signal features Nakaoka et al. (2021). Figure 7 shows the implementation of the discriminative signal network.

In the discriminator network, the generator G is initialized and given a reverberation signal data vector as input. The generator generates signals based on the mapping of the input vector, creating generated data. The discriminator network then judges and identifies the generated data, producing a classification probability that results in a judgment (true or false). During discriminator training, actual data is also inputted to train the discriminator. The de-reverberation signal is labeled 1 (effective reverberation suppression), while the signal without effective reverberation suppression is labeled 0. The loss $L_D$is generated based on the generated result. In this paper, a non-densely connected network is designed to avoid redundancy and too many parameters in the signal processing network. The Desenet is designed without dense connections. Sparse connections are used instead, maintaining the same effect.

When the signal generated by the generator network meets the requirements of the signal-to-reverberation ratio, the generator network will stop training, and the final signal generator is the underwater reverberation suppression model.

4.5 Loss functions

This section presents the calculation of reverberation suppression loss for RSWGAN-GP. Due to the sinusoidal signal characteristics and the influence of phase difference, the original WGAN-GP calculation method cannot converge the training results. Therefore, a new method is used to preprocess the loss calculation before training.

This paper gives the main parameters of the generation network and the discriminant network used in the paper. The main parameters of the network will be described in Tables 1, 2.

In Table 1, the signal input is 102400×1, and the data mapped into the same scale by the generation network is also 102400×1. In Table 2, the signal input is 102400×1, and the Classification probabilities are formed after the convolution and linear layer processing. The output scale is 1×1, which is used as the output of the discriminator in GAN.

This paper uses the signal-to-reverberation ratio(SRR) to evaluate the signal after reverberation suppression. The SRR will be used to indicate the degree of signal suppression.

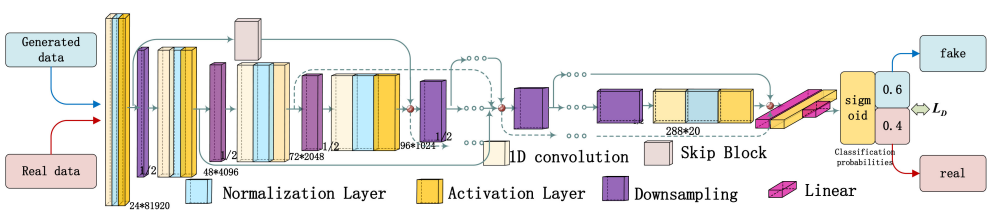


FIGURE 7
Discriminator network.

TABLE 1 Generation network.

Num	Layer	Act./Norm.	Output shape	Num	Layer	Act./Norm.	Output shape
0	Signal Latent vector	–	102400×1	14	upsample Conv1d	Datacrop12 LReLU/BN	50×576 50×288
1	Conv1d downsample	LReLU/BN -	102400×24 51200×24	15	upsample Conv1d	Datacrop11 LReLU/BN	100×522 100×264
2	Conv1d downsample	LReLU/BN -	51200×48 26500×48	16	upsample Conv1d	Datacrop10 LReLU/BN	200×504 200×240
3	Conv1d downsample	LReLU/BN -	26500×96 12800×96	17	upsample Conv1d	Datacrop9 LReLU/BN	400×456 400×216
4	Conv1d downsample	LReLU/BN -	12800×120 6400×120	18	upsample Conv1d	Datacrop8 LReLU/BN	800×408 800×192
5	Conv1d downsample	LReLU/BN -	6400×144 3200×144	19	upsample Conv1d	Datacrop7 LReLU/BN	1600×360 1600×168
6	Conv1d downsample	LReLU/BN -	3200×168 1600×168	20	upsample Conv1d	Datacrop6 LReLU/BN	3200×312 3200×144
7	Conv1d downsample	LReLU/BN -	1600×192 800×192	21	upsample Conv1d	Datacrop5 LReLU/BN	6400×576 6400×288
8	Conv1d downsample	LReLU/BN -	800×216 400×216	22	upsample Conv1d	Datacrop4 LReLU/BN	6400×264 6400×120
9	Conv1d downsample	LReLU/BN -	400×240 200×240	23	upsample Conv1d	Datacrop3 LReLU/BN	12800×216 12800×96
10	Conv1d downsample	LReLU/BN -	200×264 100×264	24	upsample Conv1d	Datacrop2 LReLU/BN	25600×168 25600×72
11	Conv1d downsample	LReLU/BN -	100×288 50×288	25	upsample Conv1d	Datacrop1 LReLU/BN	51200×120 51200×48
12	Conv1d downsample	LReLU/BN -	50×288 25×288	26	upsample Conv1d	Datacrop0 LReLU/BN	102400×72 102400×24
13	Conv1d	LReLU/BN	25×288	27	Conv1d	tanh	102400×1

$$SRR = 10log_{10}(\frac{P_{Signal} - P_{reverberation}}{P_{reverberation}}) \tag{21}$$

In Eq. (21), P_{Signal} is the power of the activated sonar signal, $P_{reverberation}$ is the power of the reverberation signal, and $P_{Signal} - P_{reverberation}$ is the power of the target echo signal. SRR can be used to indicate the degree of reverberation suppression. Denote by $\mathcal{R}(\cdot)$ in the following calculations.

The reverberation suppression network approach of RSWGAN-GP introduces the Wasserstein distance on top of the reverberation suppression model of GAN, and Eq. (22) shows the distance.

$$W(P_r, P_f) = \inf_{\gamma \prod (P_r, P_f)} E_{(\hat{s}, \hat{x}), \gamma}[\| \mathcal{R}(\hat{s}) - \mathcal{R}(s') \|] \tag{22}$$

In Eq. (22), $E(\cdot)$ is the calculated expectation; $\hat{s}$ is the desired sonar signal after actual reverberation suppression, and s' is the sonar signal after raw reverberation suppression; $\prod (P_r, P_f)$ is $\mathcal{R}(\hat{s}) - \mathcal{R}(s')$, the set of joint probability distributions of the corresponding expected signal $P_r$probability distribution and the generated signal P_f; $inf(\cdot)$ is the exact bound taken down; $E_{(\hat{s}, s'), \gamma}[\| \mathcal{R}(\hat{s}) - \mathcal{R}(s') \|]$is the expected value of the relative distance of the local discharge signal under the set of joint probability distributions γ.

In Eq. (22), the lower exact expectation bound is difficult to find so that the Wasserstein distance can be converted into the Kantorovich-Rubinstein dual form. Eq. (23) shows the Wasserstein distance's dual structure.

$$W(P_r, P_f) = \sup_{\| D \|_L \leq 1} E_{\hat{s}, P_r}[D(\hat{s})] - E_{s', P_f}[D(s')] \tag{23}$$

In Eq. (23). $D(x)$ is the distance cost function of discriminator D; $\| D \|_L \leq 1$ indicates that the discriminator distance cost function satisfies the 1-Lipschitz restriction. The GP in RSWGAN-GP indicates the gradient penalty function to satisfy the 1-Lipschitz restriction, and its loss function is shown in Eq. (19).

$$\begin{cases} L_G = -E_{Z \sim P_z} \times [D(G(Z))] \\ L_D = -Ez \sim P_z \times [D(G(Z))] - E_{X \sim P_r} \times [D(\hat{s})] + GP|_{s'} \\ GP|_{s'} = \lambda E_{s' \sim P_{s'}}[(\| \nabla_{\hat{x}} D(\hat{x}|C) \|_p - 1)^2] \end{cases} . \tag{24}$$

$L_G$is the generator loss function; $L_D$is the discriminator loss function; $G(Z)$ is the suppressed sonar signal generated by the generator; $P_Z$is the prior distribution of the input sonar signal Z; $GP|_{s'}$ is the gradient penalty term; λ is the canonical term coefficient; and $\| \cdot \|_P$ is the P-parameter.

TABLE 2 Discrimination network.

Num	Layer	Act./Norm.	Output shape
0	Signal Latent vector	–	102400×1
1	Conv1d Downsample	LReLU/BN -	102400×24 51200×24
2	Conv1d Downsample	LReLU/BN -	51200×48 26500×48
3	Conv1d Downsamplee	LReLU/BN -	26500×96 12800×96
4	Conv1d Downsample	LReLU/BN DataAdd	12800×120 6400×120
5	Conv1d Downsample	LReLU/BN DataAdd	6400×144 3200×144
6	Conv1d Downsample	LReLU/BN DataAdd	3200×168 1600×168
7	Conv1d Downsample	LReLU/BN Dataadd4	1600×192 800×192
8	Conv1d Downsample	LReLU/BN DataAdd	800×216 400×216
9	Conv1d Downsample	LReLU/BN DataAdd	400×240 200×240
10	Conv1d Downsample	LReLU/BN Dataadd7	200×264 100×264
11	Conv1d Downsample	LReLU/BN DataAdd	100×288 50×288
12	Conv1d Downsamplee	LReLU/BN DataAdd	50×288 25×288
13	Conv1d Downsample	LReLU/BN DataAdd	25×1 13×1
14	Linear	sigmoid	1×1

5 Experimental verification

This section validates the advancedness of the proposed model and method. For this purpose, we conducted marine experiments to verify that RSWGAN-GP can effectively suppress the reverberation of sonar signals and collect actual data in the field for verification. RSWGAN-GP, with other excellent reverberation suppression methods, compared to prove the effectiveness of the reverberation suppression method proposed. The following sections explain the detailed description and summary of the experiments.

In the experiment, the active sonar with 30-element with a self-receiving function is used to transmit a continuous wave (CW) signal with a $fs = 250kHz$ sampling rate and a $f = 30kHz$ frequency. The size of the training data set is 6000 data samples in total, among which 2520 actual sonar data samples are obtained through experiments, and 3480 data samples are generated by the underwater active sonar reverberation simulation method. The simulation data and actual data are randomly arranged, and the training set and the verification set are in a 5:1 ratio. The actual data are used to verify the reverberation suppression effect of the model obtained at the end of training. The experiments were conducted in one of the bays in Qingdao. Figure 8 shows the experimental scenario.

Figure 8A is a schematic diagram of the relative position of the active sonar and the target during the test. In the figure, $x_{h1} = 200m$ is the shortest distance from the relative position, and $x_{h2} = 300m$ is the longest distance from the relative position, active sonar is at the same position as the target at sea level $d_s = d_t \in [5, 10]$. In Figure 8B, the experiment in the harbor can minimize the error caused by the ship's swing caused by wind and waves. The narrowest point of the harbor exit is 200m, and the target ship tows the target at a distance of 200-300m from the active sonar. Figures 8A, B show the experimental active sonar equipment and the detected target, respectively. The detected target is a cylindrical object with an internal cavity with a diameter of 533mm and a length of 3m. The essential experimental data of the validation method are obtained in sea trials.

The data are collected and used as the validation set to validate each epoch in the training process. The active sonar of the signal is subjected to reverberation suppression, and the signal change during training is shown in Figure 9.

In Figure 9, the time domain diagram shows that the model is trained by mixing simulation data with actual acquisition data, and the trained model is used to process the experimental results of the active sonar signal data of the ocean experiment. The figure shows the results of sonar signal processing of model pairs produced by different iterations. In the Epoch 1-50 iteration training process, the processed sonar signal still has the phenomenon of signal chipping. Still, it shows the state of strong reverberation, and the target position cannot be visually observed from the time domain. Starting from Epoch 60-100, the chipping phenomenon of the echo signal disappears, the correct sinusoidal signal can be restored, and the neural network can already find the desired target feature state. From Epoch 110 onwards, the target can be precisely located, and in subsequent training iterations, the target echo feature can be highlighted while suppressing the strong reverberation state during sonar propagation. When the model training iteration is above Epoch 210, the target echo signal can be observed macroscopically from the time domain diagram. Reverberation changes in the middle state of the network as shown below:

The above-processed feature map, which results from the processing of the network $X_{En}^{N} = X_{De}^{N}$, results in a 288×25 feature array. Feature array is where the network will generate features, which will then be upsampled to recover the signal. In Figure 10, some features of the signal increase and decrease as the epoch increases, representing anti-reverberation operations. From Figure 10A, it can be seen that the convolution features of the signal at the beginning are dispersed to each corner of the array. After iteration, some features weaken, as shown in Figure 10B, which is reflected in the signal that the reverberation part begins to weaken, and then the features disappear, as shown in Figure 10C. Finally, the target echo features are enhanced, as shown in Figure 10D, and the corresponding display in the echo signal explored by sonar is the enhancement of the echo signal.

During the training process, convergence is achieved by continuously correcting errors. The loss curve changes during the training process of RSWGAN-GP, including the generator loss curve and the discriminator loss curve, as shown in Figure 11.

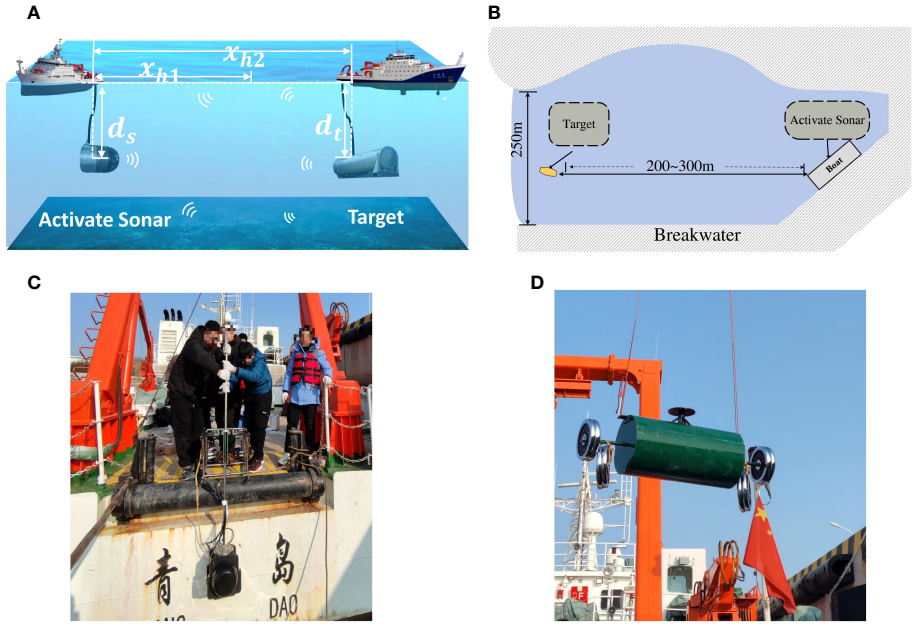


FIGURE 8
Experimental environment and equipment. **(A)** Sonar and target underwater deployment status. **(B)** Sonar and target experiment terrain environment and experiment method. **(C)** Active sonar equipment during the experiment. **(D)** Target equipment during the experiment.

Figure 10 shows the training loss curves of RSWGAN-GP under various weight decay and learning rate(lr) settings. Figure 10A is the line chart of the generator loss variation, and Figure 10B is the curve chart of the discriminator loss variation. By setting different parameter values, rapid convergence of the network can be achieved when weight decay and learning rate are set to 0.00001 and 0.0001, respectively. As shown in the figure, in the WGAN-GP network, the generator loss continuously approaches 0, and the discriminator loss continuously approaches 0. The loss of the generator shows a rising trend, and the loss of the discriminator shows a decreasing trend. During training, the generator and discriminator are in a state of mutual competition, and their loss values should fluctuate up and down. The trend shown in the figure represents the main direction.

For applications where target detection will be performed after signal processing, the difference between before and after signal processing is shown in Figure 12.

Figure 12 shows the target position detection map after nonlinear processors in matching fields on CW signal processing Sun and Li (2019), where the target indicates the target'0s location, which is about 204 m—the comparison between the original data and the data after RSWGAN-GP processing is shown. In the

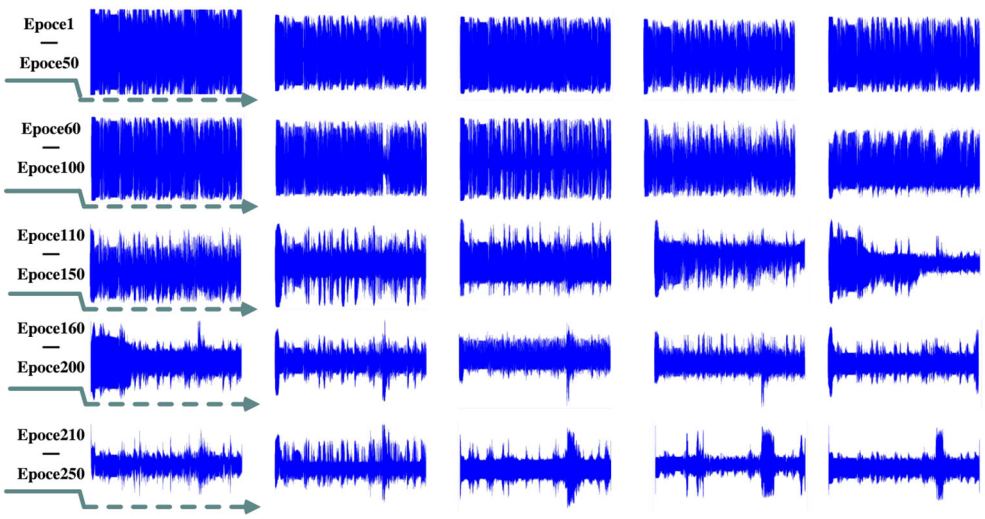


FIGURE 9
Active sonar reverberation suppression changing graph.

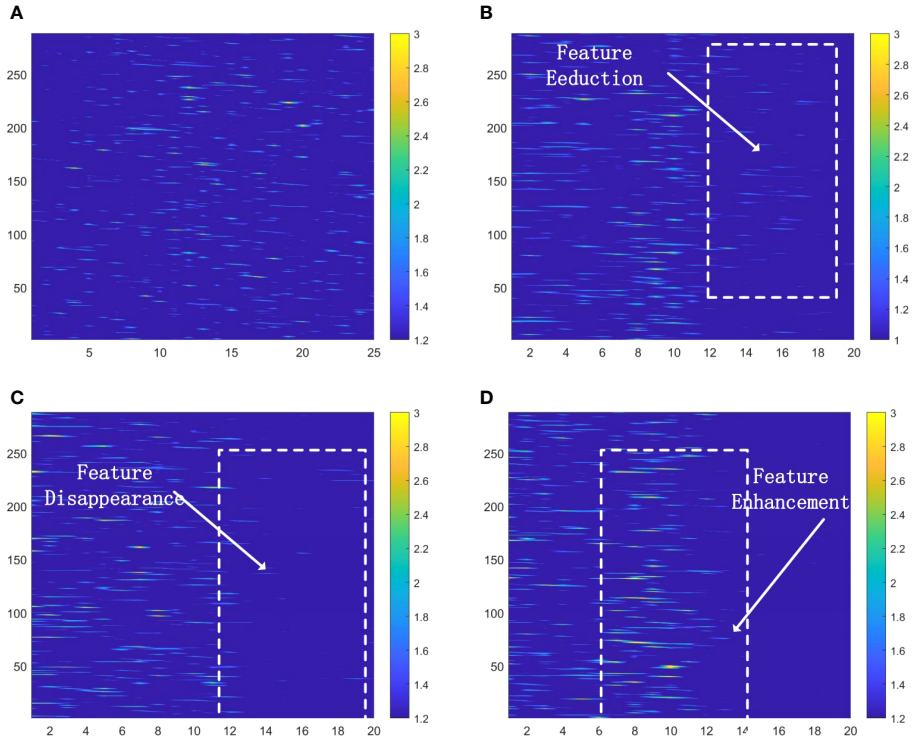


FIGURE 10
The intermediate state of the network changes. **(A)** Initial network intermediate layer processing result. **(B)** After 50 iterations. **(C)** After 100 iterations. **(D)** After 150 iterations.

original data, before the target position, there is the influence of reverberation on its judgment. When making a prediction, the highest correlation position changes and the highest point in front is judged as the target position. When comparing the red line with the blue line, it is evident that the signal processed by RSWGAN-GP can reduce the influence of reverberation when making target judgments, increasing the success rate of target judgments. Here, four contrasting points are selected, as shown in Table 3.

Among the data points, the first three are reverberation points, and the comparison shows that the reverberation is well suppressed, and the fourth is the target point. The SRR is improved after echo suppression. The data comparison can prove that the well-trained model can suppress the reverberation well, making the correlation increase by 0.022 and the reverberation part decrease by 0.3 on average.

There are many excellent algorithms in water acoustics reverberation suppression, here will use collected data for various methods to compare with the method proposed in this paper. The comparison results after processing the sonar signal are shown in Figure 13.

In Figure 13A, the Original signal is unprocessed, the reverberation power is 0.398, the signal echo power is 0.432, and its SRR is 0.361dB, the highlighted red part indicates the position of the target echo, which cannot be effectively identified from the figure. 2D-AR PreWhitener is used to eliminate correlation and frequency correlation in the signal Li et al. (2008) so that the signal is flatter in the frequency domain, and the processed echo sonar signal shows the position of the target echo. The least mean square filter (LMS) processes the echo signal Kim et al. (2000), and the reverberation component is suppressed, decreasing amplitude. Adaptive fractional Fourier transforms (FrFt) for suppressing reverberation Yu et al. (2022), Although the reverberation is partially suppressed, the actual impact is not apparent. The PCI-SVM reverberation suppression method combines Principal Component Inversion (PCI) and Support Vector Machine (SVM) techniques Wang et al. (2021). This method selects suitable and effective feature values through SVM to extract the main features for reverberation suppression. The figure shows that this method performs better than the previous ones. We present the impact of the RSWGAN-GP method proposed in this paper. Compared to the excellent methods, the reverberation component is effectively suppressed, and the echo component is more visible and prominent. The detailed data comparison is presented in Table 4.

Figure 13B compares SRR results processed by different methods under different SINR environments. In actual use, the reverberation signal processing SINR is between 0 dB and 10 dB. In order to reflect the processing ability of a high reverberation environment, the signal is mainly concentrated between -10 dB and 10 dB (data comparison in the middle of the two blue colors in the figure). As can be seen from the figure, RSWGAN-GP and PCI-SVM methods are significantly superior to other methods. At the same time, RSWGAN-GP is 3 dB higher than the PCI-SVM method at SINR -10 dB, and the advantages become more evident as SINR increases. The advantage of RSWGAN-GP is that it can learn many high reverberation data to improve the processing ability of high-reverberation data. In Figures 13A, B, we compare different

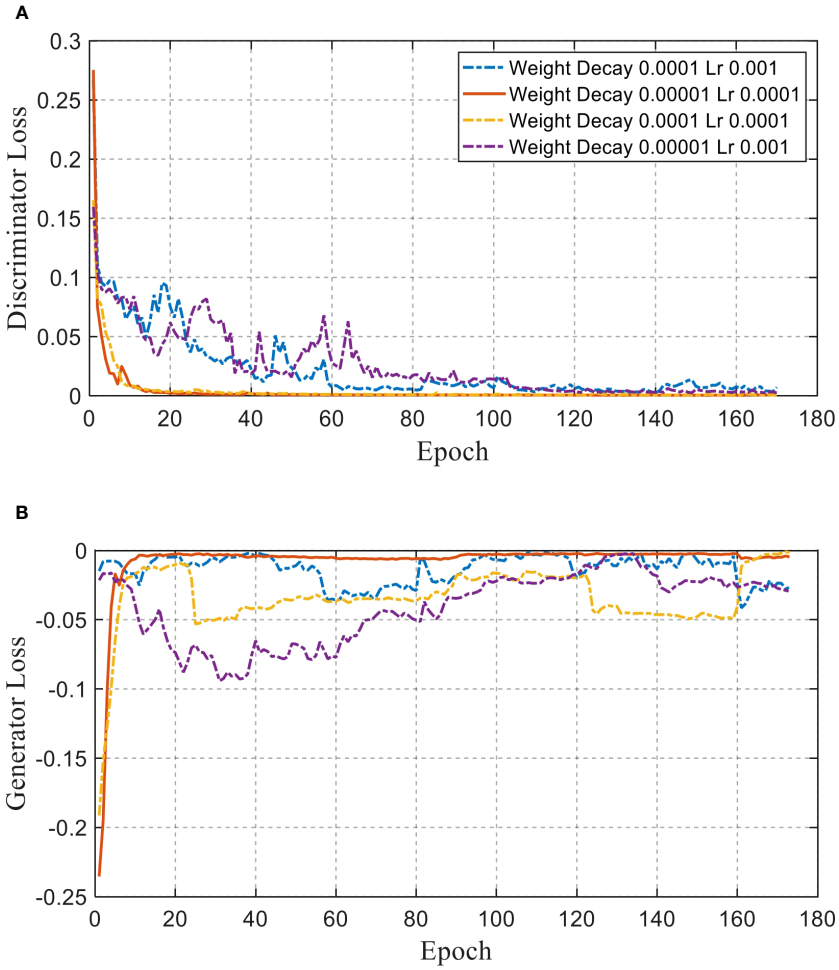


FIGURE 11
Results for different parameters. **(A)** The discriminator loss varies with epoch. **(B)** The generator loss varies with epoch.

reverberation suppression methods to demonstrate the effectiveness of RSWGAN-GP in suppressing reverberation.

Figure 14 shows the processing results and time-frequency distribution of raw signal, AR pre-whitening, LMS, FrFt, PCI-SVM and RSWGAN-GP, respectively. Experimental results show that the algorithm can effectively suppress reverberation and extract target echo components under high SRR conditions. At the same time, the time-frequency structure of the target highlight echo remains unchanged. The processing results of the original signal distribution are shown in Figure 14A. It can be seen that there is

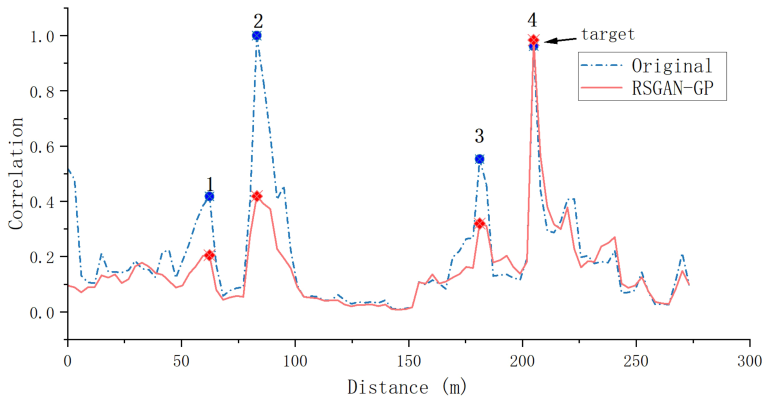


FIGURE 12
Matching field result.

TABLE 3 Target detection correlation comparison.

Num	Before treatment	After treatment	Difference
1	0.417	0.205	-0.212
2	1	0.419	-0.581
3	0.553	0.32	-0.233
4	0.962	0.984	0.022

some substantial reverberation interference near the target echo, but its energy is weaker than the target echo. The signal processed by 2D-AR PreWhitener, LMS, FrFt and PCI-SVM is shown in Figures 14B–E, and the target echo energy is still not obvious. As shown in Figure 14F, the RSWAGN-GP processing results show that the overall reverberation background has been effectively removed, and only a tiny part of weak background interference remains near the target echo.

The superiority of the proposed method can be seen in Figures 12, 13, and the detailed parameters are listed here for comparison, including reverberation amplitude, echo amplitude, reverberation power, echo power and SRR. The specific parameters are shown in Table 4.

The findings in Table 4 demonstrate that the RSWGAN-GP method proposed in this study is the most efficient technique for attenuating reverberation compared to the other methods evaluated. With a remarkable improvement in the reverberation ratio by 15dB, RSWGAN-GP significantly enhances speech quality from 0.83dB to 16.79 dB. The PCI-SVM algorithm comes in a close second, mainly when the optimal rank is 42, as determined by the SVM classification experiment, resulting in a significant improvement in the SRR by around 13dB. The 2D-AR PreWhitener, LMS, and FrFt techniques also improve the reverberant environment by -0.024dB, -2.68dB, and 0.12dB, respectively. Nevertheless, their ultimate effects are less substantial than those of RSWGAN-GP and PCI-SVM.

The effectiveness of the RSWGAN-GP method proposed in this paper for suppressing reverberation in sonar signals is demonstrated through experiments. Using experimental data from the ocean during training shows that features are extracted and amplified during the training iterations. By comparing the results of the algorithms, it is evident that this method can significantly improve signal quality. The improved SRR is 15.169 dB, demonstrating the method's effectiveness and superiority for reverberation suppression.

6 Conclusion

Reverberation suppression of echo signals is a crucial issue in active sonar systems. This paper presents a novel RSWGAN-GP method for suppressing reverberation in sonar signals using the generative adversarial network. This reverberation suppression

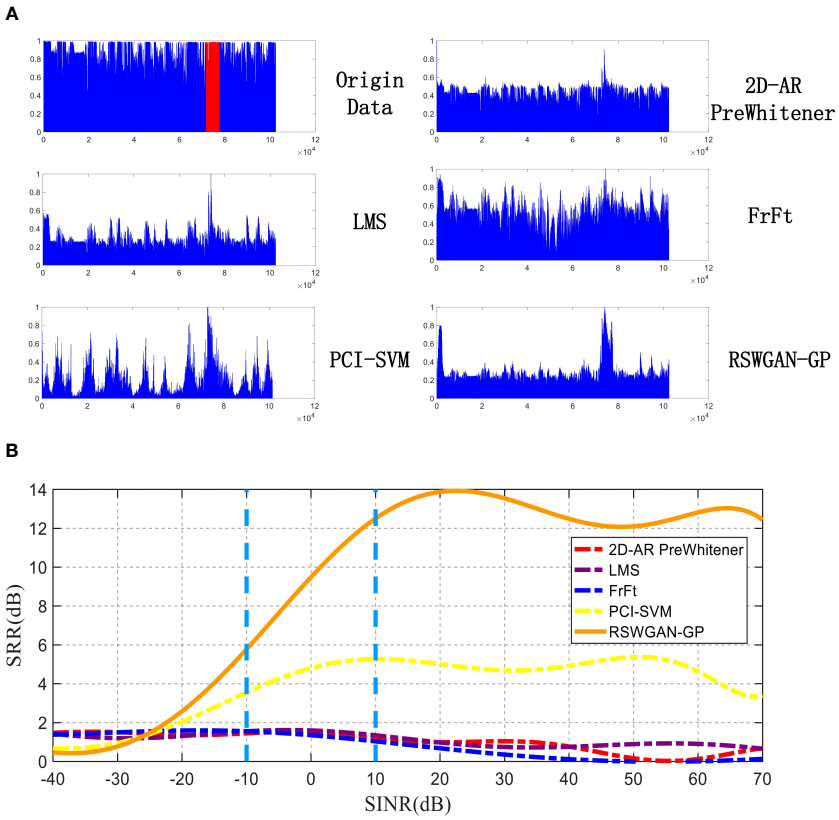


FIGURE 13
Sonar signal comparison of reverberation suppression methods. **(A)** Original signal diagram and signal diagram processed by 2D-AR Prewhitening, LMS, Frft, PCI-SVM, RSWGAN-GP methods. **(B)** SRR changes with SINR after model reverberation suppression.

TABLE 4 Experimental data comparison table.

	Reverberation power	Echo Power	SRR(dB)	Improve SRR(dB)
Original Data	0.398	0.432	0.361	0
2D-AR PreWhitener	0.096	0.087	0.35	-0.011
LMS	0.036	0.051	1.525	1.164
FrFt	0.086	0.094	0.414	0.053
PCI-SVM	0.0105	0.0562	7.2925	6.932
RSWGAN-GP	0.01	0.1265	11.021	10.659

network employs a one-dimensional convolutional network to process the signal content. RSWGAN-GP refers to U-net and DenseNet, using skip network structure and dense connection network to suppress sonar signal reverberation efficiently. Comparison using actual collected data demonstrates the effectiveness of the proposed method, which can effectively suppress the active sonar reverberation signal, improving SRR by approximately 10 dB, better than other methods.

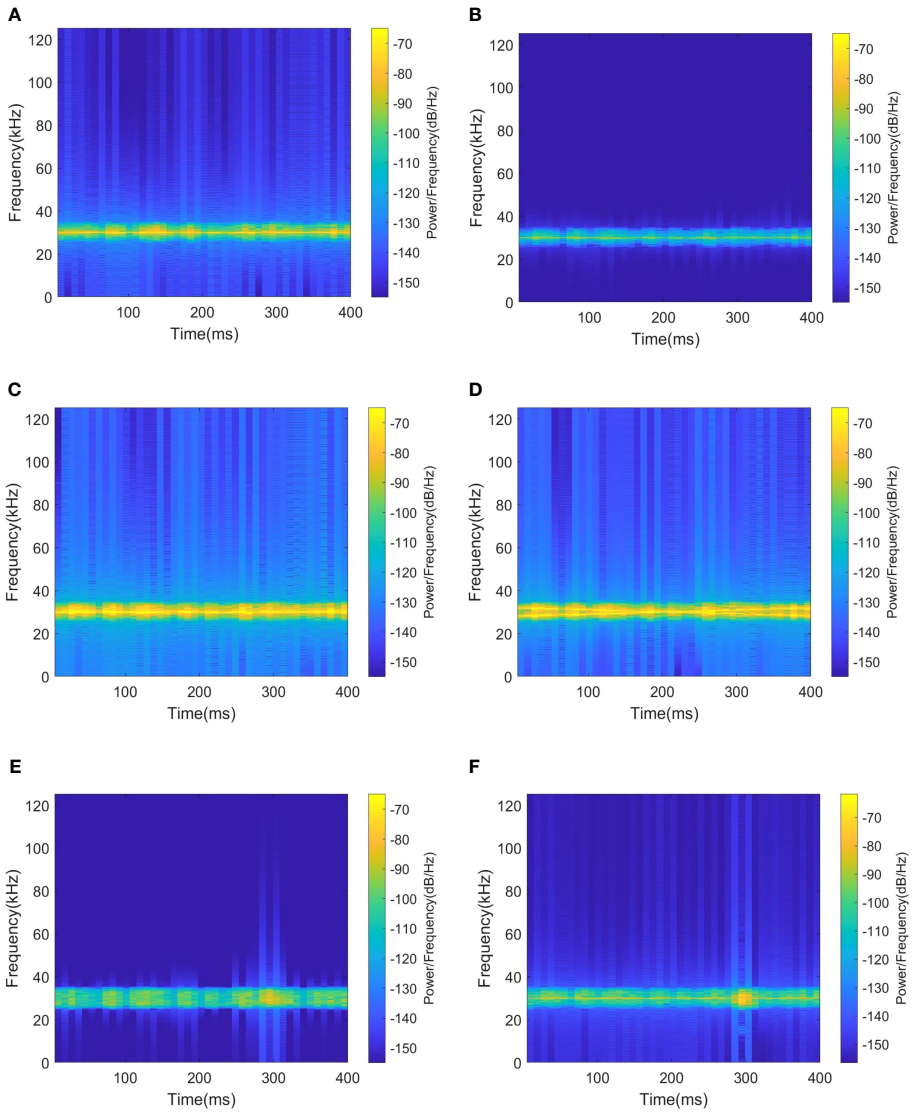


FIGURE 14
Time-frequency distributions. **(A)** Signal original state .**(B)** After 2D-AR PreWhitener processing. **(C)** After LMS processing. **(D)** After FrFt processing. **(E)** After PCI-SVM processing. **(F)** After RSWGAN-GP processing.

Through several experiments, it was discovered that the processing model trained by RSWGAN-GP has specific requirements for the transmission pulse width. The transmission pulse width of the signal needs to be adjusted for different detection distances, but the model's pulse width for echo signal processing is not sensitive to the width. However, this relationship may be specific to the dataset used. In future research, efforts will be made to improve the model's generalization capabilities to process signals with different pulse widths efficiently.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

ZW: Writing – original draft. HZ: Conceptualization, Supervision, Writing – review & editing. WH: Conceptualization, Formal Analysis, Methodology, Resources, Supervision, Writing – review & editing. XC: Methodology, Supervision, Writing – review & editing. NT: Data curation, Supervision, Writing – review & editing. YA: Data curation, Methodology, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Natural Science Foundation of China (Grant No. 91938204, 41527901, 61701462 and 62271459), the Marine S&T fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No.2018SDKJ0210), National Defense Science and Technology Innovation special Zone Project: Marine Science and Technology Collaborative Innovation Center (22-05-CXZX-04-01-02), Open Studio for Marine High Frequency Communications.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abrahamyan, L., Ziatchin, V., Chen, Y., and Deligiannis, N. (2021). Bias loss for mobile neural networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 1, 6556–6566. doi: 10.1109/ICCV48922.2021.00649

Ashraf, H., Jeong, Y., and Lee, C. H. (2021). Underwater ambient-noise removing gan based on magnitude and phase spectra. *IEEE Access* 9, 24513–24530. doi: 10.1109/ACCESS.2021.3051263

Bharathi, B. M. R., and Mohanty, A. (2019). Time delay estimation in reverberant and low snr environment by emd based maximum likelihood method. *Measurement* 137, 655–663. doi: 10.1016/j.measurement.2019.01.096

Chalapathy, R., Menon, A. K., and Chawla, S. (2017). "Robust, deep and inductive anomaly detection," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia*, (Skopje, Macedonia: Springer), September 18–22, 2017. 36–51, *Proceedings, Part I 10*Springer.

Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., and Willsky, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optimization* 21, 572–596. doi: 10.1137/090761793

Chen, J., Liu, C., Xie, J., An, J., and Huang, N. (2022). Time–frequency mask-aware bidirectional lstm: A deep learning approach for underwater acoustic signal separation. *Sensors* 22, 5598. doi: 10.3390/s22155598

Chen, L., Fang, S., and An, L. (2018). "Method of spatially correlated wideband ambient noise simulation for underwater acoustic array," in *In INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Chicago, IL: Institute of Noise Control Engineering), Vol. 258. 3160–3169.

Costas, J. P. (1984). A study of a class of detection waveforms having nearly ideal range—doppler ambiguity properties. *Proc. IEEE* 72, 996–1009. doi: 10.1109/PROC.1984.12967

Cox, H., and Lai, H. (1994). Geometric comb waveforms for reverberation suppression. *In Proc. 1994 28th Asilomar Conf. Signals Syst. Comput. (IEEE)* 2, 1185–1189. doi: 10.1109/ACSSC.1994.471646

Diniz, P. S., Da Silva, E. A., and Netto, S. L. (2010). *Digital signal processing: system analysis and design* (UK: Cambridge University Press).

Dong, H.-W., and Yang, Y.-H. (2018). Convolutional generative adversarial networks with binary neurons for polyphonic music generation. *arXiv preprint arXiv:1804.09399*. doi: 10.48550/arXiv.1804.09399

Faure, P. (1964). Theoretical model of reverberation noise. *J. Acoustical Soc. America* 36, 259–266. doi: 10.1121/1.1918943

Freburger, B., and Tufts, D. (1997). "Rapidly adaptive signal detection using the principal component inverse (pci) method," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)* (Pacific Grove, CA: IEEE), Vol. 1. 765–769, IEEE.

Gao, Y., Chen, Y., Wang, F., and He, Y. (2020). "Recognition method for underwater acoustic target based on dcgan and densenet," in *2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC)* (Beijing, China: IEEE). 215–221.

Han, K., Wang, Y., Wang, D., Woods, W. S., Merks, I., and Zhang, T. (2015). Learning spectral mapping for speech dereverberation and denoising. *IEEE/ACM Trans. Audio Speech Lang. Process.* 23, 982–992. doi: 10.1109/TASLP.2015.2416653

Hao, Y., Wu, X., Wang, H., He, X., Hao, C., Wang, Z., et al. (2023). Underwater reverberation suppression via attention and cepstrum analysis-guided network. *J. Mar. Sci. Eng.* 11, 313. doi: 10.3390/jmse11020313

Hodges, R. P. (2011). *Underwater acoustics: Analysis, design and performance of sonar* (United Kingdom: John Wiley & Sons).

Hu, N., Rao, X., Zhao, J., Wu, S., Wang, M., Wang, Y., et al. (2023). A shallow seafloor reverberation simulation method based on generative adversarial networks. *Appl. Sci.* 13, 595. doi: 10.3390/app13010595

Huang, T., and Wang, T. (2019). Research on analyzing and processing methods of ocean sonar signals. *J. Coast. Res.* 94, 208–212. doi: 10.2112/SI94-044.1

Innami, S., and Kasai, H. (2012). Nmf-based environmental sound source separation using time-variant gain features. *Comput. Mathematics Appl.* 64, 1333–1342. doi: 10.1016/j.camwa.2012.03.077

Jiang, K.-Y., Cai, Z.-M., and Lu, Z.-B. (2007). Forward and backward prediction model based signal detection in sea bottom reverberation. *Acta ELECTONICA Sin.* 35, 1766.

Kay, S., and Salisbury, J. (1990). Improved active sonar detection using autoregressive prewhiteners. *J. Acoustical Soc. America* 87, 1603–1611. doi: 10.1121/1.399408

Kim, K. M., Lee, C., and Youn, D. H. (2000). Adaptive processing technique for enhanced cfar detecting performance in active sonar systems. *IEEE Trans. Aerospace Electronic Syst.* 36, 693–700. doi: 10.1109/7.845261

Koh, S., Chia, C. S., and Tan, B. A. (2020). "Underwater signal denoising using deep learning approach," in *Global Oceans 2020: SINGAPORE–US Gulf Coast* (Biloxi, MS, USA: IEEE), 1–6.

Li, W., Ma, X., Zhu, Y., Yang, J., and Hou, C. (2008). Detection in reverberation using space time adaptive prewhiteners. *J. acoustical Soc. America* 124, EL236–EL242. doi: 10.1121/1.2963187

LiXiukun, X., and Mu, W.. (2015). Blind separation of unde rwater target echoes in reverberation back ground. *Journal of Harbin Engineering University* 36, 62.

Ma, L., Gulliver, T. A., Zhao, A., Zeng, C., and Wang, K. (2021). An underwater bistatic positioning system based on an acoustic vector sensor and experimental investigation. *Appl. Acoustics* 171, 107558. doi: 10.1016/j.apacoust.2020.107558

Mejjaoli, H., and Omri, S. (2020). Spectral theorems associated with the directional short-time fourier transform. *J. Pseudo-Differential Operators Appl.* 11, 15–54. doi: 10.1007/s11868-019-00308-z

Nakaoka, S., Li, L., Inoue, S., and Makino, S. (2021). "Teacher-student learning for low-latency online speech enhancement using wave-u-net," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Toronto, ON, Canada: IEEE). 661–665.

Neasham, J. A., Menon, R., and Hinton, O. R. (2007). "Broadband, ultra-sparse array processing for low complexity multibeam sonar imaging," in *OCEANS 2007-Europe* (Aberdeen, UK: IEEE), 1–6.

Ozaktas, H. M., Arikan, O., Kutay, M. A., and Bozdagt, G. (1996). Digital computation of the fractional fourier transform. *IEEE Trans. Signal Process.* 44, 2141–2150. doi: 10.1109/78.536672

Qian, W., and Cao, F. (2019). Adaptive algorithms for low-rank and sparse matrix recovery with truncated nuclear norm. *Int. J. Mach. Learn. Cybernetics* 10, 1341–1355. doi: 10.1007/s13042-018-0814-9

Ratnarajah, A., Ananthabhotla, I., Ithapu, V. K., Hoffmann, P., Manocha, D., and Calamia, P. (2023). "Towards improved room impulse response estimation for speech recognition," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (Rhodes Island, Greece: IEEE), 1–5.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany (Springer), October 5-9, 2015. 234–241, Proceedings, Part III 18Springer.

Singer, A. C., Nelson, J. K., and Kozat, S. S. (2009). Signal processing for underwater acoustic communications. *IEEE Commun. Magazine* 47, 90–96. doi: 10.1109/MCOM.2009.4752683

Song, Y., He, B., and Liu, P. (2019). Real-time object detection for auvs using self-cascaded convolutional neural networks. *IEEE J. Oceanic Eng.* 46, 56–67. doi: 10.1109/JOE.2019.2950974

Sun, X., and Li, R. (2019). Robust adaptive beamforming method for active sonar in single snapshot. *MATEC Web Conferences (EDP Sciences)* 283, 03006. doi: 10.1051/matecconf/201928303006

Sun, Q., Wang, H., Shen, X., Ning, W., and Fu, X. (2010). Research on the statistical modeling and simulation for interface reverberation. *IEEE* 9, 566–570. doi: 10.1109/ICCSIT.2010.5563616

Wang, X., Du, J., and Wang, Y. (2017). "A maximum likelihood approach to deep neural network based speech dereverberation," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. (Kuala Lumpur, Malaysia: IEEE), 155–158.

Wang, M., Wu, S., Guo, S., and Peng, D. (2021). Study on an anti-reverberation method based on pci-svm. *Appl. Acoustics* 182, 108189. doi: 10.1016/j.apacoust.2021.108189

Ward, S. (2001). "The use of sinusoidal frequency modulated pulses for low-doppler detection," in *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No. 01CH37295)* (Honolulu, HI, USA: IEEE), Vol. 4. 2147–2151.

Weiss, A., Singer, A. C., and Wornell, G. W. (2023). "Towards robust data-driven underwater acoustic localization: A deep cnn solution with performance guarantees for model mismatch," in *ICASSP 2023- 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (Rhodes Island, Greece: IEEE), 1–5.

Widrow, B., Mantey, P., Griffiths, L., and Goode, B. (1967). Adaptive antenna systems. *Proc. IEEE* 55, 2143–2159. doi: 10.1109/PROC.1967.6092

Wu, K., Cen, F., and Cai, H. (2008). "Svr-based approach to improve active sonar detection in reverberation," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. (Hong Kong, China: IEEE), 563–568.

Wu, Y., Li, X., and Wang, Y. (2018). Extraction and classification of acoustic scattering from underwater target based on wigner-ville distribution. *Appl. Acoustics* 138, 52–59. doi: 10.1016/j.apacoust.2018.03.026

Wuth, J., Stern, R. M., and Yoma, N. B. (2020). Non causal deep learning based dereverberation. *arXiv preprint arXiv:2009.02832*. doi: 10.48550/arXiv.2009.02832

Xiao, X., Wang, W., Ren, Q., Gerstoft, P., and Ma, L. (2021). Underwater acoustic target recognition using attention-based deep neural network. *JASA Express Lett.* 1, 106001. doi: 10.1121/10.0006299

Yin, Z., Zhang, S., Sun, R., Ding, Y., and Guo, Y. (2023). "Sonar image target detection based on deep learning," in *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*. (Ballar, India: IEEE), 1–9.

Yu, L., Ma, F., Lim, E., Cheng, E., and White, L. B. (2018). Rational-orthogonal-wavelet-based active sonar pulse and detector design. *IEEE J. Oceanic Eng.* 44, 167–178. doi: 10.1109/JOE.2018.2801158

Yu, G., Sun, J., and Li, X. (2022). A signal enhancement method based on the reverberation statistical information. *EURASIP J. Adv. Signal Process.* 2022, 1–13. doi: 10.1186/s13634-022-00902-2

Zhan, F., Zhu, H., and Lu, S. (2019). Spatial fusion gan for image synthesis. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 1, 3653–3662. doi: 10.1109/CVPR.2019.00377

Zhang, N., Wen, Z., Hou, X., and Wen, W. (2017). "Digital automatic gain control design with large dynamic range in wireless communication receivers," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*. (Chengdu, China: IEEE), 1402–1406.

Zhang, X.-W., Yang, D.-D., Guo, J.-X., and Zuo, L. (2019). "Weak moving target detection based on short-time fourier transform in sea clutter," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. (Wuxi, China: IEEE), 415–419.

Zhang, X., Yang, P., and Sun, H. (2023a). An omega-k algorithm for multireceiver synthetic aperture sonar. *Electron. Lett.* 59, e12859. doi: 10.1049/ell2.12859

Zhang, X., Yang, P., and Zhou, M. (2023b). Multireceiver sas imagery with generalized pca. *IEEE Geosci. Remote Sens. Lett.* 20. doi: 10.1109/LGRS.2023.3286180

Zhao, Y., Wang, D., Xu, B., and Zhang, T. (2018). "Late reverberation suppression using recurrent neural networks with long short-term memory," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (Calgary, AB, Canada: IEEE), 5434–5438.

Zhu, Y., Duan, R., and Yang, K. (2022). Robust shallow water reverberation reduction methods based on low-rank and sparsity decomposition. *J. Acoustical Soc. America* 151, 2826–2842. doi: 10.1121/10.0010353

Zhu, G., and Sun, H. (2008). "Improved support vectors machine for signal detection in non-reverberation," in *2008 IEEE Ultrasonics Symposium*. (Beijing, China: IEEE), 1508–1511.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 24 October 2023
DOI 10.3389/fmars.2023.1253105

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Jiahua Zhu,
National University of Defense Technology, China
Shaohua Hong,
Xiamen University, China

*CORRESPONDENCE
Jinsong Tang
jinsongtang_wh@163.com

RECEIVED 04 July 2023
ACCEPTED 29 August 2023
PUBLISHED 24 October 2023

CITATION
Ning M, Zhong H, Li H, Ma M, Dai L and Tang J (2023) A wide-beam NCS algorithm for multi-receiver SAS based on azimuth spectrum superposition.
Front. Mar. Sci. 10:1253105.
doi: 10.3389/fmars.2023.1253105

A wide-beam NCS algorithm for multi-receiver SAS based on azimuth spectrum superposition

Mingqiang Ning[1], Heping Zhong[1], Han Li[1], Mengbo Ma[2], Lili Dai[3] and Jinsong Tang[1]*

[1]Naval Institute of Underwater Acoustic Technology, Naval University of Engineering, Wuhan, China, [2]Intelligent Game and Decision Lab, Academy of Military Sciences, Beijing, China, [3]Chinese People's Liberation Army, 91917 Unit, Beijing, China

The existing multi-receiver synthetic aperture sonar (SAS) imaging algorithms are suitable for narrow-beam width, which will lead to a decrease in imaging quality under wide-beam condition and are not in line with the development needs of SAS. We propose a non-linear chirp scaling algorithm (NCSA) for wide beam multi-receiver SAS. Firstly, the point target reference spectrum (PTRS) of each receiver is obtained by the Lagrange inversion theorem (LIT), and then the under-sampled signal in the azimuth frequency domain is obtained through azimuth spectrum extension; Then, considering the cubic term of range frequency in the PTRS and the linear variation of equivalent frequency modulation slope with range, each receiver is imaged using the NCSA, and coherent superposition is performed in the azimuth frequency domain to eliminate spectrum aliasing caused by azimuth spectrum extension; Finally, the azimuth inverse transform is performed on the superimposed signal to obtain the focusing imaging. Computer simulation experiments and field data verify that this method is superior to the existing SAS imaging algorithm, improving the quality of wide-beam imaging, avoiding the interpolation operation of the traditional range-Doppler algorithm, and saving computation cost.

KEYWORDS

multi-receiver, synthetic aperture sonar, Lagrange inversion theorem, non-linear chirp scaling, azimuth spectrum superposition

1 Introduction

Synthetic Aperture Sonar (SAS) has played a very important role in ocean exploration, and its functions are constantly expanding, requiring high resolution, long detection distance, and strong detection capabilities for buried objects(Zhang and Tan, 2018; Tan et al., 2019; Ma et al., 2020; Zhang et al., 2021a; Tian et al., 2022; Zhang et al., 2023). According to the characteristics of underwater sound wave propagation, the lower the frequency of the transmitted signal, the stronger the detection distance and buried object detection ability; According to the definition for azimuth resolution, higher azimuth resolution can be achieved by using smaller transmitter (Marx et al., 2000; Zhang et al.,

2023; Zhu et al., 2023), waveform diversity (Zhu et al., 2023), and advanced synthetic aperture processing (Zhang and Ying, 2022; Zhang et al., 2022b; Zhang et al., 2023). Both low frequency and high-resolution imaging will involve wide-beam imaging algorithms. It is necessary to research wide-beam imaging algorithms to improve imaging resolution, detection distance, and the capability of buried objects.

Although SAS technology originated from synthetic aperture radar (SAR)(Qian et al., 2021; Li et al., 2022), the low speed of sound velocity underwater results in two main differences between wide-beam SAS and wide-beam SAR: low sound velocity under water causes azimuth moving distance during sending and receiving (AMDSR) to be not negligible, making the commonly used stop-and-hop assumption (Bonifant, 1999; Wilkinson, 2001; Callow, 2003) not applicable to SAS, A more complex non-stop-and-hop assumption must be used (Zhang et al., 2022); In addition, low sound velocity underwater can also cause a contradiction between the farthest detection distance and pulse repetition frequency (PRF) (Xu et al., 2003; Wang et al., 2015). To solve this contradiction, SAS normally adopts a multi-receiver configuration. This type of SAS is commonly referred to as multi-receiver SAS (Zhang et al., 2021a; Yang and Liu, 2022; Zhang et al., 2022a).

At present, imaging algorithms for wide-beam multi-receiver SAS can be divided into two categories: point by point imaging algorithms and line by line imaging algorithms. The point by point imaging algorithms is inefficient, which is a common problem of the two-dimensional time-domain imaging algorithms. Although the fast point by point imaging algorithms (Liu et al., 2009; Giardina, 2012; Duan et al., 2017; Synnes et al., 2017; Zhang and Yang, 2022) avoid this problem to some extent, compared with the line by line imaging algorithm, the computation efficiency is still too low. The line by line imaging algorithms use interpolation or Chirp Scaling (Raney et al., 1994; Wang et al., 2009; Liao and Liu, 2017; Zhang and Yang, 2019; Li et al., 2021; Huang and Yang, 2022) operation in the range Doppler domain or two-dimensional frequency domain to realize the range cell migration correction (RCMC) of all targets in the scene, so as to obtain higher efficiency than point by point imaging algorithms. Although the algorithm efficiency has improved, the imaging quality under wide beam conditions will decrease, so there are few line by line wide beam imaging algorithms. A commonly used method (Zhang et al., 2014) uses the range-Doppler algorithm (RDA) (Jiang et al., 2004; Tian et al., 2016; Zhang et al., 2019) to process the echo signal for each receiver by using method of series reversion (MSR), and then carries out coherent stacking for accurate wide-beam imaging, we called it the RDA-MSR. However, the computation load of individual receiver imaging is high, and the interpolation of RDA-MSR will increase the computation load more significantly, which is not conducive to real-time imaging. Moreover, the derivation of the point target reference spectrum (PTRS) in the RDA-MSR is not accurate enough, and as the beam width increases, the PTRS error will also increase. Therefore, the actual applicable beam width is not large; A research (Zhang, 2014) used the macro range cell migration correction (MRCMC) between different receivers, thereby transforming the multi-receiver SAS into the traditional monostatic SAS model, we called it the RDA-MRCMC. Although this method improves processing efficiency, the phase error caused by the increase in beamwidth and bandwidth may exceed $\pi/4$, thereby affecting imaging quality (Wu et al., 2016; Zhang et al., 2018; Zhang et al., 2018; Wu et al., 2019; Ma et al., 2023). The poor performance of the above methods is a limitation of RDA itself. (Zhang et al., 2021b) used the NCSA to develop a wide beam imaging algorithm, but the PTRS obtained by PCA (Bellettini and Pinto, 2002; Gough et al., 2004; Zhang et al., 2023) is not accurate enough.

This paper has three main contributions: firstly, this paper uses LIT to obtain the most accurate PTRS, which has smaller errors compared to the MSR in (Zhang et al., 2014) and is more suitable for wide beam imaging; Secondly, this paper uses the NCSA to preserve the range frequency to the cubic term of the PTRS obtained by LIT, and the frequency modulation slope varies with range, making it more suitable for wide band and wide swath model; Thirdly, the NCSA avoids interpolation and has higher computational efficiency compared to the RDA-MSR, while also achieving better imaging results compared to the RDA-MRCMC, balancing imaging quality and efficiency.

This paper is organized as follows: multi-receiver SAS model establishment and approximation, imaging algorithm derivation, and algorithm validation. In SAS model establishment and approximation, we use Lagrange inversion theorem (LIT)(Xiong et al., 2011; Vu et al., 2014; Zhang et al., 2017) to avoid tedious algebraic processing and the lengthy expression of stationary phase point, thereby obtaining the more accurate PTRS of each receiver compared to the RDA-MSR and the RDA-MRCMC, and then extending the azimuth spectrum to obtain under-sampling azimuth frequency domain signals. In imaging algorithm derivation, firstly, to adapt to the wide-beam condition, the cubic term of range frequency in the PTRS and the linear variation of equivalent frequency modulation slope with range are considered; Then, each receiver is imaged using the NCSA and then coherent superposition is performed in the azimuth frequency domain to eliminate the impact of spectrum aliasing caused by azimuth under-sampling. Compared with the RDA-MSR, the proposed method avoids interpolation and has high computational efficiency; Finally, fusing superimposed signal to obtain the focusing result. In algorithm validation, we use simulation data and field data to verify the effectiveness of the proposed method. The results show that this method saves computation costs, and the simulation imaging results of point targets under wide-beam are better than the existing methods, improving the imaging quality.

2 Establishment and approximation of multiple receiver SAS model

2.1 Accurate range history of point targets

The relative position between receivers and transmitter is shown in Figure 1, with the direction of the platform moving forward as the positive direction, and the transmitter in the middle of all the receivers. r is the range, x is the azimuth; the distance between the ith receiver and the transmitter is d_i; the time elapsed

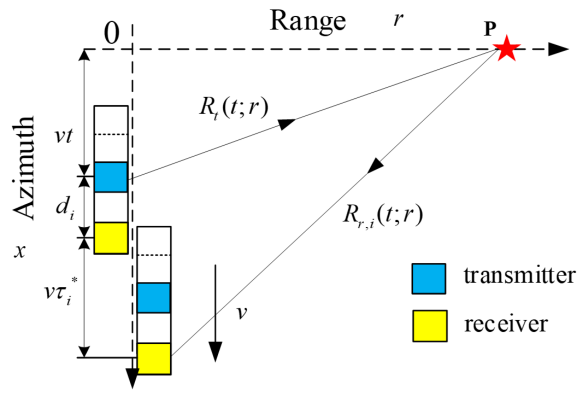


FIGURE 1
Geometric diagram of multi-receiver SAS.

between the transmission of a pulse and its reception by the ith receiver is τ_i^*. $R_t(t;r)$ is the propagation path of the transmitting signal, and $R_{r,i}(t;r)$ is the propagation path of the ith receiver. To illustrate the geometric relationship, the SAS at different times in the Figure 1 are not on the same straight line, but in reality, they are on the same straight line and move along the x-axis in the positive direction at velocity v.

During the process of moving at v m/s, the transmitter transmits linear frequency modulation (LFM) signal at a fixed pulse repetition frequency (PRF) in orthogonal mode simultaneously. At time t, for the point target $P(r,0)$, according to the geometric relationship, the path from the transmitter to the ith receiver scattered by the point target $P(r,0)$ can be expressed as follows:

$$R_i^*(t;r) = R_t(t;r) + R_{r,i}(t;r) = \sqrt{r^2 + (vt)^2} + \sqrt{r^2 + (vt + d_i + v\tau_i^*(t;r))^2} \quad (1)$$

where, $\tau_i^*(t;r)$ is the time interval between signal transmission and reception, so the propagation path of sound waves can be written as:

$$R_i^*(t;r) = c \cdot \tau_i^*(t;r) \quad (2)$$

The exact expression of $\tau_i^*(t;r)$ can be obtained by combining (1) and (2), and $\tau_i^*(t;r)$ can be expressed as

$$\tau_i^*(t;r) = \frac{1}{c^2 - v^2} \cdot \left\{ v(vt + d_i) + c\sqrt{(vt)^2 + r^2} + \sqrt{\left[v(vt + d_i) + c\sqrt{(vt)^2 + r^2}\right]^2 + (c^2 - v^2)\left[(2(vt)d_i + d_i^2\right]} \right\} \quad (3)$$

2.2 Error analysis of AMDSR

Under the narrow-beam assumption, it is generally approximated τ_i^* as a range variance $2r/c$, resulting in the error of AMDSR is

$$\Delta x = (\tau_i^* - 2r/c)v \quad (4)$$

According to the system parameters shown in Table 1, different receivers have different baseline lengths relative to the transmitter.

We analyzed the receiver with the maximum baseline length and calculated Δx across the whole swath under different beam widths. The results are shown in Figure 2.

As shown in Figure 2, Δx is maximum at the edge of the beam and increases with range and beam width. The maximum value of narrow-beam SAS is shown in. Figure 2A is 0.0009m, far less than half the length of a receiver (0.025m); the maximum value of wide-beam SAS shown in Figure 2B is 0.007m, which can be compared with the half-length of receiver, which may lead to the problem of azimuth non-uniform sampling. Therefore, under the condition of wide-beam, τ_i^* cannot be approximated to $2r/c$, and the azimuth variance must be considered, which means that the range history $R_i^*(t;r)$ must adopt a more accurate form.

2.3 Point target echo response model

The accurate range history has a complex form and cannot obtain an analytical expression for the PTRS. The current wide beam algorithms generally use the MSR (Neo et al., 2007; Wu et al., 2016; Zhang et al., 2021), which approximates the accurate range history using Taylor expansion. For example, reference (Zhang et al., 2014) preserves the fourth order term of Taylor expansion, as follows

$$R_i(t;r) = k_{0,i} + k_{1,i}t + k_{2,i}t^2 + k_{3,i}t^3 + k_{4,i}t^4 \quad (5)$$

where,

$$k_{n,i} = \frac{1}{n!}\frac{\mathrm{d}^n R_i^*(t;r)}{\mathrm{d}t^n}\bigg|_{t=0} \quad (n = 0,1,2,3,4) \quad (6)$$

The range history error obtained using the parameters shown in Table 1 is shown in Figure 3, ϵ is range error, λ is the wavelength of signal.

It can be seen that under the beam width shown in Table 1, the four-order expansions of the range history can no longer meet the requirement that the maximum range history error is less than (1/

TABLE 1 MADOM SAS simulation parameters.

Parameters	Values	Units
Beam width	25.18	°
Center frequency	20	kHz
Signal bandwidth	10	kHz
Transmitter size	0.15	m
Receiver size	0.075	m
Number of receivers	40	-
SAS platform velocity	5.0	m/s
Swath	[7.5, 300]	m

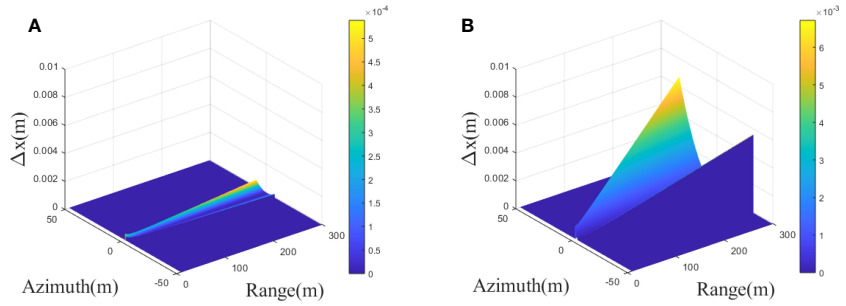


FIGURE 2
Approximation Error of AMDSR **(A)** narrow-beam **(B)** wide-beam.

16)λ, and the phase error caused cannot be ignored. Meanwhile, preserving different expansion terms has different range history error. In theory, the higher the expansion terms, the more accurate the range history, and the smaller the range history error. We plot the curve of the maximum range history error within the whole beam as a function of the number of Taylor expansions, as shown in Figure 4. It can be seen that the range history error does not decrease indefinitely with the number of expansions. The reason why the error cannot be infinitely reduced is that (Zhang et al., 2014) used narrow beam approximation. When the expansion reaches 10, the range history error is 0.216λ, which is still greater than (1/16)λ. Therefore, this research uses the LIT to derive the PTRS using the original range history, which does not make any approximation to the range history.

Let f_0 be the carrier frequency of the transmitting signal; K_r is the frequency modulation slope of the transmitted signal; $w_r(\cdot)$ is the envelope of the transmitted signal; $w_a(\cdot)$ is the analytical expression for the antenna pattern; A_0 is the signal's amplitude, it is independent with the imaging quality, so we ignore it in the next context. After demodulation, the baseband form of the ith echo signal can be expressed as

$$ss_i(\tau, t; r) = A_0 \cdot w_r\left(\tau - \frac{R_i(t;r)}{c}\right) \cdot w_a(t) \cdot \exp\left(-j\frac{2\pi f_0 R_i(t;r)}{c}\right) \cdot \exp\left(j\pi K_r\left(\tau - \frac{R_i(t;r)}{c}\right)^2\right) \quad (7)$$

To obtain each receiver's PTRS, the Principle of Stationary Phase (PSP) is performed to simplify range FFT on the baseband signal, and the range spectrum signal is obtained:

$$Ss_i(f_r, t; r) = A_0 A_1 W_r(f_r) w_a(t) \exp\left(-j\frac{2\pi(f_0 + f_r)R_i(t;r)}{c}\right) \exp\left(-j\pi\frac{f_r^2}{K_r}\right) \quad (8)$$

where, $W_r(f_r) = w_r(\frac{f_r}{K_r})$, and $W_r(\cdot)$ is the range frequency envelope, then perform azimuth FFT on (8) to obtain

$$SS_i(f_r, f_a; r) = \int_{-\infty}^{+\infty} Ss_i(f_r, t; r) \exp(-j2\pi f_a t) dt \quad (9)$$

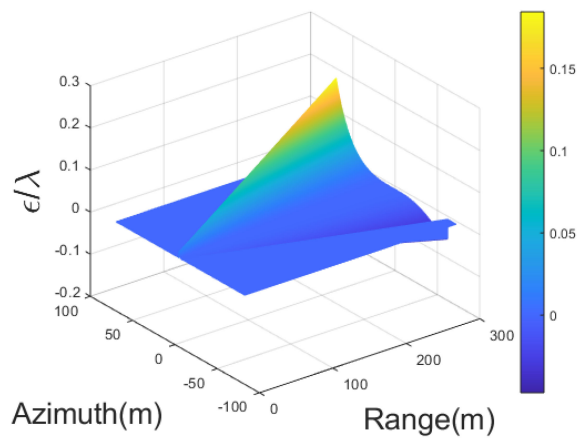


FIGURE 3
Range history error after retaining 4 terms.

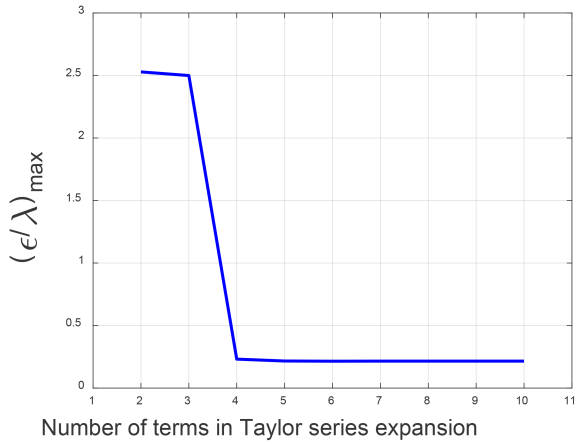


FIGURE 4
Maximum range history error vs the number of terms.

According to PSP, if the derivative of the phase in the integral expression with respect to t is 0 and the phase stationary time is assumed to be $\tilde{t}$, the following equation can be obtained

$$f_a = -\frac{(f_0 + f_r)}{c} \cdot \frac{dR_i(\tilde{t};r)}{dt} \tag{10}$$

From (10), it can be seen that f_a is a function of $\tilde{t}$. According to the LIT (Zhang et al., 2017), if f_a is close-form at a certain value $\tilde{t}_0$ and $f_a(\tilde{t}_0) \neq 0$, which is different from the approximation of $R_i(\tilde{t};r)$ in MSR, we can directly obtain the close-form solution of $\tilde{t}$.

$$\tilde{t} = \sum_{n=1}^{3} \lim_{\tilde{t}\to 0} \frac{[f_a(\tilde{t}) - f_a(0)]^n}{n!} \frac{d^{n-1}}{dt^{n-1}} \left[\frac{\tilde{t} - 0}{f_a(\tilde{t}) - f_a(0)}\right]^n \tag{11}$$

Take $\tilde{t}_0 = 0$ and $n = 3$, and bring (10) into (11), then there is

$$\tilde{t}_i = -\frac{1}{R_i^{(2)}(0;r)}\left(\frac{f_a - f_a(0)}{f_0 + f_r}c\right) - \frac{R_i^{(3)}(0;r)}{2(R_i^{(2)}(0;r))^3}\left(\frac{f_a - f_a(0)}{f_0 + f_r}c\right)^2 + \left(\frac{1}{6}\frac{R_i^{(4)}(0;r)}{(R_i^{(2)}(0;r))^4} - \frac{1}{2}\frac{(R_i^{(3)}(0;r))^2}{(R_i^{(2)}(0;r))^5}\right)\left(\frac{f_a - f_a(0)}{f_0 + f_r}c\right)^3 \tag{12}$$

The PTRS of the ith receiver is obtained by integrating (12) into the phase of (9), and we can get

$$SS_i(f_r, f_a; r) = A \cdot W_r(f_r) \cdot W_a(f_a) \cdot \exp\left(j\varphi_i(f_r, f_a; r)\right) \tag{13}$$

where,

$$\varphi_i(f_r, f_a; r) = -\frac{2\pi(f_0 + f_r)R_i(\tilde{t}_i;r)}{c} - \pi\frac{f_r^{\,2}}{K_r} - 2\pi f_a \tilde{t}_i \tag{14}$$

In order to analyze the phase errors between the PTRS obtained by different methods and the accurate PTRS, we selected the receiver 40 as the analysis object according to the parameters of the wide-beam SAS system shown in Table 1. We obtained the PTRS's phase errors of the MSR and LIT at 3 point targets at different ranges, as shown in Figures 5, 6A–C. represent the phase errors of point targets at ranges of 50m, 150m, and 250m, respectively. Comparing Figures 5, 6, it can be seen that the phase error increases with the range and the azimuth frequency. However, the phase error of the LIT is always significantly smaller than MSR. The maximum value of phase error of LIT in $(0m, 250m)$ is 0.019rad, which is much smaller than $\pi/4$ (Ning et al., 2023). Therefore, under the wide-beam SAS system parameter, the method proposed in this paper can meet the requirements of high-resolution imaging across the whole swath.

Expanding (14) to a series of f_r. To apply the method in this research to the wide beam condition, we reserve the cubic term to obtain:

$$\varphi_i(f_a, f_r; r) = \varphi_{az,i}(f_a; r) + \varphi_{rcmc,i}(f_a; r)f_r + \varphi_{rc}f_r^2 + \varphi_{src,i}(f_a; r)f_r^2 + \varphi_{cubic,i}(f_a; r)f_r^3 \tag{15}$$

where, $\varphi_{az,i}(f_a; r)$ is the azimuth modulation term; $\varphi_{rcmc,i}(f_a; r)$ is the range migration term; φ_{rc} is the range frequency modulation term; $\varphi_{src,i}(f_a; r)$ is the second range compression term; $\varphi_{cubic,i}(f_a; r)$ is the third-order coupling term of range and azimuth. According to the definitions of $R_{rd,i}(f_a; r)$ and $Km_i(f_a; r)$, we can get

$$R_{rd,i}(f_a; r) = \frac{\varphi_{rcmc,i}(f_a; r)}{-2\pi} \cdot c \tag{16}$$

$$Km_i(f_a; r) = \frac{\pi}{\varphi_{src,i}(f_a; r) + \varphi_{rc}} \tag{17}$$

$R_{rd,i}(f_a; r)$ is the range migration curve of the ith receiver in the range-Doppler domain; $Km_i(f_a; r)$ is the equivalent frequency modulation slope of the range compression filter for the ith receiver. Bring (16) and (17) into (15), $SS_i(f_r, f_a; r)$ can be rewritten as

$$SS_i(f_r, f_a; r) = W_r(f_r)W_a(f_a)\exp\left(+j\varphi_{az,i}(f_a; r)\right)\exp\left(-j\frac{2\pi R_{rd,i}(f_a; r)}{c}f_r\right)\exp\left(+j\frac{\pi}{Km_i(f_a; r)}f_r^2\right)\exp\left(+j\varphi_{cubic,i}(f_a; r)f_r^3\right) \tag{18}$$

The different receiver has different coefficients such as $R_{rd,i}$ and Km_i, which means that for the same point target, the different receiver has different point target echo responses. Therefore, it is necessary to perform matching filtering processing separately for each receiver.

3 Imaging algorithm derivation

The flowchart of the imaging algorithm is shown in Figure 7, where FFT represents the Fast Fourier Transform; IFFT is Inverse Fast Fourier Transform. The algorithm includes six-fold FFT/IFFT, one-fold azimuth spectrum extension, six-fold phase multiplication,

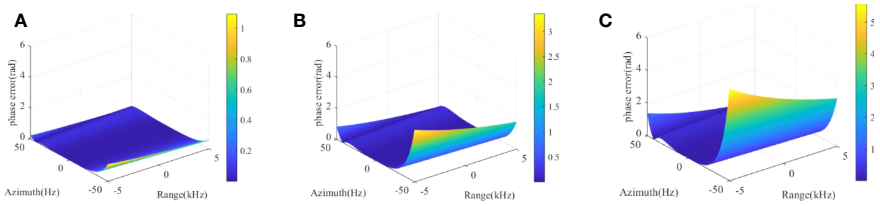


FIGURE 5
Phase error of PTRS of MSR at different ranges **(A)** r=50m **(B)** r=150 m **(C)** r=250m.

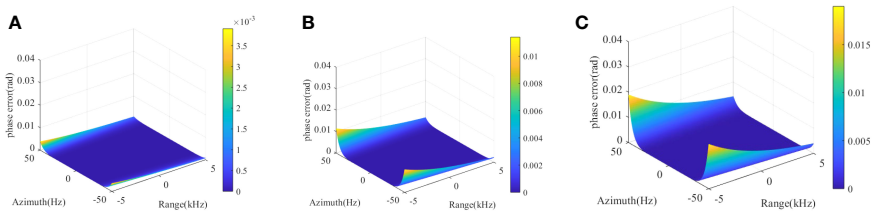


FIGURE 6
Phase error of PTRS of LIT at different ranges **(A)** r=50m **(B)** r=150 m **(C)** r=250m.

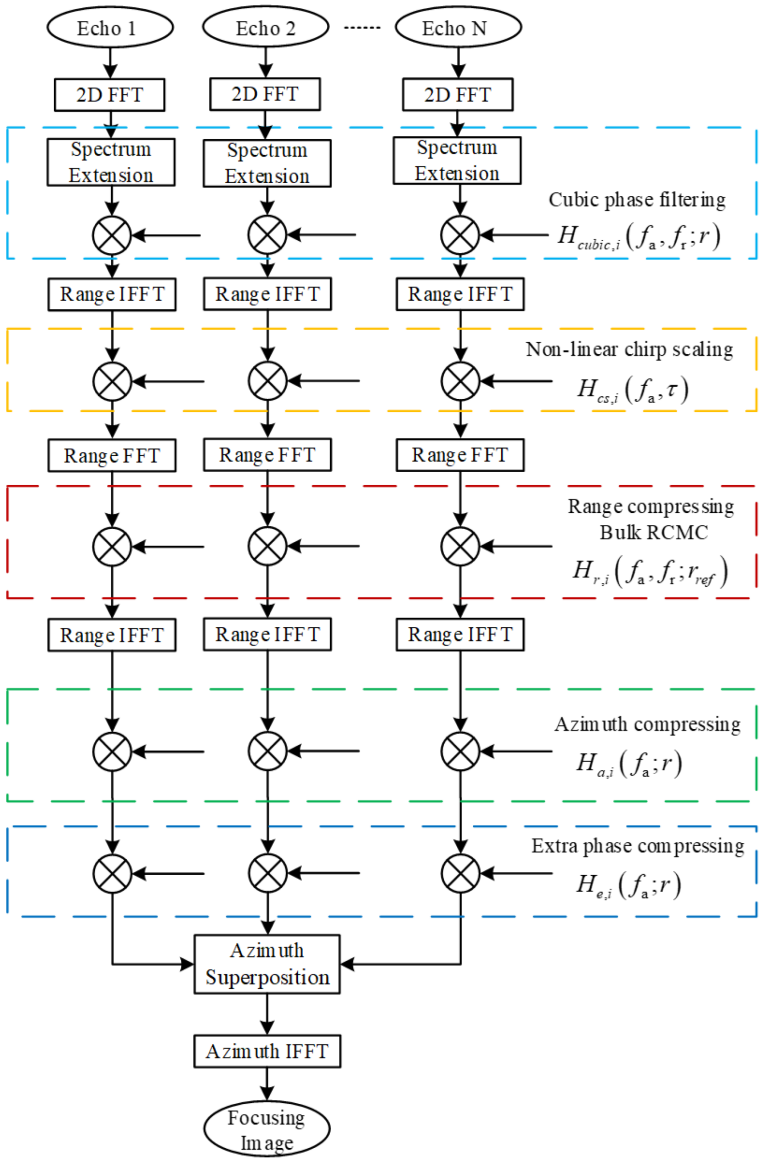


FIGURE 7
Algorithm flowchart.

and one-fold azimuth spectrum superposition. The azimuth spectrum extension is used to increase the length of one receiver's data to meet the requirements of signal processing, but the processed signal is highly under-sampled in the azimuth frequency, which can cause azimuth spectrum aliasing. The azimuth spectrum coherent superposition is used to suppress the azimuth spectrum aliasing caused by azimuth under-sampling.

3.1 Approximation of equivalent frequency modulation slope

Compared to traditional RDA and CSA, this paper considers the linear relationship of Km_i with r and approximates Km_i at the reference range

$$Km_i(f_a;r) = Km_{i,r_{ref}} + Ks_{i,r_{ref}} \cdot (r - r_{ref}) \tag{19}$$

where $Ks_{i,r_{ref}}$ is the first derivative of $Km_i(f_a;r)$ at r_{ref}, expressed as

$$Ks_{i,r_{ref}} = -\pi \cdot \frac{\varphi'_{src,i}(f_a;r_{ref})}{(\varphi_{src,i}(f_a;r_{ref}) + \varphi_{rc})^2} \tag{20}$$

where, r_{ref} generally takes the center position of the whole swath.

3.2 Cubic phase filtering

In order to eliminate the influence of the cubic term of f_r, a cubic phase filter is performed on $SS_i(f_a,f_r;r)$ in the 2D frequency domain. The expression of the cubic phase filter is

$$H_{cubic,}i(f_r,f_a;r) = \exp(j2\pi Y_i(f_a)f_r^3/3) \tag{21}$$

where, $Y_i(f_a)$ is the coefficient of the cubic phase filter that varies with the azimuth frequency f_a, with the aim of filtering out the cubic phase of the range frequency f_r and the cubic phase error generated by subsequent nonlinear Chirp Scaling. Since $\varphi_{cubic,i}(f_a;r)$ is weakly dependent on the range (Neo et al., 2008), $\varphi_{cubic,i}(f_a;r_{ref})$ can substitute $\varphi_{cubic,i}(f_a;r)$, and $Ym_i(f_a)$ is the error after the cubic phase filtering

$$Y_i(f_a) = Ym_i(f_a) - \frac{3}{2\pi}\varphi_{cubic,i}(f_a;r_{ref}) \tag{22}$$

The obtained PTRS after cubic phase filtering on (7) is

$$\begin{aligned} SS_{c,i}(f_r,f_a;r) &= SS_i(f_r,f_a;r) \cdot H_{cubic,}i(f_r,f_a;r) \\ &= W_r(f_r)W_a(f_a) \times \exp(j\varphi_{az,i}(f_a;r)) \times \exp\left(-j\frac{2\pi R_{rd,i}(f_a;r)}{c}f_r\right) \times \exp\left(j\frac{\pi}{Km_i(f_a;r)}f_r^2\right) \times \exp\left(j\frac{2\pi Ym_i f_r^3}{3}\right) \end{aligned} \tag{23}$$

3.3 Non-linear chirp scaling

Using PSP for range IFFT, the signal obtained in the range-Doppler domain is

$$\begin{aligned} sS_i(\tau,f_a;r) = {} & \omega_r\left(\frac{Km_i}{K_r}\left(\tau - \frac{2r}{c\gamma_i}\right)\right)W_a(f_a)\exp(\\ & + j\varphi_{az,i}(f_a;r))\exp(-j\pi Km_i(\tau \\ & - \tau_d)^2)\exp\left(-j\frac{2\pi Ym_i Km_i^3(\tau-\tau_d)^3}{3}\right) \end{aligned} \tag{24}$$

where $\gamma_i = \frac{2}{R'_{rd,i}(f_a;r_{ref})}$.

Let the expression of the non-linear chirp scaling equation be

$$\begin{aligned} H_{cs,i}(\tau,f_a) = {} & \exp(-j\pi q_{2,i}(\tau-\tau_{ref})^2) \\ & \cdot \exp\left(-j\frac{2\pi q_{3,i}(\tau-\tau_{ref})^3}{3}\right) \end{aligned} \tag{25}$$

where, $q_{2,i}$ and $q_{3,i}$ are undetermined coefficients, which will be solved later in the text.

Perform non-linear Chirp Scaling on $sS_i(\tau,f_a;r)$ to obtain

$$sS_{cs,i}(\tau,f_a;r) = sS_i(\tau,f_a;r) \times H_{cs,i}(\tau,f_a) \tag{26}$$

To obtain the coefficients of the cubic phase filter and the scaling equation, the range IFFT of $sS_{cs,i}(\tau,f_a;r)$ is performed to the 2D frequency domain, and then the series expansion of f_r is performed, retaining the cubic term. After the above operation, we can obtain

$$\begin{aligned} SS_{cs,i}(f_r f_a;r) = {} & W_r(f_r)\cdot W_a(f_a)\exp\left(j\frac{2\pi}{3}\frac{(Km_i^3 Ym_i+q_{3,i})f_r^3}{(Km_i+q_{2,i})^3}\right)\exp\left(j\frac{\pi}{Km_i+q_{2,i}}f_r^2\right)\exp\left(j\frac{2\pi Km_i(\tau_d-\tau_{ref})(Km_i^2Ym_iq_{2,i}-q_{3,i})}{Km_i+q_{2,i})^3}f_r^2\right) \\ & \exp\left(j-\frac{2\pi(\tau_d Km_i+q_{2,i}\tau_{ref})}{Km_i+q_{2,i}}f_r\right)\exp\left(j\frac{2\pi Km_i^2(\tau_d-\tau_{ref})^2(Km_iYm_iq_{2,i}^2+q_{3,i})}{(Km_i+q_{2,i})^3}f_r\right)\exp(j\varphi_{az,i}(f_a;r))\exp\left(-j\frac{\pi Km_iq_{2,i}(\tau_d-\tau_{ref})^2}{Km_i+q_{2,i}}\right) \\ & \exp\left(j\frac{2\pi Km_i^3(\tau_d-\tau_{ref})^3(Ym_iq_{2,i}^3-q_{3,i})}{3(Km_i+q_{2,i})^3}\right) \end{aligned} \tag{27}$$

After bringing (19) and $\tau = 2r/(c\gamma_i)$, $\tau_{ref} = 2r_{ref}/(c\gamma_i)$, and $\Delta\tau = \tau_d - \tau_{ref}$ into (27), and then expanding the coefficients of each order of f_r into the series of $\Delta\tau$, we can obtain the three undetermined coefficients as follows

$$Y_{m,i} = \frac{Ks_{i,r_{ref}}(2\alpha_i - 1)}{2Km_i^3 r_{ref}(\alpha_i - 1)} \tag{28}$$

$$q_{2,i} = Km_{i,}r_{ref}(\alpha_i - 1) \tag{29}$$

$$q_{3,i} = \frac{(\alpha_i - 1)Ks_{i,r_{ref}}}{2} \tag{30}$$

where $\alpha_i = \frac{\gamma_i(f_{ac})}{\gamma_i(f_a)}$, the phase expression of the scaled signal obtained by bringing (28), (29) and (30) into (27) is

$$\begin{aligned} \phi_{cs,i}(f_r f_a;r) = {} & \frac{\pi f_r^2}{Km_{i,r_{ref}}\alpha_i} + \frac{f_r^3 Ks_{i,r_{ref}}\pi}{3\alpha_i(\alpha_i-1)Km_{i,ref}^3} + \left(\left(-\frac{4\pi}{\gamma_i c} + \frac{4\pi}{c\gamma_i\alpha_i}\right)r_{ref} - \frac{4\pi r}{c\gamma_i\alpha_i}\right)f_r + \varphi_{az,i}(f_a;r) \\ & - \frac{\pi Km_{i,r_{ref}}(\alpha_i-1)\Delta\tau^2}{\alpha_i} - \frac{(\alpha_i-1)Ks_{i,r_{ref}}\pi\Delta\tau^3}{3\alpha_i} \end{aligned} \tag{31}$$

3.4 Range processing

By compensating for the first, second, and bulk RCMC terms in a phase multiplication, and simultaneously completing bulk RCMC, range compression, and cubic coupling term compensation, the

phase multiplication factor is

$$H_{r,i}(f_a,f_r;r_{ref}) = \exp\left(j\left(\frac{4\pi}{\gamma_i c}-\frac{4\pi}{c\gamma_i\alpha_i}\right)r_{ref}f_r\right)\exp\left(j\frac{\pi}{Km_{i,ref}\alpha_i}f_r^2\right)\exp\left(j\frac{Ks_{i,ref}\pi}{3\alpha_i(\alpha_i-1)Km_{i,ref}^3}f_r^3\right) \tag{32}$$

Perform range IFFT, and the obtained the phase of signal in range-Doppler domain is

$$sS_{rc,i}(f_a;r) = p_r\left(\tau-\frac{2r}{c\gamma_i a_i}\right)\cdot W_a(f_a)\exp\left(j\varphi_{az,i}(f_a;r)\right)exp\left(-j\frac{(a_i-1)ks_{i,ref}\pi\Delta\tau^3}{3a_i}\right)\exp\left(-j\frac{\pi Km_{i,ref}(a_i-1)\Delta\tau^2}{a_i}\right) \tag{33}$$

where $p_r(\cdot)$ is the Sinc function.

3.5 Azimuth processing

The azimuth compressing term $H_{a,i}(f_a;r)$ and the extra phase compensation factor $H_{e,i}(f_a;r)$ are respectively

$$H_{a,i}(f_a;r) = \exp\left(-j\left(\varphi_{az,i}(f_a;r)\right)\right) \tag{34}$$

$$H_{e,i}(f_a;r) = exp\left(j\frac{\pi(a_i-1)Km_{i,r_{ref}}\Delta\tau^2}{a_i}\right)\exp\left(+j\frac{\pi(a_i-1)Ks_{i,r_{ref}}\Delta\tau^3}{3a_i}\right) \tag{35}$$

After compensation, coherent superposition in the azimuth frequency domain is performed on each receiver, and then the azimuth IFFT is performed to obtain the focused SAS image.

4 Algorithm validation

4.1 Simulation experiment

To verify the effectiveness of the algorithm proposed in this research, simulations were conducted on ideal point targets at different ranges. The positions of 10 ideal point targets are shown in Figure 8, where $P_1 \sim P_5$ are close-range targets and $P_6 \sim P_{10}$ are long-range targets. Computer CPU is Intel i7-10700@2.90G, RAM is 32 GB, Matlab version R2020a.

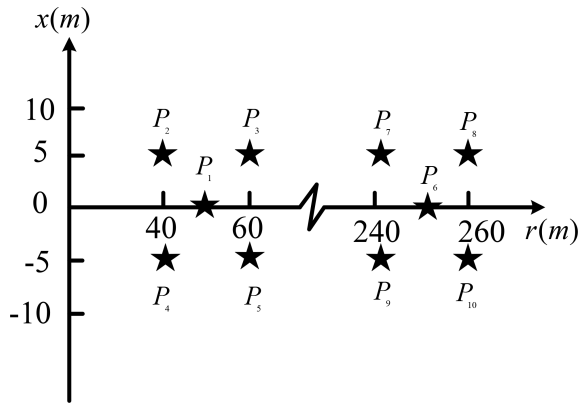


FIGURE 8
Simulation point targets setting.

In the simulation experiment, it is assumed that the distribution of the transmitter and receiver is shown in Figure 1, and the parameters are shown in Table 1. The wide-beam imaging algorithms proposed in the RDA-MSR (Zhang et al., 2014), the RDA-MRCMC (Zhang, 2014), and this research were used to image point targets shown in Figure 8. The imaging results obtained are shown in Figure 9, where Figures 9A, B are the results of the RDA-MSR; Figures 9C, D are the results of the RDA-MRCMC; Figures 9E, F are the results of the proposed method in this research. It can be seen that the RDA-MRCMC has the worst imaging performance. This is because transforming the multi-receiver SAS into a mono-static SAS model, although the imaging process was simplified, there were significant errors.

To quantitatively compare the effectiveness of different methods, we take the range and azimuth slices of point targets P_1 and P_6 as shown in Figures 10, 11, respectively, with amplitude units in dB.

The blue dashed line represents the RDA-MSR, the green dashed line represents the RDA-MRCMC, and the red solid line represents the method in this research. The impulse response width (IRW), peak side lobe ratio (PSLR), and integral side lobe ratio (ISLR) of the range slice and azimuth slice were measured, and the results are shown in Table 2.

From Table 2, it can be seen that the imaging effect of the proposed method is similar to that of the RDA-MSR, but the interpolation operation in this method is not efficient enough. We calculated the time required for imaging the point target echo signal in the scene shown in Figure 8, as shown in Table 3, our method avoids interpolation and saves about half of the time of the RDA-MSR and demonstrates the advantages of computation cost.

4.2 Field test

To further validate the effectiveness of our method, imaging was performed on the data obtained from a sea trial of ChinSAS in 2017 (Zhang et al., 2012). The parameters of ChinSAS-150 are as follows: carrier frequency is 75kHz, transmitter's length is 0.16m, receiver's length is 0.08m, the signal bandwidth is 20kHz, the total number of receivers participating in imaging is 37, SAS platform speed is 2.5m/s, and size of imaging block is 40m(azimuth)×50m(range). Based on the comprehensive analysis of the above parameters, the system operates in a narrow-beam scenario. Comparing Figures 12A–C, it is not difficult to find that the imaging results of all methods are almost identical, but the proposed method is faster than the RDA-MSR. This demonstrates the effectiveness of our method in practical applications. Due to the lack of publicly available field data on wide-beam multi-receiver SAS in China, the advantages of this method in wide-beam imaging still need further verification. We will next carry out the development of low-frequency wide-beam SAS and verify its practicality with the method proposed in this research as soon as possible.

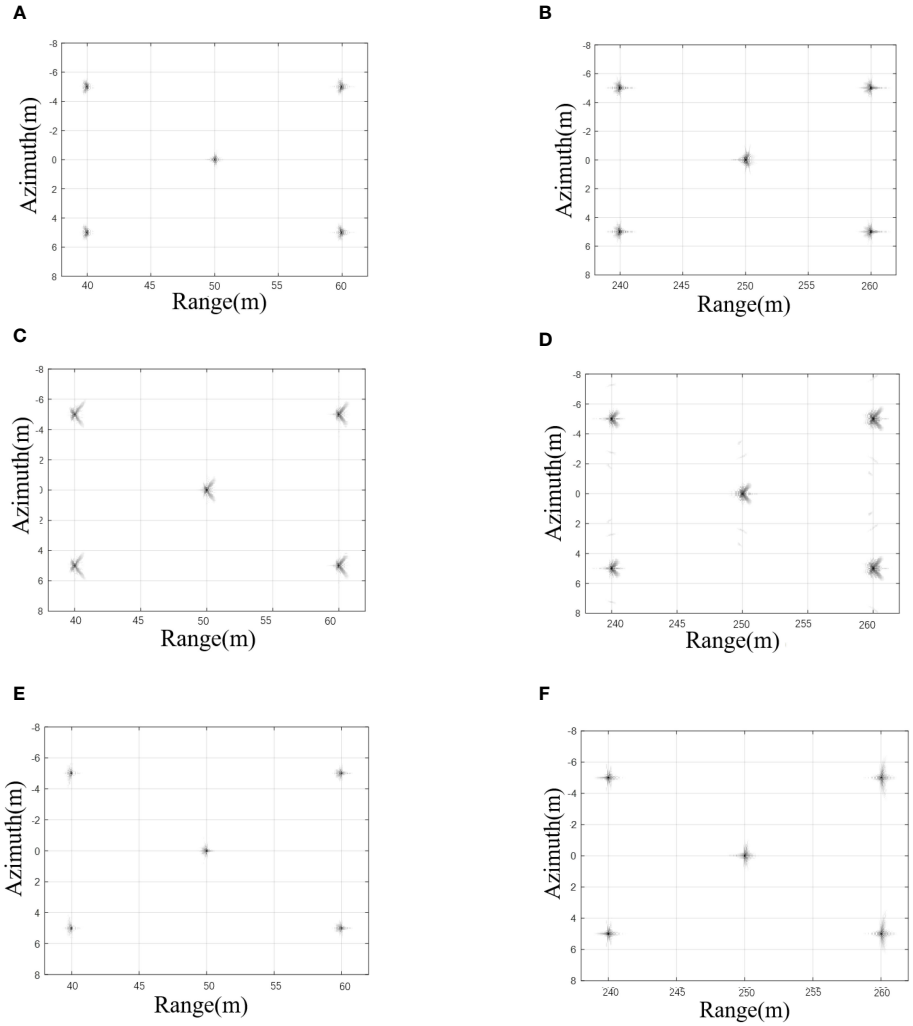


FIGURE 9
Imaging results of different methods **(A)** $P_1 \sim P_5$ (RDA-MSR) **(B)** $P_6 \sim P_{10}$ (RDA-MSR) **(C)** $P_1 \sim P_5$ (RDA-MRCMC) **(D)** $P_6 \sim P_{10}$ (the RDA-MRCMC) **(E)** $P_1 \sim P_5$ (the proposed method) **(F)** $P_6 \sim P_{10}$ (the proposed method).

We recorded the operation time of the field data imaging under different methods, as shown in Table 4. It can be seen from the Table 4 that the proposed method takes two-thirds of the time required for RDA-MSR. Although the proposed method takes longer than RDA-MRCMC, this method will have better focusing result under wide beam conditions, so it is a compromise between computation load and imaging quality.

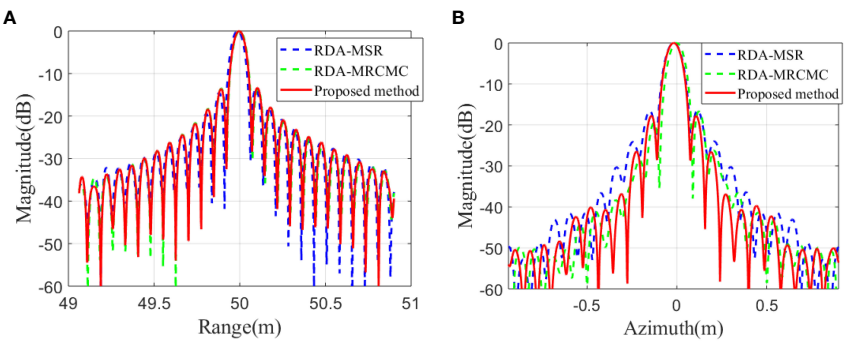


FIGURE 10
Slices of P_1 **(A)** Range slice **(B)** Azimuth slice.

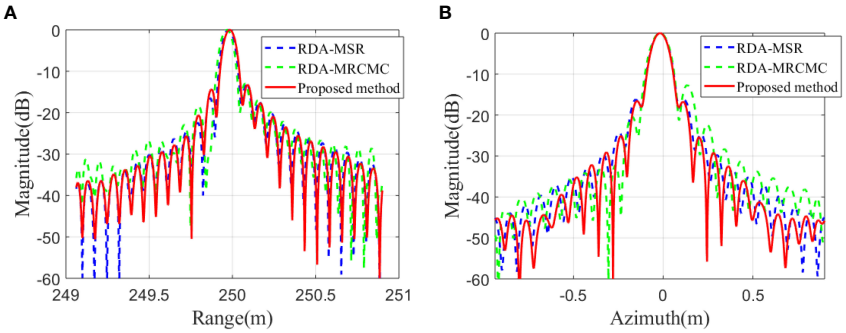


FIGURE 11
Slices of P_6 **(A)** Range slice **(B)** Azimuth slice.

TABLE 2 Quality parameters of different methods.

Methods	Indicators	P1		P6	
		Range	Azimuth	Range	Azimuth
The accurate RDA	PSLR(dB)	-13.11	-16.47	-12.82	-16.20
	ISLR(dB)	**-10.66**	-14.80	-11.04	-14.85
	IRW(cm)	7.79	8.32	7.60	9.06
The rough RDA	PSLR(dB)	-13.09	-16.66	-12.97	-12.69
	ISLR(dB)	-10.41	-16.99	**-11.15**	-13.69
	IRW(cm)	7. 76	7.89	7.61	9.80
The proposed method	PSLR(dB)	**-13.15**	**-17.76**	**-13.25**	**-16.58**
	ISLR(dB)	-10.51	**-17.05**	-10.94	**-15.62**
	IRW(cm)	**7.75**	**7.89**	**7.54**	**8.74**

Bold values represent the optimal results.

TABLE 3 Time cost of different methods under simulation.

Methods	The RDA-MSR	The RDA-MRCMC	The proposed method
Time(s)	302.16	109.75	167.78

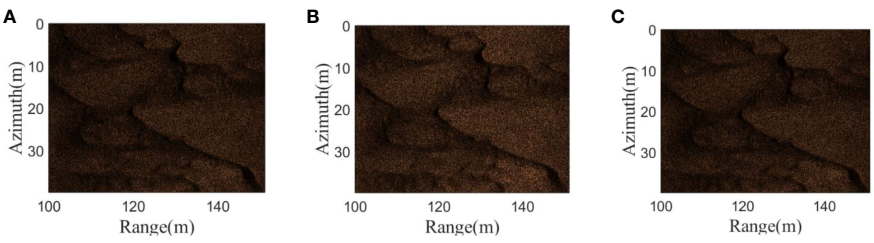


FIGURE 12
Comparison of results from different methods when processing field data **(A)** the RDA-MSR **(B)** the RDA-MRCMC **(C)** the proposed method.

TABLE 4 Time cost of different methods under field data.

Methods	The RDA-MSR	The RDA-MRCMC	The proposed method
Time(s)	155.95	56.55	92.76

5 Conclusion

This research proposes a NCSA for multi-receiver SAS based on azimuth spectrum superposition, which adopts the more accurate PTRS based on LIT and the NCS algorithm to solve the problem of poor imaging quality of existing wide-beam multi-receiver SAS. The algorithm provided in this study provides theoretical support for the future development of low-frequency wide-beam SAS.

Data availability statement

The code used in this study has been deposited in a publicly accessible repository, which can be found here: https://gitee.com/freepoet/ncsa.

Author contributions

MN wrote the program and the original manuscript. HZ provided fundings and field test data to support this research. HL translated and polished this paper. MM and LD provided valuable insights for this research and result analysis. JT proposed this problem and the methodology. All authors contributed to this paper and approved the submitted manuscript.

Funding

This research was supported in part by the National Natural Science Foundation of China under Grant 42176187 and Grant 62301592.

Acknowledgments

Special thanks to the fundings that provided support for this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bellettini, A., and Pinto, M. A. (2002). Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna. *IEEE J. Oceanic Eng.* 27 (4), 780–789. doi: 10.1109/JOE.2002.805096

Bonifant, W. W. (1999). *Interferometic synthetic aperture sonar processing* (Atlanta, Georgia: George Institute of Technology). Master.

Callow, H. J. (2003). *Signal Processing for Synthetic Aperture Sonar Image Enhancement* (Christchurch, New Zealand: Electrical and Electronic Engineering, University of Canterbury). Ph.D.

Duan, J., Huang, Y., and Liu, J. (2017). Application of FFBP algorithm to SAS imaging in non-linear trajectory with non-uniform speeds. *J. Appl. Acoustics* 36 (1), 6.

Giardina, P. E. (2012). *Interferometric Synthetic Aperture Sonar Signal Processing for Autonomous Underwater Vehicles Operating in Shallow Water* (New Orleans, Louisiana: University of New Orleans).

Gough, P., Hayes, M., and Wilkinson, D. (2004). An efficient image reconstruction algorithm for a multiple hydrophone array synthetic aperture sonar.

Huang, P., and Yang, P. (2022). Synthetic aperture imagery for high-resolution imaging sonar. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1049761

Jiang, X., Sun, C., and Feng, J. (2004). "A novel image reconstruction algorithm for synthetic aperture sonar with single transmitter and multiple-receiver configuration," in *Oceans '04 MTS/IEEE Techno-Ocean '04*, 9–12. Kobe, Japan: IEEE. Cat. No.04CH37600.

Li, C., Zhang, H., and Deng, Y. (2021). Focus improvement of airborne high-Squint bistatic SAR data using modified azimuth NLCS algorithm based on lagrange inversion theorem. *Remote Sens.* 13 (10), 19165. doi: 10.3390/rs13101916

Li, Y., Zhang, T., Mei, H., Quan, Y., and Xing, M. (2022). Focusing translational-variant bistatic forward- looking SAR data using the modified omega-K algorithm. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16. doi: 10.1109/TGRS.2021.3063780

Liao, Y., and Liu, Q. H. (2017). Modified chirp scaling algorithm for circular trace scanning synthetic aperture radar. *IEEE Trans. Geosci. Remote Sens.* 55 (12), 7081–7091. doi: 10.1109/TGRS.2017.2740063

Liu, W., Zhang, C., and Liu, J. (2009). Application of FFBP algorithm to synthetic aperture sonar imaging. *Tech. Acoustics* 28 (5), 572–576.

Ma, M., Tang, J., Wu, H., Zhang, P., and Ning, M. (2023). CZT algorithm for the doppler scale signal model of multireceiver SAS based on shear theorem. *IEEE Trans. Geosci. Remote Sens.* 61, 1–12. doi: 10.1109/TGRS.2023.3234572

Ma, M., Tang, J., and Zhong, H. (2020). CZT algorithm for multiple-receiver synthetic aperture sonar. *IEEE Access* 8, 1902–1909. doi: 10.1109/ACCESS.2019.2962314

Marx, D., Nelson, M., Chang, E., Gillespie, W., Putney, A., and Warman, K. (2000). "An introduction to synthetic aperture sonar," in *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing (Cat. No.00TH8496)*, Pittsburgh, PA, USA: IEEE. 16–16.

Neo, Y. L., Wong, F., and Cumming, I. G. (2007). A two-dimensional spectrum for bistatic SAR processing using series reversion. *IEEE Geosci. Remote Sens. Lett.* 4 (1), 93–96. doi: 10.1109/LGRS.2006.885862

Neo, Y. L., Wong, F. H., and Cumming, I. G. (2008). Processing of azimuth-invariant bistatic SAR data using the range doppler algorithm. *IEEE Trans. Geosci. Remote Sens.* 46 (1), 14–21. doi: 10.1109/TGRS.2007.909090

Ning, M., Zhong, H., Wu, H., Zhang, J., and Tang, J. (2023). "Range doppler algorithm for wide-beam multi-receiver synthetic aperture sonar considering differential range curvature," in *2023 4th International Conference on Computer Engineering and Application (ICCEA)*, Hangzhou, China: IEEE, 7-9 April 2023.

Qian, Y., Kuang, H., Zhang, Y., and Zhang, Y. (2021). "Modified generalized omega-K algorithm for low earth orbit high resolution spotlight spaceborne SAR focusing," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium: IEEE, 11-16 July 2021.

Raney, R. K., Runge, H., Bamler, R., Cumming, I. G., and Wong, F. H. (1994). Precision SAR processing using chirp scaling. *IEEE Trans. Geosci. Remote Sens.* 32 (4), 786–799. doi: 10.1109/36.298008

Synnes, S. A. V., Hunter, A. J., Hansen, R. E., Sæbø, T. O., Callow, H. J., van Vossen, R., et al. (2017). Wideband synthetic aperture sonar backprojection with maximization of wave number domain support. *IEEE J. Oceanic Eng.* 42 (4), 880–891. doi: 10.1109/JOE.2016.2614717

Tan, C., Zhang, X., Yang, P., and Sun, M. (2019). A novel sub-bottom profiler and signal processor. *Sensors (Basel Switzerland)* 19, 5052. doi: 10.3390/s19225052

Tian, Z., Tang, J., Zhong, H., and Zhang, S. (2016). Extended range doppler algorithm for multiple-receiver synthetic aperture sonar based on exact analytical two-dimensional spectrum. *IEEE J. Oceanic Eng.* 41 (1), 164–174. doi: 10.1109/JOE.2015.2402053

Tian, Z., Zhong, H., Tang, J., and Zhang, J. (2022). Azimuth-invariant motion compensation and imaging chirp scaling algorithm for multiple-receiver synthetic aperture sonar. *IEEE Access* 10, 114060–114076. doi: 10.1109/ACCESS.2022.3218279

Vu, V. T., Sjögren, T. K., and Pettersson, M. I. (2014). Two-dimensional spectrum for biSAR derivation based on lagrange inversion theorem. *IEEE Geosci. Remote Sens. Lett.* 11 (7), 1210–1214. doi: 10.1109/LGRS.2013.2289735

Wang, R., Loffeld, O., Nies, H., Knedlik, S., and Ender, J. H. G. (2009). Chirp-scaling algorithm for bistatic SAR data in the constant-offset configuration. *IEEE Trans. Geosci. Remote Sens.* 47 (3), 952–964. doi: 10.1109/TGRS.2008.2006275

Wang, X., Zhang, X., and Zhu, S. (2015). "Upsampling based back projection imaging algorithm for multi-receiver synthetic aperture sonar," in *2015 International Industrial Informatics and Computer Engineering Conference*. Xi'an, Shaanxi, China: ATLANTIS PRESS.

Wilkinson, D. R. (2001). *Efficient Image Reconstruction Techniques for a Multiple-Receiver Synthetic Aperture Sonar* (Christchurch, New Zealand: Electrical and Electronic Engineering, University of Canterbury). Master.

Wu, H., Tang, J., and Zhong, H. (2019). Moderate squint imaging algorithm for the multiple-hydrophone SAS with receiving hydrophone dependence. *IET Radar Sonar Navigation* 13(1), 139-147. doi: 10.1049/iet-rsn.2018.5055

Wu, J., Pu, W., Huang, Y., Yang, J., and Yang, H. (2016). "An Omega-K algorithm for bistatic forward-looking SAR based on spectrum modeling and optimization," in *2016 CIE International Conference on Radar (RADAR)*, Guangzhou, China: IEEE, 10-13 Oct. 2016.

Xiong, T., Xing, M., Wang, Y., Guo, R., Sheng, J., and Bao, Z. (2011). Using derivatives of an implicit function to obtain the stationary phase of the two-dimensional spectrum for bistatic SAR imaging. *IEEE Geosci. Remote Sens. Lett.* 8 (6), 1165–1169. doi: 10.1109/LGRS.2011.2159090

Xu, J., Tang, J., and Zhang, C. (2003). Multi-aperture synthetic aperture sonar lmaging algorithm. *Signal Process.* 19 (2), 4. doi: 10.3969/j.issn.1003-0530.2003.02.016

Yang, P., and Liu, J. (2022). "Effect of non-unifrom sampling on sonar focusing," in *2022 14th International Conference on Communication Software and Networks (ICCSN)*, Chongqing, China: IEEE, 10-12 June 2022.

Zhang, X. (2014). Study on multi-receiver synthetic aperture sonar imagery and motion compensation algorithm. 36 (7).

Zhang, X., Dai, X., and Yang, Bo. (2018). Fast imaging algorithm for the multiple receiver synthetic aperture sonars. *IET Radar Sonar Navigation* 6, 74303-74319. doi: 10.1049/iet-rsn.2018.5040

Zhang, X., Huang, H., Ying, W., Wang, H., and Xiao, J. (2017). An indirect range-Doppler algorithm for multireceiver synthetic aperture sonar based on lagrange inversion theorem. *IEEE Trans. Geosci. Remote Sens.* 55 (6), 3572–35875. doi: 10.1109/TGRS.2017.2676339

Zhang, T., Li, Y., Song, X., Zhang, T., Wu, C., and Sun, Z. (2021). "A modified omega-K algorithm for bistatic forward-looking SAR data imaging," in *2021 CIE International Conference on Radar (Radar)*, Haikou, Hainan: IEEE, 15-19 Dec. 2021.

Zhang, S., Sheng, J., and Xing, M. (2018). A novel focus approach for squint mode multi-channel in azimuth high-resolution and wide-swath SAR imaging processing. *IEEE Access* 6, 74303–74319. doi: 10.1109/ACCESS.2018.2873739

Zhang, X., and Tan, C. (2018). "A comparison of PCA based imaging methods for the multireceiver SAS," in *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Chongqing, China: IEEE, 8-11 Oct. 2018.

Zhang, S., Tang, J., Chen, M., and Bai, S. (2012). Development and sea trial of interferometric synthetic aperture sonar. *Tech. Acoustics* 31 (2), 7. doi: 10.3969/j.issn1000-3630.2012.02.010

Zhang, X., Tang, J., Zhang, S., Bai, S., and Zhong, H. (2014). Four-order polynomial based range-Doppler algorithm for multi-receiver synhetic aperture sonar. *J. Electron. Inf. Technol.*, 36(7). doi: 10.3724/SP.J.1146.2013.01317

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021a). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., and Yang, P. (2019). Imaging algorithm for multireceiver synthetic aperture sonar. *J. Electrical Eng. Technol.* 14 (1), 471–4785. doi: 10.1007/s42835-018-00046-0

Zhang, X., and Yang, P. (2022). Back projection algorithm for multi-Receiver synthetic aperture sonar based on two interpolators. *J. Mar. Sci. Eng.* 10 (6), 7185. doi: 10.3390/jmse10060718

Zhang, X., Yang, P., and Dai, X. (2019). Focusing multireceiver SAS data based on the fourth-Order legendre expansion. *Circuits systems Signal Process.* 38 (6), 2607–26295. doi: 10.1007/s00034-018-0982-6

Zhang, X., Yang, P., Feng, X., and Sun, H. (2022). Efficient imaging method for multireceiver SAS. *IET Radar Sonar Navigation* 16 (9), 1470-1483. doi: 10.1049/rsn2.12274

Zhang, X., Yang, P., Huang, P., Sun, H., and Ying, W. (2021b). Wide-bandwidth signal-based multireceiver SAS imagery using extended chirp scaling algorithm. *IET Radar Sonar Navigation* 16, 531–541. doi: 10.1049/rsn2.12200

Zhang, X., Yang, P., and Miao, S. (2022a). Experiment results of a novel sub-bottom profiler using synthetic aperture technique. *Curr. Sci.* 122, 461–464. doi: 10.18520/cs/v122/i4/461-464

Zhang, X., Yang, P., and Sun, H. (2022b). Frequency-domain multireceiver synthetic aperture sonar imagery with Chebyshev polynomials. *Electron. Lett.* 58 (25), 995-998. doi: 10.1049/ell2.12513

Zhang, X., Yang, P., and Sun, H. (2023). An omega-k algorithm for multireceiver synthetic aperture sonar. *Electron. Lett.* 59 (13). doi: 10.1049/ell2.12859

Zhang, X., Yang, P., and Zhou, M. (2023). Multireceiver SAS imagery with generalized PCA. *IEEE Geosci. Remote Sens. Lett.* 20, 1–5. doi: 10.1109/LGRS.2023.3286180

Zhang, X., and Ying, W. (2022). Influence of the element beam pattern on synthetic aperture sonar imaging. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics Inf. Sci. Wuhan Univ.* 47, 133–140. doi: 10.13203/j.whugis20190148

Zhu, J., Song, Y., Jiang, N., Xie, Z., Fan, C., and Huang, X. (2023). Enhanced doppler resolution and sidelobe suppression performance for golay complementary waveforms. *Remote Sens.* 15 (9), 2452. doi: 10.3390/rs15092452

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 06 November 2023
DOI 10.3389/fmars.2023.1259864

OPEN ACCESS

EDITED BY
Arata Kaneko,
Hiroshima University, Japan

REVIEWED BY
Hong Zheng,
Nanjing University of Information Science and Technology, China
Hiroyuki Matsumoto,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Ju Lin,
Ocean University of China, China
Ze-Nan Zhu,
Ministry of Natural Resources, China

*CORRESPONDENCE
Da Lianglong
liangld@126.com

RECEIVED 17 July 2023
ACCEPTED 16 October 2023
PUBLISHED 06 November 2023

CITATION
Baolong C, Jingyi L, Wuhong G and Lianglong D (2023) Enhancing ocean environment prediction in Yellow Sea through targeted observation using ocean acoustic tomography.
Front. Mar. Sci. 10:1259864.
doi: 10.3389/fmars.2023.1259864

Enhancing ocean environment prediction in Yellow Sea through targeted observation using ocean acoustic tomography

Cui Baolong[1,2], Liu Jingyi[1,2], Guo Wuhong[1,2] and Da Lianglong[1,2]*

[1]Naval Submarine Academy, Qingdao, China, [2]Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Ocean Acoustic Tomography (OAT) is an efficient and economical marine acoustic observation technique. Targeted observation is an appealing procedure to reduce the uncertainty of ocean environment prediction through additional observation. This study aimed to assess the validity of OAT as an observation method for targeted observation. OAT based on Niche Genetic Algorithm was employed to extract sound speed and temperature profiles from acoustic transmission time, utilizing data from the 2019 Yellow Sea experiment. The inversion results were compared with measurement data, which are found to be accurate and reliable. To further evaluate OAT as targeted observation method, the vertical bias structure of OAT was added on synchronous measurement data in the sensitive area of targeted observation to simulate OAT observation in sensitive area. This simulated data was then incorporated into a 3D-Var assimilation system to improve the short-term prediction of the target region. Comparing the predictions derived with the measurement data at the verification time, it shows that the simulated OAT observation improved the quality of target region prediction, indicating that OAT can be an effective observation method for targeted observation. An Observing System Simulation Experiment was conducted to assess the impact of OAT characteristics on prediction improvement. The results show that both adding observation nodes and extending the observation duration have positive effects, while extending the observation duration performs better.

KEYWORDS

ocean environment prediction, targeted observation, ocean acoustic tomography, niche genetic algorithm, observation system simulation experiment

1 Introduction

Ocean acoustic tomography (OAT) is a marine remote sensing technique by utilizing the sound field generated from measured properties (Worcester, 2019). This method extracts acoustic characteristics, such as Sound Speed Profile (SSP), through the analysis of travel time or other acoustic signals. Corresponding marine environment characteristics is inverted through the ocean-acoustic coupled relationship. The concept of OAT was initially proposed by Munk and Wunsch (Munk and Wunsch, 1979; Munk et al, 1995), aiming to investigate mesoscale phenomena such as vortices, convection, and internal waves. An advantage of OAT over other methods is its ability to facilitate long-term, large-area, and cost-effective ocean monitoring, taking advantage of the characteristics that acoustic signals transmit over long distances and acoustic propagation is sensitive to the marine environment (Dushaw et al, 2001).

Numerous OAT experiments have been conducted since the 1980s, showcasing the versatility and potential of this technique. RTE83 experiment (DeFerrari and Nguyen, 1986; Howe, 1987) validated the feasibility of flow velocity inversion using a single source-receiver pair of acoustic nodes in a range-dependent environment, achieving success at a distance of 300km in Atlantic Ocean. In 1988-1989, Greenland Sea Tomography experiment (Jin et al, 1993) became the first to employ mobile nodes to estimate the effect of sea ice on acoustic pulses. SLICE89 experiment (Howe et al, 1991) combined acoustic tomography with ocean models to enable ocean forecasting on a scale of 1000-2000km. AMODE experiment (Dushaw et al, 2001) conducted in 1991 utilized mobile nodes to measure eddy currents in Northwest Atlantic. Acoustic Thermometry of Ocean Climate (ATOC) experiment, organized by International Ocean Research Association (Dushaw and Worester, 2001; Dushaw, 1999), stands out as a remarkable achievement. This experiment incorporated vertical line arrays, submarine receiving arrays, and US Army's SOSUS system to receive low-frequency acoustic signals propagating over basin distances. Its purpose was to monitor long-term temperature changes and global warming as indicators of climate trends. In 2001, ASIAEX experiment (Duda et al, 2004), conducted in collaboration with various countries and organizations, focused on the seas surrounding China. Its primary objective was to investigate the interaction mechanism between the acoustic field and the water bodies. It is shown that the mutual correlation function and Green's function of marine environmental noise have the similarity of arrival time structure, based on which some scholars proposed the idea of using marine environmental noise for passive acoustic tomography, and the idea was realised by experimental observation (Gasparini et al, 1997; Fried et al, 2013; Li et al, 2019). The presence of mesoscale processes such as ocean fronts/vortices in the oceans has led to the development of acoustic tomography for horizontally varying environmental (Carrière and Hermand, 2008; Yang et al, 2022);. In recent years, coastal acoustic tomography technology has made significant progress, particularly in monitoring semi-enclosed environments such as ports and bays (Yamoaka et al, 2002; Zhu et al, 2010; Zhu et al, 2013). Additionally, acoustic tomography has been applied to observe mesoscale phenomena such as internal waves (Lynch et al, 1996; Dahl et al, 2004; Li et al, 2014) and Kuroshio current (Yuan et al, 1999; Lebedev et al, 2003; Huang et al, 2013; Taniguchi et al, 2023). Currently, experimental research primarily emphasizes coastal velocity inversion, with limited studies focusing on marine environment inversion, especially in the context of oceanic environment prediction.

The essence of acoustic tomography lies in the recognition of acoustic signal propagation time and structure. Several mainstream methods are commonly used in this field, including ray travel time tomography (Munk et al, 1995), matched-peak tomography (Skarsoulis et al, 1996), modal travel time tomography (Shang, 1989), modal-phase tomography, and modal-horizontal-refraction tomography (Shang et al, 2000). Ray travel time tomography, being the most classic and widely used method, employs matching filters to measure the travel time. Matched-peak tomography locates the maximum peak value in the arrival pattern and analyzes the peak structure of the signal to determine the travel time accurately. Modal travel time tomography, on the other hand, relies on the principles of normal mode theory to identify the arrival time. Normal wave phase tomography and horizontal refraction tomography, which are similar, replace the normal wave travel time with the normal wave phase or horizontal refraction angle. These substitutions are then substituted into algorithms to obtain the desired travel time information. Regarding the acquisition algorithms of travel time, two common approaches are utilized: the perturbation method (Munk et al, 1995) and the matching field method (Taroudakis and Markaki, 1997). The perturbation method assumes that the difference between the theoretical calculation and measured propagation delays is proportional to the difference in sound velocity. However, this method tends to be less accurate in complex and non-linear marine environments. In contrast, the matching field method aims to obtain the optimal solution that corresponds to the measured values through acoustic and marine models. The effectiveness of this method relies on the accuracy of the model and the efficiency of the optimization algorithm employed.

Targeted observation, also known as adaptive observation, is a strategy approach aimed at reducing numerical prediction uncertainty through employing additional observations. In this strategy, the goal is to improve the prediction quality of a specific area, referred to as the target region, at a designated verification time. To achieve this, additional observations are deployed within sensitive areas to acquire additional information. This additional information is subsequently assimilated into the ocean model to refine Initial Conditions (ICs) and improve the prediction accuracy (Rabier et al, 1996; Rabier et al, 1996; Snyder, 1996; Mu, 2013). Originally introduced in atmospheric studies, targeted observation has undergone validation through a series of field experiments such as FASTEX (Joly et al, 1999), NOPREX (Langland et al, 1999), and WSRP (Szunyogh et al, 2000). Recognizing its potential, World Meteorological Organization (WMO) proposed The Observing System Research and Predictability Experiment (THORPEX), which integrated targeted observation concepts into a scientific framework for improving global high-impact weather prediction (Parsons et al, 2017). More recently, the concept of targeted observation has been extended to oceanic prediction studies,

although the focus has primarily been on large-scale ocean phenomena, such as Indian Ocean Dipole (Feng et al, 2016) and Kuroshio (Kramer et al, 2012; Wang et al, 2013; Zhang et al, 2017). However, there remains a scarcity of researches related to the acoustic field within targeted observation studies.

The identification of sensitive areas, a crucial aspect of targeted observation, relies on two types of algorithms. The first type is based on ensemble prediction techniques, such as Ensemble Kalman Filter (EnKF) (Hamill and Snyder, 2002) and Ensemble Transform Kalman Filter (ETKF) (Bishop et al, 2001). These algorithms specifically focus on calculating the reduction in forecast error covariance resulting from different observation configurations (Wei et al, 2008; Zhang et al, 2015; Feng et al, 2019; Thiruvengadam et al, 2021). The second type of algorithm is based on adjoint mode techniques, which include approaches such as Singular Vectors (SV) (Buizza and Montani, 1999), adjoint sensitivity (Baker and Daley, 2000), and Conditional Nonlinear Optimal Perturbation (CNOP) (Mu et al, 2003). CNOP extends the concept of SV to nonlinear systems, focusing on identifying the initial perturbation that exhibits most rapid growth in the forecast. Targeted observation based on CNOP has demonstrated its wide applicability in high-impact weather events prediction and air-sea coupling events prediction (Dushaw et al, 2001; Duan and Hu, 2015; Duan and Mu, 2018; Chan et al., 2022; Liu et al, 2023).

A field experiment was conducted at Yellow Sea of China in August 2019, comprising two main components: an OAT experiment and a targeted observation experiment. In this study, OAT experiment data served as the foundation for validating the effectiveness of OAT in accurately inverting the vertical speed and temperature structure using Niche Genetic Algorithm (NGA). To simulate the OAT observation for targeted observation, the bias structure was extracted and incorporated into the measurements within the sensitive area of targeted observation. Subsequently, the simulated observations were integrated into a 3D-Var assimilation model to improve the short term (7 days) prediction accuracy of the target region. Thus, considering the large-area coverage and long-term observation capabilities characteristics of OAT, an Observing System Simulation Experiments (OSSE) was deployed to investigate the impact of increasing the observation area and extending the observation time on the prediction quality.

2 Materials and methods

2.1 Ocean model

Regional Ocean Modeling System (ROMS), specifically the Rutgers version, is employed in this study to simulate the thermocline distribution and circulation structure of Yellow Sea. The ROMS model is an open-source ocean model based on 3D non-linear oblique pressure equations employing techniques as split-explicit, free-surface, topography following-coordinate (Shchepetkin and McWilliams, 2005). The model domain covers geographical extent from 23.7°N to 41.3°N and 117°E to 132.5°E, with a horizontal resolution of 1/24° and 32 vertical levels. To initiate the model, a cold start is performed, and the integration is carried out for 25 model years. Topography data of the model domain is from ETOPO2 dataset. The initial temperature and salinity data are derived from HYCOM+NCODA multiyear averaged (1998-2018) reanalysis data. Initial current velocities and sea surface height are set to zero. Surface forcing factors, including wind stress, heat flux, and water exchange, are obtained from multiyear averaged (1998–2018) ECMWF Re-Analysis-interim data. For the open boundaries, the forcing condition of the model is driven from the multiyear averaged monthly HYCOM + NCODA reanalysis data. Further details on the model setup and validation can be found in references Hu et al. (2021) and Liu et al. (2021).

In addition to the climatology run, a hindcast run is conducted based on the results obtained. For the analysis presented in this study, daily-averaged temperature profile data from the hindcast run are utilized.

2.2 Acoustic tomography algorithm

Due to the nonlinearities of ocean and acoustic models, the parsing solution of SSP may not be feasible. Therefore, SSP solution requires the implementation of a suitable searching algorithm. In this study, NGA (Malfoud, 1995) is employed as an effective approach to obtain optimal search speed and prevent premature convergence.

NGA adopts a crossover algorithm that aims to reduce the uncertainty of individual offspring while maintaining diverse populations. Parents and offspring are preserved and compete with each other, leading to increased selection pressure. The fundamental concept is to calculate Hamming distance between every two individuals. If Hamming distance is below a specific threshold, individual with lower fitness level is penalized, making it more likely to be eliminated during the evolutionary process. Consequently, individuals are dispersed in the constrained space at a certain distance, ensuring the diversity of the population is maintained. NGA process can be summarized as follows (Figure 1):

1) Calculate Empirical Orthogonal Function (Shen et al, 1999) and determine the coefficient range based on Ocean-Acoustic Coupling Model (OACM) (Da et al, 2015) and the measured sound velocity profiles, treating them as the sample group;
2) Generate a population of M individuals randomly within the range of EOF coefficients, considering the specified operation precision;
3) Calculate the fitness of each individual as follows:

$$F = \frac{1}{\sum_k [t_k - \tau_k]^2} \quad (1)$$

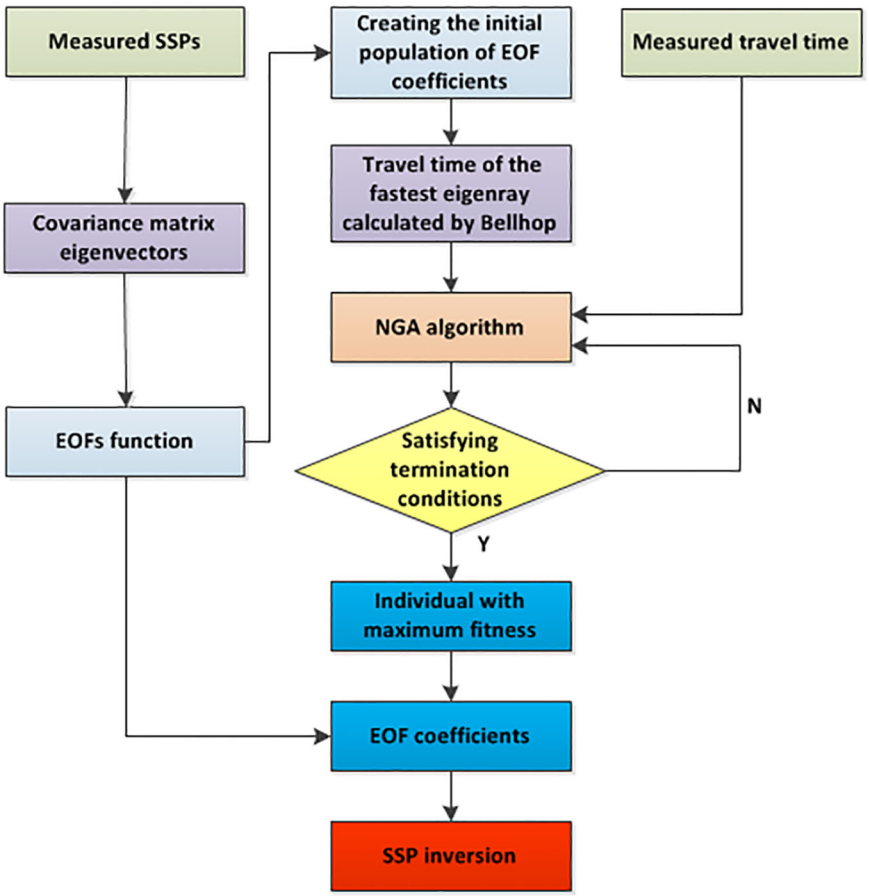


FIGURE 1
Schematic diagram of Niche Genetic Algorithm.

where t_k is the calculated value of the travel time of the fastest characteristic sound rays received by each hydrophone using Bellhop model; τ_k is the travel time measured during the experiment (k=1, 2, 3…, K), and K is the number of the hydrophones;

4) Sort the individuals in descending order according to their fitness F_i , and mark the first N $(N<M)$ individuals;
5) Apply selection, crossover, and mutation operations to the population of M individuals;
6) Execute a niche elimination operation by combine the M individuals obtained in Step 5 with the first N individuals from Step 4, resulting in a new population of M+N individuals. Calculate Hamming distance between each pair of individuals (X_i and X_j) in the new population according to the following equation:

$$\left\|\boldsymbol{X_i} - \boldsymbol{X_j}\right\| = \sqrt{\sum_{k=1}^{K}(\boldsymbol{x_{ik}} - \boldsymbol{x_{jk}})^2} \tag{2}$$

$$\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{M} + \boldsymbol{N} - \boldsymbol{1}$$

$$\boldsymbol{j} = \boldsymbol{i} + \boldsymbol{1}, \ldots, \boldsymbol{M} + \boldsymbol{N}$$

where x_{ik} represents the k-th variable of the i-th individual. When Hamming distance is less than L, the individuals with lower fitness in X_i and X_j are subject to a penalty function to reduce their fitness values;

7) From the population of M + N individuals, select the first M individuals with higher fitness values to generate the new population. If the termination condition has been met, the result is considered as the final output of NGA. Otherwise, repeat Step 3-6.

2.3 Assimilation method

The observation data from targeted observation is incorporated into 3D-Var system to improve the ICs. Numerical simulations of ocean circulation patterns are assimilated with data the ocean

environment observed by a wide range of instruments, guided by the statistical Bayesian conditional probability theory, to produce new numerical results. Such assimilated numerical results for the ocean environment contain both the extrapolation of the thermodynamic equations of ocean dynamics and the observed scenarios of the real state of the ocean environment. The numerical ocean model compensates for the shortcomings of the observations, which are always scattered and relatively sparse, and the observations control the uncertainties brought about by the non-linearities in the ocean dynamics and thermodynamic equations.

3D-Var aims to achieve an optimal state solution by minimizing the cost function. The equation of the cost function is as follows:

$$J[x] = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[L(H(x) - y_0)]^T O^{-1}[L(H(x) - y_0)] \tag{3}$$

where x is the analysis variable; x_b is the background field; y_0 is the observation value; B is the background error covariance; O is the observation error covariance; H is the observation operator; O^{-1} is the inverse matrix of the corresponding matrix; $(x - x_b)^T$ is the transpose of the corresponding matrix; and L is the filter operator. In this study, "Analysis variable" refers to the vertical temperature profile result from assimilation. "Background field" refers to the prediction of temperature profile obtained from the ROMS model. "Observation value" refers to the XBT measured temperature profile. The observation update residuals are collected and spatially filtered by the filtering operator L, and the results are fed back to the grid point where the state x is located. L can be calculated as follows:

$$L_{ij} = \mathbf{\Omega}\frac{(a, b_{ij})}{\sum_{j=1}^{K}\mathbf{\Omega}(a, b_{ij})} \tag{4}$$

$$\mathbf{\Omega}(a, b_{ij}) = \begin{cases} -\frac{1}{4}\left(\frac{b}{a}\right)^5 + \frac{1}{2}\left(\frac{b}{a}\right)^4 + \frac{5}{8}\left(\frac{b}{a}\right)^3 - \frac{5}{3}\left(\frac{b}{a}\right)^2 + 1 & 0 \le b \le a \\ \frac{1}{12}\left(\frac{b}{a}\right)^5 - \frac{1}{2}\left(\frac{b}{a}\right)^4 + \frac{5}{8}\left(\frac{b}{a}\right)^3 + \frac{5}{3}\left(\frac{b}{a}\right)^2 - 5\left(\frac{b}{a}\right) + 4 - \frac{2}{3}\left(\frac{b}{a}\right)^{-1} & a < b \le 2a \\ 0 & b > 2a \end{cases}$$

where a is the characteristic distance of the observation response; b is the distance between the observation point and the model grid point; and K is the total number of observations. Parameter a determines the scale of the multiscale method, and also the reduction ratio of each level of the scale grid to the original pattern grid when the grid varies.

The process of assimilation can be summarized as follows,

1) Observation: quality control of acquired data and production of observation data sets;
2) Assimilation: the observation dataset and model results are fed into the assimilation system module, which performs scale-by-scale 3D variational assimilation after grid transformation.
3) Forecasting: the assimilation results are substituted into the ocean model as initial values to obtain new numerical forecasts.

In this study, the process of 'Observation - Assimilation - Forecasting' is repeated with the number of observation cycles.

3 Experiment

An experiment was conducted in August 2019 on the northwest continental slope of the Yellow Sea with the objective of improving short-term (7-day) thermal structure predictions during the summer season. The experiment consisted of two main components: OAT and targeted observation sections.

3.1 Targeted observation experiment

The targeted observation experiment was conducted from 18th to 25th August with the aim of improving short-term (7-day) thermal structure predictions. The experiment focused on a selected target region, denoted by a red box in Figure 2, which is located near the margin of Yellow Sea Cold Water Mass (YSCWM). In this region, Vertical Thermal Structure (VTS) is influenced by various dynamic processes, as well as complex topography. Consequently, the prediction of VTS in this region is associated with significant uncertainties (Hu et al, 2021). To determine the sensitive areas within the target region, an adjoint-free CNOP algorithm was employed. The identified sensitive areas were found to be oriented northeast to southwest, extending from the northeast towards the target region. These sensitive areas are likely influenced by the southwestward background currents.

In the target region, a total of 5 buoys were deployed to gather data for the experiment. These buoys were composed of temperature loggers and pressure–temperature–conductivity loggers, enabling the collection of temperature profile at a vertical interval of 2m. The sampling interval of loggers is 10 minutes. The collected data from the buoys in the target region were utilized for validation purposes. Furthermore, shipboard temperature, conductivity, and depth measurements were conducted, resulting in 21 temperature profiles measured within the targeted region. In the sensitive region identified through CNOP (green area in Figure 2), eXpendable Bathy Thermographs (XBT) were employed to collect temperature profiles 4 times a day (4:30-7:30, 10:30- 13:30, 16:30-19:30, 22:30-01:30) along predesigned routes (i.e., triangles in Figure 2). The data acquired from XBT in the time-varying sensitive area were then substituted into cycle data assimilation process to refine the prediction of the target region at the 7-th day following XBT deployment (verification time). The refined prediction obtained from this assimilation, as well as the basic prediction, were both compared against the data measured by the buoys in the target region. These comparisons served to verify the effectiveness of the targeted observation approach. The experimental results demonstrated that observations within the identified sensitive area, which aimed at reducing initial errors, led to a more significant improvement in VTS prediction of the target region at the verification time compared to similar actions

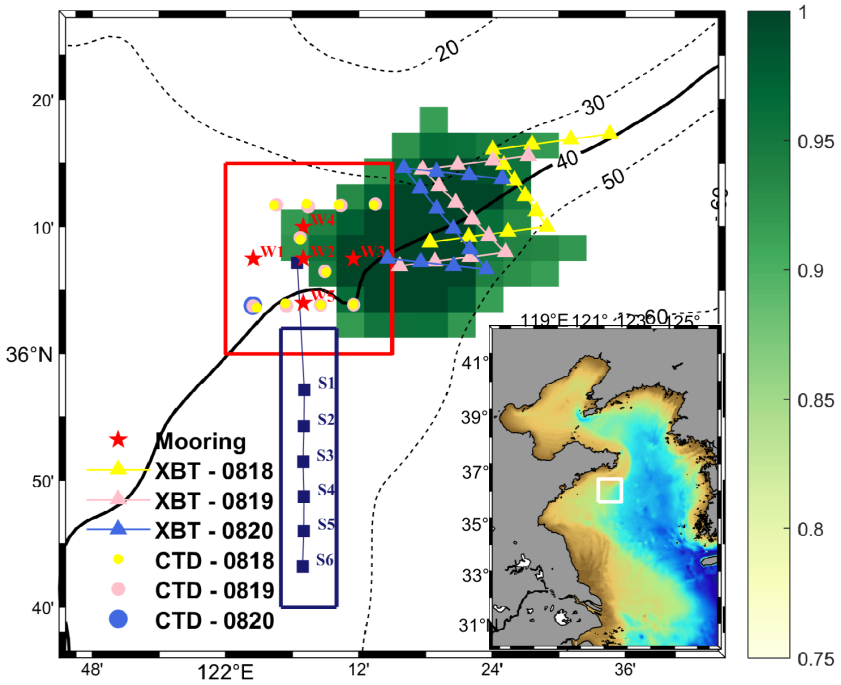


FIGURE 2
Schematic diagram of the experiment. The red open rectangle indicates the location of the target region of targeted observation experiment. The green area is the sensitive area in which the data were obtained on 20 August during CNOP, and which is extended from northeast to southwest towards the target region. The red closed stars indicate five temperature profile buoy stations, carried out during Aug 18-25. The yellow, grey and blue closed triangles locate thirty-six XBT locations, obtained on Aug 18, 19 and 20, respectively, and the yellow, grey and blue closed circles locate twenty-one shipboard CTD stations, obtained at the same date allocation. The violet closed squares, accompanied by a vertical array of S1, S2, S3, S4, S5 and S6, indicates 6 fixed-depth explosive acoustic source locations for OAT experiment on Aug 20. The lower right figure shows the position of the ocean model domain, in which the white open rectangle indicates the position of the experiment.

conducted solely within the target region itself (Hu et al, 2021; Liu et al, 2021). This study focuses on the application of acoustic tomography in targeted observation, and as such, the conclusions of the targeted observation experiment will not be repeated.

3.2 OAT experiment

OAT experiment was designed to validate the feasibility of acoustic tomography for the inversion of the marine environment. The experiment was carried out on August 25 in the southern region of the targeted observation experiment, as denoted by the blue box in Figure 2. NGA algorithm was employed to invert SSP using acoustic travel time data. Subsequently, based on OACM, the corresponding temperature profile was calculated through the inversion of SSP obtained from NGA algorithm.

A launching ship was employed during the experiment to deploy fixed-depth explosive as the acoustic source. The launching ship moved away from the receiving ship and followed a predetermined trajectory from point S1 (approximately 10 nautical miles away from the receiving ship) to point S6 (approximately 22.5 nautical miles). At intervals of 2.5 nautical miles along this trajectory, the launching ship came to a halt and dropped 3 kinds of bombs at controlled depths: 7, 25, and 35m. The depths of explosions and distances between launching and receiving ships are shown in Table 1. To capture the acoustic signals generated by the explosions, a standard hydrophone was fixed at a depth of 10m on the stern of the launching ship. This hydrophone was utilized to record the explosion time and the corresponding source level.

On the receiving ship, a vertical array comprising 15-element hydrophones was deployed on the port aft deck. The hydrophone array spanned depths ranging from 5 to 33m, with a uniform interval of 2m. The receiving ship remained anchored at a fixed position throughout the experiment, enabling the recording of the acoustic signals. The schematic diagram of OAT experiment is shown in Figure 3. Examples of signals received by hydrophones are shown in Figure 4.

Both the launching and receiving vessels were equipped with a multi-channel hydroacoustic signal synchronization acquisition system. This system facilitated the acquisition of the explosion sound source signals from the launching ship and the hydroacoustic signals recorded by the hydrophone array on the receiving ship. Importantly, the embedded a GPS module, enabling the acquisition of precise GPS clock information and position data for real-time synchronization of the explosion sound source signals. This synchronization ensured accurate temporal alignment between the recorded acoustic signals and facilitated reliable analysis of the acoustic data obtained during the experiment.

10.3389/fmars.2023.1259864

TABLE 1 Depths of explosions and distances between launching and receiving ships.

Explosion serial number	Depth of explosion (m)	Location	Range (km)	Average travel time (s)
1	7	S1	18.65	12.3457
2	35		18.68	12.3389
3	25		18.72	12.3983
4	35		18.72	12.4057
5	7	S2	23.95	15.8577
6	25		23.95	15.8925
7	35		23.96	15.8875
8	25	S3	29.14	19.3528
9	35		29.13	19.3142
10	7	S4	34.20	22.7262
11	35		34.20	22.7157
12	7	S5	39.27	27.1309
13	25		39.27	26.0844
14	35		39.27	26.1230
15	7	S6	44.45	29.5747
			44.42	
16	25			29.5574
17	35		44.45	29.5425

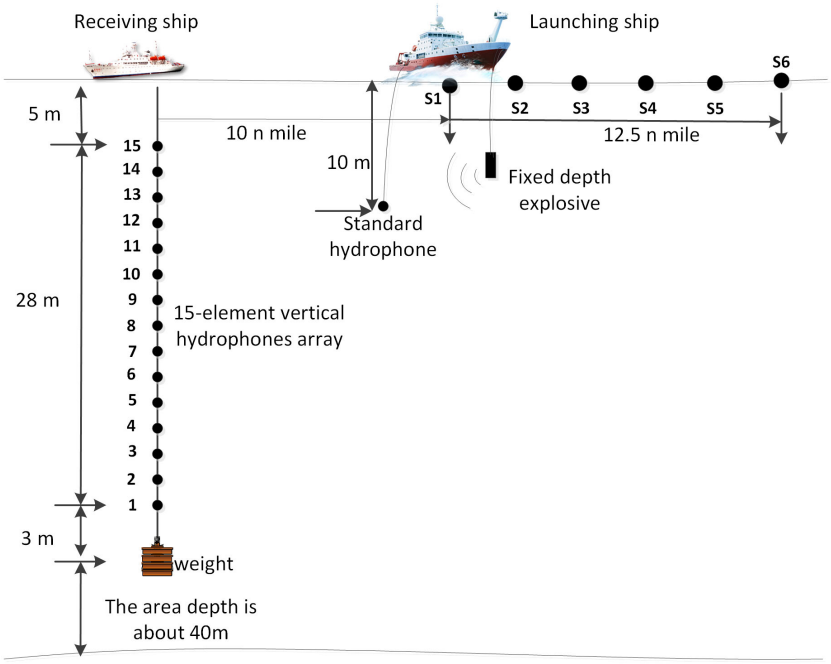


FIGURE 3
Schematic diagram of OAT experiment. Launching ships dropped fixed-depth explosive at specific ranges, while the explosion time were recorded using a hydrophone fixed at the stern. Receiving ship recorded acoustic signals with a 15-element vertical hydrophones array at a fixed location.

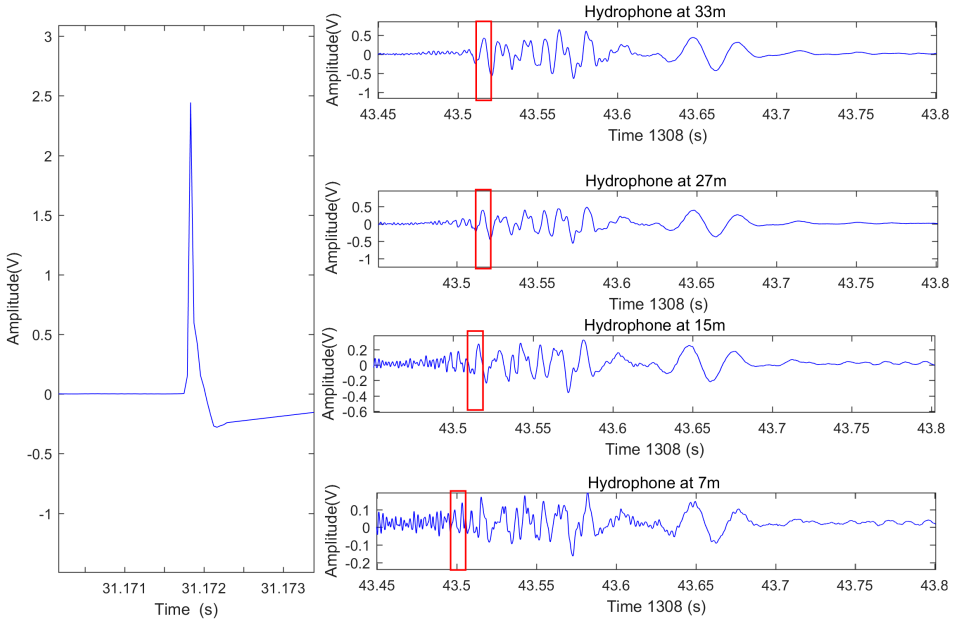


FIGURE 4
Examples of signals received by hydrophones on the launching and receiving ships. The figure on the left is the signal example received by the hydrophone on the launching ship. The figures on the right are the signal examples received by the receiving ship at point S1, 33m, 27m, 15m, and 7m depth hydrophones from top to bottom. The horizontal axis is the time axis, 1308 means 13:08 p.m., and the values under the horizontal axis represent the corresponding seconds.

4 Result

4.1 Ocean environment inversion based on OAT

The steps of inverting the ocean environment in section 2.2 were executed as follows:

1) A total of 29 SSPs were measured by the launching and receiving ships during the OAT experiment. The covariance matrix eigenvectors and eigenvalues of these SSPs were computed. As the largest three eigenvalues accounted for more than 95% of the sum of all the eigenvalues, so the 3-order EOFs was used. The eigenfunctions corresponding to top 3 eigenvalues are the EOF functions. The EOF coefficients of every SSP were calculated;
2) The EOF coefficients obtained above were augmented with a normal perturbation to generate an initial population with a population size of 500. The population size of 500 was found in the simulations to cover the variable space of the EOF coefficients well and achieve as large a diversity of populations as possible;
3) The positioning and timing information is obtained through synchronous GPS data collected by the standard hydrophone on the launching ship and the hydrophone array on the receiving ship. The travel time of the acoustic signal is calculated using the first wave peak of the signal received by the receiving ship and the signal pulse received by the launching ship. A matched filter algorithm was used to determine the travel times of different eigenrays and the eigenray with the min travel time is consider as the fastest eigenray. The time obtained above is considered as the travel time of the fastest eigenray (Table 1). It's worth mentioning that there is a distance (30m) between the bomb launching point and the hydrophone on the launching ship. So, an extra delay is added on travel time. Considering that the depth of 0-15m is a homogeneous layer and the speed of sound is 1535m/s, it is necessary to add another 0.0195s to the travel time. In this way, the τ_k in equation (1) is calculated. Subsequently, the BELLHOP acoustic model is employed to calculate the intrinsic acoustic propagation delay from the sound source to each hydrophone i.e., t_k in equation (1). Thus, the fitness F_i of the individuals of each population was calculated according to equation (1);
4) The 500 individuals were sorted by F_i and the top 100 individuals were labelled, i.e., the crowding factor was set to 1/5 (De Jong, 1975; Zhang, 2013; Cui et al, 2021);
5) Selection, crossover and mutation operations of genetic algorithm were performed on all 500 individuals to produce the next generation. The maximum number of genetic generations was set to be 40, the mating probability

to be 0.5, the mutation probability to be 0.2. The parameters are the result of a comprehensive consideration after simulation, which takes into account both the need to traverse the entire search space and also the computational efficiency;

6) The Hemming distances between the 500 individuals generated in the step 5 and the 100 individuals labeled in the step 4 were calculated according to equation (2). The less adapted of the two individuals within specific distance was penalized. Thus, 500 individuals (out of 500 + 100 individuals above) with higher fitness F_i are the next generation.

The above steps were repeated until the fitness function of the optimal individual satisfied the termination condition, then the corresponding SSP of the optimal individual was output as the inversion result.

The SSP obtained from the experiments of 35m bombs at 6 release points (S1-S6) and measured data are selected as samples and shown in Figure 5. The measured data are XBT measurement from launching and receiving vessels during the experiment. Due to the limited availability of salinity data (only 2 CTD measurements per buoy), the salinity profile is assimilated with measured data based on ROMS dataset, shown in Figure 6A. The average SSP of OAT is shown in Figure 6B. Consequently, the temperature profile is extracted using OACM and illustrated in Figure 6C. Figure 6 reveals that the biases primarily originate in the thermocline depth, while the biases in the sea surface mixed layer and deeper layers are relatively smaller. Root Mean Square Error (RMSE) for SSP and temperature profile is calculated to be 1.07m/s and 0.40°C, respectively. Specifically, RMSE in thermocline depth (15-40m) is 1.21m/s and 0.47°C. These results are considered accurate, taking into account the limited number of blast sources and the duration of the experiment. The findings suggest that NGA algorithm-based OAT can reliably invert the marine environment in Yellow Sea.

4.2 Application of OAT in targeted observation

As OAT experiment was not conducted within the sensitive area of targeted observation, a simulation experiment was employed to verify the impact of OAT on ocean environment prediction.

Considering the variation characteristics of temperature in the OAT experiment area and the sensitive area are quite different, it's irrational to assimilate OAT inversion result as targeted observation data directly. In this study, the temperature bias obtained from the acoustic tomography inversion and the temperature measurement data from XBT in the sensitive area were combined to simulate the acoustic tomography inversion data within the sensitive area. The vertical bias structure is related only to the OAT inversion method, but not the region. Thus, the simulated observations of "truth + bias" avoid the influence of the bias in different regions on the results. These simulated OAT observation data were then brought

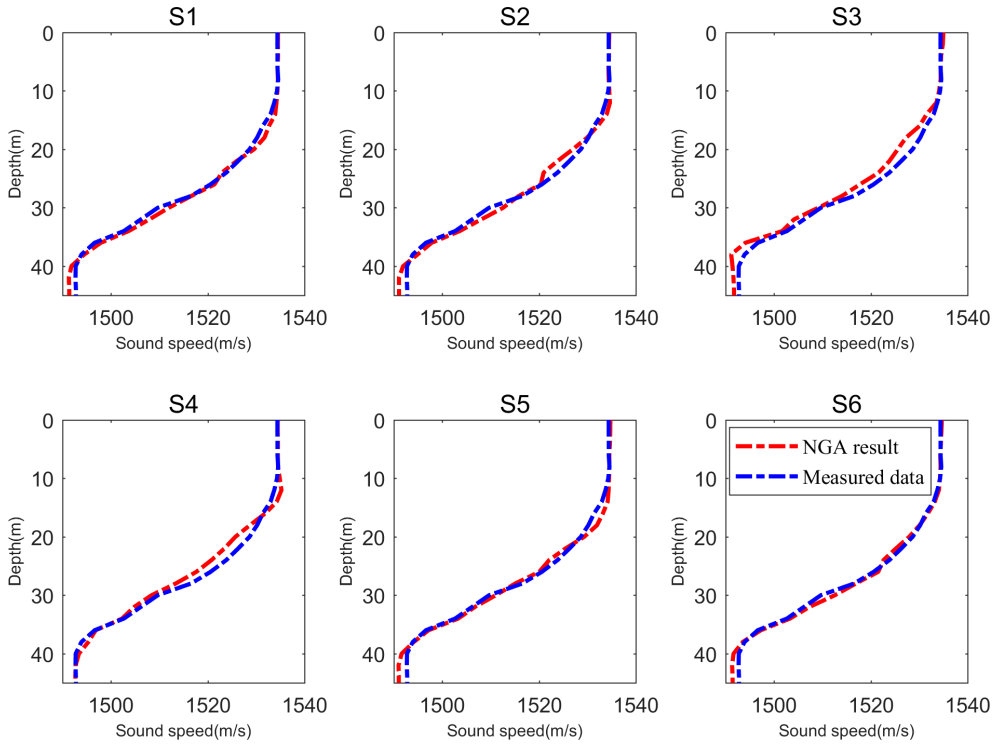


FIGURE 5
Comparisons of NGA results and measured data obtained from the experiments of 35m bombs at 6 release points (S1-S6).

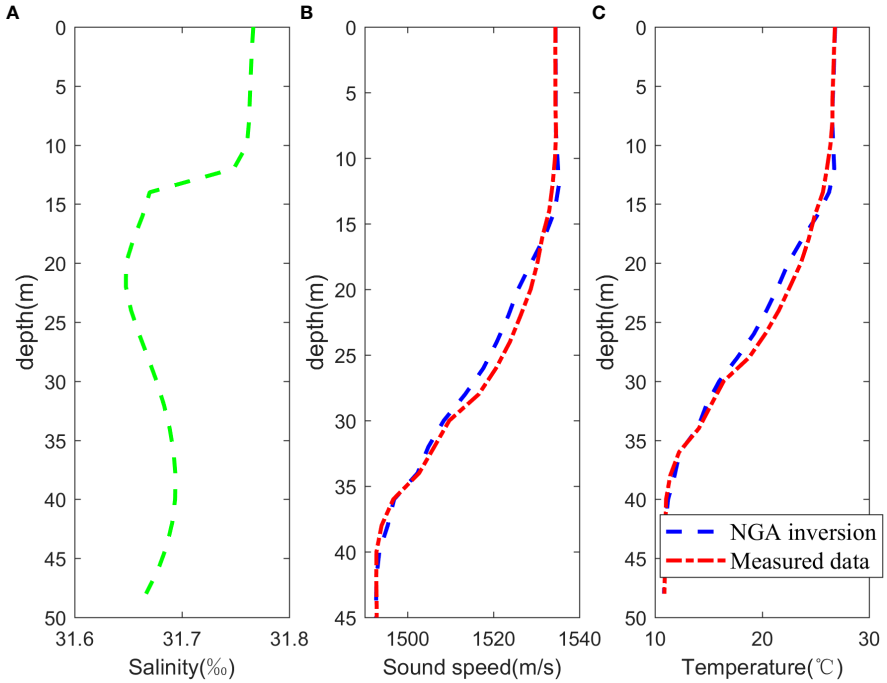


FIGURE 6
Results of OAT inversion. **(A)** Salinity profile after assimilation with measured data based on ROMS dataset, **(B)** Comparison of average SSP obtained from OAT and measured data, **(C)** Comparison of average temperature profiles obtained from OAT and measured data.

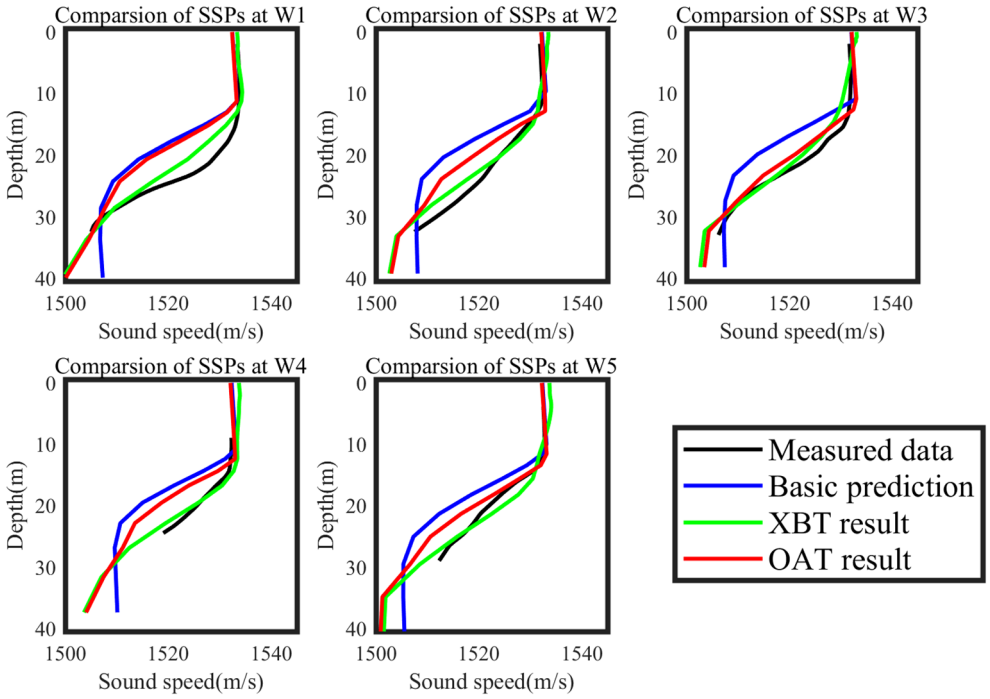


FIGURE 7
Comparisons of sound speed profile at different buoy station at verification time (day 7), including measured data (black line), basic prediction (blue line), assimilation results from XBT measured data (green line) and assimilation results from OAT inversion simulation data (red line).

TABLE 2 RMSEs of basic prediction, experimental data, and OAT data compared with measured data in the target region at 7-th day.

Position	W1	W2	W3	W4	W5
Basic prediction (m/s)	7.31	6.43	5.59	6.88	5.3
Exp result (m/s)	2.44	2.06	1.51	1.85	2.2
OAT result (m/s)	3.82	3.90	1.93	2.96	3.2

into the assimilation system to obtain updated ICs, leading to improved predictions through the use of 3D-Var and ROMS methods. The prediction results at the verification time were compared between the basic prediction, the measured data from buoys in the target region (noted by red stars in Figure 2), the assimilation of XBT measured data, and the assimilation of the simulated OAT assimilation. The comparison results are shown in Figure 7.

The results demonstrate that both the simulated OAT observation and Exp observation(XBT result) significantly improve the accuracy of temperature profile forecasts compared to the basic prediction at the verification time. These results are closer to the measured data, particularly in surface and thermocline layers. It should be noted that the comparison does not include the deeper layers since the buoy depths do not reach the seafloor. RMSEs of the basic prediction, Exp, and OAT simulation data compared to the measured data at 5 buoys are calculated and presented in Table 2. Overall, Exp yields more accurate results than OAT simulation data, primarily due to the introduced bias of OAT inversion. On average, RMSE of XBT prediction is reduced by 68.1%, while RMSE of OAT prediction is decreased by 49.9% comparing with basic prediction. On the other hand, setting VTS at W3 buoy as an example, the RMSEs of these predictions along with a Ctrl Run, which is the result of assimilation based on observation in the target region, are shown in Figure 8. From Figure 8, the bias of XBT and OAT results change on a similar trajectory. At the verification time, the RMSE in the target region was greatly reduced by experiment with deploying XBT observation and simulated OAT observation in the identified sensitive area (XBT result and OAT result) than that of experiment with observations being deployed in the verification area itself (Ctrl Run). These findings demonstrate that OAT can serve as a reliable observation method for targeted observation.

The simulation experiment described above provides validation regarding the influence of OAT data from XBT locations. However, the unique characteristics of acoustic tomography, including its large-area coverage and long-term observation capabilities under low-cost conditions, necessitate further verification of OAT's influence on prediction using OSSE. For the OSSE, two sets of predictions with different ICs and same driving conditions are selected: "True Run" and "Ctrl Run". "True Run" and "Ctrl Run" are predictions from same boundary and driving conditions, but different initial conditions. "True Run" is regarded as the real ocean measured data. "Ctrl Run" is regarded as the basic prediction

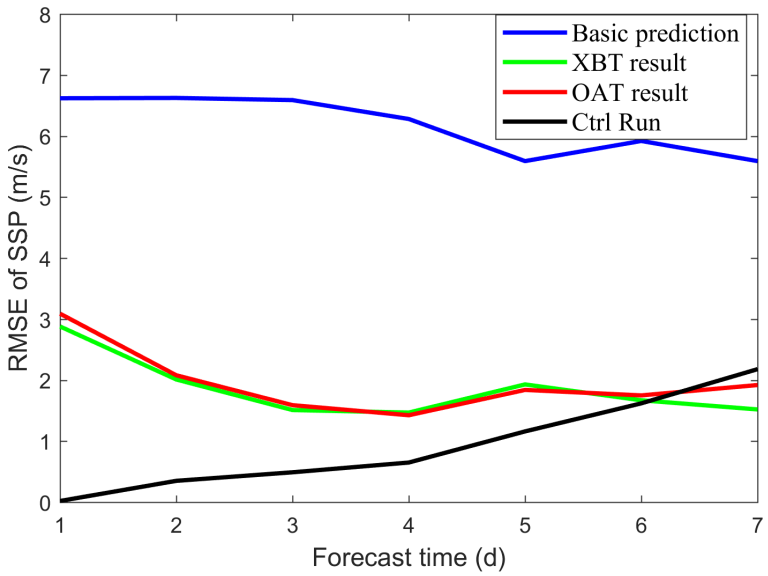


FIGURE 8
Temporal evolution of vertical integration RMSEs of SSP at W3 station, including basic prediction (blue line), assimilation results from XBT measured data in sensitive area (green line), OAT inversion simulation data in sensitive area (red line) and measured data in the target region (Ctrl Run-black line).

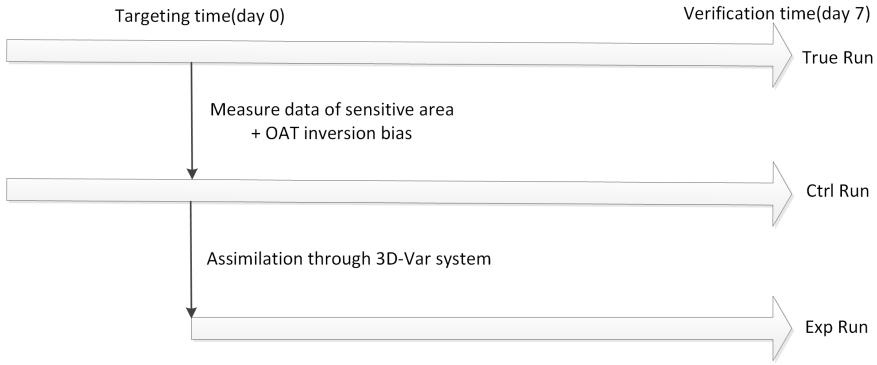


FIGURE 9
Schematic diagram of Observing System Simulation Experiments (OSSE). Targeted observation data from the sensitive area at the targeting time is extracted from "True Run". This data is then combined with OAT inversion bias and assimilated with data from "Ctrl Run" to generate "Exp Run".

without additional observation. Targeted observation data from the sensitive area, extracted from "True Run", is combined with OAT inversion bias to simulate OAT targeted observation data. This OAT targeted observation data is then assimilated with data from "Ctrl Run" using 3D-Var system, resulting in the generation of "Exp Run". By comparing the predictions of "Ctrl Run" and "Exp Run" with "True Run" at the verification time, the impact of OAT as a targeted observation method on prediction can be analyzed (Figure 9). To further investigate the effect of OAT observations on prediction quality under different conditions, various experiment setups were employed. EXP1 replicates the same OAT observation condition as XBT measurement. Observations were carried out at the locations of the triangular markers of the three Z-lines in the sensitive area of Figure 2 from day 1 to day 3. EXP2 simulates observation on all ocean model nodes in the sensitive area, meaning the observation area is 4 times the area of EXP1,

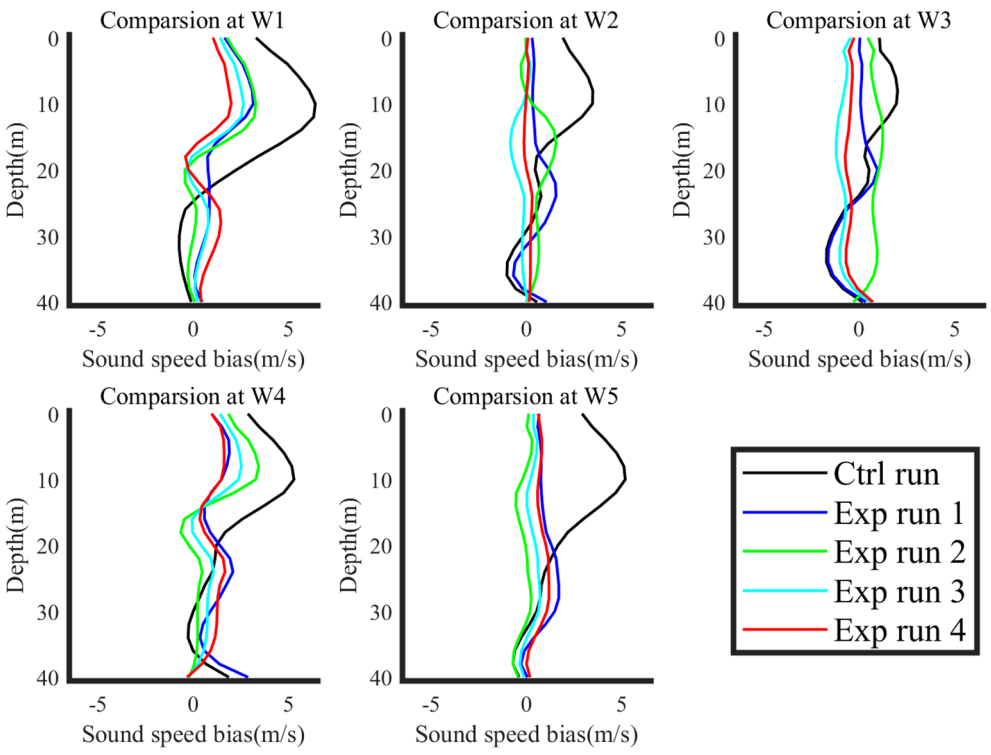


FIGURE 10
Results of OSSE. Vertical sound speed bias structures of different experiment conditions are configured, including Ctrl Run (black line), EXP1 for basic OAT observation (blue line), EXP2 for larger observation area (green line), EXP3 for extended observation time (black line), and EXP4 for both a larger observation (cyan line) area and an extended observation time (red line).

TABLE 3 Setup and results of OSSE.

Experiments	Description	RMSE of SSP (m/s)	Improvement percentage on 'Ctrl Run'
Ctrl Run	Control Run	2.48	---
EXP1	Basic OAT observation	1.07	53.6%
EXP2	4 times observation area (cover the whole sensitive area of day 0)	1.02	58.6%
EXP3	Extended observation time from 3 days to 5 days	0.86	65.2%
EXP4	4 times observation area and extended observation time from 3 days to 5 days	0.78	68.5%

while the observation keeps the same(day 1-3). EXP3 takes observation in the same area as EXP1 while the observations time is extended, which indicates that observations are located on the triangular locations from day 1 to day 5. EXP4 is the combination of EXP2 and EXP3, i.e., observations are carried out on all ocean model nodes in the sensitive area from day 1 to day 5. The results of these experiments are depicted in Figure 10 and Table 3.

The results of EXP1 reinforce the finding that OAT can improve prediction quality, thereby validating its utility as a targeted observation method. EXP2 and EXP3 demonstrate that increasing the observation area and extending the observation time further improve the prediction quality. However, it is observed that the enhancement in prediction quality is more pronounced with an extended observation time. This phenomenon can be attributed to the fact that both the horizontal resolution of the ocean model employed and the assimilation radius of the 3D-Var system exceed the observation spacing. Consequently, expanding the observation area may not yield enough additional valuable environmental information. Conversely, extending the observation time not only provides more observations but also reduces the time interval between the final observations and the verification time. EXP4 results reveal that the combination of increasing the observation area and extending the observation time improves the prediction quality for maximum. However, improvement achieved through this approach is not notably different from that achieved solely by extending the observation time. Furthermore, extending the observation time is more cost-effective and logistically feasible compared to deploying additional observation nodes during sea trials. Consequently, prolonged observations duration emerges as an efficient and economical approach to observation, thereby offering reference for the implementation of acoustic tomography targeted observation projects.

5 Summary

OAT is a cost-effective, long-term, and wide-area ocean monitoring method that obtains acoustic signals to invert marine environment characteristics. In this study, the validity of OAT for the marine environment inversion was verified using data collected during the 2019 Yellow Sea experiment. The OAT inversion biases were incorporated into measurements obtained from the sensitive area, identified by CNOP method, to simulate OAT observations from the sensitive area. These simulated OAT observations were substituted into a 3D-Var assimilation system to improve the quality of ICs and subsequently enhance the short term (7-day) prediction of the target region. These findings confirm the effectiveness of OAT as a targeted observation method. Considering the large-scale and long-duration nature of OAT, OSSE method was employed to further test the impact of OAT on prediction quality. Specifically, the effects of adding observation nodes and extending the observation duration were examined. The results show that both approaches and their combination have positive effects in reducing prediction uncertainty. However, it was found that extending the observation duration is a more efficient strategy.

This study aims to verify the feasibility of acoustic tomography as a targeted observation method in a simulated environment using actual measurement data. It is important to note that the findings of this study are yet to be validated in sea trials. Additionally, most existing acoustic tomography observation methods utilize fixed or submerged buoys, while the sensitive area for targeted observation changes with time. Therefore, it is crucial to investigate optimal selection strategies for observation nodes that can yield the highest improvement in prediction quality. Additionally, it is necessary to examine the effect of parameter variations in the ocean model and assimilation model on acoustic tomography and its corresponding targeted observations. Understanding the interrelationships and contribution of these parameters to the prediction quality requires further investigation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CB: Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. LJ: Data curation, Formal Analysis, Writing – review & editing. GW: Funding acquisition, Investigation, Resources, Writing – review & editing. DL: Funding acquisition, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research received support by National Basic Research Program of China (Grant Number 2019-JCJQ-ZD-149-00) and Taishan Scholars Program.

Acknowledgments

We are grateful to Liu Kun for his help in ocean model building, Hu Huiqin for her assistance in assimilation system, the captains Bing Liu, Xiufeng Chen and all the crews of the R/V KeXue No. 3 and R/V ChuangXin No. 2 for their cooperation in collecting the observation data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baker, N. L., and Daley, R. (2000). Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q.J.R. Meteorol. Soc* 126 (565), 1431–1454. doi: 10.1002/qj.49712656511

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. Part I: theoretical aspects. *Monthly Weather Review.* 129 (3), 420–436. doi: 10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2

Buizza, R., and Montani, A. (1999). Targeting observations using singular vectors. *J. Atmospheric Sci.* 56 (17), 2965–2985. doi: 10.1175/1520-0469(1999)056<2965:TOUSV>2.0.CO;2

Carrière, O., and Hermand, J. P. (2008). A sequential Bayesian approach to vertical slice tomography of a shallow water environment. *J. Acoustical Soc. America.* 123 (5_Supplement), 3339–3339. doi: 10.1121/1.2933874

Chan, P. W., Han, W., Mak, B., Qin, X., Liu, Y., Yin, R., et al. (2022).). Ground-space-sky observing system experiment during tropical cyclone mulan in August 2022. *Adv.Atmos. Sci.* 40 (2), 194–200. doi: 10.1007/s00376-022-2267-z

Cui, B., Xu, G., Da, L., and Guo, W. (2021). Shallow sea sound speed profile inversion based on niche genetic algorithm. *J. Appl. Acoustics* 40 (2), 1–8. doi: 10.11684/j.issn.1000-310X.2021.02

Da, L., Guo, W., Zhao, J., and Fan, P. (2015). Capture uncertainty of underwater environment by ocean-acoustic coupled model. *Acta Acustica.* 40 (3), 477–486. doi: 10.15949/j.cnki.0371-0025.201

Dahl, P. H., Zhang, R., Miller, J. H., Bartek, L. R., Peng, Z., Ramp, S. R., et al. (2004). Overview of results from the asian seas international acoustics experiment in the east China sea. *IEEE J. Oceanic Engineering.* 29 (4), 920–928. doi: 10.1109/JOE.2005.843159

DeFerrari, H. A., and Nguyen, H. B. (1986).). Acoustic reciprocal transmission experiments, Florida Straits. *J. Acoust Soc. Am.* 79 (2), 299–315. doi: 10.1121/1.393569

De Jong, K. A. (1975). *An analysis of the bechavior of a class of genetic adaptive systems* (MI, United States: University of Michigan). Ph. D Dissertation.

Duan, W., and Hu, J. (2015). The initial errors that induce a significant "spring predictability barrier" for El Nino ˜ events and their implications for target observation: results from an earth system model. *Clim. Dynam.* 46 (11–12), 3599–3615. doi: 10.1007/s00382-015-2789-5

Duan, W. S., and Mu, M. (2018). Predictability of el nino-southern ˜ Oscillation events. *OXFORD research encyclopedia, CLIMATE SCIENCE.* (United States: Oxford University). Available at: https://climatescience.oxfordre.com. doi: 10.1093/acrefore/9780190228620.013.80

Duda, T. F., Lynch, J. F., Newhall, A. E., Wu, L., and Chiu, C. S. (2004). Fluctuation of 400-hz sound intensity in the 2001 ASIAEX south China sea experiment. *IEEE J. Oceanic Engineering.* 29 (4), 1264–1279. doi: 10.1109/joe.2004.836997

Dushaw, B. D. (1999). "The Acoustic Thermometry of Ocean Climate (ATOC) Project: Towards depth-averaged temperature maps of the North Pacific Ocean," in *Proceeding of the International Symposium on Acoustic Tomography and Thermometry*, Tokyo, Japan. doi: 10.1121/1.413035

Dushaw, B., Forbes, A., Gaillard, F., Gavrilov, A., Gould, J., Howe, B., et al. (2001). "Observing the ocean in the 2000's: A strategy for the role of acoustic tomography in ocean climate observation," in *Observing the oceans in the 21st century* (University of Washington,United States: GODAE Project Office and Bureau of Meteorology).

Dushaw, B. D., and Worester, P. F. (2001). "Acoustic remote sensing of the North Pacofic on gyre and regional scales," in *Pacific CLIVAR international Pacific Implementation Workshop, International Pacific Research Center at the University of Hawaii*, Honolulu, Hawaii.

Feng, R., Duan, W., and Mu, M. (2016). Estimating observing locations for advancing beyond the winter predictability barrier of Indian Ocean dipole event predictions. *Climate Dynamics.* 48 (3-4), 1173–1185. doi: 10.1007/s00382-016-3134-3

Feng, Y., Min, J., Zhuang, X., and Wang, S. (2019). Ensemble sensitivity analysis-based ensemble transform with 3D rescaling initialization method for storm-scale ensemble forecast. *Atmosphere.* 10, 24. doi: 10.3390/atmos10010024

Fried, S., Walker, S., Hodgkiss, W., and Kuperman, W. (2013). Measuring the effect of ambient noise directionality and split-beam processing on the convergence of the cross-correlation function. *J. Acoustical Soc. America* 134 (3), 1824–1832. doi: 10.1121/1.4816490

Gasparini, O., Camporeale, C., and Crise, A. (1997). Introducing passive matched field acoustic tomography. *Nuovo Cimento- Societa Italiana di Fisica Sezione C.* 20 (4), 497–520.

Hamill, T. M., and Snyder, C. (2002). Using improved background-error covariances from an ensemble kalman filter for adaptive observations. *Monthly Weather Review.* 130 (6), 1552–1572. doi: 10.1175/1520-0493(2002)130

Howe, B. M. (1987). Multiple receivers in single vertical slice ocean acoustic tomography experiments. *J. Geophysical Research: Oceans.* 92 (C9), 81–86. doi: 10.1029/JC092iC09p09479

Howe, B. M., Mercer, J. A., Spindel, R. C., Worcester, P. F., Hildebrand, J. A., Hodgkiss, W. S., et al. (1991). "Slice89: A single slice tomography experiment," in *Ocean variability & Acoustic propagation*. Eds. J. Potter and A. Warn-Varnas (Dordrecht: Springer). doi: 10.1007/978-94-011-3312-8_6

Hu, H., Liu, J., Da, L., Guo, W., Liu, K., and Cui, B. (2021). Identification of the sensitive area for targeted observation to improve vertical thermal structure prediction in summer in the Yellow Sea. *Acta Oceanologica Sinica.* 40 (7), 77–87. doi: 10.1007/s13131-021-1738-x

Huang, C.-F., Yang, T. C., Liu, J.-Y., and Schindall, J. (2013). Acoustic mapping of ocean currents using networked distributed sensors. *J. Acoustical Soc America* 134, 2090–2105doi: 10.1121/1.4817835

Jin, G., Lynch, J., Pawlowicz, R., and Wadhams, P. (1993). Effects of sea ice cover on acoustic ray travel times, with applications to the Greenland Sea Tomography Experiment. *J. Acoust Soc. Am.* 94 (2), 1044–1057. doi: 10.1121/1.406951

Joly, A., Browning, K. A., Bessemoulin, P., Cammas, J., Caniaux, G., Chalon, J., et al. (1999). Overview of the field phase of the fronts and Atlantic Storm-Track EXperiment (FASTEX) project. *Quarterly Journal of the Royal Meteorological Society.* 125 (561), 3131–3163. doi: 10.1002/qj.49712556103

Kramer, W., Dijkstra, H. A., Pierini, S., and van Leeuwen, P. J. (2012). Measuring the impact of observations on the predictability of the kuroshio extension in a shallow-water model. *J. Phys. Oceanography.* 42 (1), 3–17. doi: 10.1175/JPO-D-11-014.1

Langland, R., Tóth, Z., Gelaro, R., Szunyogh, I., Shapiro, M., Majumdar, S., et al. (1999). The north pacific experiment (NORPEX-98): targeted observations for improved north american weather forecasts. *Bull. Amer Meteorol Soc* 80, 1363–1384. doi: 10.1175/1520-0477(1999)080<1363:TNPENT>2.0.CO;2

Lebedev, K. V., Yaremchuk, M., Mitsudera, H., Nakano, I., and Yuan, G. (2003). Monitoring the kuroshio extension with dynamically constrained synthesis of the acoustic tomography, satellite altimeter and in *situ* data. *J. Oceanography.* 59 (6), 751–763. doi: 10.1023/B:JOCE.0000009568.06949.c5

Li, F., Guo, X., Hu, T., and Ma, L. (2014). Acoustic travel-time perturbations due to shallow-water internal waves in the Yellow Sea. *J. Comput. Acoustics.* 22, 1–11. doi: 10.1142/S0218396X14400037

Li, F., Yang, X., Zhang, Y., Luo, W., and Gan, W. (2019). Passive ocean acoustic tomography in shallow water. *J. Acoustical Soc. America.* 145 (5), 2823–2830. doi: 10.1121/1.5099350

Liu, K., Guo, W., Da, L., Liu, J., Hu, H., and Cui, B. (2021). Improving the thermal structure predictions in the Yellow Sea by conducting targeted observations in the CNOP-identified sensitive areas. *Sci. Rep.* 11 (1), 19518. doi: 10.1038/s41598-021-98994-7

Liu, J. Y., Liu, K., Guo, W. H., Liang, P., and Da, L. L. (2023). Optimal initial errors related to the prediction of the vertical thermal structure and their application to targeted observation: A 3-day hindcast case study in the northern South China Sea. *Deep Sea Res. Part I: Oceanographic Res. Papers.* 200, 104146. doi: 10.1016/j.dsr.2023.104146

Lynch, J. F., Jin, G., Pawlowicz, R., Ray, D., Plueddemann, A. J., Chiu, C. S., et al. (1996). Acoustic travel-time perturbations due to shallow-water internal waves and internal tides in the Barents Sea Polar Front: Theory and experiment. *J. Acoust Soc. Am.* 99 (2), 803–821. doi: 10.1121/1.414657

Malfoud, S. W. (1995). *Niche methods for genetic algorithms* (Illinois: University of Illinois, Urbana-Champain).

Mu, M. (2013). Methods, current status, and prospect of targeted observation. *Sci. China Earth Sci.* 56 (12), 1997–2005. doi: 10.1007/s11430-013-4727-x

Mu, M., Duan, W. S., and Wang, B. (2003). Conditional nonlinear optimal perturbation and its applications. *Nonlin. Processes Geophys.* 10 (6), 493–501. doi: 10.5194/npg-10-493-2003

Munk, W., Worcester, P. F., and Wunsch, C. (1995). *Ocean acoustic tomography* (Cambridge: Cambridge University Press).

Munk, W., and Wunsch, C. (1979). Ocean acoustic tomography: a scheme for large scale monitoring. *Deep Sea Res. Part A. Oceanographic Res. Papers.* 26 (2), 123–161. doi: 10.1016/0198-0149(79)90073-6

Parsons, D., Beland, M., Burridge, D., Bougeault, P., Brunet, G., Caughey, J., et al. (2017). THORPEX research and the science of prediction. *Bull. Amer Meteorol Soc* 98 (4), 807–830. doi: 10.1175/bams-d-14-00025.1

Rabier, F., Klinker, E., Coutie, P. R., and Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteor. Soc* 122, 121–150. doi: 10.1002/qj.49712252906

Shang, E. C. (1989). Ocean acoustic tomography based on adiabatic mode theory. *J. Acoust Soc. Am.* 85-4, 1531–1537. doi: 10.1121/1.397355

Shang, E., Voronovich, A., Wang, Y., Naugolnykh, K., and Ostrovsky, L. (2000). New Schemes of ocean acoustic tomography. *J. Comp. Acoust.* 8 (3), 459–471. doi: 10.1016/S0218-396X(00)00030-3

Shchepetkin, A. F., and McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography following-coordinate oceanic model. *Ocean Model.* 9 (4), 347–404. doi: 10.1016/j.ocemod.2004.08.002

Shen, Y., Ma, Y., Du, Q., and Jiang, X. (1999). Feasibility of description of the sound speed profile in shallow water *via* empirical orthogonal function (EOF). *Acta Acustica.* 18 (20), 21–25.

Skarsoulis, E. K., Athanassoulis, G. A., and Send, U. (1996). Ocean acoustic tomography based on peak arrivals. *J. Acoust Soc. Am.* 100 (2), 797–813. doi: 10.1121/1.416212

Snyder, C. (1996). Summary of an informal workshop on adaptive observations and FASTEX. *Bull. Amer Meteorol Soc* 77 (5), 953–961. doi: 10.1175/1520-0477-77.5.953

Szunyogh, I., Toth, Z., Morss, R. E., Majumdar, S. J., Etherton, B. J., and Bishop, C. H. (2000). The effect of targeted dropsonde observations during the 1999 winter storm reconnaissance program. *Monthly Weather Review.* 128 (10), 3520–3537. doi: 10.1175/1520-0493(2000)128<3520:TEOTDO>2.0.CO;2

Taniguchi, N., Mutsuda, H., Arai, M., Sakuno, Y., Hamada, K., Takahashi, T., et al. (2023). Reconstruction of horizontal tidal current fields in a shallow water with model-oriented coastal acoustic tomography. *Front. Mar. Sci.* 10, 1–17. doi: 10.3389/fmars.2023.1112592

Taroudakis, M. I., and Markaki, M. G. (1997). On the use of matched-field processing and hybrid algorithms for vertical slice tomography. *J. Acoust Soc. Am.* 102 (2), 885~895. doi: 10.1121/1.419955

Thiruvengadam, P., Indu, J., and Ghosh, S. (2021). Radar reflectivity and radial velocity assimilation in a Hybrid ETKF-3DVAR System for Prediction of a Heavy Convective Rainfall. *Q. J. R. Meteorological Society.* 147, 1–17. doi: 10.1002/qj.4021

Wang, Q., Mu, M., and Dijkstra, H. (2013). The similarity between optimal precursor and optimally growing initial error in prediction of Kuroshio large meander and its application to targeted observation. *J. Geophysical Research: Oceans.* 118, 869–884. doi: 10.1002/jgrc.20084

Wei, M., Tóth, Z., Wobus, R., and Zhu, Y. (2008).). Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP global operational forecast system. *Tellus A* 60, 62–79. doi: 10.1111/j.1600-0870.2007.00273.x

Worcester, P. F. (2019). Tomography. *1st edition of Encyclopedia of Ocean Sciences.* 6, 2969–2986. doi: 10.1016/B978-0-12-409548-9.11591-X

Yamoaka, H., Kaneko, A., Jae-Hun, P., Hong, Z., Gohda, N., Takano, T., et al. (2002). Coastal acoustic tomography system and its field application. *IEEE J. Oceanic Engineering.* 27 (2), 283–295. doi: 10.1109/joe.2002.1002483

Yang, S. S., Li, Z. L., and He, L. (2022). Range dependent sound speed profile inversion in the northern area of the South China Sea. *Acta ACUSTICA* 47 (3), 339–347. doi: 10.15949/j.cnki.0371-0025.2022.03.010

Yuan, G., Nakano, I., Fujimori, H., Nakamura, T., Kamoshida, T., and Kaya, A. (1999). Tomographic measurements of the Kuroshio Extension Meander and its associated eddies. *Geophysical Res. Letters.* 26 (1), 79–82. doi: 10.1029/1998GL900253

Zhang, W. (2013). Inversion of sound speed profile in three-dimensional shallow water. PH.D dissertation*Harbin Eng. University.* 42–45.

Zhang, H., Chen, J., Zhi, X., and Wang, Y. (2015). A comparison of ETKF and downscaling in a regional ensemble prediction system. *Atmosphere* 6, 341–360. doi: 10.3390/atmos6030341

Zhang, K., Mu, M., and Wang, Q. (2017). Identifying the sensitive area in adaptive observation for predicting the upstream Kuroshio transport variation in a 3-D ocean model. *Sci. China Earth Sci.* 60 (5), 866–875. doi: 10.1007/s11430-016-9020-8

Zhu, X., Kaneko, A., Wu, Q., Zhang, C., Taniguchi, N., and Gohda, N. (2013). Mapping tidal current structures in zhitouyang bay, China, using coastal acoustic tomography. *IEEE J. Oceanic Engineering.* 38 (2), 285296. doi: 10.1109/JOE.2012.2223911

Zhu, X. H., Wu, Q., Zheng, H., Liao, G., and Zhang, C. (2010). "The Chinese Coastal Acoustic Tomography system and its application to the Luotou Channel, China," in *2010 3rd International Congress on Image and Signal Processing* Yantai, China, 3890–3894. doi: 10.1109/CISP.2010.5647342

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 09 November 2023
DOI 10.3389/fmars.2023.1302077

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Jialun Chen,
University of Western Australia, Australia
Wouter Wittebol,
Eindhoven University of Technology, Netherlands
Sartaj Khan,
Harbin University, China
Irfan Hussain,
Khalifa University, United Arab Emirates

*CORRESPONDENCE
Zhichao Lv
lvzhichao@hrbeu.edu.cn

RECEIVED 26 September 2023
ACCEPTED 25 October 2023
PUBLISHED 09 November 2023

CITATION
Du L, Wang Z, Lv Z, Wang L and Han D (2023) Research on underwater acoustic field prediction method based on physics-informed neural network.
Front. Mar. Sci. 10:1302077.
doi: 10.3389/fmars.2023.1302077

Research on underwater acoustic field prediction method based on physics-informed neural network

Libin Du, Zhengkai Wang, Zhichao Lv*, Lei Wang and Dongyue Han

College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China

In the field of underwater acoustic field prediction, numerical simulation methods and machine learning techniques are two commonly used methods. However, the numerical simulation method requires grid division. The machine learning method can only sometimes analyze the physical significance of the model. To address these problems, this paper proposes an underwater acoustic field prediction method based on a physics-informed neural network (UAFP-PINN). Firstly, a loss function incorporating physical constraints is introduced, incorporating the Helmholtz equation that describes the characteristics of the underwater acoustic field. This loss function is a foundation for establishing the underwater acoustic field prediction model using a physics-informed neural network. The model takes the coordinate information of the acoustic field point as input and employs a fully connected deep neural network to output the predicted values of the coordinates. The predicted value is refined using the loss function with physical information, ensuring the trained model possesses clear physical significance. Finally, the proposed prediction model is analyzed and validated in two dimensions: the two-dimensional acoustic field and the three-dimensional acoustic field. The results show that the mean square error between the prediction and simulation values of the two-dimensional model is only 0.01. The proposed model can effectively predict the distribution of the two-dimensional underwater sound field, and the model can also predict the sound field in the three-dimensional space.

KEYWORDS

underwater acoustic field, prediction model, neural network, physical constraints, physics-informed neural network

1 Introduction

High-precision underwater acoustic field model is of great significance for underwater acoustic communication, sonar effectiveness evaluation, underwater target recognition and location, etc. Establishing a high-precision underwater acoustic field prediction model is one of the important research contents in underwater acoustic field. For example, the establishment of highly accurate underwater acoustic field can help synthetic aperture sonar (SAS) to obtain higher resolution sonar images. Zhang (2023) proposed a new method to simulate the original SAS echo. The transmitted signal was Fourier transformed and multiplied by the phase shift of the delay, and the spectrum of the echo signal was accurately obtained. Yang et al. (2023) proposed a multi-receiver SAS imaging algorithm based on Loffeld Bistatic formula (LBF). Zhang et al. (2021) proposed a multi-receiver SAS image processing method and proved that under certain conditions, the bistable formula of Loffeld can be simplified to the same formula as the spectrum based on phase center approximation. Zhang et al. (2023) proposed a SAS imaging algorithm by rerepresenting the Loffeld bistable formula (LBF), which includes quasi-monostable (QM) and multi-receiver deformed (MD) phases, as range-variant phase and range-invariant phase. In the process of SAS signal transmission, there will be attenuation, and the establishment of high-precision underwater acoustic field can compensate the attenuation signal accordingly. At present,the numerical simulation and the machine learning are common methods to forecast the underwater acoustic field. The numerical simulation method mainly uses ray method, normal mode method, parabola method, beam integration method (Belibassakis et al., 2014) to establish physical models and calculate underwater acoustic field. Kiryanov et al. (2015) established a random non-uniform wave field model for evaluating sound velocity field based on the results of deep-sea acoustic long-range propagation test. Miller (1954) introduced the coupled mode to extend the solution range of the differential equation to the number of waveguides dependent on the distance. For the normal mode method, the finite element method is usually used to build the acoustic field model, and the KRAKEN model is widely used to build the acoustic field by finite element as a representative model. Zhou and Luo (2021) established a finite element model for predicting underwater acoustic field based on Cartesian coordinate system in a two-dimensional environment, whose universality is better than that of KRAKEN model. Teng et al. (2010) used the boundary element method to simulate the acoustic field around two kinds of underwater communication transducers, and the prediction results are generally applicable. The spectral method is a high precision method for solving differential equations, and it also plays an important role in promoting the calculation of underwater acoustic field. Tu et al. (2022) used spectral method and coupled modes to solve the acoustic field of underwater linear source. In this paper, Chebyshev-Tau spectral method was used to solve the horizontal wave number of irrelevant segments in the approximate range, and a global matrix was constructed to solve the coupling coefficient of the acoustic field and synthesize the complete acoustic field. Tu et al. (2021) used Chebyshev-Tau spectral method to construct the normal mode model of underwater acoustic field, and converted the relevant differential equations into a complex matrix eigenvalue problem formed by orthogonal basis with Chebyshev polynomials to solve the horizontal beam. Tu et al. (2020) used Chebyshev-Tau spectral method to solve the normal mode model and parabolic equation, and the solution accuracy was higher than that of the finite element method. Although the numerical simulation method can directly forecast the underwater acoustic field by using the physical rules, the numerical solution often needs to divide the regular grid to simplify the model calculation, and it is difficult to predict the acoustic field model with irregular boundaries. With the development of computer hardware, the neural network, which is one of the important methods in machine learning, has been used more and more to predict underwater acoustic field. Ahmed et al. (2021) established a machine learning model to predict the sound velocity profile in deep water and shallow water. The accuracy of this model reached 99.99% and the prediction effect was better than the acoustic field model forecasted by the equation. Based on the self-defined loss function, He et al. (2022) constructed a single output joint neural network and a multi-output neural network with physical constraints to accurately forecast the beam and feature function of the underwater acoustic field.

Machine learning method has greatly improved the accuracy of underwater acoustic field prediction, but there are some obvious problems. First of all, the model trained by the neural network does not have a clear physical meaning, and it has poor adaptability to different environments. Secondly, the neural network needs a large amount of historical data as support to ensure that the trained model has a high accuracy. Figure 1 (Karniadakis et al., 2021) shows the relationship between data volume and physical parameters in the model prediction problem. In case 1, assuming clear physical laws and boundary conditions are known, the corresponding

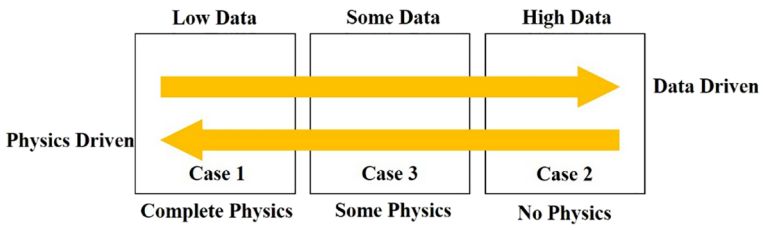


FIGURE 1
The relationship between data volume and physical rules.

problem can be solved according to physical rules. In this case, the numerical simulation method can be used to predict the underwater acoustic field. In case 2, only a large amount of data is known but the specific physical rules are not clear, and machine learning can be used to predict the acoustic field problem. The real underwater acoustic field prediction is a problem in case 3: there are sufficient data but some parameters in the physical rules are not clear, which cannot be solved directly by the physical rules.

Physics-informed neural network (PINN) is a new kind of neural network, which is used to solve the problem of case 3 in Figure 1. It essentially trains the neural network with physical equations as constraints, so that the prediction model can meet certain physical rules. The method has been applied to geophysics, fluid mechanics, plasma dynamics, high dimensional system problems, quantum chemistry, materials science and other fields closely related to physics. Zhu et al. (2021) introduced a deep learning framework for inversion of seismic data. This paper combined DNN and numerical partial differential equation solvers to solve problems such as seismic wave velocity estimation, fault rupture imaging, seismic location and source time function inversion. Raissi et al. (2019) combined Navier-Stokes equations with deep learning to build a model based on physics-informed neural network and predict pressure distributions in incompressible fluids. Shukla et al. (2020) used physics-informed neural network to detect cracks on the surface of materials, designed a trained PINN to solve the problem of identification and characterization of cracks on the surface of metal plates, and solved the acoustic wave equation using measured ultrasonic surface acoustic wave data with a frequency of 5 MHz. Wu et al. (2022) introduced the Helmholtz equation and its corresponding boundary conditions into neural networks to establish physics-informed neural networks describing acoustic problems. These neural network algorithms can not only reflect the distribution of training data samples, but also follow the physical laws described by partial differential equations. Pfau et al. (2020) combined the wave function of Fermi-Dirac statistics with deep learning networks to calculate the solution of the multi-electron Schrodinger equation. Rotskoff et al. (2022) used PINN method to solve the high-dimensional problem and gave the results of the probability distribution in the 144-dimensional Allen-Cahn type system, indicating that the method is effective for high-dimensional systems, but its adaptability needs to be optimized for more complex systems. Zhang et al. (2022) used deep neural network to modify the displacement factor of surrounding rock of Verruijt-Booker solution, and constructed the correlation between the surface settlement and the spatial position of tunnel excavation face. Then, the physics equations of the corrected solutions were used to construct PINN, and the results were better than those of DNN alone. Zou et al. (2023) designed a PINN model to solve the seismic wave equation. Du et al. (2023) used the three-dimensional function equation and other physical rules to form a loss function, and trained the neural network by minimizing the loss function. The final output satisfied the function equation and the result was better than the traditional calculation result.

In the underwater acoustic field, wave theory is usually used to describe underwater acoustic propagation. In this paper, the underwater sound propagation equation is derived based the framework of wave theory, which is used as constraint to train the deep neural network, and finally the underwater acoustic field prediction model with practical physical significance is obtained. The specific arrangement of this paper is as follows:

1) Deriving the Helmholtz equation of underwater acoustic propagation in homogeneous medium and establishing the model of underwater acoustic propagation based on the Helmholtz equation.
2) Designing a fully connected deep neural network. Introducing Helmholtz equation into the training process of neural network. Establishing a physics-informed neural network based on the Helmholtz equation.
3) Adjusting different training parameters of neural network, analyzing model training efficiency and prediction accuracy, and finding the best network design parameters.

The rest of the paper is organized as follows: in Chapter 2, the Helmholtz equation describing the distribution of sound pressure in underwater acoustic field is derived. In Chapter 3, the structure of underwater acoustic field prediction physics-informed neural network (UAFP-PINN) is described in detail. In Chapter 4, UAFP-PINN is used to forecast the 2D and 3D underwater acoustic fields, and the prediction results are analyzed in detail. Finally, the conclusion and summary is mentioned in Chapter 5.

2 Theory

2.1 Helmholtz equation

Wave theory is a strict mathematical method, which can be used to derive the Helmholtz equation describing the law of underwater sound propagation. For ideal fluids, the wave equation for sound pressure can be written as follows (Jensen et al., 2011):

$$\rho \nabla \cdot \left(\frac{1}{\rho} \nabla p \right) - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \tag{1}$$

In the above formula, p is the sound pressure value of the acoustic field, ρ is the density of the medium, c is the speed of sound in the medium, and both density and speed of sound are functions of space and time. ∇ is a Hamiltonian operator. To simplify the calculation, assuming that the density does not vary with space (Jensen et al., 2011), Formula 1 can be simplified to the following formula:

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \tag{2}$$

Formula 2 is the wave equation in a homogeneous medium, which can be approximated to the ocean acoustic field in a homogeneous medium for a smaller scale ocean acoustic field model. ∇^2 stands for Laplace operator. For simple harmonic wave, $\frac{\partial^2}{\partial t^2} = -\omega^2$, ω is radiant frequency, introducing the potential function $\Psi = \frac{p}{\sqrt{\rho}}$, Formula 2 can be written as the following formula (Liu et al., 2019):

$$\nabla^2 \Psi + k^2(x,y,z)\Psi = 0 \tag{3}$$

In Formula 3, Ψ is the potential function, k is the wave number in the medium, which is calculated by the formula $k = \frac{\omega}{c}$. The density in a uniform medium is a constant, and it can be seen from the potential function formula that there is a linear relationship between the sound pressure and the potential function, so the sound pressure also satisfies Formula 3. The Helmholtz equation describing the sound pressure can be written as follows:

$$\nabla^2 p + k^2(x,y,z)p = 0 \tag{4}$$

The Formula 4 describes the sound pressure relationship between adjacent positions of sound waves in a uniform medium. The beam k in the medium is a position function of space. The equation belongs to the partial differential equation with variable coefficient. In order to simplify the calculation, the density ρ and the sound velocity c of the medium are regarded as constant value. In this paper, k is a fixed constant in the model presented.

2.2 Physics-informed neural network

Most physical laws can be expressed in the form of partial differential equations, but it is difficult to find specific analytical solutions of higher-order partial differential equations, which are usually approximated by various methods. The superiority of neural network is that it is a universal approximator. If the neural network has at least one nonlinear hidden layer, as long as the network has a sufficient number of neurons, it can fully approximate the continuous function defined on any compact subset in theory.

Neural network is a data-driven approximation tool, and its obvious disadvantage is that it needs a large amount of historical data for training. The trained model reflects the characteristics of the data dimension, and cannot clearly represent the physical characteristics of the result. In order to solve these defects of neural networks, the training process of neural networks can incorporate partial differential equations describing physical laws to constrain this model, so that the training results contain corresponding physical characteristics. This kind of neural network is called physics-informed neural network(PINN), and its general structure is shown in Figure 2 (Karniadakis et al., 2021).

As shown in Figure 2, PINN consists of two parts: the deep neural network prediction part and the partial differential equation constraint part. Using the location (x, y, z) as input, the predicted value P in the region Ω is predicted after passing through the fully connected layer. The mean square error is calculated as the loss function 1, denoted as $LOSS_1$ in Figure 2. The predicted value is put into the pre-set partial differential equation and its loss function $LOSS_2$ is calculated. Finally, two kinds of loss functions are combined to train the deep neural network as constraints.

An optimizer is an algorithm used to optimize the model parameters in deep learning, which updates the model parameters according to the gradient information of the loss function, so that the model can gradually approximate the optimal results. The optimizers commonly used in neural networks are stochastic gradient descent (SGD), Adam, AdaGrad and RMSProp. Two optimizers, SGD and Adam, are used to train the PINN in this paper. SGD is one of the most basic optimizers in neural networks. Adam is an optimizer that combines momentum method and adaptive learning rate adjustment, which is a commonly used optimizer in neural networks.

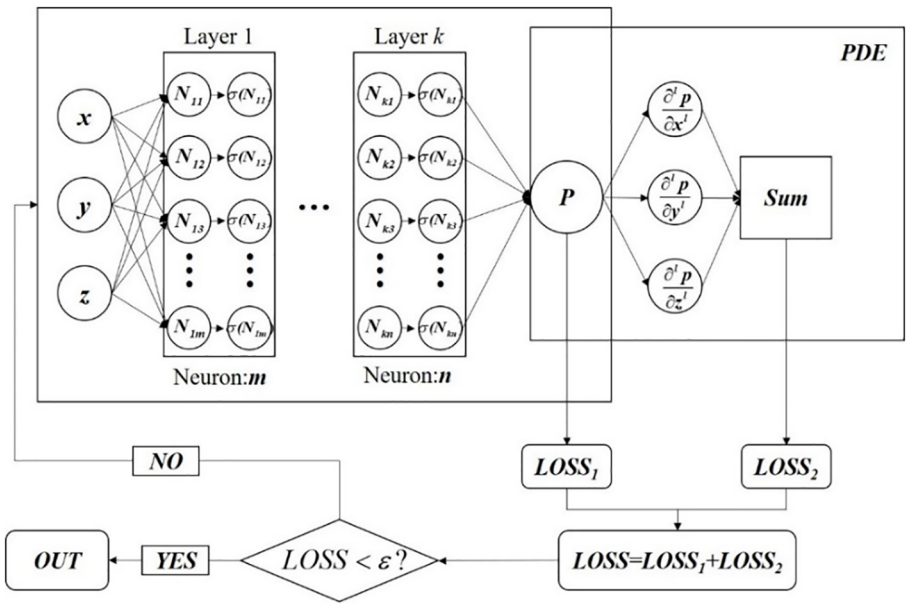


FIGURE 2
General structure of PINN.

3 UAFP-PINN

This section introduces the underwater acoustic field prediction model based on physics-informed neural network(UAFP-PINN), builds a fully connected deep neural network with six hidden layers, and uses Helmholtz equation to construct the loss function in the neural network and add it to the training of the neural network. This section introduces the specific content of UAFP-PINN model from three parts: model structure, loss function based on Helmholtz equation and model activation function.

3.1 Frame of prediction

The neural network takes the position coordinate of sound pressure P as the input and the corresponding sound pressure value as the output for training. The network consists of one input layer, six hidden layers and one output layer. In order to verify the difference between two-dimensional and three-dimensional model, two-dimensional and three-dimensional physics-informed neural network is established respectively, and the models are trained using (x, z, p) and (x, y, z, p) as inputs respectively.

The input layer of the neural network is the coordinate information of sound pressure, the output layer is the predicted sound pressure, and there are six hidden layers in this neural network. The number of neurons in the hidden layer was (8, 16, 32, 64, 128, 256) and each neuron is connected by full connection. The i neuron in layer $l-1$ and the j neuron in layer l are connected by weighting parameters w_{ji}^{l}. Each neuron trains the model through input weighting parameters w_{ji}^{l} and bias terms b^{l} in layer $l-1$. Figure 3 shows the computational relationship between the two related neurons. In the feedforward model, Formula 5 shows the output of the k neuron in the next layer l (Bishop and Nasrabadi, 2006). σ is the activation function, which is covered in the third part of this section.

$$u_k^l = \sigma\left(\sum_{j=1}^{N_{l-1}} w_{kj}^l u_j^{l-1} + b_k^l\right) \tag{5}$$

3.2 Loss function

The key of physics-informed neural network is to train the neural network with physical partial differential equation which describes the state of object. The traditional neural network usually use the mean square error of predicted and simulated values to evaluate the training results. In this study, a Helmholtz equation describing underwater sound propagation is added as another loss function. The loss function of the mean square error and the loss function of the physical constraint are used as constraints to train the model. The loss function of mean square error is denoted as $LOSS_1$ and the loss function of physical constraint is denoted as $LOSS_2$ (Borrel-Jensen et al., 2021).

The reference formula of loss function $LOSS_1$ is the formula for calculating mean square error, and the specific content is shown in Formula 6:

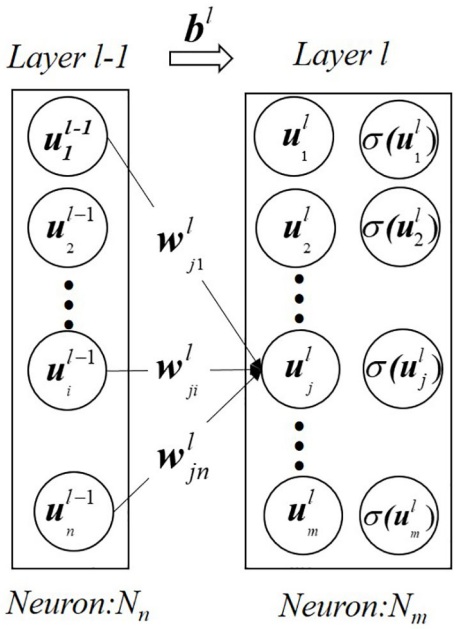


FIGURE 3
Computational relationship between adjacent neurons.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|P_i - T_i|^2 \tag{6}$$

In the above formula, MSE represents the mean square error, n is the total number of samples, P is the predicted value, and T is the true value. Formula 6 is one of the important indicators to measure the accuracy and precision of the prediction model.

For the prediction model in this paper, the sound pressure value predicted by the neural network is denoted as $\boldsymbol{p}^{pre}$, the corresponding simulated sound pressure value is denoted as $\boldsymbol{p}^{tra}$, and the number of samples is denoted as N, then the mean square error loss function $LOSS_1$ of the neural network is shown as Formula 7:

$$LOSS_1 = \frac{1}{N}\sum_{i=1}^{N}(p_i^{pre} - p_i^{tra})^2 \tag{7}$$

The mean square error loss function $LOSS_1$ represents the degree of similarity between the predicted value and the simulated value. Traditional neural networks use this loss function to continuously approximate the predicted value to the simulated value. In essence, the model trained by means of the mean square error loss function represents the characteristics of the data dimension.

The sound pressure value $\boldsymbol{p}^{pre}$ predicted by the neural network is a function of spatial coordinates, and the Laplacian operator of the sound pressure p in formula 4 can be expressed as:

$$\nabla^2 p = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{8}$$

According to Formula 4, k is the medium beam and the calculation formula is $k = \frac{\omega}{c}$, ω is the radiation frequency, c is the medium sound speed. Bringing Formula 8 into Formula 4 gives

the Helmholtz equation with the predicted values.

$$L = \frac{\partial^2 p^{pre}}{\partial x^2} + \frac{\partial^2 p^{pre}}{\partial y^2} + \frac{\partial^2 p^{pre}}{\partial z^2} + k^2 p^{pre} \tag{9}$$

Formula 9 describes the Helmholtz equation of the predicted value of the neural network, which is a vector, and defines the square of the 2-norm of this vector as the loss function $LOSS_2$ of the physical constraint (Song et al., 2022), from which the expression of the loss function of the physical constraint can be obtained as:

$$LOSS_2 = \left\| \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial \boldsymbol{x}^2} + \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial y^2} + \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial z^2} + k^2 p^{pre} \right\|_2^2 \tag{10}$$

The physical constraint loss function LOSS2 represents the physical characteristics of the predicted value and brings the predicted value into the Helmholtz equation describing the underwater sound field. The model trained with the loss function 2 represents the characteristics of the physical dimension.

In order to make the trained neural network have both data characteristics and physical characteristics, the mean square error loss function $LOSS_1$ and physical constraint loss function $LOSS_2$ will be combined in this paper. In order to make the model better fitting effect and have strong physical interpretability, the two loss functions will be summed with the same weight. As a whole, the $LOSS$ function $LOSS$ trains the model. The model trained by the $LOSS$ function has clear physical interpretability. The calculation formula of the loss function is as follows:

$$LOSS = \frac{1}{N}\sum_{i=1}^{N}(p_i^{pre} - p_i^{tra})^2 + \left\| \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial x^2} + \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial y^2} + \frac{\partial^2 \boldsymbol{p}^{pre}}{\partial z^2} + k^2 \boldsymbol{p}^{pre} \right\|_2^2 \tag{11}$$

3.3 Activation function

As an important parameter in deep neural network training, the activation function (σ) has great influence on the training efficiency and prediction accuracy of the neural network. The activation functions are mainly used to introduce nonlinear properties that enables neural networks to learn and represent complex nonlinear relationships. The activation function is typically applied to each neuron in a neural network, converts the input signal to nonlinear and passes the transformed result to the next layer. In the training of neural network, common activation functions mainly include tangent activation function ($Tanh(x)$), sine activation function ($\sin(x)$), Relu function ($\mathrm{Relu}(x)$), and arctangent activation function ($Atan(x)$). The images of these four activation functions are shown in Figure 4.

The tangent function is more commonly used in cases where the neuronal output has negative values, such as symmetric centralized data. Using tangential activation functions can help neural network introduce nonlinear transformations so that neural network can learn and represent more complex patterns and relationships. The tangent activation function outputs a negative value when the input is negative and a positive value when the input is positive. This makes the tangent activation function more suitable for processing data with positive and negative symmetries.

The sine function is a nonlinear activation function that maps the input values to an output range between -1 and 1. Sine activation functions have nonlinear properties, which can help neural network model learn and represent nonlinear patterns and relationships.

The Relu function is one of the widely used activation functions in deep learning, especially in the hidden layer. Its main advantages are computational efficiency and avoiding gradient saturation problems. Relu function passes positive values and truncate negative values to zero, which makes Relu sparsely active, that is, only some neurons are activated while others are zero. Sparse activation can provide higher model representation and help to reduce the computational load and complexity of the model.

The arctangent function can help mitigate gradient vanishing or gradient explosion problems in some cases because it has a gentler gradient as the input approaches the boundary, and these problems can affect the model's learning ability and convergence.

In this study, the tangent function, the sine function, the Relu function and the arctangent function (Al-Safwan et al., 2021; Song et al., 2022) are used to predict the model. Different activation functions are selected to observe the decline of the model's loss function, and the effect of different activation functions is evaluated according to the model prediction effect. Finally, we select the activation function that best fits PINN model.

4 Experiment

4.1 Data

In order to verify the feasibility of the physics-informed neural network, an ocean environment model is established using COMSOL software. A point sound source is placed at the edge of the ocean environment to simulate the excitation conditions of the underwater acoustic field, and the effectiveness of the physics-information neural network is verified according to the acoustic field data.

The test area is 30 meters long, 10 meters wide and 10 meters high, and the test point sound source is located at coordinates (0,5,5). The upper boundary of the area is the air-sea interface, which can be approximated as an absolute soft boundary, and the sound pressure values on the boundary are satisfied the condition $p(x, y, z) = 0$; the lower boundary of the area is a hard submarine interface, which can be approximated as an absolute hard boundary, and the sound pressure values on the boundary are satisfied the formula $\frac{\partial p}{\partial z} = 0$. The surrounding boundary is a perfectly matched layer (Chen et al., 2013). Due to the small scale of the area, the density of seawater in the area can be approximately constant, the average density of seawater is 1025 kg/m^3, and the sound velocity in the seawater medium is 1500 m/s. The structure of the area model is shown in Figure 5.

In order to avoid the influence of reverberation on the acoustic field, the point sound source with a frequency of 100 Hz is selected in this paper, the sound wave is a sine wave, and the amplitude is selected as one. Figure 6 shows the spatial acoustic field distribution at 0.1s drawn by COMSOL according to the above conditions, and

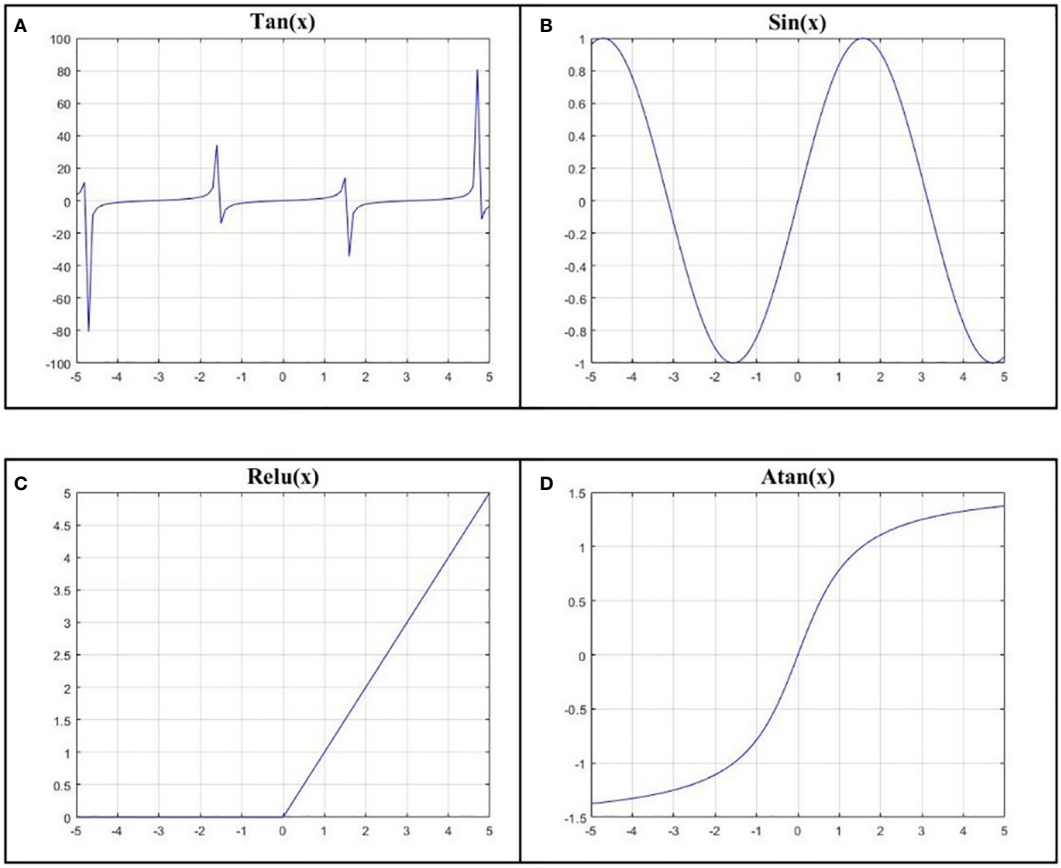


FIGURE 4
(A) is the tangent activation function, **(B)** is the sine activation function, **(C)** is the Relu activation function, and **(D)** is the arctangent activation function.

Figure 7 shows the sound pressure distribution at XZ-plane when coordinate y is five meters.

In this paper, the sound pressure data in 2-dimensional plane and 3-dimensional space are predicted respectively. In this study, prediction data and test data are separated. A total of 28,100 sets of simulated values are collected in the 2-dimensional plane data training set and 112,400 sets of simulated values are collected in the test set. A set of 2-dimensional plane training data is collected every 0.1 m. A total of 352,500 sets of simulated values are collected in the 3-dimensional training set and 982,150 sets of simulated values are collected in the test set. A set of 3-dimensional training data is collected every 0.2 m.

4.2 Introduction to experimental environment

The experimental environment will affect the predicted rate, so this section describes the hardware configuration for the experiment. The GPU is NVIDIA GeForce GT370, the CPU is Intel i7-9700, the operating system is Windows10, and the memory is 64 GB. This paper establishes a prediction model based on Python language, and uses Pytorch framework to establish a neural network. The compiler uses Pycharm2018.

4.3 Hyper parameter setting

In the experiment, the adaptive moment estimation (Adam) optimizer and stochastic gradient Descent (SGD) are used to analyze the influence of the optimizer on the prediction accuracy of the model. For this optimization process, the first-order momentum factor, second-order momentum factor and Fuzz factor in Adam are configured as 0.9, 0.999 and 0.0000001, respectively. The initial learning rate is set to 0.001, the weight attenuation factor is set to 0.0005, and 1/10 of the total training data is used for a batch. Before the actual test, a small batch of test data was used for training, and it is found that the model could converge within 100 times. Therefore, the number of iterations of the 2-dimensional model is set to 500 epochs and the number of iterations of the 3-dimensional model was set to 250 epochs. Finally, in order to ensure that the weight of data-driven and physical constraints is the same, the two loss functions are summed with the same proportional coefficient 1 and combined into

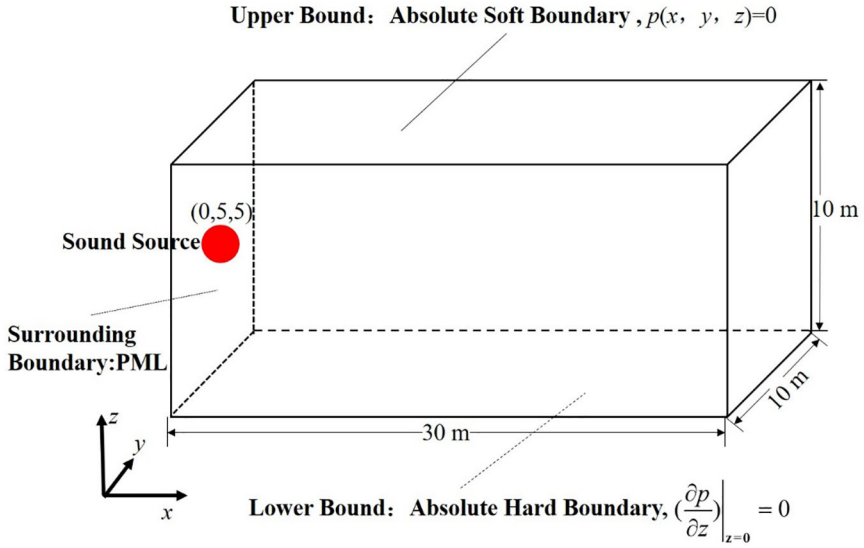


FIGURE 5
Area model structure.

an overall loss function to train the model. For the specific loss function, see Formula 11.

4.4 Results and analysis

In order to verify the effectiveness of the underwater acoustic field prediction model, 2D underwater acoustic prediction model based on physics-informed neural network (2D UAFP-PINN) and 3D underwater acoustic prediction model based on physics-information neural network (3D UAFP-PINN) are established by using 2D and 3D acoustic field data. The two models are used to predict the acoustic field data of the test location, adjust different optimizers and activation functions to analyze the optimal model parameters, and finally evaluate the model by analyzing the statistical characteristics between the predicted values and the simulated values.

4.4.1 2D UAFP-PINN

In this section, a 2D underwater acoustic prediction model based on physics-informed neural network (2D UAFP-PINN) is established, and the effects of different activation functions and optimizers on the prediction accuracy of the model are analyzed. Model parameters are as follows: there are 28100 sets of training data and 112,400 sets of test data;The training iteration epochs are 500 times, the data of each training is 1/10 of the total training data.

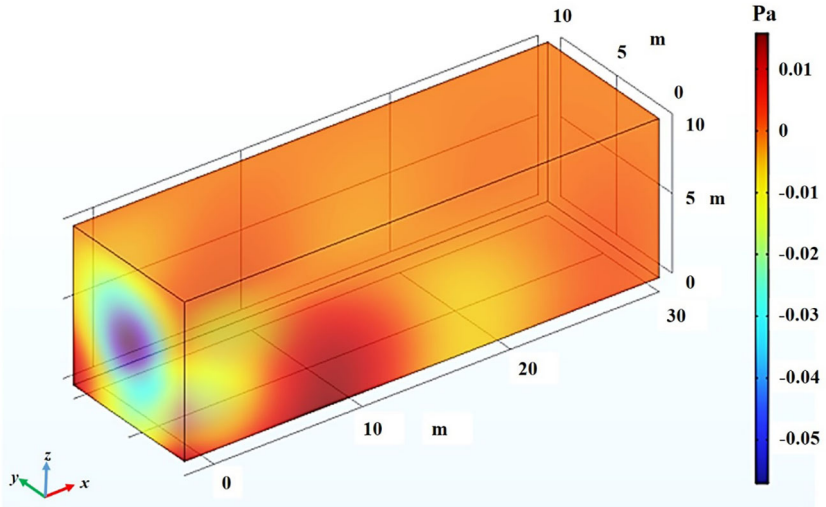


FIGURE 6
Sound pressure distribution at t= 0.1s.

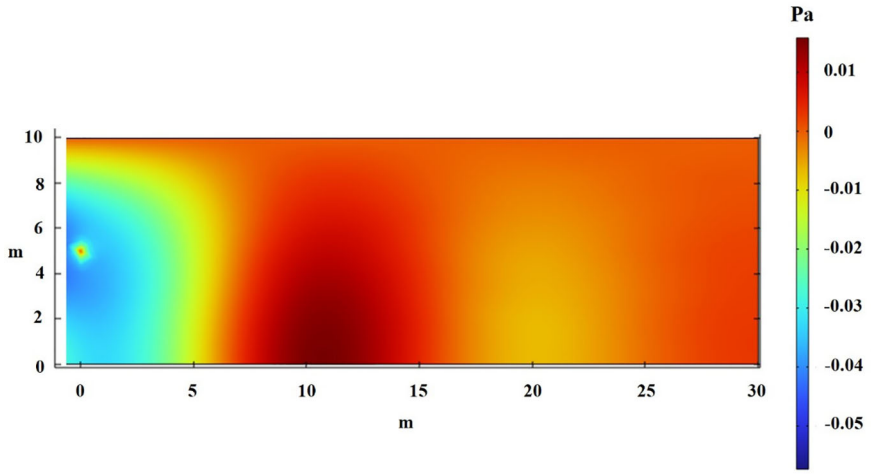


FIGURE 7
Sound pressure distribution in the XZ-plane at y=5 m.

It selects different activation functions and optimizers to train the model and gets the curve of loss function with training times. Figure 8 shows the change curves of the different loss functions using the two optimizers. Since the loss function reached the optimal trend after about 100 training times, only the results of the first 100 training times are shown in the figure to make it clearer.

As can be seen from Figure 8, when the Relu activation function is combined with the Adam optimizer, the loss function drops to the lowest values, reaching 1.09, and the minimum trend is reached when the training times are about 13 times. The model convergence speed is faster than other activation functions. Therefore, using the Adam optimizer, the loss function decreases faster than the SGD optimizer, indicating that the Adam optimizer is more suitable for the training of 2D UAFP-PINN model. In summary, it can be seen that using the Adam optimizer and Relu activation function is the best choice for the 2D UAFP-PINN model.

In order to verify the prediction effect of this model, 112,400 sets of data simulated by COMSOL are selected as simulation values to evaluate this model. In this paper, the validity of the forecast results is analyzed from four perspectives: R-squared (R^2), mean square error (MSE), mean absolute error (MAE) and absolute error distribution. R-squared is a common regression model evaluation metric used to measure the model's ability to explain the target variable. The value range of R-squared is between zero and one, when it is closer to one indicates that the model has a better ability to explain the target variable, and when it is closer to zero indicates that the model has a worse ability to explain the target variable. The expression of R-squared is as follows:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{12}$$

Where y_i are the simulated values of the test set, $\hat{y}_i$ are the predicted values by PINN, $\bar{y}$ are the mean of the simulated values. *SSE* represents the sum of squares of residuals, which is the sum of squares of the difference between the predicted values and the

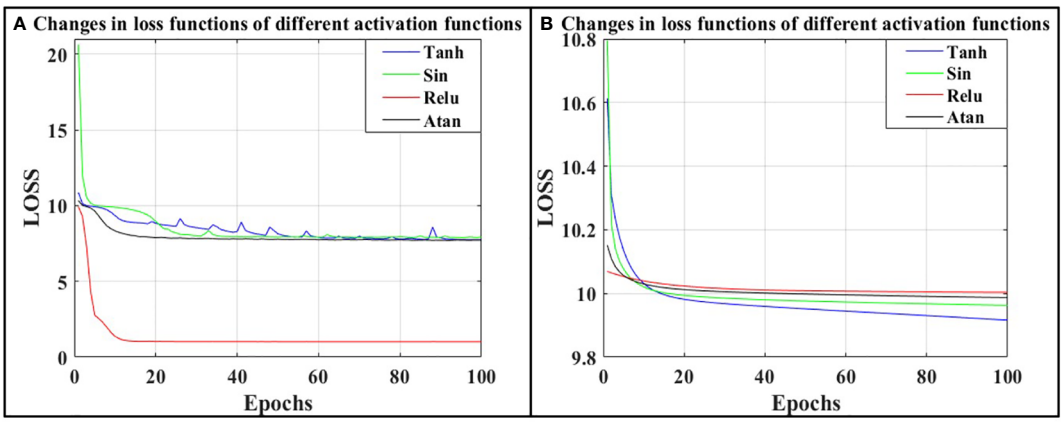


FIGURE 8
(A) shows the different loss function changes under the Adam optimizer, **(B)** shows the different loss function changes under the SGD optimizer.

simulated values. *SST* represents the total sum of squares, which is the sum of squares of the difference between the predicted values and the mean of the simulated values.

The formula for calculating the mean square error can be referred to formula 6 in Section 3.2 of the article. The formula for calculating the mean absolute error is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - T_i| \tag{13}$$

In the above formula, *MAE* represents the mean absolute error, n is the total number of samples, *P* is the predicted value, and *T* is the true value.

Table 1 shows the results of R-squared, mean square error and absolute mean values error of predicted values and simulated values of different activation functions under the Adam optimizer, and Figure 9 shows the absolute error distribution of predicted and simulated values of different activation functions. Table 1 and Figure 9 show the statistical characteristics between the predicted values and the simulated values.

The R-squared values represents the correlation between the predicted values and the simulated values, and the larger the value, the stronger the correlation between the predicted values and the simulated values. It can be seen from Table 1 that the model using Relu activation function for prediction has the strongest correlation with the simulated values, that the R-squared value is 0.98953.The mean square error between the predicted values and the simulated values is only 0.01047 Pa when the model uses Relu activation function, and the mean square error of other activation functions are all around 0.75.The mean absolute error between the predicted values and the simulated values is only 0.06759 Pa when the model uses Relu activation function, and other activation functions' mean absolute error are all around 0.67 Pa. The data predicted by the Relu activation function is very close to the simulated values. Figure 9 shows the distribution of absolute error between the predicted and simulated values of several activation functions. It can be analyzed from the figure that the absolute error of the data predicted by the Relu activation function is distributed within 0.05 Pa, and the data with an absolute error higher than 0.3 Pa is basically not distributed, while the error distribution of the other three activation functions are basically similar, and most of them are concentrated within 1.5 Pa. It can be analyzed that the prediction accuracy is much lower than that of the Relu activation function. Therefore, the best activation function and optimizer for this model are Relu and Adam.

TABLE 1 Statistical results of different activation functions.

Activation function	R-squared	MSE	MAE
Tanh	0.24107	0.75893	0.68727
Sin	0.23773	0.76223	0.67919
Relu	0.98953	0.01047	0.06759
Atan	0.23879	0.76121	0.67166

4.4.2 3D UAFP-PINN

In this section, a 3D underwater acoustic prediction model based on physics-informed neural network (3D UAFP-PINN) is established, and the effects of different activation functions and optimizers on the prediction accuracy of the model are analyzed. Model parameters are as follows: there are 352500 pieces of training data and 982150 sets of test data; The training iteration epochs are 250 times, the data of each training is 1/10 of the total training data. It selects different activation functions and optimizers to train the model and gets the curve of loss function with training times. Figure 10 shows the variation curves of the different loss functions using the two optimizers. The optimal trend reached by the loss function after about 20 training sessions. To make it clearer, Adam only shows the results of the first 100 training sessions in the figure, while SGD only shows the results of the first 20 training sessions.

It can be seen from Figure 10 that when Relu activation function and Adam optimizer are used, the loss function decreases to the lowest degree, reaching 6.94, and reaches the lowest trend when the training times are about 10 times. The model convergent speed is faster than other activation functions. The loss function has the best decreasing effect when the optimizer chooses Adam. It can be concluded that the loss function reduction effect using the Adam optimizer is slightly better than that of the SGD optimizer. In summary, it can be seen that using the Adam optimizer and the Relu activation function is the best choice for the PINN framework. However, compared with the two-dimensional training model, with the increase of data dimension, the training complexity greatly increases, and the gap between the optimization effect of the optimizer and the activation function on the network is also significantly reduced, which indicates that with the increase of data dimension, it is necessary to appropriately increase the network complexity to represent the features of higher-dimensional data. Simply changing the activation function and the optimizer does not make the model convergence better.

In order to verify the prediction effect of the model, 982,152 pieces of data were selected to evaluate the model. Table 2 shows the results of R-squared, mean square error and absolute values error of the predicted values and the simulated values using different activation functions under the Adam optimizer. Figure 11 shows the absolute error distribution between the predicted values and the simulated values of different activation functions, which is used to visually display the error distribution of the predicted values.

As can be seen from Table 2, the largest R-squared value is the result predicted by Relu activation function, which reaches 0.47823, and the predicted values have a relatively high correlation with the simulated values. The correlations of the other three functions are very low. From the mean square error and the mean absolute error, it can be seen that the prediction effect of Relu activation function is much better than that of other activation functions. From the absolute error distribution in Figure 11, the absolute error range of the four activation functions is basically the same, but the absolute error of Relu activation function is mostly concentrated within 0.5 Pa, and the absolute error distribution is the largest around 0.25 Pa, and there is a small peak around 1.0 Pa. However, compared with the prediction results of other activation functions, the prediction effect of Relu

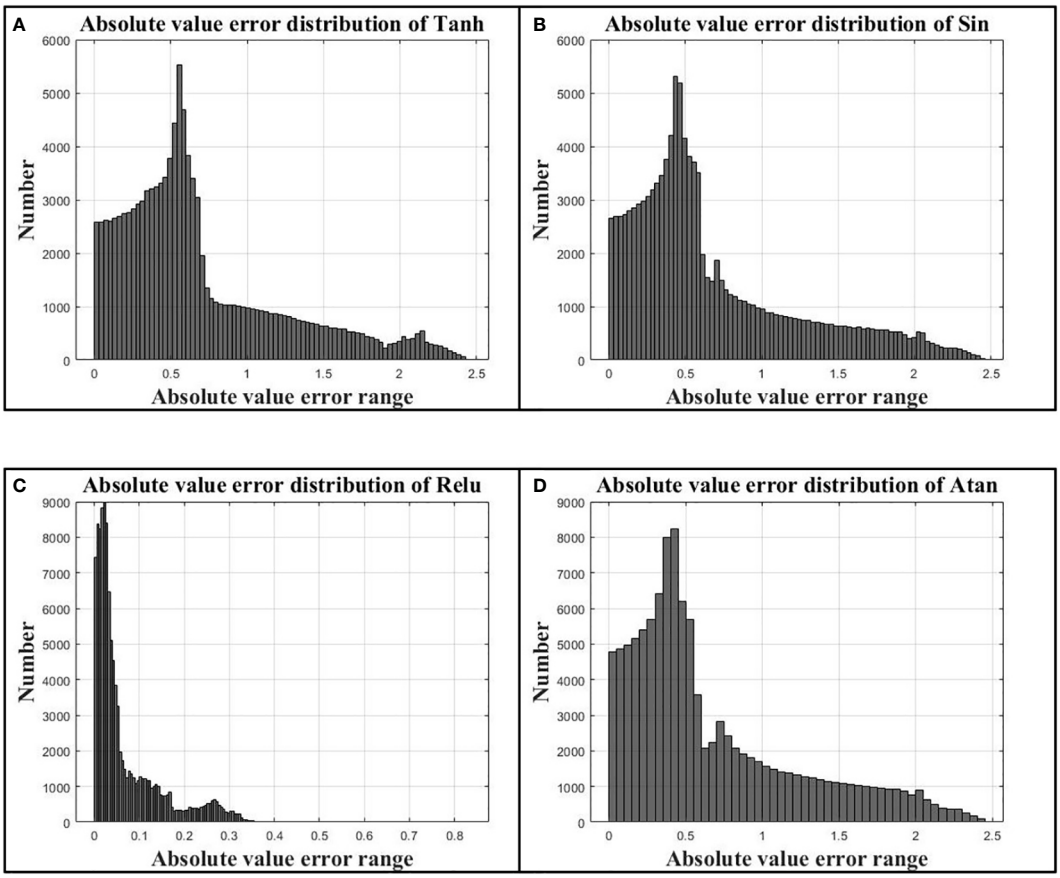


FIGURE 9
(A) is the absolute error distribution using Tanh activation function, **(B)** is the absolute error distribution using Sin activation function, **(C)** is the absolute error distribution using Relu activation function, and **(D)** is the absolute error distribution using Atan activation function.

activation function is obviously better than that of the other three activation functions.

Comparing the 2D UAFP-PINN and 3D UAFP-PINN training and forecasting results, under the condition of Adam optimizer and Relu activation function, the 2D UAFP-PINN model is much better than the 3D UAFP-PINN model, and the error difference between the two models can reach tens of times. The reason for the big difference between the two models is that the complexity of the acoustic field will also increase with the increase of the dimension of the forecast data. Therefore, if the number of hidden layers and neurons of the 3D model remains the same as that of the 2D model, the fitting effect of the 3D model will have the problem of

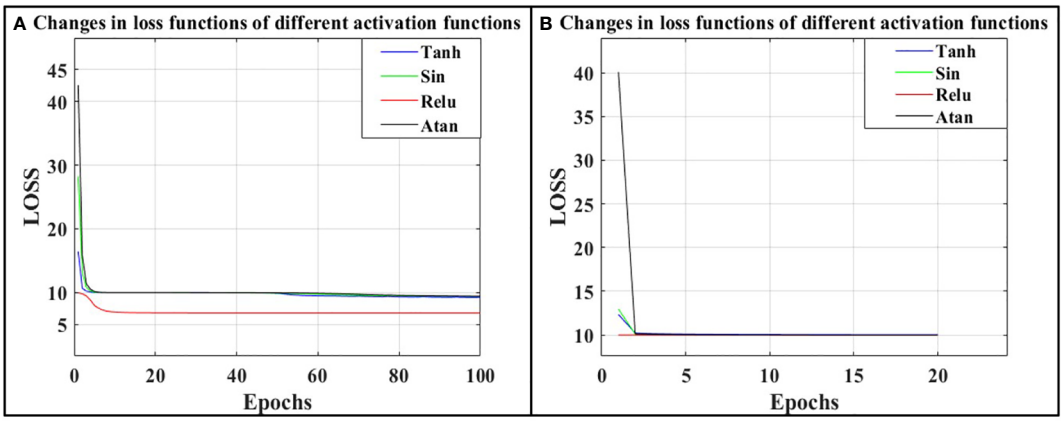


FIGURE 10
(A) shows the different loss function changes under the Adam optimizer, **(B)** shows the different loss function changes under the SGD optimizer.

TABLE 2 Statistical results of different activation functions.

Activation function	R-squared	MSE	MAE
Tanh	0.10428	0.89572	0.70253
Sin	0.09045	0.90955	0.70488
Relu	0.47823	0.52177	0.51988
Atan	0.09411	0.90589	0.70217

underfitting. If the dimension of the prediction model is increased, the number of hidden layers and neurons of the model should be increased to match the corresponding complexity and prevent the problem of underfitting.

5 Conclusion

Compared with the numerical method to solve the underwater acoustic field, PINN has the advantage that it can handle the acoustic field of different media and irregular shape models. PINN does not use a regular network to predict the acoustic field. It can be predicted at any point in the input region if the position is known, and there is no limit to the irregular shape of the model. For the three-dimensional acoustic field, the calculation cost of the numerical method will increase sharply due to the addition of one dimension of the data, and PINN can quickly forecast the high-dimensional acoustic field space.

In addition to PINN, there are machine learning-based methods to predict acoustic fields. Onasami et al. (2021) used deep neural networks and long and short time memory networks to model underwater acoustic channels, and established a data-driven underwater acoustic channel model. The acoustic field prediction model based on machine learning is mainly a data-driven method, which needs a lot of training data to support, and has certain timeliness. The underwater acoustic field is time-varying, and it is often difficult to predict the time-varying underwater acoustic field when the model is trained using only historical data. The advantage of PINN is that new constraint variables can be added via partial differential equations, and it has good environmental adaptability. According to the experimental data, the convergence rate of the model loss function is fast.

The experimental results show that PINN using Relu activation function and Adam optimizer can effectively predict the underwater acoustic field. The model is constrained by the Helmholtz equation describing the underwater acoustic field and combined with the excellent model approximation characteristics of the neural network. It can realize the acoustic field prediction in the case of

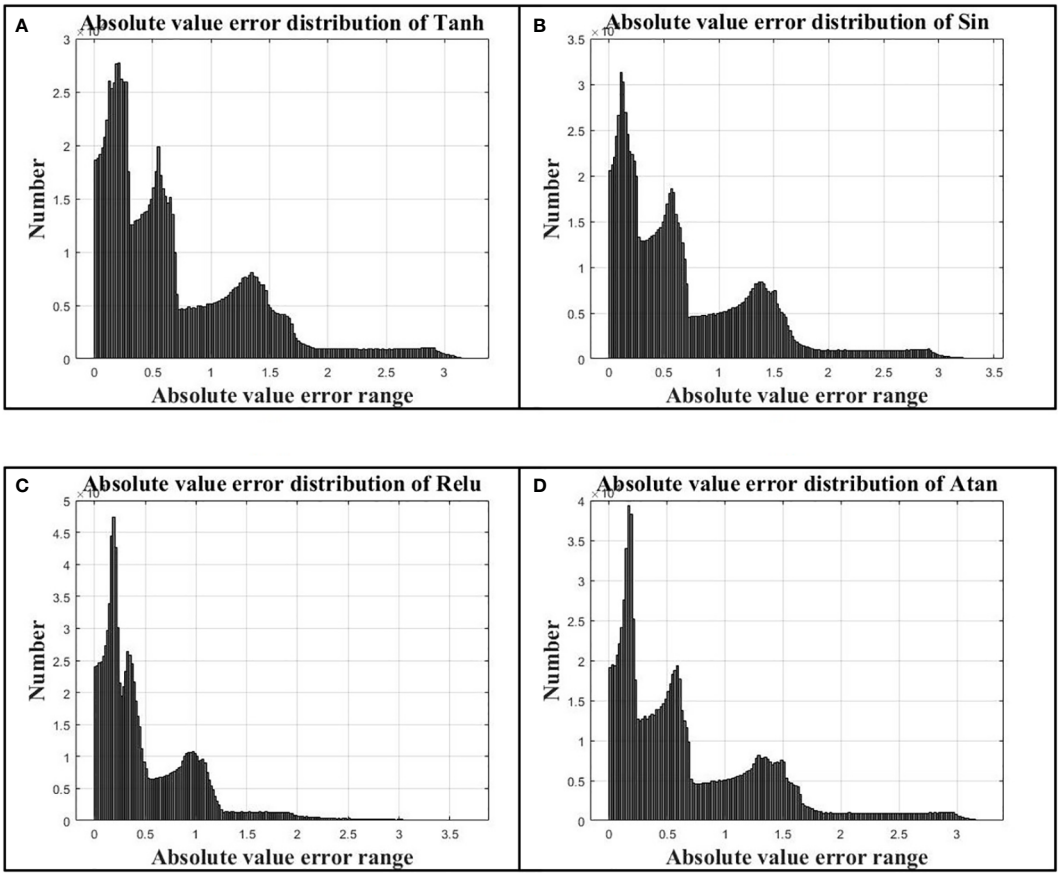


FIGURE 11
(A) is the absolute error distribution using Tanh activation function, **(B)** is the absolute error distribution using Sin activation function, **(C)** is the absolute error distribution using Relu activation function, and **(D)** is the absolute error distribution using Atan activation function.

small samples. The Helmholtz equation, which describes the underwater acoustic field, gives the parameters that affect the acoustic field, such as medium density, medium sound velocity, sound source position, vibration frequency, etc. These parameters cannot be given directly in the loss function as constraints to be trained. For the underwater acoustic field of large-scale medium, the sound velocity and the density of medium change with space, so it is difficult for the network to predict the underwater acoustic field of large-scale medium. The main limitation of the underwater acoustic field prediction model proposed in this paper is that the scale of trained model is small. If the source frequency, medium density, and medium sound velocity change, the new network model needs to be retrained separately for these changing conditions. To solve this problem, the source position, medium density, sound velocity and different boundary conditions can be used as new inputs to train the network model together with the coordinate information. Song et al. (2022) and Alkhalifah et al. (2020) used the similar method to generate wave field solutions of multiple seismic sources with one network, and solved the problem of seismic field adaptation of different seismic sources. In addition, time can also be used as input to add time constraint term to Helmholtz equation, and it can establish a kind of physics-informed neural network for spatial-time cooperative prediction. Finally, the combination of transfer learning and PINN is a research direction to solve the problem of underwater sound field model prediction in different scenes.

Using PINN to predict underwater acoustic field, it is necessary to adjust the structure and training amount of prediction network according to the complexity of acoustic field. By comparing the prediction results of 2D UAFP-PINN and 3D UAFP-PINN models in this paper, the following conclusions can be drawn: with the increase of model dimensions, the complexity of model prediction will increase accordingly, and simply changing the activation function and optimizer cannot effectively improve the prediction accuracy of the model. For acoustic fields with more complex dimensions, it is necessary to increase the complexity of the model with more neurons and hidden layers to adapt to more complex physical environments, so as to achieve better prediction results.

In this paper, it establishes physics-informed neural network to forecast underwater acoustic field. By analyzing several activation functions and the accuracy of the results predicted by the optimizer, it is found that the Relu activation function and the Adam optimizer can accurately predict the sound pressure value of the two-dimensional acoustic field. For three-dimensional space, the accuracy of PINN prediction is lower than the two-dimensional acoustic field prediction model, because the complexity of the problem increases with the increase of the dimension of acoustic field. Therefore, it is also necessary to adjust the number of hidden layers and the number of neurons in the network structure. The two-dimensional and three-dimensional neural network structure in this paper is the same as that of neurons, and subsequent work can be verified in this direction. Compared with the numerical method, this method can adapt to different media environments, has certain physical characteristics, and the prediction accuracy can be improved by adjusting the network structure and parameters, so it is an effective method for underwater acoustic field prediction.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LD: Funding acquisition, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. ZW: Conceptualization, Data curation, Formal Analysis, Methodology, Resources, Visualization, Writing – original draft, Writing – review & editing. ZL: Funding acquisition, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. LW: Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. DH: Investigation, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by several funding sources, including the Shandong Province “Double-Hundred” TalentPlan (WST2020002), the National Key R&D programs (2022YFC2808003;2023YFE0201900) and the Qingdao Natural Science Foundation funded project "Research on Underwater Acoustic Communication Technology of Underwater Unmanned Platform Under Acoustic Compatibility" (23-2-1-100-zyyd-ch).

Acknowledgments

The authors would like to thank the teachers and students of the Underwater Acoustic Laboratory of Shandong University of Science and Technology for their support of this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

References

Ahmed, A. B. R., Younis, M., and De Leon, M. „. H. (2021). "Machine learning based sound speed prediction for underwater networking applications," in *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. (Pafos, Cyprus: IEEE), 436–442. doi: 10.1109/DCOSS52077.2021.00074

Alkhalifah, T., Song, C., and Waheed, U. (2020). "Machine learned Green's functions that approximately satisfy the wave equation," in *the SEG International Exposition and Annual Meeting*. (SEG). doi: 10.1190/segam2020-3421468.1

Al-Safwan, A., Song, C., and Waheed, U. (2021). "Is it time to swish? Comparing activation functions in solving the Helmholtz equation using PINNS," in *82nd EAGE Annual Conference & Exhibition*. (European Association of Geoscientists & Engineers), 2021 (1), 1–5. doi: 10.3997/2214-4609.202113254

Belibassakis, K. A., Athanassoulis, G. A., Papathanasiou, T. K., Filopoulos, S. P., and Markolefas, S. (2014). Acoustic wave propagation in inhomogeneous,layered waveguides based on modal expansions and hp-FEM. *J.Wave Motion.* 51 (6), 1021–1043. doi: 10.1016/j.wavemoti.2014.04.002

Bishop, C. M., and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (New York: Springer).

Borrel-Jensen, N., Engsig-Karup, A. P., and Jeong, C. H. (2021). Physics-informed neural networks for one-dimensional sound field predictions with parameterized sources and impedance boundaries. *J.JASA Express Lett.* 1 (12), 122402. doi: 10.1121/10.0009057

Chen, Z., Cheng, D., Feng, W., and Wu, T. (2013). An optimal 9-point finite difference scheme for the Helmholtz equation with Pml. *J.International J. Numerical Anal. Modeling.* 10 (2), 389–410.

Du, G., Tan, J., Song, P., Xie, C., and Wang, S. (2023). 3D traveltime calculation of first arrival wave using physics-informed neural network. *J.Oil Geophysical Prospecting.* 58 (01), 9–20. doi: 10.13810/j.cnki.issn.1000-7210.2023.01.010

He, J., Li, X., Gao, W., Liu, P., Wang, L., and Tang, R. (2022). "Solving differential equations with neural networks: application to the normal-mode equation of sound field under the condition of ideal shallow water waveguide," in *OCEANS 2022-Chennai* (Chennai, India: IEEE), 1–7. doi: 10.1109/OCEANSChennai45887.2022.9775421

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (2011). *Computational Ocean Acoustics* (New York: Springer).

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *J.Nat Rev. Phys.* 3, 422–440. doi: 10.1038/s42254-021-00314-5

Kiryanov, A. V., Salnikova, E. N., Salnikov, B. A., Yu, N., and Slesarev, (2015). "Modeling and study of main regularities of the formation of sound fields in randomly inhomogeneous underwater waveguides," in *2015 Proceedings of Meetings on Acoustics*, (AIP). vol. 24. doi: 10.1121/2.0000126

Liu, B., Huang, Y., Chen, W., and Lei, J. (2019). *Principles of underwater acoustics* (Beijing: Science Press).

Miller, S. E. (1954). Coupled wave theory and waveguide applications. *J.Bell System Tech. J.* 33 (3), 661–719. doi: 10.1002/j.1538-7305.1954.tb02359.x

Onasami, O., Adesina, D., and Qian, L. (2021). "Underwater acoustic communication channel modeling using deep learning," in *the 15th International Conference on Underwater Networks & Systems*. (New York: Association for Computing Machinery), 1–8. doi: 10.1145/3491315.3491323

Pfau, D., Spencer, J. S., Matthews, A. G. D. G., and Foulkes, W. M. C. (2020). Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *J. Phys. Rev. Res.* 2 (3), 33429. doi: 10.1103/PhysRevResearch.2.033429

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J.Journal Nondestructive Evaluation.* 378, 686–707. doi: 10.1016/j.jcp.2018.10.045

Rotskoff, G. M., Mitchell, A. R., and Vanden-Eijnden, E. (2022). "Active importance sampling for variational objectives dominated by rare events: Consequences for optimization and generalization," in *2nd Mathematical and Scientific Machine Learning Conference*. (Proceedings of Machine Learning Research), 145. 757–780.

Shukla, K., Di Leoni, P. C., Blackshire, J., Sparkman, D., and Karniadakis, G. E. (2020). Physics-Informed neural network for ultrasound nondestructive quantification of surface breaking cracks. *J.Journal Nondestructive Evaluation.* 39 (3), 1–20. doi: 10.1007/s10921-020-00705-1

Song, C., Alkhalifah, T., and Waheed, U. B. (2022). A versatile framework to solve the Helmholtz equation using physics-informed neural networks. *J.Geophysical J. Int.* 228 (3), 1750–1762. doi: 10.1093/gji/ggab434

Teng, D., Chen, H., and Zhu, N. (2010). "Computer simulation of sound field formed around transducer source used in underwater acoustic communication," in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*. (Chengdu, China: IEEE). V1–144-V1-148. doi: 10.1109/ICACTE.2010.5579046

Tu, H., Wang, Y., Lan, Q., Liu, W., Xiao, W., and Ma, S. (2021). A Chebyshev-Tau spectral meth-od for normal modes of underwater sound propagation with a layered marine environment. *J. J. Sound Vibration* 492, 115784. doi: 10.1016/j.jsv.2020.115784

Tu, H., Wang, Y., Liu, W., Ma, X., Xiao, W., and Lan, Q. (2020). A Chebyshev spectral method for normal mode and parabolic equation models in underwater acoustics. *J.Mathematical Problems Eng* 2020, 7461314. doi: 10.1155/2020/7461314

Tu, H., Wang, Y., Yang, C., Wang,., X., Ma, S., Xiao, W., et al. (2022). A novel algorithm to solve for an underwater line source sound field based on coupled modes and a spectral method. *J. J. Comput. Phys.* 468, 111478. doi: 10.1016/j.jcp.2022.111478

Wu, G., Wang, F., Cheng, S., and Zhang, C. (2022). Numerical simulation of forwad and inverse problems in internal sound field analysis based on Physics-Informed Neural Network. *J.Chinese J. Comput. Phys.* 39 (06), 687–698. doi: 10.19596/j.cnki.1001-246x.8520

Yang, P., Zhang, X., and Sun, H. (2023). An imaging algorithm for high-resolution imaging sonar system. *J. Multimedia Tools Appl.*, 1–17. doi: 10.1007/s11042-023-16757-0

Zhang, X. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *J.Multimed Tools Appl.*, 1–18. doi: 10.1007/s11042-023-16992-5

Zhang, Z., Pan, Q., Ji, W., and Huang, F. (2022). Prediction of surface settlements induced by shield tunnelling using physical-information neural networks. *J.Engineering Mechanics.* 39, 13. doi: 10.6052/j.issn.1000-4750.2022.04.0369

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver SAS imagery based on monostatic conversion. *J.IEEE J. Selected Topics Appl. Earth Observations Remote Sensing.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., Yang, P., and Sun, H. (2023). An omega-k algorithm for multireceiver synthetic aperture sonar. *J. Electron. Letters.* 59 (13), e12859. doi: 10.1049/ell2.12859

Zhou, Y., and Luo, W. (2021). A finite element model for underwater sound propagation in 2-D environment. *J.Journal Mar. Sci. Engineering.* 9 (9), 956. doi: 10.3390/jmse9090956

Zhu, W., Xu, K., Darve, E., and Beroza, G. C. (2021). A general approach to seismic inversion with automatic differentiation. *J.Computers Geosciences.* 151, 104751. doi: 10.1016/j.cageo.2021.104751

Zou, J., Liu, C., and Zhao, P. (2023). Research progress of physics-informed neural network in seismic wave modeling. *J.Progress Geophysics.* 38 (01), 430–448. doi: 10.6038/pg2023GG0142

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 08 December 2023
DOI 10.3389/fmars.2023.1290931

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Jia Liu,
Chinese Academy of Sciences (CAS), China
Rongxin Zhang,
Xiamen University, China
Chong Xiao Wang,
Agency for Science, Technology and Research (A*STAR), Singapore

*CORRESPONDENCE
Zhiping Xu
zhipingxu@jmu.edu.cn

RECEIVED 08 September 2023
ACCEPTED 20 November 2023
PUBLISHED 08 December 2023

CITATION
Zheng J, Zhao S, Xu Z, Zhang L and Liu J (2023) Anchor boxes adaptive optimization algorithm for maritime object detection in video surveillance. *Front. Mar. Sci.* 10:1290931. doi: 10.3389/fmars.2023.1290931

Anchor boxes adaptive optimization algorithm for maritime object detection in video surveillance

Jiachun Zheng[1], Shijia Zhao[1], Zhiping Xu[1]*, Lei Zhang[1] and Jiantao Liu[2]

[1]Xiamen Key Laboratory of Marine Intelligent Terminal R&D and Application, The School of Ocean Information Engineering, Jimei University, Xiamen, China, [2]Fujian Electronic Port Co., Ltd, Xiamen, China

With the development of the marine economy, video surveillance has become an important technical guarantee in the fields of marine engineering, marine public safety, marine supervision, and maritime traffic safety. In video surveillance, maritime object detection (MOD) is one of the most important core technologies. Affected by the size of maritime objects, distance, day and night weather, and changes in sea conditions, MOD faces challenges such as false detection, missed detection, slow detection speed, and low accuracy. However, the existing object detection algorithms usually adopt predefined anchor boxes to search and locate for objects of interest, making it difficult to adapt to maritime objects' complex features, including the varying scale and large aspect ratio difference. Therefore, this paper proposes a maritime object detection algorithm based on the improved convolutional neural network (CNN). Firstly, a differential-evolutionary-based K-means (DK-means) anchor box clustering algorithm is proposed to obtain adaptive anchor boxes to satisfy the maritime object characteristics. Secondly, an adaptive spatial feature fusion (ASFF) module is added in the neck network to enhance multi-scale feature fusion. Finally, focal loss and efficient intersection over union (IoU) loss are adopted to replace the original loss function to improve the network convergence speed. The experimental results on the Singapore maritime dataset show that our proposed algorithm improves the average precision by 7.1%, achieving 72.7%, with a detection speed of 113 frames per second, compared with You Only Look Once v5 small (YOLOv5s). Moreover, compared to other counterparts, it can achieve a better speed–accuracy balance, which is superior and feasible for the complex maritime environment.

KEYWORDS

maritime video surveillance, object detection, anchor box, You Only Look Once, adaptive spatial feature fusion

1 Introduction

With the rapid development of global economy and trade, maritime traffic is becoming heavier and denser, bringing a higher risk of maritime traffic accidents. Accurate and rapid maritime object detection is largely beneficial for maritime surveillance, thus effectively reducing the risk of maritime traffic accidents. Affected by maritime objects' distance, weather, and sea conditions, traditional maritime video surveillance mainly relies on manual methods. However, the monitors will inevitably experience visual fatigue when the monitoring behavior continues, leading to false or missed detection of maritime objects and even grave consequences. To solve these problems, an intelligent processing algorithm is adopted to detect object instance in maritime images, and it plays an increasingly important role in maritime object detection tasks.

Traditional maritime object detection algorithms generally follow a three-phase detection framework, namely, horizon detection, static-background subtraction, and foreground segmentation (Lyu et al., 2022). In the first phase, Fefilatyev et al. (2012) utilized Hough transform to detect the horizon position and thus reduced the object search space, and used threshold segmentation to obtain the maritime ship object after image registration. In the second phase, Chen et al. (2018) proposed a Gaussian mixture model to judge the pixels in the foreground part of the image, and then utilized background subtraction and adjacent frame continuity to segment the ship object. In the last phase, Chan (2021) proposed a maritime noise prior method to reduce the interference of noise on the sea surface, and thus improved the accuracy of foreground detection in complex maritime scenarios, and this method was based on a dark channel prior and observation of sea surface characteristics. From these works, it can be found that each phase of this traditional detection framework needs to be designed carefully and manually to ensure the detection performance. Furthermore, these algorithms (Fefilatyev et al., 2012; Chen et al., 2018; Chan, 2021; Zhu et al., 2023) could not efficiently extract high-dimensional semantic information; not only do they need to consume more manpower and time resources, but also the algorithms are easy to be disturbed by the complex marine environment, and they find it difficult to achieve stable maritime object detection.

Benefiting from deep learning (DL) technologies, the object detection algorithms based on the convolutional neural network (CNN) have aroused great interest of scholars. Based on the multi-layer topology structure, CNN can realize the automatic extraction of high-dimensional semantic information of images with stronger anti-interference ability (Simonyan and Zisserman, 2014; Girshick, 2015). CNN-based object detection algorithms are divided into two kinds: two-stage and one-stage algorithms. Generally, the former represented by the R-CNN series has a higher accuracy with a lower speed (Girshick, 2015; Ren et al., 2015; Sun et al., 2021). The latter represented by You Only Look Once (YOLO) series runs faster with a lower accuracy (Redmon and Farhadi, 2018; Bochkovskiy et al., 2020; Ultralytics, 2021; Wang et al., 2023). These DL-based object detection algorithms give a new direction for the research of maritime object detection. Bousetouane and Morris (2016) proposed a Fast-R-CNN-based surveillance algorithm for ship classification and detection in maritime scenarios to improve accuracy with faster speed. To improve marine object detection accuracy, Fu et al. (2021) fused a convolutional attention module in the YOLOv4 framework to enhance valid features and suppress invalid ones. Chang et al. (2022) proposed a modified YOLOv3 model with lower computation complexity through adjustment of input image size, number of convolution kernel, and detection scale, and then introduced the spatial pyramid pooling module to further improve the maritime ship detection accuracy. Recently, many useful technologies have emerged, such as multiple access (Chen et al., 2023; Xie et al., 2023), joint/separated source and channel coding (Xu et al., 2019; Xu et al., 2021; Fang et al., 2023; Xu et al., 2023), index modulation (Dai et al., 2023), and multi-receiver synthetic aperture sonar (Zhang et al., 2021; Zhang. et al., 2022; Yang, 2023; Zhang, 2023; Zhang et al., 2023a; Zhang et al., 2023b; Zhang et al., 2023c). The DL-based object detection algorithms combined with different technologies can build a better maritime object detection system to promote ocean observation.

In summary, DL-based maritime object detection algorithms are simpler, more efficient, and more robust against sea surface noise interference, compared with traditional algorithms. However, existing DL-based maritime object detection algorithms mainly focus on the improvement and optimization of CNN structures, neglecting the characteristics of maritime object instances. To solve this problem, an anchor box adaptive object detection algorithm based on the characteristics of maritime object instances is proposed for maritime video surveillance. The main contributions are as follows:

(1) A differential-evolutionary-based K-means (DK-means) anchor box clustering algorithm is proposed to generate adaptive anchor boxes to adapt for the characteristics of maritime object instances, improving the detection performance without extra computation.

(2) An adaptive spatial feature fusion (ASFF) module is added in the neck network to enhance multi-scale feature fusion to improve the detection performance.

(3) A new loss function that adopts focal loss and efficient intersection over union (IoU) loss is defined for the maritime object characteristics to improve network convergence speed.

(4) On the Singapore maritime dataset, the proposed method achieves 72.7% AP, outperforming the YOLOv5 small (YOLOv5s) by 7.1% with 113 FPS.

(5) The proposed method can perform better than the YOLOv5s in multi-scale maritime objection detection with tighter predicted bounding boxes and fewer number of redundant bounding boxes.

This paper is organized as follows: Section 2 describes the overall research of the proposed methodology in detail. The experimental results including ablation studies, performance comparison, and detection results are shown in Section 3. The paper is concluded in Section 4.

2 Proposed model and optimization

In the maritime surveillance, both real-time and accuracy of object detection need to be considered essentially. Thus, the one-stage object algorithm is chosen as the detection algorithm, which can realize speed–accuracy trade-off efficiently in the maritime object detection. Moreover, compared with other one-stage object algorithms, the YOLOv5s model is lightweight with a higher detection accuracy; thus, we choose it as our baseline model. The purpose of this study is to develop a maritime object detection model considering the characteristics of maritime object instances, and this optimized model can achieve higher accuracy while remaining lightweight.

This section details the main method of the proposed anchor boxes' adaptive objection detection algorithm. Section 2.1 describes the overall structure of our model. The details of the proposed anchor boxes' adaptive algorithm is described in Section 2.2. Sections 2.3 and 2.4 introduce the adaptive spatial feature fusion module and the loss function adopted, respectively.

2.1 Overall structure of our model

Figure 1 provides a detailed depiction of our model's structure, which is composed of three components: the backbone structure, the enhanced neck, and the head. The backbone structure is tasked with extracting features from input images using predefined anchor boxes. Then, the enhanced neck is specifically designed to augment the fusion of these features. Lastly, the head plays the role of predicting maritime objects at three different scales. In the context of object detection algorithms that employ the anchor boxes' mechanism, it is common practice to predefine nine anchor boxes of varying sizes and scales for feature maps. This strategy is implemented to ensure a high level of accuracy in object detection. To adapt to the characteristics of maritime object instances, we adopt the optimized anchor boxes (OABs) as a predefined substitute for the original ones, and the OABs are generated by the proposed DK-means algorithm.

In the training process, the input labeled images are performed by data augmentation operations to increase feature diversity at first. Then, the processed images are performed by feature extraction and subsampling operations in the backbone network part. After three subsampling stages, the backbone generates three different scale feature maps. These feature maps are fed into the enhanced neck network part to reinforce semantic information in shallow feature maps and spatial information in deep feature maps. In the enhanced neck network, the ASFF modules (Liu et al., 2019) are utilized to enhance the multi-scale feature fusion, thus improving the multi-scale detection ability. Finally, the enhanced feature maps are fed into the head network part to obtain the predicted results. The loss values are calculated by comparing them with the label values, and the network parameters are updated through gradient information. Moreover, to make the trained network parameters more consistent with the characteristics of maritime objects, the loss function used in YOLOv5s are also optimized and improved. In the model inference process, the final detection results are obtained after non-maximum suppression (NMS) operation.

2.2 Anchor boxes adaptive algorithm

The predefined anchor boxes in the original detection model are sensitive to object scale: Specifically, smaller anchor boxes are ineffective at detecting larger objects, while larger anchor boxes struggle to accurately capture smaller objects. The mismatch between the aspect ratio of anchor boxes and the objects will result in the decrease in detection accuracy. How to obtain appropriate anchor boxes that can satisfy the characteristics of the maritime object instances is the key to improving the detection accuracy of the maritime object instances.

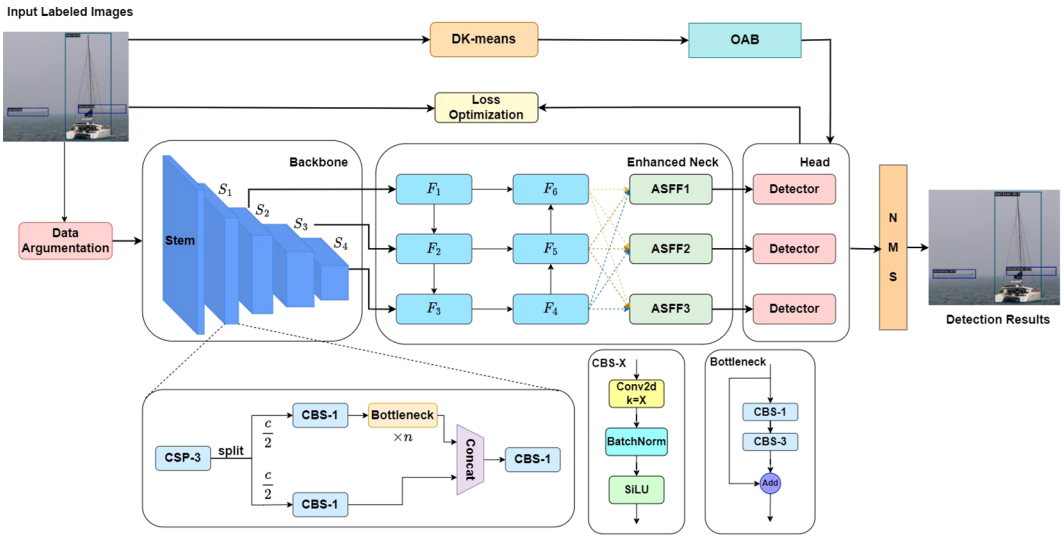


FIGURE 1
The overall structure of the detection model.

2.2.1 Characteristic analysis of maritime object instances

In this part, the Singapore marine dataset (SMD) (Prasad et al., 2017) is taken as an example to analyze the characteristics of the maritime object instances. According to the standard of the Common Objects in Context (COCO) dataset, the maritime object instances in the SMD can be classified into three classes: detection objects with an area less than 32×32 pixels are defined as small objects, detection objects with an area greater than 32 × 32 pixels and less than 96 × 96 pixels are defined as medium objects, and detection objects with an area greater than 96 × 96 pixels are defined as large objects. With these definitions, the scale distributions of the maritime object instances are shown in Figure 2A. From this figure, it can be seen that different types of maritime objects nearly have the different scales. Moreover, the scales of the different types of maritime objects are mainly medium and small. In Figure 2B, it can be found that maritime objects generally have a relatively high aspect ratio. Furthermore, the aspect ratio of the detection object can even reach 17.66 in extreme cases.

From the above analysis, it can be concluded that maritime object detection usually faces the following problems:

(1) The detection object has varying scales.

(2) The detection object has large aspect ratio difference.

2.2.2 Details of the proposed algorithm

In the YOLO series frameworks, the K-means clustering algorithm is usually adopted to generate adaptive anchor boxes, which will be used for training the detection model (Redmon and Farhadi, 2018; Bochkovskiy et al., 2020). However, influenced by the initial clustering centers, the results of the K-means clustering algorithm easily fall into the local optimal solution, making it difficult to generate the optimal predefined anchor boxes. The differential evolution (DE) algorithm (Storn and Price, 1997) is a kind of global search optimization algorithm that achieves evolution from the current population to the next generation through operations such as mutation, crossover, and selection, thus possessing the ability to search for global optimal solutions in the solution space. To solve this problem, the DE algorithm is introduced to reduce the dependence on the initial clustering centers and enhance global search ability. Through this method, the dependence of the K-means algorithm on initial values can be reduced and more robust predefined anchors can be obtained. The improved K-means clustering algorithm with the DE method is named *DK-means clustering algorithm*, which is shown in Algorithm 1. The parameters include the number of iteration N_t, the number of anchor boxes K, the population size N_p, and $\mathcal{N} \triangleq \{1,2, \ldots, N_p\}$. The sample dataset is denoted by D. The tth generation of the population is denoted by $\chi^t = \left\{X_1^t, X_2^t, \cdots, X_{N_p}^t\right\}$, and the i-th candidate is denoted by X_i^t.

The algorithm mainly includes three parts:

1) Line 1: The first part generates the initial population χ^0 according to the sample dataset.
2) Lines 2–8: The second part finds N_p candidates for group anchor boxes by the DE process including mutation and crossover operations at N_t generations.
3) Lines 9–11: The third part chooses the final optimized results for the output group anchor boxes as the clustering centers. Then, divide real anchors boxes into clusters with corresponding clustering centers $X_{i_{best}}^{t+1}$ according to the closest distance principle.
4) Line 12: Return the optimized anchor boxes as the output of this algorithm.

Remark 1: The best group anchors are chosen according to maximizing the IoU values as follows:

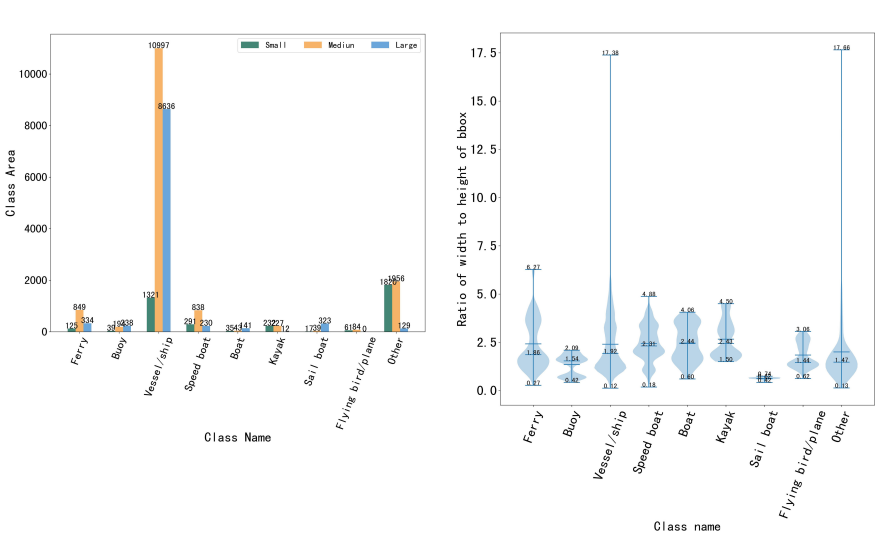


FIGURE 2
Characteristics of maritime object instances in the SMD dataset. (a) area and number of large, medium and small objects of each class and (b) width to height ratio distribution of class and bbox.

$$K_{best} = argmin_{i \in \mathcal{N}} (f(\mathcal{X}^t)), \quad \text{(Eq. 1)}$$

where the function $f(\cdot)$ returns the IoU values, and the K_{best}-th candidate in the population is the optimized results.

2.3 Adaptive spatial feature fusion optimization

In the object detection, multi-scale features can be utilized to improve the model detection performance. When an image is input into CNN, different resolution feature maps can be obtained to detect objects at different scales, according to the different downsampling rates. Shallow feature maps with high resolution are suitable for detecting small objects due to their rich detailed information and small receptive fields, and deep feature maps with low resolution are suitable for detecting large objects due to their strong semantic information and large receptive fields. Therefore, fully utilizing the semantic and detailed information of features at different scales is of great significance for improving the object detection accuracy. To fully utilize multi-scale features, an adaptive spatial feature fusion (ASFF) module is introduced to the neck module to enhance multi-scale feature fusion. Figure 3 shows details of the ASSF module. The feature maps *F*4, *F*5, and *F*6 are fused adaptively and enhanced by the ASFF module at different spatial scales, and then transferred to head module. The feature fusion process can be represented by

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \to l} + \beta_{ij}^l \cdot x_{ij}^{2 \to l} + \gamma_{ij}^l \cdot x_{ij}^{3 \to l}, \quad \text{(Eq. 5)}$$

where y_{ij}^l represents the (i, j)-th feature vector of the output feature y^l, and $x_{ij}^{n \to l}$ represents the (i, j) feature vector after adjusting nth level feature map to the same size of the lth level feature map. $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1]$ represent spatial importance weights of three different scale feature maps at the l-th level feature map, respectively, and $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1$. They can be defined by the softmax function. For example, α_{ij}^l can be calculated by

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}}, \quad \text{(Eq. 6)}$$

Input: D, N_t, K, N_p, real anchor boxes.
Output: Real anchor boxes cluster results.
1: Generate N_p candidates X_i^0 randomly from D for $i \in \mathcal{N}$, and every candidate has K anchor boxes.
Here, $X_i^t = (x_{i,1}, x_{i,2}, \cdots, x_{i,K})$.
2: **for** $t = 1 \to N_t$ **do**
3: **for** $i = 1 \to N_p$ **do**
4: Mutation operations are performed with random $r_1, r_2, r_3 \in \mathcal{N}$, and $r_1 \neq r_2 \neq r_3$:

$$V_i^t = X_{r1}^t + 0.5(X_{r2}^t - X_{r3}^t), \quad \text{(Eq. 2)}$$

where $V_i^t = (v_{i,1}^t, v_{i,2}^t, \cdots, v_{i,j}^t)$, and $j = \{1, 2, \ldots, d\}$ is the dimension number of the sample in the dataset.
5: Crossover operations are performed:

$$u_{i,j}^t = \begin{cases} v_{i,j}^t, with \ \ probability \ \ p_c, \\ x_{i,j}^t, with \ \ probability \ \ 1 - p_c; \end{cases} \quad \text{(Eq. 3)}$$

where $U_i^t = (u_{i,1}^t, u_{i,2}^t, \cdots, u_{i,j}^t)$, and $j = \{1, 2, \ldots, d\}$.
6: Fitness function calculation:

$$X_i^{t+1} = \begin{cases} U_i^t, f(U_i^t) \leq f(X_i^t) \\ X_i^t, f(U_i^t) > f(X_i^t) \end{cases} \quad \text{(Eq. 4)}$$

7: **end for** * N_p *\
8: **end for** * N_t *\
9: $i_{best} = argmax_{i \in N}(X_i^{t+1})$
10: Choose the results of $X_{i_{best}}^{t+1}$ as the clustering centers.
11: Divide real anchors boxes into clusters with corresponding clustering centers $X_{i_{best}}^{t+1}$ according to closest distance principle.
12: **return** cluster results.

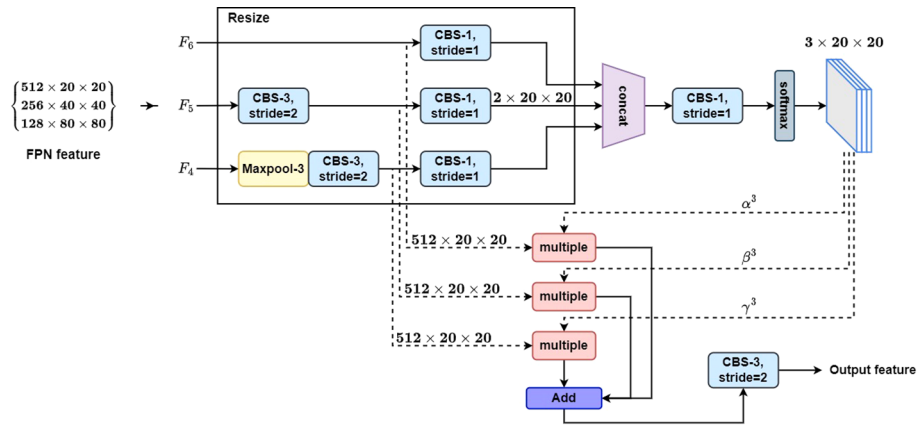


FIGURE 3
Adaptive spatial feature fusion module.

Algorithm 1. DK-means clustering algorithm.

where $\lambda^l_{\alpha_{ij}}$, $\lambda^l_{\alpha_{ij}}$ and $\lambda^l_{\beta_{ij}}$ are the control parameters of the softmax function, which can be trained and learned by 1 × 1 convolution operation.

2.4 Loss function optimization

The loss function $\mathcal{L}$ used in our improved network is a combination of three loss functions: the classification loss $\mathcal{L}_{cls}$, the localization loss $\mathcal{L}_{reg}$, and the confidence loss $\mathcal{L}_{obj}$. It is represented by

$$\mathcal{L} = \lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{reg} + \lambda_3 \mathcal{L}_{obj}, \quad \text{(Eq. 7)}$$

where λ_1, λ_2, and λ_3 represent balance factors, whose values are 1.0, 0.05, and 0.1, respectively. They are used to control the impact of different loss functions on the network training process. In the YOLOv5 model, cross entropy (CE) loss function $\mathcal{L}_{ce}$ is adopted as classification and confidence loss functions, and the complete IoU (CIoU) loss is adopted as localization loss. The CE loss function is represented by

$$\mathcal{L}_{ce}(p, y) = -y log(p) - (1 - y) log(1 - p), \quad \text{(Eq. 8)}$$

where y denotes the true value of the label category and $y \in \{\pm 1\}$, and p denotes category prediction probability when $y = 1$ and $p \in [0,1]$.

In the task of bounding box regression, the localization loss function is divided into two categories: n-norm-based and IoU-based loss (Tian et al., 2022). To reduce the sensitivity of the model to the scale changes of object and improve convergence speed of the model, the complete IoU (CIoU) loss is adopted as localization loss in the YOLOv5 model and it is expressed by

$$\mathcal{L}_{ciou}(b^{pr}, b^{gt}) = 1 - IoU + \frac{\rho^2(b^{pr}, b^{gt})}{c^2} + \lambda v, \quad \text{(Eq. 9)}$$

where IoU represents the intersection over union between the predicted bounding box and the ground truth bounding box, b^{pr} and b^{gt} represent centers of the predicted bounding box and the ground truth bounding box respectively, and c is the diagonal length of the smallest covering box. $\rho(\cdot) = \|b^{pr} - b^{gt}\|_2$ represents the Euclidean distance between the center points of two bounding boxes, λ represents the balance factor, and v represents the consistency of the aspect ratio between the predicted bounding box and the ground truth bounding box. IoU can be described as Figure 4 and is expressed by

$$IoU = \frac{|b^{pr} \cap b^{gt}|}{|b^{pr} \cup b^{gt}|}, \quad \text{(Eq. 10)}$$

and v is expressed by

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w^{pr}}{h^{pr}}\right), \quad \text{(Eq. 11)}$$

where w^{gt} and w^{pr} represent the width of the ground truth and the predicted bounding boxes respectively, and h^{gt} and h^{pr} represent the height of the ground truth and the predicted bounding boxes respectively. Thus, Equation 7 becomes

$$\mathcal{L} = \lambda_1 \mathcal{L}^{cls}_{ce} + \lambda_2 \mathcal{L}^{reg}_{ciou} + \lambda_3 \mathcal{L}^{obj}_{ce} \quad \text{(Eq. 12)}$$

where $\mathcal{L}^{cls}_{ce}$ and $\mathcal{L}^{obj}_{ce}$ represent the cross entropy loss adopted for the classification loss and the confidence loss functions, and $\mathcal{L}^{reg}_{ciou}$ represents the CIoU loss adopted for the localization loss function.

2.4.1 Focal loss function

To improve detection accuracy, the anchor box detection mechanism usually requires a dense set of the distribution of anchor boxes in images, which can easily lead to an imbalanced problem between positive and negative samples. To alleviate this imbalanced problem, a focal loss (FL) function (Lin et al., 2017) is introduced. Define p_t as

$$p_t = \begin{cases} p, if \quad y = 1, \\ 1 - p, otherwise\,. \end{cases} \quad \text{(Eq. 13)}$$

Thus, Equation 8 can be rewritten as

$$\mathcal{L}_{ce}(p_t) = -\log(p_t)\,. \quad \text{(Eq. 14)}$$

The FL function is expressed by

$$\mathcal{L}_{focal} = -\alpha_t(1 - p_t)^{\gamma}\log(p_t), \quad \text{(Eq. 15)}$$

where $(1 - p_t)^{\gamma}$ represents an adjustment factor, γ represents an adjustable focusing parameter, and α_t represents a balanced parameter. Equation 7 can be modified as

$$\mathcal{L} = \lambda_1 \mathcal{L}^{cls}_{focal} + \lambda_2 \mathcal{L}^{reg}_{ciou} + \lambda_3 \mathcal{L}^{obj}_{ce}, \quad \text{(Eq. 16)}$$

where $\mathcal{L}^{cls}_{focal}$ represents the FL function, which is adopted for the classification loss.

2.4.2 Efficient IoU loss function

From Equation 11, it can be found that v just considers the aspect ratio difference between the predicted bounding box and the ground truth bounding box, ignoring the difference between specific values of width and height. In this way, the penalty term v is almost no longer effective, as the width and height cannot be simultaneously enlarged or reduced. To make the detection model more suitable for the maritime objects with varying scales in the marine environment, we adopt the idea of the efficient IoU (EIoU) (Zhang Y.-F.et al., 2022), and the v is modified as

$$v = \frac{\rho^2(w^{pr}, w^{gt})}{C_w^2} + \frac{\rho^2(h^{pr}, h^{gt})}{C_h^2}, \quad \text{(Eq. 17)}$$

where C_w^2 and C_h^2 are the width and height of the smallest enclosing box covering the ground truth and predicted bounding boxes, and $\rho^2(w^{pr}, w^{gt})$ and $\rho^2(h^{pr}, h^{gt})$ are the width and height difference between the ground truth bounding box and the predicted bounding box, respectively. According to Equation 17, Equation 16 can be further modified as

$$\mathcal{L} = \lambda_1 \mathcal{L}^{cls}_{focal} + \lambda_2 \mathcal{L}^{reg}_{eiou} + \lambda_3 \mathcal{L}^{obj}_{ce}, \quad \text{(Eq. 18)}$$

where $\mathcal{L}^{reg}_{eiou}$ is the EIoU loss function, and the only difference between the EIoU used in this paper and the CIoU is the penalty term. The EIoU loss will bring faster convergence speed and better localization effect.

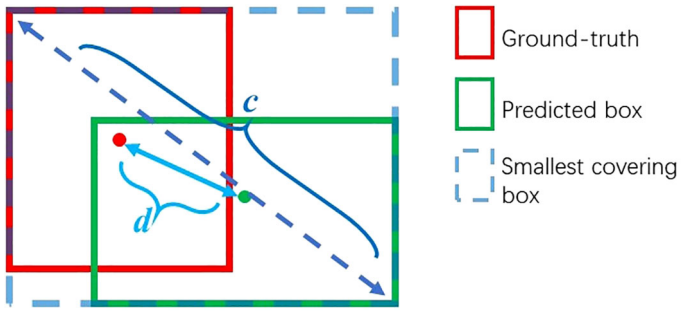


FIGURE 4
Description of IoU.

3 Experimental result and analysis

3.1 Dataset construction

In this section, the experiments adopt SMD (Prasad et al., 2017), which contains large video data with labeled bounding boxes. Detection objects in SMD include nine categories, namely, ferry, buoy, vessel/ship, speed boat, boat, kayak, sail boat, flying bird/plane, and other. In the DL-based object detection algorithms, the construction of a sizable and representative dataset is the first and important step. By sampling images from the SMD at one time per five frames, 6,350 maritime images are obtained, and each image has a resolution of 1,920 × 1,080 pixels. They are split into train, validation, and test sets at a ratio of 6:2:2 with COCO style. The dataset construction process is shown in Figure 5.

3.2 Experimental environment

The experiments are all carried out using PyTorch 1.12.1 and CUDA 11.3 on an NVIDIA RTX 3090 GPU and an Intel Core i9-10920X CPU. The details of the experimental hardware and software environment are shown in Table 1. A momentum gradient descent algorithm with a momentum value of 0.937 is adopted in the train model. In the train process, the input image is fixed at 640 × 640, the batch size is set to 32, and the total number of train epochs is set to 150. The initial learning rate is set to 0.0025 and the linear decline strategy is used as the learning rate attenuation strategy. Moreover, to maintain the stability of the model in the initial train stage, a warm-up training strategy is adopted in the first three epochs to gradually increase the learning rate from 0 to the initial learning rate. Moreover, for each stage in Figure 1 (represented as S_1,S_2,S_3,S_4), we respectively configure the number of bottlenecks as [3, 6, 9, 3], and perform downsampling operation within the first convolution layer of each stage. In addition, we adhere to the yolov5s configuration, setting the scaling factors for width and depth to 0.5 and 0.33, respectively.

3.3 Evaluation metrics

Average precision (AP) (Padilla et al., 2021) is utilized as an indicator to evaluate the accuracy of the maritime object detection algorithm. AP can be expressed by

$$\mathrm{AP} = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 P(R)\mathrm{dR}, \qquad \text{(Eq. 19)}$$

where N represents the number of object categories, and P and R represent precision and recall rate respectively, which are expressed by

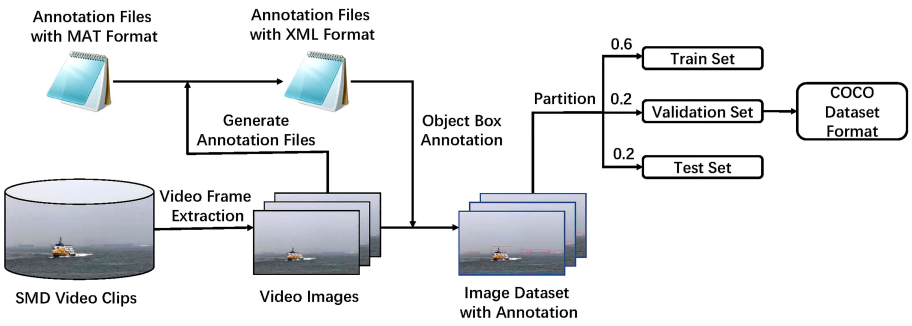


FIGURE 5
The construction process of the dataset.

TABLE 1 Experimental hardware/software environment.

Configuration	Project Model	Parameter
Hardware Environment	CPU RAM	Intel Core i9-10920X 64 GB
	GPU	NVIDIA RTX3090 (24GB)
Software Environment	System Pytorch	Ubuntu 20.04 V1.12.1
	Python	V3.8
	CUDA	V11.3

$$P = \frac{TP}{TP + FP}, \quad \text{(Eq. 20)}$$

$$R = \frac{TP}{TP + FN}, \quad \text{(Eq. 21)}$$

where TP represents the number of correct predicted positive samples, FP represents the number of negative samples predicted as positive samples, and FN represents the number of positive samples predicted as negative samples. From Equations 19–21, it can be inferred that P and R measure the model's ability to accurately predict and locate objects respectively, and AP is the comprehensive evaluation of these two indicators. Here, the AP value means that the IoU threshold is set from 0.50 to 0.95 with a 0.05 step. AP_{50} and AP_{75} mean that the IoU threshold is set to 0.5 and 0.75, respectively. AP_S, AP_M, and AP_L denote small, medium, and large ground-truth objects, respectively.

Moreover, frames per second (FPS) is used to measure model detection speed. Model size and floating point operations (FLOPs) are used for evaluating the occupied memory of the model and calculation complexity, respectively, and they are as follows:

$$\text{Params} = K_W \cdot K_H \cdot C_{in} \cdot C_{out}, \quad \text{(Eq. 22)}$$

$$\text{FLOPs} = K_W \cdot K_H \cdot C_{in} \cdot C_{out} \cdot F_W \cdot F_H, \quad \text{(Eq. 23)}$$

Where K_W and K_H represent the width and height of the convolutional kernel, respectively, C_{in} and C_{out} represent the number of input and output channels, respectively, and F_W and F_H represent the width and height of the feature map.

3.4 Detection performance comparison with different anchor box algorithms

In this part, the predefined anchor boxes, which are obtained from different clustering algorithms, are shown in Table 2 and the detection performance on the SMD dataset with different anchor boxes is shown in Table 3. From Table 3, it can be seen that YOLOv5s with predefined anchor boxes generated by the K-means and DK-means algorithm improve AP value by 1.9% and 3.3%, respectively, when compared with the original method, and the DK-means algorithm can perform better than the K-means algorithm by 1.4%. Moreover, when the threshold of IoU increases to 0.75, the DK-means algorithm can achieve 6.0% AP value improvement, compared with the original method. Furthermore, in the DK-means method, the small object detection performance AP_S is improved by 6.5% and the medium object detection performance AP_M is improved by 4.3% with a slight large object detection performance loss, when compared with the original method.

3.5 Ablation studies

This section presents the ablation studies to illustrate the effects of FL function, ASFF module, and the DK-means algorithm in the detection model. The ablation studies are shown in Table 4, where ✓ and ✗ denote the detection model with or without relevant modules or algorithm, respectively. The first line in this table show the YOLOv5s (baseline model) without any improvement. When single EIoU, FL function, or the ASFF module is adopted, the AP values are increased by approximately 0.4%, 2.2%, or 1.7% with improvement of multi-scale detection ability, respectively. As shown in the last line in this table, when the DK-mean algorithm is further adopted, the AP value is increased by 3.3%. Compared to the standard detection model, the optimized scheme can achieve 4.8% improvement and the small and medium object detection performances are also increased by 10.9% and 4.8% respectively.

3.6 Detection performance comparison with other object techniques

In this part, the proposed detection model performance comparison with other object detection techniques is shown in Table 5. The comparison models include YOLOX-s (Ge et al., 2021), YOLOv8-s (Ultralytics, 2023), YOLOv7-tiny (Wang et al., 2023), YOLOv5s (Ultralytics, 2021), YOLOv5m (Ultralytics, 2021), and YOLOv5l (Ultralytics, 2021). From Table 5, it can be seen that the proposed detection model can achieve best results between accuracy and detection speed. The AP value of our model is 0.6%, 1.6%, 1.6%, 3.7%, 4.8%. and 5.7% higher than that of YOLOv5l, YOLOv5m, YOLOv8s, YOLOX-s, YOLOv5s, and YOLOv7-tiny, respectively, which means that our proposed model has the best accuracy among

TABLE 2 The results of different anchor box algorithms.

Algorithm	80 × 80	40 × 40	20 × 20
	(10,13)	(30,61)	(116,90)
Original method	(16,30)	(62,45)	(156,198)
	(33,23)	(59,119)	(373, 326)
	(11,9)	(54,15)	(46,53)
K-means	(22,8)	(34,26)	(149,52)
	(20,15)	(87,23)	(263,94)
	(11,9)	(45,13)	(47, 39)
DK-means	(19, 7)	(31,22)	(135,41)
	(17,12)	(69,18)	(189,70)

TABLE 3 The experimental results of different anchor box algorithms on the SMD test dataset.

Algorithms	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Original method	67.9%	96.1%	73.3%	52.2%	61.8%	**87.3%**
K-means	69.8%	97.2%	78.1%	54.3%	64.4%	86.8%
DK-means	**71.2%**	**97.5%**	**79.3%**	**58.7%**	**66.1%**	86.6%

The bold values mean that the best values in the relevant columns of the table.

TABLE 4 Ablation experiments.

EIoU	FL function	ASFF module	DK-means	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
✗	✗	✗	✗	67.9%	96.1%	73.3%	52.2%	61.8%	87.3%
✓	✗	✗	✗	68.3%	96.2%	75.7%	54.8%	61.1%	87.4%
✗	✓	✗	✗	70.1%	96.2%	76.9%	56.8%	64.1%	87.5%
✗	✗	✓	✗	69.6%	97.0%	74.7%	55.0%	62.9%	87.8%
✗	✗	✗	✓	71.2%	97.5%	79.3%	58.7%	66.1%	86.6%
✓	✓	✗	✗	69.9%	96.6%	77.3%	60.0%	63.7%	87.7%
✗	✓	✓	✗	71.1%	97.2%	77.4%	58.6%	64.5%	**88.5%**
✓	✓	✓	✓	**72.7%**	**97.7%**	**81.1%**	**63.1%**	**66.6%**	87.0%

The bold values mean that the best values in the relevant columns of the table.

these counterparts. The proposed model's FPS value is 113 and detects faster than YOLOv51, YOLOX-s, and YOLOv5s. Moreover, the model size of the proposed model is smaller than YOLOv5m and YOLOv5l, and the FLOPs of the proposed model are fewer than YOLOX-s, YOLOv5m, YOLOv51, and YOLOv8s. All these show that the proposed model can achieve satisfactory results among accuracy, detection speed, model size, and calculation complexity.

The detection results on the SeaShips dataset (Shao et al., 2018) are shown in Table 6. The SeaShips is a large dataset dedicated to maritime shipping detection, and it includes 31,455 images with 7,000 open-source images. We also divided the open-source part of the Seaships dataset in a 6:2:2 ratio with a COCO format, which is consistent with the processing method of the SMD dataset. Here, the proposed method is compared with the baseline YOLOv5s. From this table, we can see that our proposed model performs better than the YOLOv5s in many aspects. These means that our model can perform well on other datasets.

TABLE 5 The experi|mental results with different object detection algorithms on the SMD dataset.

Methods	AP	Model Size	FLOPs	FPS
YOLOX-s (Ge et al., 2021)	69.0%	8.97M	13.40G	86
YOLOv8s (Ultralytics, 2023)	71.1%	11.14M	14.28G	182
YOLOv7-tiny (Wang et al., 2023)	67.0%	**6.23M**	**6.89G**	**143**
YOLOv5s (Ultralytics, 2021)	67.9%	7.24M	8.27G	131
YOLOv5m (Ultralytics, 2021)	71.1%	21.19M	24.53G	104
YOLOv5l (Ultralytics, 2021)	72.1%	46.56M	54.65G	82
Proposed model	**72.7%**	10.14M	10.81G	113

The bold values mean that the best values in the relevant columns of the table.

3.7 Detection results on the SMD dataset

Figure 6 shows the detection results of the proposed model in different weather and light intensity conditions on the SMD dataset.

Figure 7 shows the detection result comparisons between YOLOv5s and the proposed model. The subfigures in the first column are the detection results of YOLOv5s, and those in the second column are the detection results of the proposed model. From this figure, it can be seen that the proposed detection model can achieve more accurate detection of maritime object instances with tighter predicted bounding boxes and fewer number of redundant bounding boxes.

4 Conclusions

In this paper, we propose a maritime object detection algorithm for maritime video surveillance. At first, a DE-based

TABLE 6 The experimental results on the Seaships dataset.

Methods	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv5s (Ultralytics, 2021)	68.2%	97.7%	81.7%	–	51.2%	69.3%
Proposed model	80.1%	98.9%	92.7%	–	61.5%	81.3%

FIGURE 6
Detection results of the proposed model in various environment conditions.

K-means anchor box clustering algorithm, considering the maritime object characteristics, is proposed to realize the adaptive anchor boxes. Then, to enhance the multi-scale feature fusion, the neck network adopts the ASFF module. Lastly, the loss function integrates the focal loss and efficient IoU loss is defined to alleviate the samples' imbalanced problem and consider the varying scales of the maritime objects. All consider the complexity characteristic of maritime objects. The ablation studies show that the proposed algorithm meets the multi-scale maritime object detection performance. The experimental results show that AP can reach 72.7%, which is 4.8% higher than YOLOv5s, and better than YOLOv5m and YOLOv5l; this algorithm does not occupy high additional computational resources, and its inference speed can reach 113 FPS, which can achieve better speed–accuracy balance.

How to make the proposed model lightweight for resource-constrained devices with less detection accuracy loss will be a meaningful topic for future research.

FIGURE 7
Detection result comparisons of different object detection algorithms.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JZ: Writing – original draft, Writing – review & editing, Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization. SZ: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization. ZX: Writing – original draft, Writing – review & editing, Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization. LZ: Data curation, Visualization, Writing – review & editing. JL: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Xiamen Ocean and Fishery Development Special Fund Project (No. 21CZB013HJ15), the Xiamen Key Laboratory of Marine Intelligent Terminal R&D and Application (No. B18208), the Fund Project of Jimei University (No. ZP2020042), the Fujian Province Young and Middle-aged Teacher Education Research Project (No. JAT220182), Jimei University Startup Research Project (No. ZQ2022039) and the Scientific Research Foundation of Jimei University (No. XJ2022000201).

Conflict of interest

Author JL is employed by Fujian Electronic Port Co., Ltd, Xiamen, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Bousetouane, F., and Morris, B. (2016). "Fast cnn surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 242–248 (IEEE).

Chan, Y.-T. (2021). Maritime filtering for images and videos. *Signal Processing: Image Communication* 99, 116477.

Chang, L., Chen, Y.-T., Wang, J.-H., and Chang, Y.-L. (2022). Modified yolov3 for ship detection with visible and infrared images. *Electronics* 11, 739. doi: 10.3390/electronics11050739

Chen, P., Shi, L., Fang, Y., Lau, F. C., and Cheng, J. (2023). Rate-diverse multiple access over Gaussian channels. *IEEE Trans. Wireless Commun*. doi: 10.1109/TWC.2022.3233798

Chen, Z., Yang, J., Chen, Z., and Kang, Z. (2018). Ship target detection algorithm for maritime surveillance video based on gaussian mixture model. *J. Physics: Conf. Ser.* 1098, 012021.

Dai, L., Fang, Y., Guan, Y. L., and Guizani, M. (2023). Design of protograph LDPC-coded MIMO-VLC systems with generalized spatial modulation. *China Communication*.

Fang, Y., Zhuo, J., Ma, H., Mumtaz, S., and Li, Y. (2023). Design and analysis of a new indexmodulation-aided DCSK system with frequency-and-time resources. *IEEE Trans. Vehicular Technol.*

Fefilatyev, S., Goldgof, D., Shreve, M., and Lembke, C. (2012). Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system. *Ocean Eng.* 54, 1–12. doi: 10.1016/j.oceaneng.2012.06.028

Fu, H., Song, G., and Wang, Y. (2021). Improved yolov4 marine target detection combined with cbam. *Symmetry* 13, 623. doi: 10.3390/sym13040623

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint*.

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*. 2980–2988.

Liu, S., Huang, D., and Wang, Y. (2019). Learning spatial fusion for single-shot object detection. *arXiv preprint arXiv:1911.09516*.

Lyu, H., Shao, Z., Cheng, T., Yin, Y., and Gao, X. (2022). "Sea-surface object detection based on electro-optical sensors: A review," in *IEEE Intelligent Transportation Systems Magazine*. 2–27.

Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., and Da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics* 10, 279. doi: 10.3390/electronics10030279

Prasad, D. K., Rajan, D., Rachmawati, L., Rajabally, E., and Quek, C. (2017). Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Trans. Intelligent Transportation Syst.* 18, 1993–2016. doi: 10.1109/TITS.2016.2634580

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28.

Shao, Z., Wu, W., Wang, Z., Du, W., and Li, C. (2018). Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Trans. multimedia* 20, 2593–2604. doi: 10.1109/TMM.2018.2865686

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Storn, R., and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J. Global optimization* 11, 341–359. doi: 10.1023/A:1008202821328

Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., et al. (2021). "Sparse r-cnn: End-to-end object detection with learnable proposals," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14454–14463.

Tian, Y., Su, D., Lauria, S., and Liu, X. (2022). Recent advances on loss functions in deep learning for computer vision. *Neurocomputing* 497, 129–158. doi: 10.1016/j.neucom.2022.04.127

Ultralytics. (2021). *YOLOv5*. Available at: https://github.com/ultralytics/yolov5.

Ultralytics. (2023). *YOLOv8*. Available at: https://github.com/ultralytics/ultralytics.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023). "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7464–7475.

Xie, K., Cai, G., Kaddoum, G., and He, J. (2023). Performance analysis and resource allocation of STAR-RIS aided wireless-powered NOMA system. *IEEE Trans. Commun.* doi: 10.1109/TCOMM.2023.3292471

Xu, Z., Song, D., Zheng, J., and Wang, L. (2023). Joint-decoding-complexity-oriented collaborative design for joint source-channel coding system based on double protograph-LDPC codes. *Sci. China Inf. Sci.* 66, 189301. doi: 10.1007/s11432-022-3765-2

Xu, Z., Wang, L., and Chen, G. (2021). Joint coding/decoding optimization for DC-BICM system: Collaborative design. *IEEE Commun. Lett.* 25, 2487–2491. doi: 10.1109/LCOMM.2021.3081678

Xu, Z., Wang, L., Hong, S., Lau, F. C., and Sham, C.-W. (2019). Joint shuffled scheduling decoding algorithm for DP-LDPC codes-based JSCC systems. *IEEE Wireless Commun. Lett.* 8, 1696–1699. doi: 10.1109/LWC.2019.2937766

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Appl.*, 1–17. doi: 10.1007/s11042-023-16757-0

Zhang, M., Cai, G., and He, J. (2023a). Performance analysis of MAC protocols for single-cell LoRa network with power control. *IEEE Commun. Lett.* doi: 10.1109/LCOMM.2023.3314030

Zhang, X. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *Multimedia Tools Appl.* doi: 10.1007/s11042-023-16992-5

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver sas imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., Yang, P., Huang, P., Sun, H., and Ying, W. (2022). Wide-bandwidth signal-based multireceiver sas imagery using extended chirp scaling algorithm. *IET Radar Sonar Navigation* 16, 531–541. doi: 10.1049/rsn2.12200

Zhang, X., Yang, P., and Sun, H. (2023b). An omega-k algorithm for multireceiver synthetic aperture sonar. *Electron. Lett.* 59, e12859. doi: 10.1049/ell2.12859

Zhang, X., Yang, P., and Zhou, M. (2023c). Multireceiver SAS imagery with generalized PCA. *IEEE Geosci. Remote Sens. Lett.* doi: 10.1109/LGRS.2023.3286180

Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi: 10.1016/j.neucom.2022.07.042

Zhu, J., Song, Y., Jiang, N., Xie, Z., Fan, C., and Huang, X. (2023). Enhanced doppler resolution and sidelobe suppression performance for golay complementary waveforms. *Remote Sens.* 15. doi: 10.3390/rs15092452

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 12 December 2023
DOI 10.3389/fmars.2023.1306229

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Zhichao Lv,
Shandong University of Science and Technology, China
Irfan Hussain,
Khalifa University, United Arab Emirates
Zhiping Xu,
Jimei University, China

*CORRESPONDENCE
Jianxun Tang
alio@mails.guet.edu.cn

RECEIVED 03 October 2023
ACCEPTED 13 November 2023
PUBLISHED 12 December 2023

CITATION
Chen Z, Tang J, Qiu H and Chen M (2023) MGFGNet: an automatic underwater acoustic target recognition method based on the multi-gradient flow global feature enhancement network.
Front. Mar. Sci. 10:1306229.
doi: 10.3389/fmars.2023.1306229

MGFGNet: an automatic underwater acoustic target recognition method based on the multi-gradient flow global feature enhancement network

Zhe Chen[1,2], Jianxun Tang[3]*, Hongbin Qiu[1,2] and Mingsong Chen[1,3]

[1]School of Information and Communication, Guilin University of Electronic Technology, Guilin, Guangxi, China, [2]Cognitive Radio and Information Processing Key Laboratory Authorized by China's Ministry of Education Foundation, Guilin University of Electronic Technology, Guilin, Guangxi, China, [3]School of Ocean Engineering, Guilin University of Electronic Technology, Beihai, Guangxi, China

The recognition of underwater acoustic targets plays a crucial role in marine vessel monitoring. However, traditional underwater target recognition models suffer from limitations, including low recognition accuracy and slow prediction speed. To address these challenges, this article introduces a novel approach called the Multi-Gradient Flow Global Feature Enhancement Network (MGFGNet) for automatic recognition of underwater acoustic targets. Firstly, a new spectrogram feature fusion scheme is presented, effectively capturing both the physical and brain-inspired features of the acoustic signal. This fusion technique enhances the representation of underwater acoustic data, resulting in more accurate recognition results. Moreover, MGFGNet utilizes the multi-gradient flow network and incorporates a multi-dimensional feature enhancement technique to achieve fast and precise end-to-end recognition. Finally, a loss function is introduced to mitigate the influence of unbalanced data sets on model recognition performance using Taylor series. This further enhances model recognition performance. Experimental evaluations were conducted on the DeepShip dataset to assess the performance of our proposed method. The results demonstrate the superiority of MGFGNet, achieving a recognition rate of 99.1%, which significantly surpasses conventional methods. Furthermore, MGFGNet exhibits improved efficiency compared to the widely used ResNet18 model, reducing the parameter count by 51.28% and enhancing prediction speed by 33.9%. Additionally, we evaluated the generalization capability of our model using the ShipsEar dataset, where MGFGNet achieves a recognition rate of 99.5%, indicating its superior performance when applied to unbalanced data. The promising results obtained in this study highlight the potential of MGFGNet in practical applications.

KEYWORDS

underwater acoustic target recognition, underwater acoustic signal processing, feature enhancement, deep learning, feature fusion

1 Introduction

With the development of artificial intelligence, there is an increasing focus on utilizing AI-based methods to address research challenges in aquaculture. Fisheries and aquaculture constitute a global industry valued at $200 billion (Gladju et al., 2022). As this industry continues to expand, traditional processes involving essential technologies such as aquaculture environment monitoring, feeding, and fish behavior surveillance (Wu et al., 2022) incur significant costs. Hence, the urgent need arises to employ artificial intelligence technologies to enhance the economic, social, and environmental sustainability of the fish supply chain (Lim, 2022). AI-based aquaculture technologies primarily encompass environmental monitoring, intelligent feeding, biological behavior monitoring, and fishing vessel motion tracking (Setiyowati et al., 2022).

Environmental monitoring relies on water quality management systems to control the health of aquaculture water, preventing widespread diseases or issues such as slow growth in fish fry due to water quality problems (Hu et al., 2022). Koparan et al. (2018) developed an intelligent unmanned aerial vehicle to continuously monitor the water quality of a 1.1-hectare pond through intelligent sampling and analysis. Given that feed costs constitute over 60% of aquaculture expenses (Boyd et al., 2022), effective control of feed distribution is crucial. Lim and Whye, (2023) proposed a system that monitors fish behavior by detecting water wave vibrations caused by competitive feeding, thereby assessing fish hunger levels and significantly reducing feed consumption.

Biological behavior monitoring encompasses various aspects. Ahmed et al. (2022) and Darapaneni et al. (2022) employed computer vision and underwater optical imaging techniques, respectively, to obtain underwater images of fish activities for disease detection and prevention before widespread mortality. Fishing activities require strict control over timing and quantity globally. Bradley et al. (2019) and Kritzer, (2020) integrated automatic identification with artificial intelligence technology, utilizing underwater acoustic target recognition systems to track fishing vessel movements in real-time and predict their fishing activities, ensuring legitimacy.

In summary, due to the rapid development of computer vision technology effectively addressing the first three issues in aquaculture, our research focus shifts towards utilizing underwater acoustic target recognition technology for vessel motion monitoring.

Underwater acoustic target recognition involves collecting target radiated noise using hydrophones, analyzing and processing the data to discern target types (Ma et al., 2022). It holds significant importance in maritime vessel monitoring and underwater vehicle detection. Acoustic target recognition models typically consist of two modules: feature extraction and feature classification (Hong et al., 2021), and research in this field revolves around these modules.

Traditional methods of underwater acoustic target feature extraction can be categorized into signal physics-based and brain-like computing methods (Zhu et al., 2023). Signal physics-based methods rely on basic characteristics, temporal features, and non-Gaussian characteristics of underwater acoustic signals (Yao X. et al., 2023). This includes time-domain features like zero-crossing distribution, frequency-domain features like cepstral analysis (Zhu et al., 2022), and joint time-frequency domain features such as wavelet transforms (Han et al., 2022; Liu et al., 2022; Tian et al., 2023). Brain-like computing features for underwater acoustic signals include Mel-frequency cepstral coefficients (MFCC) simulating nonlinear processing of the human ear (Di et al., 2023) and Gammatone filtering simulating peripheral auditory processing (Zhou et al., 2022). Traditional classifier models include case-based reasoning (Ali et al., 2018) and perceptron neural networks (Linka and Kuhl, 2023). While traditional methods provide explicit directional analysis based on the physical meaning of underwater acoustic signals, they depend on prior knowledge and exhibit poor model generalization (Xiao et al., 2021).

Deep learning models, including Convolutional Neural Networks (CNN) (Yao Q. et al., 2023), provide new solutions for underwater acoustic target recognition (Jin and Zeng, 2023). Wang and Zeng (2015). demonstrated the feasibility of CNN models in underwater acoustic target recognition by testing them on three different measured acoustic targets. Studies have validated the applicability of deep learning in feature extraction. Huang et al. (2021) used autoassociative neural networks (AANN) to directly process mixed time-domain information of raw acoustic data without prior information, filtering ocean background noise, and obtaining effective spectral features of underwater acoustic targets. Additionally, research on deep learning-based classifiers is active. Li J. et al. (2022) designed AResNet to enhance feature extraction capability by increasing the width of the ResNet (He et al., 2016) residual network and incorporating channel attention mechanisms. Yang S. et al. (2023) developed LW-SEResNet10 to improve target recognition accuracy by reducing the number of ResNet residual structures and adding attention mechanisms. These classifiers operate similarly, performing feature extraction first and then inputting the features to obtain classification results.

Despite the advantages of existing deep learning-based underwater acoustic target recognition models in addressing some shortcomings of traditional methods, several challenges persist:

1. Existing models have independent feature extraction and classifiers (Zhufeng et al., 2022), failing to meet end-to-end underwater acoustic target recognition requirements.
2. Current feature extraction methods primarily use two-dimensional feature methods based on signal physics or brain-like computing features or their fusion methods (Li J. et al., 2022; Yang S. et al., 2023), overlooking the high-dimensional features of underwater acoustic data, resulting in insufficient representation capabilities of fused features.
3. Current classifiers mainly enhance feature extraction capabilities by stacking convolutional layers (Ji et al., 2023). However, due to the mixture of ocean environmental noise and partial information of underwater acoustic target features (Xu et al., 2019), standard convolutional operations tend to lose some effective features of underwater acoustic targets and erroneously retain ocean environmental noise (Li J. et al.,

2022), reducing the capability to extract effective features in underwater acoustic target recognition models. Thus, the model's parameter quantity and its recognition performance cannot achieve an effective balance, failing to meet the requirements of fast recognition speed and high accuracy in underwater acoustic target recognition.

4. As underwater acoustic data collection requires substantial financial and labor support, most existing publicly available underwater acoustic datasets exhibit imbalances in sample quantities across categories (Zhou et al., 2021). When training deep learning-based target recognition models, this can lead to overfitting phenomena (Li B. et al., 2022), suppressing model recognition performance.

To address these issues, we propose a novel underwater acoustic target automatic recognition network model based on a multi-gradient flow global feature enhancement network, referred to as MGFGNet.

Contributions of this work include:

1. Introducing a high-dimensional feature fusion method based on signal analysis and brain-like features.
2. Proposing a multi-gradient network to reduce model parameters and enhance feature extraction capabilities.
3. Presenting an adaptive feature fusion and enhancement module to enrich the physical, channel, and contextual information of pre-existing features.
4. Inventing a loss function, adding only three hyperparameters, and transforming the multi-classification task into multiple binary classification tasks, significantly improving the model's ability to suppress sample imbalances and recognition accuracy.

The following outlines the general structural framework of the remaining content in this article. Section 2 provides a detailed exposition of the Ship Radiated Noise Classification Method, known as MGFGNet. In Section 3, qualitative and quantitative experiments are conducted to compare MGFGNet with existing advanced underwater acoustic target recognition models, followed by an analysis of the experimental findings. Finally, Section 4 serves as the conclusion of this article.

2 Methods

This section primarily delineates MGFGNet. Section 2.1 provides an overview of its architectural framework. Sections 2.2 through 2.5 subsequently delve into its Feature Extraction and Fusion Module (FEFM), the Multi-gradient Flow Block with Attention (Multi-grad Block), the Context Augmentation and Fusion Module (CAFM), and the dynamic classification loss function known as Taylor-MCE Loss.

2.1 Proposed model

MGFGNet comprises two core modules: FEFM and the MGFGNet classifier. Figure 1 illustrates its detailed architecture.

FEFM utilizes various feature extraction algorithms based on signal analysis and brain-like features to extract multidimensional features from vessel radiated noise signals. Subsequently, multiple three-dimensional features are fused using the proposed feature fusion method to form high-dimensional fused features, which serve as inputs to the MGFGNet network.

The MGFGNet classifier primarily consists of the Multi-grad Block module and the CAFM module. The Multi-grad Block utilizes a multi-gradient flow network and residual modules to rapidly extract deep abstract features with different receptive fields from underwater acoustic target signals while reducing model parameters. Simultaneously, it leverages the multi-head self-attention mechanism (MHSA) (Han et al., 2021) to enhance the model's focus on foreground information, aiming to preserve the spatiotemporal characteristics of target line spectra in the acoustic energy spectrogram. This enhances the model's ability to extract effective information from sonar signals.

The CAFM module uses dilated convolutions with different dilation rates to adaptively fuse and enhance contextual information with a broad range of receptive fields, enriching the feature representation of physical, channel, and contextual information extracted by the preceding module. Finally, the Taylor-MCE Loss is employed to calculate prediction loss, addressing the issue of suppressing model recognition performance on imbalanced datasets. The Taylor-MCE Loss incorporates Taylor series (Gonzalez and Miikkulainen, 2021)

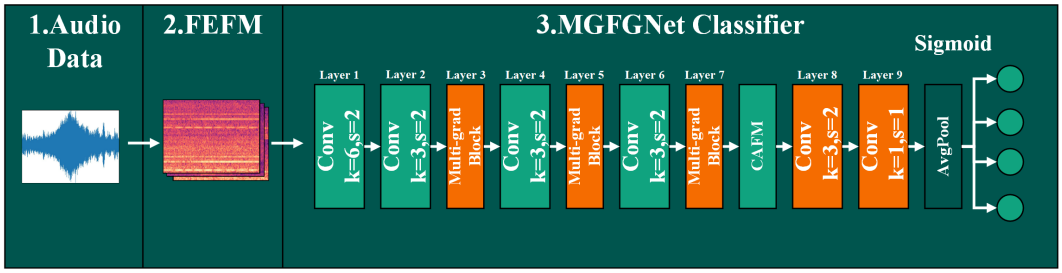


FIGURE 1
MGFGNet model architecture.

into binary cross-entropy loss (BCE) (Ruby and Yendapalli, 2020), including two components: one suppresses imbalances in sample components, and the other is a low-order term of the perturbation factor aimed at enhancing model recognition accuracy. Additionally, it transforms the multi-class classification task into multiple independent binary classification tasks.

2.2 Feature extraction and fusion module

Although deep learning-based feature extraction methods can capture more profound abstract features compared to traditional signal processing methods, they also come with a substantial increase in computational costs (Aggarwal et al., 2022). Vessel radiated noise primarily consists of mechanical noise, hydrodynamic noise, and propeller noise (Yang et al., 2019). Additionally, different feature extraction methods express distinct signal characteristics, and using multiple features for fusion can yield improved recognition results (Li Y. et al., 2022). Therefore, this paper, based on the generation mechanism of ship radiated noise, employs a fusion feature extraction method grounded in signal physical characteristics and brain-like features to represent underwater acoustic signals in multiple dimensions.

The fusion features in this paper mainly comprise energy-enhanced features from three types of features: CQT (Singh et al., 2022), delta MFCC (Nouhaila et al., 2022), and double delta MFCC (Nouhaila et al., 2022).

Firstly, since vessel radiated noise carries a significant amount of valid information in the low-frequency subband (Zhang et al., 2023), CQT provides better frequency resolution in the low-frequency subband (Mateo and Talavera, 2020). Hence, CQT is utilized as one of the feature extraction methods.

Secondly, MFCC, as a static feature, can not only eliminate ocean background noise but also effectively represent the spectral information of underwater acoustic targets. However, it lacks dynamic temporal signal features (Yang S. et al., 2023). To introduce temporal dynamic information, this paper performs local estimation of the differential operation along the time axis for the MFCC feature, obtaining delta MFCC and double-delta MFCC features. Both of these feature extraction methods are incorporated into the extraction of underwater acoustic target features.

Furthermore, as the single-channel feature information (graphically represented as a grayscale image) formed by these feature extraction methods can only express three-dimensional information of underwater sound, such as time, frequency, and energy domains, this paper expands the single-channel energy domain digital features of the above feature extraction methods into three-channel energy domain features using a color space representation. The detailed expansion method is described as follows.

Finally, this feature extraction and fusion module are embedded in the front end of the target recognition network, significantly reducing the computational burden of the classifier while achieving end-to-end target recognition.

Figure 2 illustrates our raised feature extraction method. Its process consists of four main parts:

1. In the first step, CQT features and MFCC features are extracted.

2.2.1 CQT extraction process

In the feature extraction process, the frame length is 2048 and the frame overlap is the portion between two frames of size 75% of the frame length, then using a Hanning window with a window size equal to the frame length for each frame signal.

The CQT transform of a finite length sequence x(n) is

$$X^{CQT}(k) = \frac{1}{N_k} \sum_{n=0}^{N_k-1} x(n) w_{N_k}(n) e^{-j\frac{2\pi Q}{N_k}n} \tag{1}$$

where $w_{N_k}(n)$ is a Hanning window of length N_k; Q is a constant factor in the CQT; k is the CQT frequency number, and the value of N_k is related to the value of k.

$$Q = \frac{1}{2^{\frac{1}{b}} - 1}, \tag{2}$$

where b is the number of frequency spectral lines, the

$$f_k = f_{\min} \times 2^{\frac{k}{b}}, k = 0, 1, \ldots, K-1, \tag{3}$$

$$N_k = \left\lceil Q \frac{f_s}{f_k} \right\rceil, k = 0, 1, \ldots, K-1, \tag{4}$$

where CQT information are stored in a matrix X^{CQT} (k,n), $f_{\min} = 1$, $f_s = 22050$. Since the sampling rate of the raw underwater acoustic data is 22050Hz for 5s, the shape of the CQT is 128×216.

2.2.2 MFCC extraction process

In the feature extraction process, the frame length and frame overlap are set to be the same as in the CQT extraction process. A Hanning window with a window size equal to the frame length is

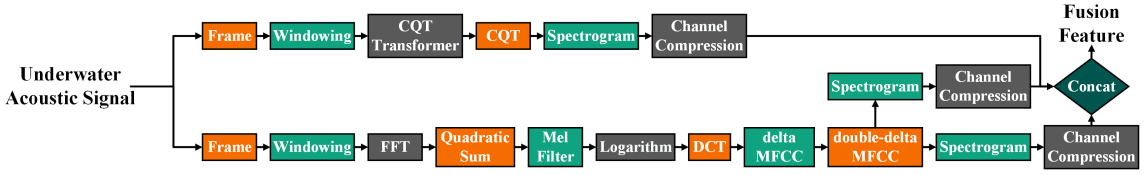


FIGURE 2
Feature extraction process for fused features.

then used for each frame. The short-time Fourier transform is then used to filter the noise and the sum of squares is used to obtain the power spectrum. Then 128 Mel filter banks were used to filter the information of each frame and logarithm was obtained to obtain Mel spectrum. Finally, MFCC was obtained by logarithm fitting the Mel spectrum to human hearing and discrete cosine transform (DCT). Since the sampling rate of the raw underwater acoustic data is a 5s signal at 22050Hz, the shape of the MFCC is 128×216.

2. The second step focuses on the extraction of delta MFCC and double-delta MFCC features by adding delta features and double-delta features to the MFCC features.

3. The third step focuses on transforming the above three features into spectrograms based on the size of 512, 12 and 0 for Hop length, bins per octave and tuning, respectively, with a preset image size of 3 × 640 × 480 for per image. Figure 3 illustrates the time-domain waveform diagram of radiated noise of a ship in the Deepship (Irfan et al., 2021) dataset and the spectrum diagram of CQT, delta MFCC and double-delta MFCC.

4. In the fourth step, the spectral graphs of CQT, delta MFCC and double-delta MFCC are fused respectively in channel dimension. The detailed fusion process is as follows.

2.2.3 Feature compression

as a result of the image pixel values reflected the important degree of information, so each spectrum diagram of three channel dimension values together to form a characteristic picture of 640 x 480.

2.2.4 Feature range mapping

Since the original pixel size range of each channel dimension is 0-255, the pixel value range of the feature map at this time is 0-765. To facilitate input for subsequent model calculations, map it to the range [0,255].

2.2.5 Feature fusion

Finally, the mapped features are in the order of CQT and two MFCC-derived features from top to bottom in the channel dimension to form a fusion feature with a shape of 3×640×480.

The formula of the fusion process above is expressed as:

$$T' = Map(concat(\sum_{j=0}^{2} T_j^{CQT}, \sum_{j=0}^{2} T_j^{deltaMFCC}, \sum_{j=0}^{2} T_j^{double-deltaMFCC})) \quad (5)$$

Where T_j^{CQT} represents the feature map of the J-th layer in the channel dimension of the CQT spectral graph feature matrix, $T_j^{deltaMFCC}$ and $T_j^{double-deltaMFCC}$ have the same meaning. Concat represents connecting matrices in the channel dimension. Map represents the range mapping of matrix data, a matrix T with data range of (x_{min},x_{max}), mapping its data to the range of (y_{min},y_{max}), and the mapped matrix is

$$Map = \frac{y_{max} - y_{min}}{x_{max} - x_{min}} \times (T - T_{x_{min}}) + T_{y_{min}}, \quad (6)$$

Where $T_{x_{min}}$, and $T_{y_{min}}$ both represent a constant matrix with the same latitude as T, and its content is the value represented by the Angle symbol.

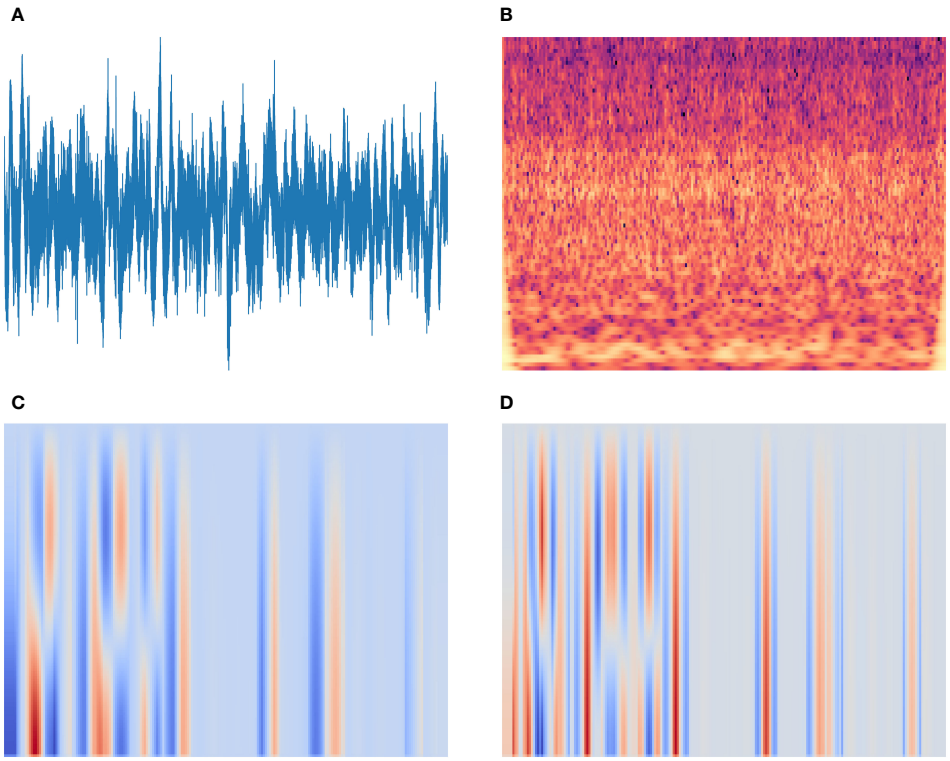


FIGURE 3
Spectrogram feature plots of radiated noise from a ship in the Deepship database: **(A)** Time-domain waveform; **(B)** CQT; **(C)** Delta MFCC; **(D)** Double-delta MFCC.

2.3 Multi-gradient flow with attention block

Existing models primarily increase the depth of the network to enhance feature extraction capabilities, but this leads to an increase in parameters while also losing a substantial amount of valuable information (Ji et al., 2023). In order to reduce the model's parameter count and enhance its ability to extract multidimensional features, this paper, inspired by the Cross Stage Partial Network (CSPNet) (Wang et al., 2020), which efficiently extracts effective feature information to alleviate model complexity, proposes the Multi-gradient flow bottleneck with attention Block (Multi-grad Block).

The Multi-grad Block concatenates multiple residual modules (Resblocks) to form a multi-gradient flow network. This structure enables the rapid acquisition of target information and gradient flow information from different receptive fields, accelerating the model's feature extraction speed while reducing the model's parameter count. Since traditional convolution operations lack sufficient discrimination between the spectra of multiple target lines and ocean background noise during feature extraction (Li J. et al., 2022), MHSA is introduced in the Resblock to increase the model's focus on targets rather than background noise or other irrelevant elements (Han et al., 2021). The detailed model structure is illustrated in Figure 4.

The detailed calculation process for the MHSA is as follows. MHSA is calculated as follows.

$$MH(A,B,C) = Concat(H_1, H_2, \ldots, H_h)W^O, \tag{7}$$

where A, B, C denote the query vector, key vector and value vector respectively, H_i illustrates the output of the i-th head, h is the number of headers, and W° is the output transformation matrix. The output of each header $head_i$ can be expressed as

$$head_i = Attention(QW_i^A, KW_i^B, VW_i^C) \tag{8}$$

where W_i^A, W_i^B, W_i^C are the A, B, and C transformation matrices for the i-th header, respectively, and Attention is a self-attentive calculation function with the following equation.

$$Attention(A_h, B_h, C_h) = soft\max\left(\frac{A_h B_h^T}{\sqrt{d_k}}\right)C_h, \tag{9}$$

Where d_k is the dimension of the key vector, softmax function mainly performs normalization, calculates the weight of each key vector, then multiplies the weight by the value vector, and finally performs weighted summation to get the attention output.

2.4 Context augmentation and fusion module

Due to the complex distribution of targets in the hybrid spectrogram generated by the feature extraction and fusion module of the original underwater acoustic signal, there are numerous small targets locally and larger, medium-sized targets globally (Wang B. et al., 2023). Using a single receptive field cannot fully capture the multidimensional features of the original signal, which reduces target classification accuracy (Wang Z. et al., 2023). To address these issues, this article introduces the Context Augmentation and Fusion Module (CAFM).

CAFM, as depicted in Figure 5, employs dilated convolution with varying rates to extract feature information from different receptive fields effectively (Gao et al., 2023). It enhances and fuses the multidimensional feature information obtained from the preceding gradient flow feature extraction module. Here's a breakdown of its structure:

1. The effective feature information obtained from the pre-gradient flow feature extraction module is rapidly processed using dilated convolution with three distinct rate values.

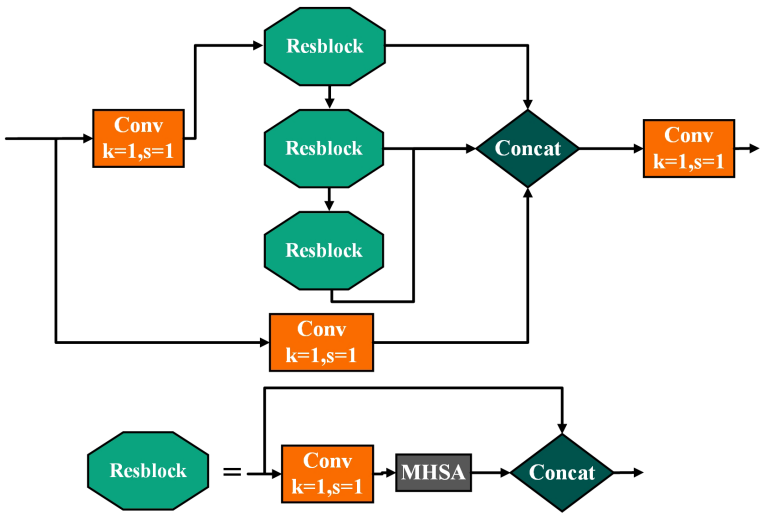


FIGURE 4
Model structure of Multi-grad Block.

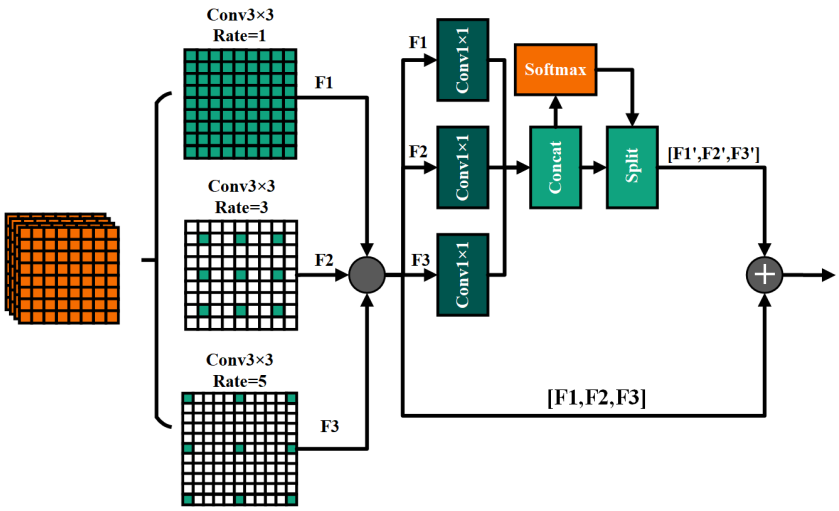


FIGURE 5
CAFM operation flows.

2. Target feature information is subsequently enhanced separately by the adaptive feature enhancement module and the cascade computing module.
3. The effective features derived from the adaptive feature enhancement module and the cascade computing module are then weighted and fused.

The former approach initially employs 1×1 convolution to compress and decrease the dimension of the pre-feature maps to single-channel feature maps. It then concatenates the feature maps in increasing rate order and calculates the weights for each channel using softmax. Finally, it enhances the channel dimension features through softmax-weighted multiplication.

The latter approach concatenates the feature maps obtained via expansion convolution at different rates to create a new feature map.

2.5 Taylor-MCE Loss

Existing mainstream classification loss functions primarily encompass the Cross-entropy Loss (CE) (Ho and Wookey, 2019) and its variations tailored for specific classification tasks. These adaptations include log loss (LL) (Lin et al., 2022) and BCE (Ruby and Yendapalli, 2020) for binary classification and focal loss (FL) (Lin et al., 2017) and categorical cross-entropy (CCE) (Ho and Wookey, 2019) for multi-class classification. However, the presence of a severe class imbalance among categories in underwater acoustic datasets poses a significant challenge (Zhou et al., 2021). Utilizing the aforementioned classification loss functions often leads to model overfitting (Leng et al., 2022), subsequently impacting recognition accuracy.

To tackle this challenge, this article introduces a novel loss function termed the Taylor-MCE Loss (Multiple Cross-Entropy Joint Loss Function based on Taylor Series). The Taylor-MCE Loss combines the polynomial terms derived from Taylor series expansion with BCE, FL, and low-order perturbation factors. It then transforms the multi-class classification task into a set of independent binary classification tasks, effectively resolving the issue of sample imbalance within the dataset and significantly enhancing the model's recognition performance.

The detailed design process is as follows:

1. Selection of the base loss function

Multi-class classification aims to calculate the likelihood of an object belonging to multiple categories, while binary classification seeks to identify whether an object is a specific category (e.g., discerning whether an object is a dog or not). Although these tasks may seem to differ only in the number of predicted categories, they have fundamental distinctions. In standard multi-class classification, CCE serves as the loss function, primarily relying on softmax to calculate the likelihood of an object belonging to multiple categories and selecting the category with the highest probability as the prediction. In contrast, binary classification tasks primarily employ BCE as the loss function, using sigmoid (output values between 0 and 1) to determine whether an object is closer to category 0 or category 1.

To select a more suitable base loss function and assess whether binary classification loss functions can be adapted for multi-class tasks, we conducted experiments in multi-class target recognition. The application of binary classification loss functions in multi-class tasks involved treating each category as an independent binary classification task. During our experiments, we made an intriguing observation: when inter-class sample sizes were balanced, CCE exhibited stable performance. However, in cases of sample imbalance, the use of BCE for multi-class tasks resulted in a significant improvement in accuracy compared to CCE (refer to Table 1 for details).

Treating each category as an individual binary classification task ensured that predictions for each category were mutually exclusive and independent (Ruby and Yendapalli, 2020), thereby addressing an issue. The problem when using CCE was that multiple categories

were predicted simultaneously [mutually exclusive but not independent (Ho and Wookey, 2019)]. Models employing CCE often favored categories with larger sample sizes, potentially overshadowing smaller categories during training (Leng et al., 2022). Additionally, BCE's core function was to enhance foreground weights while suppressing backgrounds (considering all other categories as backgrounds when predicting a single category) (Ruby and Yendapalli, 2020). This effectively balanced feature acquisition for different categories.

2. Exploring the relationship between loss functions using taylor series

The mutual constraints imposed by multiple categories can slow down model convergence. While combining multiple loss functions can enhance convergence speed and recognition accuracy (Li et al., 2019), it can also increase computational complexity. To minimize computational overhead while mitigating the impact of imbalanced datasets on the model, we drew inspiration from the Taylor series (Gonzalez and Miikkulainen, 2021) and explored the mathematical properties of BCE's polynomial form and loss functions designed to address imbalanced datasets. Our goal was to introduce minimal perturbation terms that retained the essential functionality of the loss function.

Since BCE can be represented as:

$$L_{BCE}(a,b) = -b_i \log(a) - (1-b_i)\log(1-a), \tag{10}$$

where $b_i \in \{0,1\}$ represents labels, and a represents predicted probabilities, and BCE is a special form of CCE, assuming

TABLE 1 Recognition accuracy, convergence time and number of parameters of CAFM at different locations of MGFGNet.

Model	Model convergence time (hours)	Parameters (M)	Accuracy
MGFGNet (-)	0.766	**5.576**	0.968
MGFGNet (+)	0.617	5.742	0.985
MGFGNet (1)	0.635	5.742	0.985
MGFGNet (2)	0.642	5.742	0.985
MGFGNet (3)	0.654	5.742	0.987
MGFGNet (4)	0.661	5.742	0.986
MGFGNet (5)	0.673	5.742	0.986
MGFGNet (6)	0.677	5.742	0.989
MGFGNet (7)	0.692	5.742	**0.991**
MGFGNet (8)	0.711	5.754	0.988

Bold font indicates the best-performing values within their respective columns.

$$a_t = \begin{cases} a, b = 1 \\ 1-a, otherwise, \end{cases} \tag{11}$$

CCE can be expressed as:

$$L_{CCE}(a,b) = -\log(a_t) \tag{12}$$

Applying Taylor series to CCE, the expression becomes:

$$L_{CCE} = -\log(a_t) = \sum_{i=1}^{\infty} \frac{1}{i}(1-a_t)^i \tag{13}$$

By observing the relationship between the Taylor expansion of CCE and FL, it is apparent that FL is equivalent to a horizontal shift (modulation factor) c of CCE. This is expressed as:

$$L_{FL} = -(1-a_t)^c \log(a_t) = (1-a_t)^c L_{CCE} \tag{14}$$

BCE is a special form of CCE; therefore, their physical properties are fundamentally consistent, differing mainly in the prediction process. To enhance the model's ability to address imbalanced datasets, we introduced an element to strengthen the suppression of imbalanced samples within the original BCE. This addition involved increasing the horizontal offset, resulting in the loss function:

$$\begin{aligned} L_{Taylor-MCE'} &= \alpha_1 L_{BCE} + \alpha_2 (1-a_t)^c L_{BCE} \\ &= [\alpha_1 + \alpha_2 (1-a_t)^c] L_{BCE} \end{aligned} \tag{15}$$

where $\alpha_1 + \alpha_2 = 1$ represents a scaling factor.

3. Analyzing the impact of gradient on loss functions

To enhance model recognition accuracy with minimal computational overhead, we compared the gradients of various loss functions and evaluated the influence of low-order and high-order terms on model recognition accuracy. The gradients of the aforementioned two loss functions (Eqs. 13 and 14) are expressed as follows:

$$-\frac{dL_{CCE}}{da_t} = \sum_{i=1}^{\infty}(1-a_t)^{i-1} = 1 + (1-a_t) + (1-a_t)^2 + \ldots \tag{16}$$

$$\begin{aligned} -\frac{dL_{FL}}{da_t} &= \sum_{i=1}^{\infty}(1+\frac{c}{i})(1-a_t)^{i+c-1} \\ &= (1+c)(1-a_t) + (1+\frac{c}{2})(1-a_t)^{1+c} + \ldots \end{aligned} \tag{17}$$

From the equations, it is evident that CCE possesses a fixed gradient term of 1. As i surpasses 1 and a_i approaches 1, the ith gradient tends towards zero. FL exhibits similar characteristics but introduces an additional perturbation factor (c). Consequently, the coefficients of high-order, low-order, and high-order terms collectively influence the outcomes of the loss function. The high-order parts primarily serve to suppress model errors, while the low-order components play a crucial role in fine-tuning the model to reach correct conclusions (Zhang et al., 2023). Therefore, we introduce a perturbation factor into the low-order term coefficients of CCE to enhance the model's recognition performance.

In summary, to mitigate the impact of sample imbalance on the model while minimizing the increase in parameter complexity, we

propose the Taylor-MCE Loss. The expression is as follows:

$$L_{Taylor-MCE} = \alpha_1 L_{BCE} + \alpha_2 (1 - a_t)^c L_{BCE} + \beta_1 (1 - a_t) \quad (18)$$

where $\beta_1 \in [-1,\infty)$ represents the perturbation factor.

3 Experimentation and analysis

To evaluate the performance of MGFGNet in a real underwater environment, we employ authentic underwater acoustic public datasets for both qualitative and quantitative comparisons. These comparisons involve MGFGNet and various versions with varying network depth and width of mainstream existing underwater acoustic target recognition models, including ResNet and EfficientNet (Mateo and Talavera, 2020).

3.1 Experimental dataset

3.1.1 Deepship

To assess the model's performance under ideal conditions, this study employed the Deepship dataset (Irfan et al., 2021), comprised of underwater acoustic data from vessels recorded by Northwestern Polytechnical University in the marine environment beneath the sea surface at depths ranging from 141 to 147 meters in the Georgia Strait Delta from 2016 to 2018. The data and time labels for this dataset were obtained by deploying sensors to locate vessel positions. Only singular vessel signals within a 2-kilometer range of the sonar device were considered, and recording ceased whenever a vessel exceeded this range. The dataset encompasses data from 265 vessels, including Cargo ships, Passenger Ships, Oil Tankers, and Tugs.

The data underwent preprocessing, with all WAV format audio files standardized to a 22,050Hz sample rate. Additionally, the underwater acoustic data were segmented into 5-second units, resulting in over 30,000 labeled sound samples. Recognizing that the model's recognition accuracy is proportional to the sample size of the training set, a significant number of samples were allocated for model training to mitigate the risk of overfitting. To prevent substantial fluctuations in the model's recognition accuracy due to a small sample size, a portion of the data was reserved for validating and testing the model's performance. Consequently, for optimal model parameter training, a large portion of the data was allocated to model training, with only a small amount used for validation and testing, following an 8:1:1 split ratio for the training, validation, and test sets. Table 2 provides details of the dataset division.

3.1.2 ShipsEar

So as to assess the model's capacity to adapt to diverse maritime environments, emphasizing its generalization capability, this study incorporated an additional authentic dataset of ship radiated noise collected in a real-world marine setting. The data collection took place along the Atlantic coast of Spain and encompasses recordings from 11 distinct ship types. These 11 ship categories were subsequently classified into four classes based on ship categorization, with the actual ocean background noise measurements, taken within these four categories, amalgamated to construct a five-class underwater acoustic dataset.

The dataset encompasses a total of 90 audio recordings, with individual recording durations varying from 15 seconds to 10 minutes. To ensure experimental precision, the "ShipsEar" dataset (Santos-Domínguez et al., 2016) underwent preprocessing identical to that applied to the "Deepship" dataset. A comprehensive class distribution is outlined in Table 3.

It is conspicuous that in the "Deepship" dataset, class proportions for classes 1-4 approximate ratios of 1:1.2:1.15:1.06. Conversely, the "ShipsEar" dataset presents imbalanced class proportions for classes 1-5, displaying a ratio of approximately 1.64:1.34:3.76:2.17:1. Consequently, when compared to the

TABLE 2 Details of the four categories in the Deepship dataset after pre-processing.

Class Number	Target	Total	Training set	Validation set	Testing set
1	Cargo Ship	7621	6097	762	762
2	Passenger Ship	9211	7369	921	921
3	Oil Tanker	8776	7022	877	877
4	Tug	8085	6467	809	809

TABLE 3 Details of the five categories of the ShipsEar dataset after pre-processing.

Class Number	Target	Total	Training set	Validation set	Testing set
1	Fishing boats, Trawlers, Mussel boats, Tugboats, Drafgers	369	296	37	36
2	Motoboats, Pilot boats, Sailboats	301	241	30	30
3	Passenger ferries	843	675	84	84
4	Ocean liner, Ro-Ro vessels	486	389	49	48
5	Background noise recordings	224	180	22	22

"Deepship" dataset, the "ShipsEar" dataset not only illustrates class imbalance but also contains significantly fewer samples, representing approximately 1/15 of the "Deepship" dataset. Such a dataset is highly susceptible to overfitting during the training process due to its limited sample size. Additionally, class imbalance can lead to notably reduced accuracy in recognizing classes with fewer samples.

3.1.3 SCTD

Synthetic Aperture Sonar (SAS) images (Huang and Yang, 2022; Wang and Huang, 2023; Yang, 2023; Zhang, 2023), known for their high resolution, significantly aid in target recognition in underwater acoustics. In order to assess the model's performance on a high-resolution underwater acoustic image dataset, this study introduces the SCTD dataset (Zhou et al., 2021). Since the original SCTD dataset is primarily designed for target detection tasks and its structure is not conducive to underwater acoustic target recognition models, certain modifications were implemented to adapt it to the classification task. Specifically, for the aircraft, human, and shipwreck categories within SCTD, the following steps were taken:

Firstly, multiple targets within a single image were individually cropped to ensure that each final image contains only one target, aligning it with the training sample format for target recognition.

Secondly, to augment the samples and balance the representation of each category, random cropping and flipping techniques were employed.

Finally, the dataset was partitioned into training, testing, and validation sets in an 8:1:1 ratio, as detailed in Table 4.

3.2 Hyperparameter setting

During the experimental process, the underwater acoustic target recognition model, MGFGNet, employed the Adaptive Moment Estimation optimizer (Adam) (Irfan et al., 2021) to mitigate sample noise interference. For this optimization process, the first-order momentum factor, second-order momentum factor, and Fuzz factor within Adam were configured at 0.9, 0.999, and 0.0000001, respectively. The initial learning rate was set to 0.001, with a weight decay coefficient of 0.0005, and a batch size of 32 was utilized. Finally, $a_1 = 0.5$, $a_2 = 0.5$, $c = 5$ and $\beta_1 = 5$ in Taylor-MCE Loss are set. The model was trained for 120 epochs (iterations) using the aforementioned parameters. Throughout the experimental process described below, unless otherwise specified, the experiment parameters mentioned above were consistently applied.

3.3 Experimental environment and performance indicators

The experiments were conducted on the PyTorch platform, running on the Windows 10. The hardware setup employed for these experiments is detailed in Table 5. To mitigate the potential influence of experimental variability, a systematic approach was taken. It involved the training and testing of various models, both qualitatively and quantitatively. Subsequently, a comparative analysis of algorithmic performance was performed.

Given that Accuracy can reflect the model's recognition capability across multiple classes, while Precision and Recall can indicate the overall classification performance of the model, these three evaluation criteria are employed to assess different models. Their formulas are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{19}$$

$$precision = \frac{TP}{TP + FP}, \tag{20}$$

$$recall = \frac{TP}{TP + FN}, \tag{21}$$

where TP represents instances that were originally true positive samples and were correctly predicted as positive samples by the underwater acoustic target recognition model. TN corresponds to instances that were originally true negative samples and were accurately predicted as negative samples by the model. FP signifies instances that were originally true negative samples but were erroneously classified as positive samples by the underwater acoustic target recognition model. FN stands for instances that were originally true positive samples but were incorrectly predicted as negative samples by the model.

3.4 Ablation experiments

3.4.1 Feature ablation experiments

To confirm the representational capabilities of the feature extraction approach raised in this study for original underwater acoustic signals, Table 6 presents an extensive comparison of diverse characteristics abstraction approachs on the Deepship. This comparison encompasses the original two-dimensional features, their corresponding three-dimensional counterparts, and the three-dimensional feature fusion approach introduced in Section 2.2 within the MGFGNet model. Notably, the recognition

TABLE 4 Details of the three categories of the SCTD dataset after pre-processing.

Class Number	Target	Total	Training set	Validation set	Testing set
1	Aircraft	575	459	58	58
2	Human	546	436	55	55
3	Shipwreck	488	390	49	49

TABLE 5 Details of the hardware environment for the experiment.

Hardware name	Parameters	Number
CPU	Intel Xeon Sliver 4310	2
GPU	NVIDIA Tesla A100 80G	1
RAM	SAMSUNG RECC DDR4 32GB	8

accuracy of spectral features for each feature extraction method surpasses that of the original two-dimensional features. Delta MFCC, owing to its ability to capture temporal correlations of MFCC, exhibits higher experimental accuracy than MFCC features, albeit with a modest 0.2% increase. Similarly, double-delta MFCC records a mere 0.4% improvement over delta MFCC since it primarily focuses on local estimations along the time axis for the differential operations of MFCC. CQT features, reflecting the frequency distribution patterns of underwater acoustic targets, outperform MFCC and its derivative features in terms of classification accuracy, thereby validating the superiority of CQT features over mel-spectrogram features (Domingos et al., 2022) in underwater acoustic target recognition. The horizontal comparison of spectral feature extraction methods among various feature extraction techniques exhibits similar characteristics as mentioned above.

It's worth mentioning that the overall accuracy of the fusion feature approach proposed in Section 2.2 surpasses that of other feature extraction techniques, achieving 99.1%. This represents a substantial increase of 3.9%, 3.7%, and 3.3% over the spectral features of MFCC, delta MFCC, and double-delta MFCC, respectively. Additionally, it outperforms CQT's spectral features by 1.6%. Moreover, there is a substantial increase in recognition accuracy across all categories compared to the spectral feature extraction methods of the remaining four features, thus validating the superiority of the fusion approach based on signal processing and brain-like features proposed in this study.

So as to provide a clearer illustration of the computational cost and efficiency of the feature extraction and fusion method introduced in this paper, we conducted additional experiments to assess the performance metrics of various feature extraction techniques. The testing dataset comprised 10 sets, each containing 10 noise data samples, and the experimental results represent the average of these 10 sets. Detailed experimental data is displayed in Table 6.

It is evident that the execution time for each feature extraction method's feature mapping technique increased by only approximately 0.0003 in comparison to the original method, with a memory consumption increment of around 20 MiB. Concurrently, the execution time of the feature extraction and fusion method proposed in this paper, which integrates three original feature components, remains within the same order of magnitude as their individual runtimes, indicating minimal additional time overhead.

Moreover, the memory consumption of the proposed method in this paper remains approximately at 350 MiB, aligning with the memory usage of all other feature extraction methods. This reaffirms the superiority and efficiency of the proposed method.

3.4.2 CAFM ablation experiment

So as to comprehensively evaluate the computational cost, convergence time, and performance of the feature extraction and fusion method presented in this paper at various positions within MGFGNet, a series of experiments were conducted. The experimental results on the Deepship dataset are presented in Table 1. In the model parameter nomenclature, the suffix indicates the layer within the target recognition model as depicted in Figure 1. For instance, "MGFGNet (1)" signifies the placement of the CAFM module after Layer 1 of MGFGNet, "-" indicates the absence of the CAFM module, and "+" denotes its placement at the beginning of MGFGNet, as illustrated in Figure 1.

Firstly, the integration of the CAFM module results in a modest 2.9% increase in model parameters compared to the original model. However, it significantly expedites the model's convergence speed. Furthermore, the convergence speed varies when the CAFM module is positioned at different locations within the model, and

TABLE 6 Recognition Accuracy of MGFG model on Deepship dataset using different features and the memory consumption and efficiency of each feature extraction method.

Feature	Cargo	Passenger Ship	Tanker	Tug	all	Time consumption (s)	Memory used (MiB)
MFCC	0.542	0.671	0.670	0.843	0.683	0.00033257	348.960938
MFCC Spec	0.930	0.950	0.951	0.975	0.952	0.00060603	368.828125
delta MFCC	0.629	0.655	0.623	0.849	0.687	0.00134368	352.488281
delta MFCC Spec	0.946	0.949	0.944	0.977	0.954	0.00166959	373.804688
double-delta MFCC	0.606	0.681	0.681	0.794	0.691	0.00137352	351.417969
double-delta MFCC Spec	0.946	0.957	0.957	0.97	0.958	0.00169537	372.640625
CQT	0.765	0.767	0.771	0.865	0.791	0.00353861	352.429688
CQT Spec	0.973	0.973	0.973	0.984	0.975	0.00356747	372.406250
Fusion Feature	0.929	0.929	0.977	0.993	0.957	0.00381105	357.812500
Fusion Feature of Spec	**0.984**	**0.985**	**0.994**	**1**	**0.991**	0.00384105	376.367188

Bold font indicates the best-performing values within their respective columns.

the speed is directly proportional to the sequence of the CAFM within the model. This is primarily due to the enhanced discriminability between background and target foreground in the feature maps when this module is applied, resulting in accelerated model convergence speed. Notably, after introducing the CAFM, the convergence time consistently remains between 0.6 and 0.7 hours, confirming the model's stability. In this experiment, convergence is defined as the point at which the loss remains unchanged in the thousandths place for three consecutive iterations.

Additionally, the incorporation of the CAFM module leads to a minimum 1.7% enhancement in recognition accuracy within MGFGNet, validating the CAFM module's capacity to boost model recognition accuracy through feature fusion and enhancement. However, the placement of the CAFM module also exerts an impact on recognition accuracy. For example, when the CAFM module is positioned at the head of MGFGNet and after Layer 1-2, the model exhibits improved recognition accuracy due to the fusion of multiscale acoustic target information and enhanced channel features. However, when the CAFM is placed at Layer 1, it leads to a rapid extraction of raw input features through a large convolutional kernel (kernel size of 6), resulting in the loss of significant valuable features and, consequently, inhibiting recognition accuracy. Furthermore, there is no subsequent feature enhancement in the feature extraction process, causing lower recognition accuracy compared to when the CAFM is placed after Layer 3-6.

Conversely, placing the CAFM module after Layer 3-6 introduces the Multi-gradient Block in front of the CAFM module, enriching the fused and enhanced features with a substantial amount of multi-gradient flow contextual information compared to the original information. This, in turn, enhances target feature information, leading to improved recognition accuracy. The highest recognition accuracy is achieved when the CAFM module is placed after Layer 7, as the model has undergone all the Multi-gradient Blocks by this stage, resulting in feature maps rich in multi-gradient flow, physical features, and numerous feature details. When the CAFM module is employed for feature fusion and enhancement at this stage, it effectively increases the importance of target information, thereby enhancing recognition accuracy.

However, due to the feature enhancement process preserving a substantial amount of suppressed background features, direct utilization of these feature maps for predictions can compromise experimental accuracy (Hu et al., 2018; Hou et al., 2021). Therefore, after employing the feature enhancement module, it is necessary to conduct further feature extraction on the enhanced feature maps using convolutional or feature extraction modules. This step helps discard numerous non-target features. For instance, attention mechanisms (AM) (Yang S. et al., 2023) and channel attention modules (CAM) (Li J. et al., 2022) both serve as feature enhancement modules. Ablation experiments have demonstrated that utilizing feature extraction or convolutional modules after feature enhancement enhances model recognition accuracy (Li J. et al., 2022; Yang S. et al., 2023). This substantiates why placing CAFM after Layer 7 results in higher recognition accuracy compared to after Layer 8.

3.4.3 Classification loss function ablation experiments

To assess the impact of the Taylor-MCE Loss on MGFGNet, this paper compared the recognition results of MGFGNet with various loss versions, including BCE, CCE, FL, and Taylor-MCE Loss, utilizing the Deepship dataset. The numbers 1, 2, and 3 following Taylor-MCE represent $\alpha_1 L_{BCE}$, $\alpha_2(1-a_t)^c L_{BCE}$, and $\beta_1(1-a_t)$, respectively. Notably, α_1 and α_2 have real values only when coexisting; otherwise, both are set to 1. A comprehensive summary of the experimental results is presented in Table 7.

Firstly, it is evident that Taylor-MCE Loss outperforms CCE, FL, and BCE in terms of recognition accuracy, demonstrating improvements of 2.4%, 2.2%, and 1.9%, respectively. The recognition accuracy of CCE and FL is quite similar. FL is derived from CCE through lateral shifting, aimed at mitigating the issue of sample imbalance. However, within the context of the Deepship dataset, where various classes exhibit a good balance, its effectiveness in addressing sample imbalance is reduced, resulting in a modest improvement of 0.2% compared to CCE. BCE, serving as a special form of CCE for binary classification, achieves a recognition accuracy improvement of 0.5%. This is primarily because BCE transforms multi-class classification into multiple binary classification tasks, where the predictions for each class are mutually exclusive and independent. This approach addresses a problem present in CCE where multiple classes are predicted simultaneously, leading the model to favor classes with larger sample sizes. This imbalance gradually drowns out smaller classes during training, providing a key rationale for choosing BCE as the base loss function for Taylor-MCE Loss. Taylor-MCE (1,3), inclusive of low-order perturbation terms ($\beta_1(1-a_t)$), contributes to the model's improved recognition accuracy, resulting in a significant advantage over Taylor-MCE (1,2), which only encompasses the component for addressing imbalance ($\alpha_2(1-a_t)^c L_{BCE}$). This finding reinforces the conclusion that low-order terms enhance recognition accuracy (Zhang et al., 2023). Taylor-MCE (2,3) achieves similar recognition accuracy to (1,3), primarily due to the relatively balanced distribution of class samples in the Deepship dataset, rendering the influence of (2,3) insufficient to significantly alter recognition accuracy.

TABLE 7 Recognition accuracy of MGFGNet with different classification loss functions on Deepship and ShipsEar.

Loss Function	Accuracy (Deepship)	Accuracy (ShipsEar)
CCE	0.967	0.937
FL	0.969	0.953
BCE	0.972	0.959
Taylor-MCE (1,2)	0.977	0.982
Taylor-MCE (1,3)	0.985	0.972
Taylor-MCE (2,3)	0.983	0.976
Taylor-MCE	**0.991**	**0.995**

Bold font indicates the best-performing values within their respective columns.

As the Deepship dataset comprises a substantial number of samples with a relatively balanced class distribution, it does not effectively validate the loss function's ability to suppress small samples and enhance recognition accuracy in unbalanced datasets. To further confirm the adaptability of Taylor-MCE Loss to imbalanced, small-sample underwater sound datasets, we conducted experiments using different classification loss functions on the ShipsEar dataset, characterized by class imbalance and limited sample sizes. A detailed overview of the experimental results is provided in Table 7. The unique sample characteristics of ShipsEar, featuring fewer samples and imbalanced class distributions, result in notable differences in model recognition accuracy when employing various loss functions. CCE, due to its lack of optimization for class imbalance, exhibits lower recognition accuracy compared to other loss functions. Both FL and BCE, which address class imbalance using different approaches (FL introduces horizontal shifting on top of CCE, while BCE transforms multi-class into multiple binary classification tasks to mitigate imbalance), yield similar and significantly improved recognition accuracy compared to CCE. In contrast, the results of the various versions of Taylor-MCE Loss are entirely opposite to those observed in the Deepship dataset. Given that the ShipsEar dataset has fewer samples and imbalanced class distributions, it necessitates substantial suppression of the imbalance component. When utilizing only the low-order perturbation term to enhance recognition accuracy, specifically Taylor-MCE (1,2), its recognition accuracy surpasses BCE by 2.3%, compared to the mere 0.9% improvement. Taylor-MCE effectively balances recognition accuracy and mitigates model overfitting attributed to class imbalance during training, ultimately yielding a recognition accuracy of 99.5%. This figure is 5.8%, 4.2%, and 3.6% higher than CCE, FL, and BCE, respectively.

These experiments affirm the adaptability of Taylor-MCE Loss to small-sample, imbalanced datasets, significantly enhancing model recognition accuracy.

3.5 Performance analysis

In this section, we compare the performance of MGFGNet with existing state-of-the-art target recognition models [such as ResNet (He et al., 2016), EfficientNet (Koonce, 2021), DenseNet (Iandola et al., 2014), etc.] under the same experimental conditions, examining various aspects.

3.5.1 Model identification accuracy and parameter analysis

To validate whether MGFGNet outperforms existing mainstream target recognition models, we trained and validated MGFGNet and other mainstream models under the experimental conditions described in Sections 3.2 and 3.3. The parameters of each model and their experimental accuracy on the Deepship test set are presented in Table 8. It is noteworthy that, to reduce the training time for various models, we modified the training epochs for all models on the Deepship dataset to 90. This decision is supported by the observation, as depicted in Figure 6, that MGFGNet exhibits a tendency toward convergence in loss before 90 epochs, with the optimal model being formed around the 71st epoch.

TABLE 8 Details of the number of parameters and the recognition accuracy on the Deepship dataset for various models.

Model	MFCC	delta MFCC	double-delta MFCC	CQT	Fusion Feature	Parameters (M)
ResNet18 (He et al., 2016)	0.939	0.942	0.947	0.963	0.970	11.7
ResNet34 (He et al., 2016)	0.929	0.937	0.942	0.966	0.971	21.8
ResNet50 (He et al., 2016)	0.921	0.933	0.937	0.952	0.965	25.6
ResNet101 (He et al., 2016)	0.913	0.931	0.937	0.947	0.953	44.5
EfficientNet_b0 (Koonce, 2021)	0.931	0.941	0.949	0.964	0.971	**5.3**
EfficientNet_b1 (Koonce, 2021)	0.930	0.938	0.945	0.967	0.968	7.8
EfficientNet_b2 (Koonce, 2021)	0.917	0.935	0.939	0.955	0.959	9.1
EfficientNet_b3 (Koonce, 2021)	0.915	0.931	0.934	0.945	0.951	12.2
DenseNet (Iandola et al., 2014)	0.866	0.871	0.878	0.913	0.931	1.1
CSPDenseNet (Wang et al., 2020)	0.889	0.896	0.913	0.937	0.951	0.9
CSPResNet18 (Wang et al., 2020)	0.938	0.945	0.953	0.966	0.973	5.6
MobileNetV1 (Howard et al., 2017)	0.759	0.787	0.793	0.822	0.841	3.2
MobileNetV2 (Sandler et al., 2018)	0.876	0.888	0.893	0.907	0.921	2.2
MobileNetV3-S (Howard et al., 2019)	0.732	0.747	0.752	0.773	0.806	1.5
MobileNetV3-L (Howard et al., 2019)	0.820	0.822	0.829	0.877	0.894	4.2
ViT (Dosovitskiy et al., 2020)	0.871	0.875	0.879	0.882	0.889	86.6
MGFGNet	**0.952**	**0.954**	**0.958**	**0.975**	**0.991**	5.7

Bold font indicates the best-performing values within their respective columns.

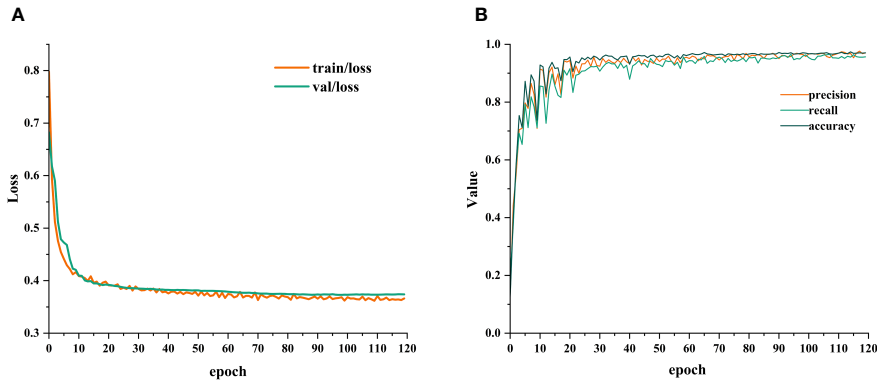


FIGURE 6
Variation of parameters during MGFGNet training: **(A)** Loss variation plot; **(B)** Precision, recall, and accuracy variation plot.

Clearly, MGFGNet demonstrates experiment accuracy superior to existing mainstream target recognition models across various feature input scenarios. This validates the robust feature extraction capability of MGFGNet in diverse experimental environments.

Furthermore, CSPNet (Wang et al., 2020) not only reduces model parameters but also effectively promotes the model's feature extraction capability. For instance, testing the original versions of ResNet18 (He et al., 2016) and DenseNet (Iandola et al., 2014), along with their versions incorporating CSPNet, reveals a noticeable reduction in parameters and an improvement in model performance under various feature inputs. MGFGNet, based on the CSPNet philosophy with the multi-gradient flow module as a primary component, successfully achieves an effective balance between recognition accuracy and parameter count.

The size of the model's parameters also influences recognition accuracy. Models with either too many or too few parameters yield suboptimal experimental accuracy. For example, ViT (Dosovitskiy et al., 2020) has significantly more parameters than other models, yet its recognition rate is lower than most models. In contrast, the MobileNet (Howard et al., 2017; Sandler et al., 2018; Howard et al., 2019) series, characterized by smaller parameter counts as lightweight models, generally exhibits lower recognition accuracy compared to other models. However, DenseNet and CSPDenseNet (Wang et al., 2020), despite having fewer parameters, achieve high recognition accuracy. This is mainly attributed to the dense connectivity in DenseNet, ensuring low-dimensional feature information and a stronger gradient flow (Iandola et al., 2014).

Within the same model, variations in recognition accuracy due to changes in model depth show a negative correlation with the number of parameters. As the number of parameters increases from ResNet18 to ResNet101 in the ResNet model, the recognition accuracy gradually decreases. A similar trend is observed in the EfficientNet (Koonce, 2021) model. However, different network models do not exhibit this phenomenon due to diverse feature extraction methods. For example, ResNet18 and EfficientNet_b3 have similar parameter counts, but ResNet18 outperforms EfficientNet_b2 in recognition accuracy across various feature extraction methods. Different versions of MobileNet do not show this phenomenon because new feature extraction or enhancement modules are introduced in each version.

3.5.2 Analysis of computational load, training time, and prediction time

To assess the training and inference efficiency of MGFGNet, this section analyzes the time consumption for training and inference of MGFGNet and its comparative models. Detailed comparative results are provided in Table 9. It is noteworthy that, to reduce the training time of the models, we maintained consistency with Section 3.5.1 and modified the training epochs for all models on the Deepship dataset to 90. The convergence definition for this experiment is when the value of the thousandth loss percentile remains unchanged for three consecutive times during the training process, indicating model convergence.

During the experiments, variations in the training and prediction times of models were observed in different operating environments. To ensure the accuracy of experimental data, each model, during its runtime, had the host free of other GPU-intensive deep learning tasks, preventing interference with the experimental results. Additionally, to mitigate random factors, all experimental data are the averages of results obtained from five repeated experiments. Floating Point Operations per Second (FLOPs) are used to measure the computational complexity of the model. Training time refers to the total time for model training and validation. Inference time denotes the total time required for predicting 3369 individual samples from the validation set of Deepship.

Notably, MGFGNet exhibits superior inference time compared to all comparative models, especially the lightweight MobileNet series commonly used in embedded systems, demonstrating its practical utility. This is primarily attributed to the inference time reduction effect of CSPNet (Wang et al., 2020). For instance, the inclusion of CSPNet in ResNet and DenseNet also significantly reduces inference time. MGFGNet, incorporating the CSPNet philosophy through the Multi-grad Block, outperforms EfficientNet_b0 in prediction time, despite having similar parameter counts and FLOPs values.

Furthermore, MGFGNet achieves convergence in the fewest epochs, indicating a faster convergence rate. This is mainly due to Taylor-Loss suppressing the rate of model variation under different numbers of input categories, thereby accelerating model

TABLE 9 Floating-point computation vs. training and predicting time.

Model	FLOPs@224(B)	Training Time (hours)	Epochs at convergence	Predicting Time(s)	Support
ResNet18	3.7	**1.075**	52	56	3369
ResNet34	7.4	1.391	68	59	3369
ResNet50	8.5	1.868	75	67	3369
ResNet101	15.9	4.975	98	104	3369
EfficientNet_b0	1.0	3.423	83	91	3369
EfficientNet_b0	1.5	5.121	91	102	3369
EfficientNet_b0	1.7	5.368	97	104	3369
EfficientNet_b0	2.4	5.753	102	107	3369
DenseNet	1.6	4.792	62	135	3369
CSPDenseNet	1.4	4.872	57	119	3369
CSPResNet18	0.5	1.397	48	42	3369
MobileNetV1	0.6	4.693	96	66	3369
MobileNetV2	0.4	4.401	72	63	3369
MobileNetV3-S	**0.1**	2.661	65	55	3369
MobileNetV3-L	0.2	3.100	78	75	3369
ViT	17.6	7.295	107	139	3369
MGFGNet	0.7	1.779	**41**	**37**	3369

Bold font indicates the best-performing values within their respective columns.

convergence. It is observed that within models of the same architecture, parameters and convergence epochs exhibit a positive correlation, as seen in ResNet and EfficientNet series. While MobileNet has a smaller parameter count, its frequent occurrence of gradient vanishing during training, mainly due to the use of depthwise separable convolution, leads to extensive time spent correcting and updating the model, resulting in an increased number of convergence epochs.

Additionally, while MGFGNet's training time is lower than that of most target recognition models, it still exceeds that of ResNet18, ResNet34, and CSPResNet18. This is mainly because the Multi-grad Block module, based on the CSPNet philosophy, invented in MGFGNet, reduces the number of parameters but increases the computational workload for backward gradient updates (Wang et al., 2020), thus extending the model's training time. The increased training time for CSPDenseNet and ResNet18 with CSPNet also validates this characteristic. However, since practical applications primarily require low prediction times for rapid target recognition, this drawback has minimal impact in real-world scenarios.

Finally, upon contrasting Tables 8 and 9, it becomes evident that there is no inherent correlation between the training time, model parameters, and FLOPs for the models. For instance, when compared to ResNet18, DenseNet, CSPDenseNet, and the MobileNet series all exhibit smaller parameter counts and FLOPs. However, these models demonstrate longer training times than ResNet18. A similar experimental pattern is observed between the EfficientNet and ResNet series.

3.5.3 Model stability validation

Figure 6 presents the loss variation chart as well as the precision, recall, and accuracy variation charts on the validation set during the same 120-epoch training process on the Deepship dataset.

From the loss curve, it can be observed that the network gradually stabilizes after the 60th epoch. By examining the changes in precision, recall, and accuracy on the validation set during the training process, with smooth variations and the absence of overfitting, it can be concluded that the proposed underwater acoustic target recognition model, MGFGNet, demonstrates stability.

3.5.4 Robustness analysis of models.

To assess the robustness of MGFGNet, i.e., the extent to which the model is affected by small variations in the data, we utilized spectrogram features of MFCC and its derived characteristics. Due to the high similarity between spectrograms of MFCC and its derived features (Yang S. et al., 2023), this study thoroughly compared the dependency of various models on different input conditions based on spectrogram features of MFCC and its derivatives, as illustrated in Figure 7.

It is evident that MGFGNet exhibits a relatively small disparity in experimental accuracy when considering spectrogram features of MFCC and its derived characteristics. However, there is still some improvement, indicating that MGFGNet can capture minor variations in the derived features of MFCC without causing significant predictive differences due to slight changes. This validates the robustness of the model.

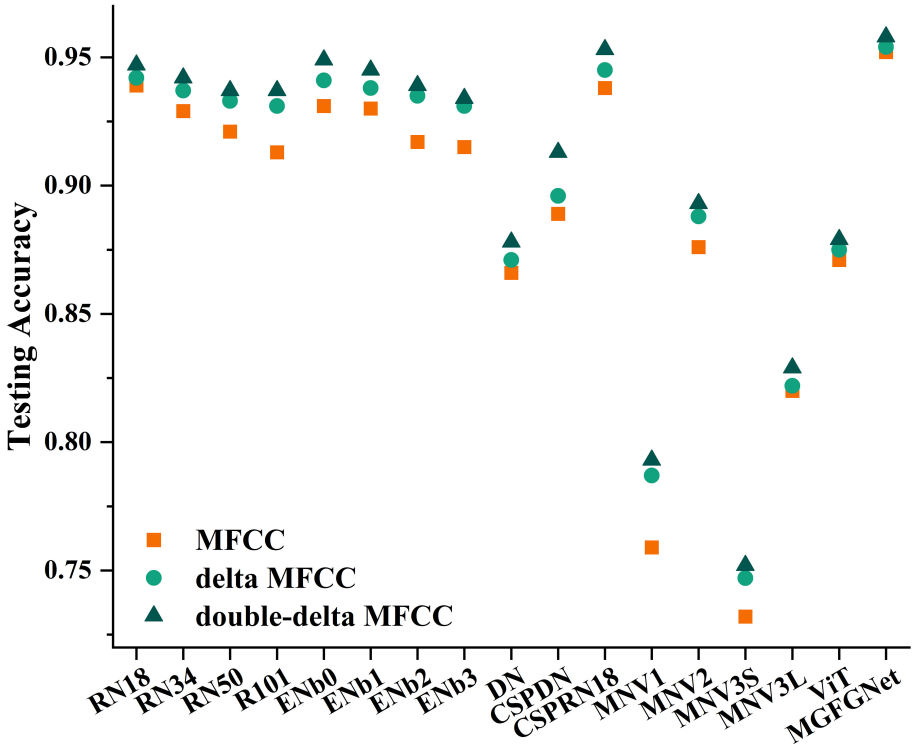


FIGURE 7
Recognition accuracy of multiple models under MFCC, delta MFCC and double-delta MFCC features.

ResNet18, DenseNet, and ViT models demonstrate comparable recognition accuracy across these three different feature extraction methods. In contrast, other models exhibit significant variations in model responses under these three feature extraction methods, indicating their reliance on features with high separability.

3.5.5 Generalizability analysis of the model

Due to varying predictive capabilities of models across different variable domains (distinct real underwater acoustic datasets), it is essential to assess the generalization performance of MGFGNet on additional datasets. This study conducts experiments placing each model under the experimental conditions defined in Sections 3.2 and 3.3, utilizing the shipsEar dataset. Detailed experimental results are presented in Table 10.

Evidently, MGFGNet exhibits a recognition accuracy surpassing all comparative models, achieving 99.5%. Furthermore, it is observed that MGFGNet achieves a 100% recognition rate for all categories except Class 1. This fact indicates a robust generalization capability of the model.

Additionally, on the shipsEar dataset, the ResNet series, Efficient series, and DenseNet also demonstrate strong performance, with recognition accuracies exceeding 93%. It is noteworthy that, with the involvement of CSPNet, DenseNet and ResNet18 show improved recognition accuracy, exceeding 96%, validating their enhancement in model feature extraction capabilities (Wang et al., 2020).

Finally, the MobileNet series performs poorly, with MGFGNet surpassing the highest recognition accuracy within its series, MobileNetV2, by 12.2%, and outperforming the lowest accuracy in MobileNetV3Small by 35%.

3.5.6 Scalability analysis

In order to further validate the scalability of MGFGNet on high-resolution sonar images, this study conducted experimental analyses, comparing MGFGNet with 16 other underwater acoustic target recognition models on the high-resolution sonar dataset SCTD. The recognition accuracy of each model is depicted in Figure 8.

Clearly, MGFGNet's recognition rate continues to surpass that of all other models, confirming the model's scalability. Due to the interference of underwater background noise, which results in poor separability between targets and background in sonar images (Huang and Yang, 2022), the multi-gradient flow model proposed in MGFGNet, based on CSPNet and MHSA, enhances the model's attention to targets (Wang et al., 2020; Han et al., 2021), ensuring the retention of a substantial amount of relevant target information during the feature extraction process. Additionally, further feature enhancement and fusion through CAFM contribute to an improved distinction between target foreground and background, effectively enhancing the model's recognition accuracy.

It is noteworthy that, as indicated in the performance and parameter analysis in Section 4.2, under the same model architecture, depth and recognition rate exhibit a proportional relationship in underwater sonar image target recognition. For instance, the recognition rates of the ResNet series and EfficientNet

10.3389/fmars.2023.1306229

TABLE 10 Details of the models's recognition Accuracy on the shipsEar dataset.

Model	Class 1	Class 2	Class 3	Class 4	Class 5	All
ResNet18	0.944	0.933	0.952	0.979	0.955	0.954
ResNet34	0.917	0.933	0.952	0.979	0.955	0.950
ResNet50	0.917	0.933	0.952	0.979	0.955	0.950
ResNet101	0.889	0.900	0.952	0.979	0.955	0.941
EfficientNet_b0	0.889	0.933	0.988	0.958	0.909	0.950
EfficientNet_b1	0.889	0.933	0.976	0.979	0.909	0.950
EfficientNet_b2	0.861	0.933	0.988	0.958	0.909	0.945
EfficientNet_b3	0.833	0.867	0.964	0.979	0.935	0.931
DenseNet	0.944	0.933	0.976	0.958	0.864	0.950
CSPDenseNet	0.972	0.9	0.988	0.958	0.955	0.964
CSPResNet18	0.972	1	0.964	0.979	0.909	0.968
MobileNetV1	0.861	0.8	0.905	0.792	0.818	0.85
MobileNetV2	0.889	0.867	0.917	0.854	0.727	0.873
MobileNetV3-S	0.667	0.6	0.655	0.625	0.682	0.645
MobileNetV3-L	0.917	0.867	0.786	0.854	0.909	0.845
ViT	0.944	0.933	0.929	0.938	0.909	0.875
MGFGNet	**0.972**	**1**	**1**	**1**	**1**	**0.995**

Bold font indicates the best-performing values within their respective columns.

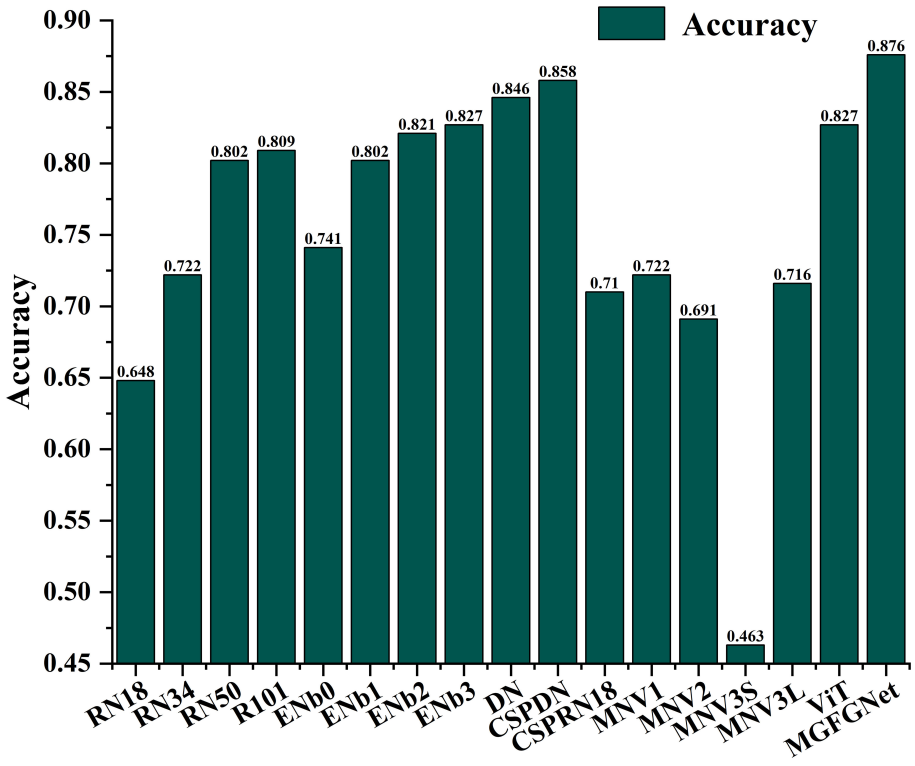


FIGURE 8
Details of the recognition Accuracy of the model on the SCTD dataset.

series validate this conclusion. Conversely, lightweight models such as MobileNet perform poorly, with recognition rates not exceeding 73%, once again confirming that MobileNet is not well-suited for underwater acoustic target recognition scenarios.

3.5.7 Computational bottleneck analysis

Due to the challenges associated with acquiring underwater acoustic datasets, the currently available open datasets are primarily limited to two ship radiated noise datasets: Deepship and ShipsEar. Given that the ShipsEar dataset comprises multiple ship types within each category and has a limited data volume, we conducted experiments with a substantial sample dataset extracted from Deepship to examine MGFGNet's recognition accuracy in relation to dataset size and to identify potential computational bottlenecks. This dataset, which was subject to preprocessing, included a total of 33,693 samples. Figure 9 presents the model accuracy of MGFGNet for various training set sizes sourced from Deepship.

The numerical values in the dataset version indicate the quantity of samples randomly chosen from each category in the Deepship dataset to form the training set for model training, while the test set configuration remained consistent with that presented in Table 2. The results clearly show that as the training set sizes for each category range from 100 to 800, the network model's recognition accuracy experiences rapid growth. Beyond the 800 mark, recognition accuracy tends to plateau, although there is still noticeable improvement as the dataset size increases. Importantly, no indications of encountering computational bottlenecks were observed.

4 Conclusion

An underwater acoustic object identification model MGFGNet based on multi-gradient flow global feature enhancement network is raised in this article. Firstly, by embedding feature extraction module into the target recognition network, the whole target recognition network forms an end-to-end model with underwater acoustic signal as input and classification result as output. Secondly, the invention of Muti-grad block uses multi-gradient flow network to obtain underwater acoustic signal features quickly and effectively, reducing the quantity of model parameters and feature extraction time. Then the CAFM module is used for multi-dimensional feature fusion and feature enhancement to improve the effective characteristic weight of underwater sound. Finally, the Taylor-MCE Loss function is introduced, which enhances model recognition accuracy and mitigates sample imbalance issues within the binary cross-entropy loss. This is achieved by incorporating low-order perturbation terms into the binary cross-entropy loss to suppress sample imbalance components. Consequently, the multi-class classification task is transformed into a set of independent binary classification tasks, effectively addressing the problem of dataset sample imbalance and improving model recognition performance.

The experimental results show that on the Deepship and ShipsEar underwater acoustic data sets, the feature extraction and fusion methods raised in this article have better ability to represent the original underwater acoustic signals. Compared with mainstream underwater acoustic target recognition models such

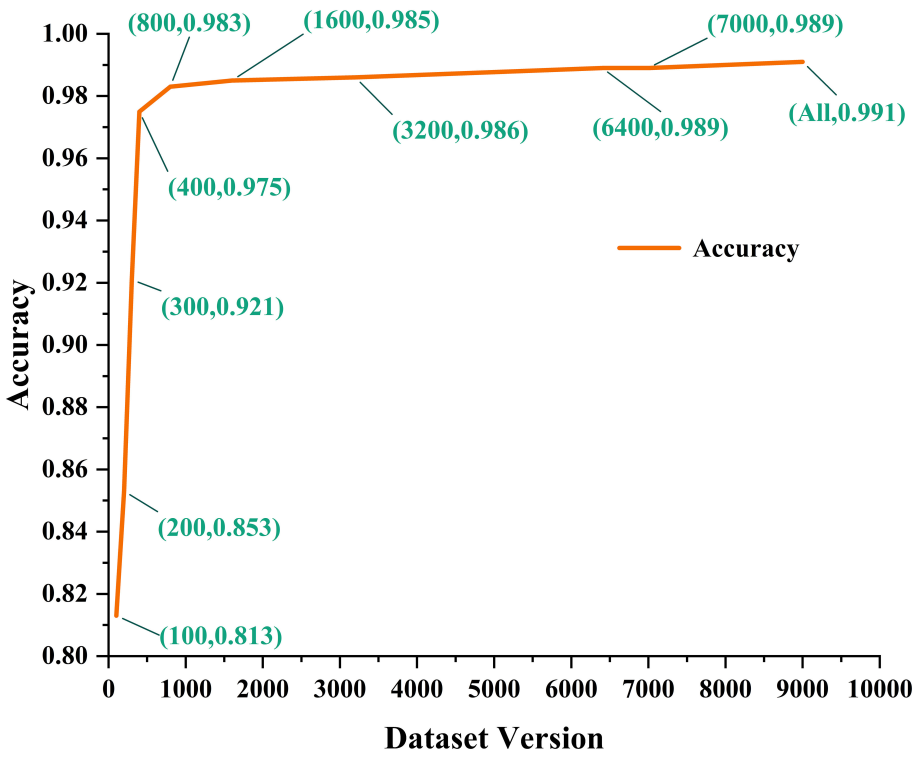


FIGURE 9
Details of the recognition Accuracy of the model on different versions of the Deepship dataset.

as ResNet and EfficientNet, the recognition accuracy of MGFGNet was greatly improved, and the inference time was greatly reduced. MGFGNet network has simple structure and few parameters, which can meet the requirements of end-to-end high precision and low latency in underwater acoustic target identification.

The experimental results of the model proposed in this article have the following potential implications for current underwater target recognition models:

Firstly, our feature extraction and fusion methods have demonstrated that traditional spectrogram-based feature extraction methods are better suited for representing the raw underwater acoustic features. This suggests that current methods for extracting underwater target features, such as those based on signal analysis and bio-inspired features, can be effectively combined with computer vision's feature enhancement techniques (e.g., channel and spatial feature enhancement methods) to further enhance feature representation.

Secondly, the design and experimentation with the Multi-grad block in our proposed classifier have shown that multi-gradient flow networks can better extract deep abstract features of the model while reducing the number of model parameters. This enables underwater target recognition models to depart from the mainstream design pattern of extracting effective features for underwater targets solely through convolution and residual network stacking.

Furthermore, the design of the CAFM in our proposed classifier has demonstrated that incorporating feature fusion and enhancement modules before the classification module in the classifier can significantly enhance the model's recognition accuracy. This enhancement may be related to feature loss during the extraction of effective features before classification and the numerical loss during the normalization process, as this process lacks specific loss control. This can lead to similarities between foreground and background values, making it difficult for the model to effectively recognize the target foreground. Subsequent research can focus on designing feature fusion and feature enhancement modules to improve the distinguishability between target foreground and background.

Lastly, the loss function designed in this paper was explored using the Taylor series, revealing factors influencing the loss function's functionality, such as lateral shifting to address sample imbalance and boosting model recognition accuracy through low-order terms in the Taylor expansion of the function. This enables future research to introduce fewer hyperparameters while gaining more benefits, providing a reference for subsequent studies and better explaining the underlying physical meaning of the loss function.

Additionally, the experimental results of the model proposed in this paper open up potential avenues for future research:

1. Deep learning-based underwater target recognition models have encountered certain bottlenecks, primarily due to their reliance on convolution and residual network stacking, which can lead to limitations in accuracy. The model design approach presented in this paper transforms traditional underwater target recognition into underwater image target recognition, broadening the model construction methods. In the future, lightweight module design methods from computer vision and ideas for feature enhancement and fusion based on the characteristics of underwater feature images can be introduced to enhance the model's efficiency and recognition accuracy, enabling real-time applications.
2. Existing underwater target recognition models mainly employ multi-class classification, but empirical evidence suggests that converting traditional multi-class tasks into multiple binary classification tasks is more suitable for underwater target recognition. Therefore, future research can delve into designing more effective underwater target recognition models based on multiple binary classification tasks that align with the physical characteristics of underwater sound.
3. Current research primarily focuses on building underwater target recognition models, with limited attention to loss function research. Traditional classification loss functions are primarily designed for object image classification and may not be highly adaptable to underwater target feature. Future research can focus on designing loss functions that better align with underwater target features based on the physical characteristics of underwater targets.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

JT: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. ZC: Writing – original draft, Writing – review & editing. HQ: Funding acquisition, Writing – original draft. MC: Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Special Program of Guangxi Science and Technology Base and Talents, grant number AD21220098, the Guangxi Natural Science Foundation, grant number 2022GXNSFDA035070, and the Innovation Project of Guangxi Graduate Education, grant number YCSW2023329.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aggarwal, A., Srivastava, A., Agarwal, A., Chahal, N., Singh, D., Alnuaim, A. A., et al. (2022). Two-way feature extraction for speech emotion recognition using deep learning. *Sensors* 22 (6), 2378. doi: 10.3390/s22062378

Ahmed, M. S., Aurpa, T. T., and Azad, M. A. K. (2022). Fish disease detection using image-based machine learning technique in aquaculture. *J. King Saud University-Computer Inf. Sci.* 34 (8), 5170–5182. doi: 10.1016/j.jksuci.2021.05.003

Ali, S., Iqbal, N., and Hafeez, Y. (2018). Towards requirement change management for global software development using case base reasoning. *Mehran Univ. Res. J. Eng. Technol.* 37, 639–652. doi: 10.22581/muet1982.1803.17

Boyd, C. E., McNevin, A. A., and Davis, R. P. (2022). The contribution of fisheries and aquaculture to the global protein supply. *Food Secur.* 14 (3), 805–827. doi: 10.1007/s12571-021-01246-9

Bradley, D., Merrifield, M., Miller, K. M., Lomonico, S., Wilson, J. R., and Gleason, M. G. (2019). Opportunities to improve fisheries management through innovative technology and advanced data systems. *Fish. Fish.* 20 (3), 564–583. doi: 10.1111/faf.12361

Darapaneni, N., Sreekanth, S., Paduri, A. R., Roche, A. S., Murugappan, V., Singha, K. K., et al. (2022). "AI based farm fish disease detection system to help micro and small fish farmers," in *2022 Interdisciplinary Research in Technology and Management (IRTM)* (Kolkata, India: IEEE), 1–5.

Di, N., Sharif, M. Z., Hu, Z., Xue, R., and Yu, B. (2023). Applicability of VGGish embedding in bee colony monitoring: comparison with MFCC in colony sound classification. *PeerJ* 11, e14696. doi: 10.7717/peerj.14696

Domingos, L. C., Santos, P. E., Skelton, P. S., Brinkworth, R. S., and Sammut, K. (2022). An investigation of preprocessing filters and deep learning methods for vessel type classification with underwater acoustic data. *IEEE Access* 10, 117582–117596. doi: 10.1109/ACCESS.2022.3220265

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv* abs/2010.11929. doi: 10.48550/arXiv.2010.11929

Gao, X., Xie, J., Chen, Z., Liu, A. A., Sun, Z., and Lyu, L. (2023). Dilated convolution-based feature refinement network for crowd localization. *ACM Trans. Multimedia Computing Commun. Appl.* 19 (6), 1–16. doi: 10.1145/3571134

Gladju, J., Kamalam, B. S., and Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. *Smart Agric. Technol.* 2, 100061. doi: 10.1016/j.atech.2022.100061

Gonzalez, S., and Miikkulainen, R. (2021). "Optimizing loss functions through multivariate taylor polynomial parameterization," in *Proceedings of the Genetic and Evolutionary Computation Conference*. (Lille, France: ACM) 305–313. doi: 10.1145/3449639.3459277

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., and Wang, Y. (2021). Transformer in transformer. *Adv. Neural Inf. Process. Syst.* 34, 15908–15919. doi: 10.48550/arXiv.2103.00112

Han, R., Jia, N., Huang, J., and Guo, S. (2022). Joint time-frequency domain equalization of MSK signal over underwater acoustic channel. *Appl. Acoustics* 189, 108597. doi: 10.1016/j.apacoust.2021.108597

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, USA: IEEE) 770–778. doi: 10.1109/CVPR.2016.90

Ho, Y., and Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access* 8, 4806–4813. doi: 10.1109/ACCESS.2019.2962617

Hong, F., Liu, C., Guo, L., Chen, F., and Feng, H. (2021). Underwater acoustic target recognition with a residual network and the optimized feature extraction method. *Appl. Sci.* 11, 1442. doi: 10.3390/app11041442

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13713–13722. doi: 10.1109/CVPR46437.2021.01350

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). Searching for mobilenetv3. *Proc. IEEE/CVF Int. Conf. Comput. Vision* (Seoul, Korea: IEEE), 1314–1324. doi: 10.1109/ICCV.2019.00140

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv. abs/1704.04861*. doi: 10.48550/arXiv.1704.04861g

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Utah, USA: IEEE) 7132–7141. doi: 10.1109/CVPR.2018.00745

Hu, W. C., Chen, L. B., Huang, B. K., and Lin, H. M. (2022). A computer vision-based intelligent fish feeding system using deep learning techniques for aquaculture. *IEEE Sens J.* 22 (7), 7185–7194. doi: 10.1109/JSEN.2022.3151777

Huang, P., and Yang, P. (2022). Synthetic aperture imagery for high-resolution imaging sonar. *Front. Mar. Sci* 9, 1049761. doi: 10.3389/fmars.2022.1049761

Huang, C., Yang, K., Yang, Q., and Zhang, H. (2021). Line spectrum extraction based on autoassociative neural networks. *JASA Express Lett.* 1, 016003. doi: 10.1121/10.0003038

Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., and Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. *ArXiv.abs/1404.1869*. doi: 10.48550/arXiv.1404.1869

Irfan, M., Jiangbin, Z., Ali, S., Iqbal, M., Masood, Z., and Hamid, U. (2021). DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Syst. Appl.* 183, 115270. doi: 10.1016/j.eswa.2021.115270

Ji, F., Ni, J., Li, G., Liu, L., and Wang, Y. (2023). Underwater acoustic target recognition based on deep residual attention convolutional neural network. *J. Mar. Sci. Eng.* 11, 1626. doi: 10.3390/jmse11081626

Jin, A., and Zeng, X. (2023). A novel deep learning method for underwater target recognition based on res-dense convolutional neural network with attention mechanism. *J. Mar. Sci. Eng.* 11, 69. doi: 10.3390/jmse11010069

Koonce, B. (2021). *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization* (Berkeley (CA): Apress).

Koparan, C., Koc, A. B., Privette, C. V., and Sawyer, C. B. (2018). *In situ* water quality measurements using an unmanned aerial vehicle (UAV) system. *Water* 10 (3), 264. doi: 10.3390/w10030264

Kritzer, J. P. (2020). Influences of at-sea fishery monitoring on science, management, and fleet dynamics. *Aquacult Fisheries* 5 (3), 107–112. doi: 10.1016/j.aaf.2019.11.005

Leng, Z., Tan, M., Liu, C., Cubuk, E. D., Shi, X., Cheng, S., et al. (2022). Polyloss: A polynomial expansion perspective of classification loss functions. *ArXiv. abs/2204.12511*. doi: 10.48550/arXiv.2204.12511

Li, Y., Gao, P., Tang, B., Yi, Y., and Zhang, J. (2022). Double feature extraction method of ship-radiated noise signal based on slope entropy and permutation entropy. *Entropy* 24, 22. doi: 10.3390/e24010022

Li, J., Wang, B., Cui, X., Li, S., and Liu, J. (2022). Underwater acoustic target recognition based on attention residual network. *Entropy* 24, 1657. doi: 10.3390/e24111657

Li, B., Yao, Y., Tan, J., Zhang, G., Yu, F., Lu, J., et al. (2022). "Equalized focal loss for dense long-tailed object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6990–6999.

Li, C., Yuan, X., Lin, C., Guo, M., Wu, W., Yan, J., et al. (2019). "Am-lfs: Automl for loss function search," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (Seoul, South Korea: IEEE) 8410–8419. doi: 10.1109/ICCV.2019.00850

Lim, L. W. K. (2022). Implementation of artificial intelligence in aquaculture and fisheries: deep learning, machine vision, big data, internet of things, robots and beyond. *J. Comput. Cogn. Eng*, 1-7. doi: 10.47852/bonviewJCCE3202803

Lim, K., and Whye, L. (2023). Blended Learning in Animal Biotechnology during Pre-COVID-19, COVID-19 and Post COVID-19 Recovery Phase Periods across the Globe: a Step Forward or Backward? *Int. J. Zool. Anim. Biol.* 6 (2), 1–5. doi: 10.23880/izab-16000451

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*. (Venice, Italy: IEEE), 2980–2988. doi: 10.1109/ICCV.2017.324

Lin, S., Zheng, H., Han, B., Li, Y., Han, C., and Li, W. (2022). Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction. *Acta Geotech.* 17, 1477–1502. doi: 10.1007/s11440-021-01440-1

Linka, K., and Kuhl, E. (2023). A new family of Constitutive Artificial Neural Networks towards automated model discovery. *Comput. Methods Appl. Mechanics Eng.* 403, 115731. doi: 10.1016/j.cma.2022.115731

Liu, Z., Peng, D., Zuo, M. J., Xia, J., and Qin, Y. (2022). Improved Hilbert–Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. *ISA Trans.* 125, 426–444. doi: 10.1016/j.isatra.2021.07.011

Ma, Y., Liu, M., Zhang, Y., Zhang, B., Xu, K., Zou, B., et al. (2022). Imbalanced underwater acoustic target recognition with trigonometric loss and attention mechanism convolutional network. *Remote Sens.* 14, 4103. doi: 10.3390/rs14164103

Mateo, C., and Talavera, J. A. (2020). Bridging the gap between the short-time Fourier transform (STFT), wavelets, the constant-Q transform and multi-resolution STFT. *Signal Image Video Process.* 14, 1535–1543. doi: 10.1007/s11760-020-01701-8

Nouhaila, B., Taoufiq, B. D., and Benayad, N. (2022). An intelligent approach based on the combination of the discrete wavelet transform, delta delta MFCC for Parkinson's disease diagnosis. *Int. J. Adv. Comput. Sci. Appl.* 13 (4), 562-571. doi: 10.14569/IJACSA.2022.0130466

Ruby, U., and Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* 9 (4), 5393-5397. doi: 10.30534/ijatcse/2020/175942020

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Salt Lake City, USA: IEEE) 4510–4520. doi: 10.1109/CVPR.2018.00474

Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A., and Pena-Gimenez, A. (2016). ShipsEar: An underwater vessel noise database. *Appl. Acoustics* 113, 64–69. doi: 10.1016/j.apacoust.2016.06.008

Setiyowati, H., Thalib, S., Setiawati, R., Nurjannah, N., and Akbariani, N. V. (2022). An aquaculture disrupted by digital technology. *Austenit* 14 (1), 12–16. doi: 10.5281/zenodo.6499775

Singh, P., Waldekar, S., Sahidullah, M., and Saha, G. (2022). Analysis of constant-Q filterbank based representations for speech emotion recognition. *Digital Signal Process.* 130, 103712. doi: 10.1016/j.dsp.2022.103712

Tian, C., Zheng, M., Zuo, W., Zhang, B., Zhang, Y., and Zhang, D. (2023). Multi-stage image denoising with the wavelet transform. *Pattern Recognition* 134, 109050. doi: 10.1016/j.patcog.2022.109050

Wang, M., and Huang, P. (2023). A multireceiver SAS imaging algorithm and optimization. *IEEE Access* 11, 75112-75120. doi: 10.1109/ACCESS.2023.3297138

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., and Yeh, I. H. (2020). "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Virtual: IEEE) 390–391. doi: 10.1109/CVPRW50498.2020.00203

Wang, Z., Wang, Z., Zeng, C., Yu, Y., and Wan, X. (2023). High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network. *Circuits Systems Signal Process.* 42 (3), 1593–1616. doi: 10.1007/s00034-022-02181-6

Wang, Q., and Zeng, X. Y. (2015). Deep learning methods and their applications in underwater targets recognition. *Tech. Acoust* 34, 138–140.

Wang, B., Zhang, W., Zhu, Y., Wu, C., and Zhang, S. (2023). An underwater acoustic target recognition method based on AMNet. *IEEE Geosci. Remote Sens. Lett.* 20, 1–5. doi: 10.1109/LGRS.2023.3235659

Wu, Y., Duan, Y., Wei, Y., An, D., and Liu, J. (2022). Application of intelligent and unmanned equipment in aquaculture: A review. *Comput. Electron. Agric.* 199, 107201. doi: 10.1016/j.compag.2022.107201

Xiao, X., Wang, W., Ren, Q., Gerstoft, P., and Ma, L. (2021). Underwater acoustic target recognition using attention-based deep neural network. *JASA Express Lett.* 1 (10), 106001. doi: 10.1121/10.0006299

Xu, J., Huang, Z., and Li, C. (2019). Advances in underwater target passive recognition using deep learning. *J. Signal Process.* 35, 1460–1475. doi: 10.16798/j.issn.1003-0530.2019.09.003

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Appl.* doi: 10.1007/s11042-023-16757-0

Yang, H., Xu, G., Li, J., Shen, S., and Yao, X. (2019). Summary of passive underwater acoustic target recognition. *Unmanned Syst. Technol.* 2, 1–7.

Yang, S., Xue, L., Hong, X., and Zeng, X. (2023). A lightweight network model based on an attention mechanism for ship-radiated noise classification. *J. Mar. Sci. Eng.* 11, 432. doi: 10.3390/jmse11020432

Yao, Q., Wang, Y., and Yang, Y. (2023). Underwater acoustic target recognition based on data augmentation and residual CNN. *Electronics* 12, 1206. doi: 10.3390/electronics12051206

Yao, X., Yang, H., and Sheng, M. (2023). Automatic modulation classification for underwater acoustic communication signals based on deep complex networks. *Entropy* 25, 318. doi: 10.3390/e25020318

Zhang, X. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *Multimedia Tools Appl.* doi: 10.1007/s11042-023-16992-5

Zhang, R., He, C., Jing, L., Zhou, C., Long, C., and Li, J. (2023). A modulation recognition system for underwater acoustic communication signals based on higher-order cumulants and deep learning. *J. Mar. Sci. Eng.* 11, 1632. doi: 10.3390/jmse11081632

Zhou, Y., Chen, S.-c., Wu, K., Ning, M.-q., Chen, H.-k., and Zhang, P. (2021). SCTD 1.0:Sonar common target detection dataset. *Comput. Sci.* 48 (11A), 334–339. doi: 10.11896/jsjkx.210100138

Zhou, C., Wu, Y., Fan, Z., Zhang, X., Wu, D., and Tao, Z. (2022). Gammatone spectral latitude features extraction for pathological voice detection and classification. *Appl. Acoustics* 185, 108417. doi: 10.1016/j.apacoust.2021.108417

Zhu, C., Cao, T., Chen, L., Dai, X., Ge, Q., and Zhao, X. (2023). High-order domain feature extraction technology for ocean acoustic observation signals: a review. *IEEE Access* 11, 17665-17683. doi: 10.1109/ACCESS.2023.3244782

Zhu, X., Dong, H., Rossi, P. S., and Landrø, M. (2022). Time-frequency fused underwater acoustic source localization based on contrastive predictive coding. *IEEE Sens J.* 22, 13299–13308. doi: 10.1109/JSEN.2022.3179405

Zhufeng, L., Xiaofang, L., Na, W., and Qingyang, Z. (2022). Present status and challenges of underwater acoustic target recognition technology: A review. *Front. Phys.* 10, 1044890. doi: 10.3389/fphy.2022.1044890

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fmars.2023.1329898

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Yanqun Wu,
National University of Defense Technology, China
Wei Guo,
National University of Defense Technology, China
Kun Ye,
Xiamen University, China

*CORRESPONDENCE
Yunfei Lv
lvyunfei@hrbeu.edu.cn
Jidan Mei
meijidan@hrbeu.edu.cn

RECEIVED 30 October 2023
ACCEPTED 18 December 2023
PUBLISHED 08 January 2024

CITATION
Lu M, Sun D, Gulliver TA, Lv Y and Mei J (2024) Broadband high-resolution direction of arrival estimation using the generalized weighted Radon transform.
Front. Mar. Sci. 10:1329898.
doi: 10.3389/fmars.2023.1329898

Broadband high-resolution direction of arrival estimation using the generalized weighted Radon transform

Mingyang Lu[1,2,3], Dajun Sun[1,2,3], T. Aaron Gulliver[4], Yunfei Lv[1,2,3]* and Jidan Mei[1,2,3]*

[1]National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin, China, [2]Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin, China, [3]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, [4]Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada

Traditional direction of arrival (DOA) estimation algorithms typically have poor spatial resolution and robustness. In this paper, we propose a broadband high-resolution DOA estimation method based on the generalized weighted Radon transform (GWRT). The array signal can be converted into the frequency-wavenumber (*f-k*) domain using the conditional wavenumber spectrum function (CWSF). Then, a linear integral mathematical model for high-resolution DOA estimation is derived by transforming the *f-k* domain into the azimuth-energy domain using the GWRT. Computer simulation and sea trials were conducted to validate the feasibility and performance of the proposed method. The results obtained indicate that the proposed method yields a lower sidelobe level and can more effectively suppress the output energy in the non-target direction when compared to the conventional beamforming (CBF), steered minimum variance (STMV), and deconvolution (DCV) methods. Further, the proposed method provides improved spatial resolution and robustness in a multi-target environment.

KEYWORDS

direction of arrival estimation, generalized weighted Radon transform, broadband signal, high-resolution, low sidelobe levels

1 Introduction

Target azimuth is an important parameter for the identification, detection, positioning, and tracking of underwater targets (Luo and Shen, 2021; Chen et al., 2023; Xie et al., 2023; Zhao et al., 2023). Array signal processing has been shown to be effective for the direction of arrival (DOA) estimation. The methods can be classified as traditional beamforming, subspace-based, deconvolution (DCV), and transform domain.

The most commonly employed traditional beamforming method is the conventional beamforming (CBF). However, CBF has the disadvantages of wide beamwidths and poor spatial resolution due to the Rayleigh limit. Several high-resolution methods have been proposed to overcome these such as the maximum entropy algorithm (Burg, 1975) and the minimum variance distortionless response (MVDR) algorithm (Capon, 1969). These methods improve the spatial resolution but suffer from poor performance when used to detect broadband signals in actual ocean environments. This is because they are sensitive to signal mismatch and cannot estimate the DOA of coherent sound sources (Somasundaram, 2012). The steered minimum variance (STMV) algorithm was proposed to overcome these problems (Krolik and Swingler, 1989). STMV has better spatial resolution for coherent acoustic sources and fast convergence, but poor robustness (Somasundaram et al., 2015). The Rayleigh limit was overcome with the subspace-based algorithm multiple signal classification (MUSIC) (Schmidt, 1986). Subsequent subspace-based algorithms such as estimation of signal parameters via rotational invariance techniques (ESPRIT) (Roy and Kailath, 1989), root-multiple signal classification (RMUSIC) (Rao and Hari, 1989), maximum likelihood (ML) (Stoica and Nehorai, 1989), and weighted subspace fitting (WSF) (Bengtsson and Ottersten, 2001) provide improved performance but have sensitivity and snapshot deficiency problems when used in practical applications (Baggeroer and Cox, 1999). Another issue is that the number of acoustic sources is usually unknown and this makes it difficult to estimate the signal and noise subspaces. Further, existing algorithms can only be used to estimate DOA for incoherent or weak-coherent acoustic sources, making detection of coherent signals difficult in the actual ocean environments.

The above algorithms are either sensitive to array element errors or limited to array aperture. Various studies on the formation structure have been introduced to improve the performance of the algorithms (Zhang et al., 2021; Zhou et al., 2022; Yang, 2023; Ye et al., 2023). Additionally, in recent years, there has been a growing focus on researching robust high-resolution beamforming algorithms. Deconvolution (DCV) algorithms have attracted widespread attention for underwater acoustic applications. DCV was initially considered with both uniform linear arrays and circular arrays (Yang, 2017; Yang, 2018). It was shown that the performance is better than CBF. The super-directivity performance of DCV with a small-sized array was verified using the SwellEx96 horizontal array (Yang, 2019). However, these DCV methods are only suitable for arrays with a shift-invariant point spread function (PSF) beam pattern, such as a horizontal line array or circular array. Therefore, new DCV methods were developed for shift-variant PSF beam patterns. A DCV method based on non-negative least squares (NNLS) and an improved NNLS method called extended Richardson-Lucy (Ex-RL) were presented which provide high resolution, robustness, and excellent array gain (Sun et al., 2019; Sun et al., 2020).

Transform domain methods were originally developed to estimate seismic wave velocity and azimuth (Cheng et al., 2018). More specifically, the frequency-wavenumber (*f-k*) power spectrum can be obtained using the space-time two-dimensional Fourier transform of the seismic signal. Then, the *f-k* power spectrum can be converted into the transform domain to extract the velocity and azimuth of the waves. While transform domain methods have been widely used in seismic exploration (Zywicki and Rix, 1999), there have been few DOA estimation applications. The least squares line fitting (LSLF) algorithm was employed to obtain the slope of the local peak-energy line in the *f-k* domain and then the sum of the points on this line was used as an estimate of the energy output of the azimuth spectrum (Li et al., 2019). However, this method is sensitive to outliers in the image since it minimizes the sum of the squares of the distances from the points to the line. Thus, the performance can be degraded significantly, particularly in low signal-to-noise ratio (SNR) environments or when there are multiple adjacent targets. In this paper, a broadband high-resolution DOA estimation method based on the generalized weighted Radon transform (GWRT) is proposed. The array signal is converted into the *f-k* domain by solving the conditional wavenumber spectrum function (CWSF) and then the mathematical relationship between the spatial distribution of broadband signal energy in the *f-k* domain and target azimuth is obtained. To improve performance, image gradient information is utilized as weights for the GWRT, and a linear integral mathematical model is derived by the GWRT processing in the *f-k* domain. The resulting model contains the complete image information in the *f-k* domain. This is then converted into the azimuth-energy domain to realize high-resolution DOA estimation. The proposed method does not require prior knowledge of the number of sources or signal pre-estimation. In addition, it is not sensitive to outliers in the image and the results in the *f-k* domain provide higher transform gain and better robustness. Both simulation and sea-trial experiments are conducted to validate the proposed method. The results obtained indicate that the proposed method has better performance and offers several advantages compared with existing approaches as follows.

1. The proposed method produces a narrow mainlobe width similar to, or better than, many commonly used high-resolution methods such as STMV and DCV.
2. The proposed method produces lower sidelobe levels than the CBF, STMV, and DCV methods.
3. The proposed method has better robustness to position errors compared to the STMV and DCV methods.
4. The proposed method exhibits good performance when there are multiple targets and when the target signal is weak.

The remainder of this paper is organized as follows. Section 2 introduces the broadband signal model and DOA estimation using the CWSF is presented. In Section 3, we derive the expression of the mathematical model for DOA estimation using GWRT. The performance of the proposed method is evaluated via simulation and compared with other DOA methods in Section 4. The results of the sea-trial experiments are given in Section 5. Finally, Section 6 provides a summary of the paper.

2 Signal modeling

We consider a linear sensor array with M receivers uniformly spaced at a distance d. The signal is assumed to be from a broadband source located at the far field of the array. This signal has a look direction θ and arrives at the array as a plane wave. Then, θ is the azimuth angle of the target, which is defined as the anticlockwise angle between the horizontal array and the target as shown in Figure 1. Therefore, the signal received by element m at time t can be expressed as

$$x_m(t) = a_m s(t) + n_m(t) \tag{1}$$

where $m = 1, 2, \cdots, M$, $s(t)$ is the incident signal, $n_m(t)$ is the noise received by element m, which is uncorrelated with $s(t)$, a_m is the array manifold factor of element m which is equal to $\exp(\mathrm{j}2\pi f(m-1)d\cos\theta/c)$, and f and c are frequency and the speed of sound, respectively.

As the actual data processing is based on a discrete-time model, the received signal should be sampled at f_s. Therefore, the signal received by element m can be expressed as

$$x_m(l) = x_m(t)|_{t=l/f_s},\ l = 1, \cdots, L \tag{2}$$

where L is the number of samples which is an integer. The discrete Fourier transform (DFT) of $x_m(l)$ gives the corresponding frequency spectrum

$$X_m(f) = \frac{1}{L}\sum_{l=1}^{L} x_m(l) e^{-\frac{j2\pi fl}{L}}. \tag{3}$$

Then, the frequency-wavenumber spectrum of the received signal can be obtained by applying the DFT to $X_m(f)$ with Q points in the spatial domain as

$$\begin{aligned} I(k,f) &= \frac{1}{Q}\sum_{m=1}^{M} X_m(f) e^{-j2\pi mk/Q} \\ &= \frac{1}{Q}\sum_{m=1}^{M}\left[S(f)e^{\left(\frac{j2\pi f(m-1)d\cos\theta}{c}\right)} + N_m(f)\right]e^{-j2\pi mk/Q} \end{aligned} \tag{4}$$

where k is the wavenumber which is an integer in the range $-Q/2 \ll k < Q/2$, Q is an integer with $Q \gg M$, $S(f)$ is the frequency spectrum of $s(t)$, $N_m(f)$ is the frequency spectrum of $n_m(t)$.

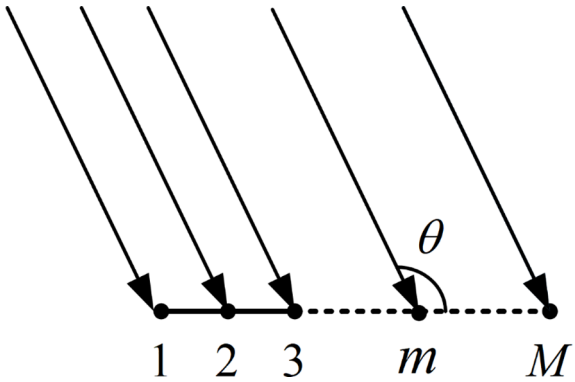


FIGURE 1
Uniform linear array geometry.

From Equation (4), the frequency-wavenumber power spectral density can be expressed as

$$\begin{aligned} \Psi(k,f) &= \lim_{Q,L\to\infty} E\{I^{\mathrm{H}}(k,f)I(k,f)\} \\ &= |S(f)|^2 \left|\frac{\sin\left[\pi M\left(fd\cos\theta/c - \frac{k}{Q}\right)\right]}{Q\sin\left[\pi\left(fd\cos\theta/c - \frac{k}{Q}\right)\right]}\right|^2 + M\frac{\sigma^2}{Q^2} \end{aligned} \tag{5}$$

where $E\{\cdot\}$ denotes expectation, superscript H denotes conjugate transpose, and σ^2 is the noise power. The frequency power spectra $\Psi(f)$ can be expressed as

$$\Psi(f) = \sum_{k=-Q/2}^{Q/2}\Psi(k,f). \tag{6}$$

To mitigate the impact of high-frequency attenuation and enhance the outcomes of high-frequency components within $\Psi(k,f)$, the CWSF (Beall et al., 1982) is employed to derive the conditional wavenumber spectral density $\Psi(k|f)$ from Equations (5) and Equations (6). $\Psi(k|f)$ can be expressed as

$$\begin{aligned} \Psi(k|f) &= \frac{\Psi(k,f)}{\Psi(f)} \\ &= \frac{|S(f)|^2}{M(|S(f)|^2+\sigma^2)}\left|\omega\left(\frac{fd\cos\theta}{c} - \frac{k}{Q}\right)\right|^2 + \frac{\sigma^2}{Q^2(|S(f)|^2+\sigma^2)} \end{aligned} \tag{7}$$

where $\omega(v) = \left|\frac{\sin(\pi Mv)}{Q\sin(\pi v)}\right|$ is a periodic function with period 1. It is symmetric about $v = 0$, and has its maximum value when $v = 0$. The first zero points of $\omega(v)$ are $v_0 = \pm 1/M$, hence the mainlobe width is $2/M$.

In this paper, we only consider $\omega(v)$ for a single cycle. From Equation (7), $\Psi(k|f)$ has its maximum value when $v = \frac{fd\cos\theta}{c} - \frac{k}{Q} = 0$, and mainlobe width $2Q/M$. Figure 2 gives $\Psi(k|f)$ for a broadband signal. This shows that the target energy is concentrated in the mainlobe, and the peak-energy points in the mainlobe are on a straight line r passing through the origin. The slope ε of the line r can be expressed as $\varepsilon = \frac{k}{f} = \frac{Qd\cos\theta}{c}$. Therefore, ε is a linear function of $\cos\theta$.

3 DOA estimation method based on GWRT

As mentioned above, θ can be accurately estimated using the slope of the line r in $\Psi(k|f)$, thus realizing high-resolution DOA estimation. However, there will be sidelobels and perhaps outliers in $\Psi(k|f)$ due to the windowing effect of the DFT and the random fluctuation noise which makes determining r difficult. To solve these issues, morphological grayscale reconstruction is used to extract regional maxima in $\Psi(k|f)$ and obtain the reconstructed matrix $\Psi_{\mathrm{new}}(k|f)$ (Vincent, 1992; Vincent, 1993). This method utilizes erosion and dilation operations based on a structuring element to reconstruct or eliminate specific regions in an image, which removes most of the outliers and significantly reduces the sidelobes.

The generalized Radon transform (GRT) is commonly used to extract information from images (Radon, 1986; Hansen and Toft,

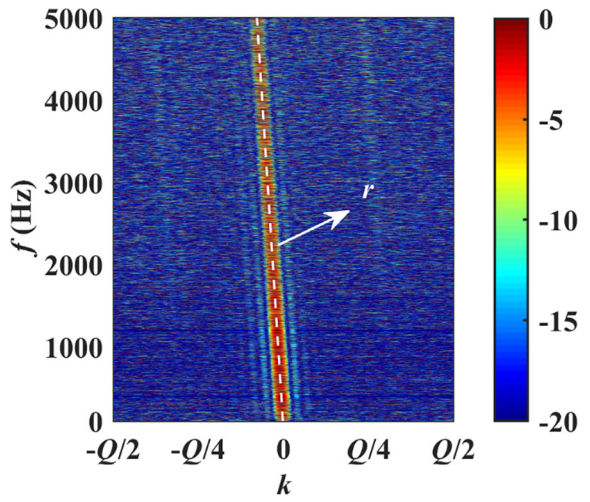


FIGURE 2
$\Psi(k|f)$ for a broadband signal.

1996; Ramm and Katsevich, 2020; Sun et al., 2021). However, it only considers amplitude information and ignores gradient information, which is not ideal. Image gradients provide the rates and directions of change for the pixels, which is useful information. Therefore, we propose a method based on the GWRT to achieve the integration of a multivariate function over a given path. Compared with the GRT, the GWRT makes full use of the image gradient information as its weights, thus providing better results. For a two-dimensional Euclidean space, the GWRT can be defined for a continuous image as (Alpatov et al., 2015)

$$U_{\text{weighted}}(\boldsymbol{p}) = \iint_{(x,y)\in c(\boldsymbol{p})} \langle \nabla \phi(x,y), \vec{n} \rangle^2 \, \mathrm{d}x\mathrm{d}y \tag{8}$$

where $\boldsymbol{p}$ is a vector containing the parameters of the line, and $c(p)$, (x, y), and $\phi(x, y)$ are, respectively, a known line, space coordinates, and the intensity of points on a line in the two-dimensional image ϕ.∇, $\langle\ \rangle$, and $\vec{n}$ in Equation (8) are, respectively, the gradient operator, scalar product operation, and the unit normal vector which is perpendicular to the line $c(p)$. As the transform will be obtained from discrete-time data, the GWRT in discrete form is used rather than the integral form. The GWRT of $\Psi_{\text{new}}(k|f)$ in discrete form can be expressed as

$$U_{\text{weighted}}(\theta) = \sum_{f_{\min}}^{f_{\max}} \sum_{\cos\theta=-1}^{\cos\theta=1} \left\langle \nabla \ \Psi_{\text{new}}\left(f, k = \varepsilon f = \frac{Qfd\cos\theta}{c}\right), \vec{n} \right\rangle^2 \tag{9}$$

where $f_{\min}$ and $f_{\max}$ are the lower and upper limits of frequency employed, respectively, with $f_{\max} \le c/2d$. Equation (9) converts $\Psi_{\text{new}}(k|f)$ into a one-dimensional matrix $U_{\text{weighted}}(\theta)$ which reflects the energy distribution related to the parameter θ. The position of the maximum value of $U_{\text{weighted}}(\theta)$ is the DOA estimation.

As mentioned above, the slope of the peak-energy line in $\Psi(k|f)$ is a function of the signal direction θ. Therefore, the lines in $\Psi(k|f)$ must be discernible so targets from different directions can be distinguished. In other words, the difference in coordinates on the k-axis for $f_{\max}$ must be equal to or greater than the mainlobe width of $\Psi(k|f)$ which means

$$\frac{f_{\max} Qd}{c} (\cos\theta_1 - \cos\theta_2) \ge \frac{2Q}{M} \tag{10}$$

where $\cos\theta_1$ and $\cos\theta_2$ represent the directions of two targets. Then, the difference in directions should satisfy

$$\cos\theta_1 - \cos\theta_2 \ge \frac{2c}{Mf_{\max}d}. \tag{11}$$

Equation (11) is a function of $f_{\max}$, so this frequency should be large to obtain high-resolution performance. Note that the mainlobe of $\Psi_{\text{new}}(k|f)$ will be much narrower than that of $\Psi(k|f)$ due to the morphological grayscale reconstruction operation. Therefore, the resolution of the GWRT will be less than $\frac{2c}{Mf_{\max}d}$, which confirms that the proposed method has high-resolution performance.

The steps of the proposed method are as follows.

1) Obtain $\Psi(k|f)$ using Equtaions (1–7).
2) Perform morphological grayscale reconstruction to obtain the matrix $\Psi_{\text{new}}(k|f)$.
3) Compute the GWRT of $\Psi_{\text{new}}(k|f)$ to convert the image information in the f-k domain into the azimuth-energy domain to realize high-resolution DOA estimation.

4 Simulation analysis

The performance of the proposed method is evaluated for a scalar towed array measurement system. Consider a line array of 32 receivers uniformly spaced at a distance d=0.25 m. The proposed method is compared with three commonly used DOA estimation methods, namely CBF, STMV, and DCV.

4.1 Single source

Consider a broadband target located in the direction of the array with $\cos\theta = 0.5$. The target signal is a broadband noise and is assumed to have random amplitude and phase. The broadband spectrum is between 1500 Hz and 3000 Hz, and f_s = 20 kHz. The direction scanning range is $\cos\theta \in [-1, 1]$, the scanning interval is 1/1800 rad, and Q=256. The SNR is 10 dB, and the noise is assumed to be isotropic and uncorrelated at the receivers. $\Psi(k|f)$ and $\Psi_{\text{new}}(k|f)$ obtained using Steps 1 and 2 in Section 3 are shown in Figures 3A, B, respectively.

Comparing Figures 3A, B reveals that $\Psi_{\text{new}}(k|f)$ is more prominent and the mainlobe width is narrower. This confirms the improvement due to morphological grayscale reconstruction. The DOA estimation results obtained from the GWRT of $\Psi_{\text{new}}(k|f)$ are given in Figures 3C, D for SNRs 10 dB and -10 dB, respectively. These results show that all methods can accurately estimate the direction of the target for both SNR values. The sidelobe levels increase as the SNR decreases, but the proposed method still exhibits the lowest sidelobe levels. The GWRT method also has a narrower mainlobe width than the CBF and STMV methods, and it is similar to that of the DCV method. Therefore, the GWRT method has the advantages of lower sidelobe levels and narrower mainlobe which will result in better performance.

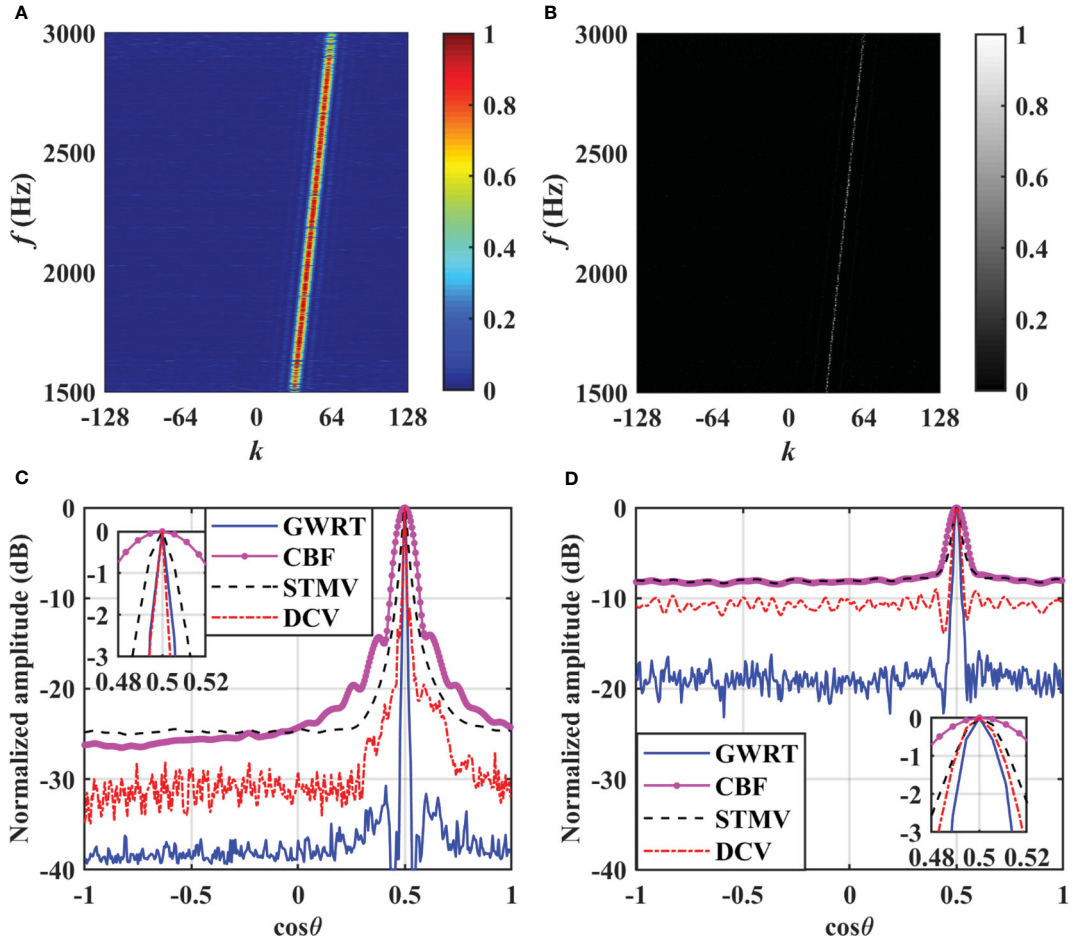


FIGURE 3
DOA estimation results for a single source. **(A)** $\Psi(k|f)$. **(B)** $\Psi_{new}(k|f)$. **(C)** SNR=10 dB. **(D)** SNR= -10 dB.

4.2 Sensitivity to position errors

Sensitivity to signal mismatch is an important consideration for DOA estimation methods. In this section, the performance degradation due to signal mismatch is evaluated with random position errors for the receivers. Figure 4A gives the actual receiver positions in the line array (denoted by *) and the erroneous positions (denoted by o). The position errors have a mean of 0.04 m which can be considered worst case. It is assumed that these errors are unknown and DOA estimation is conducted assuming a straight line array. The other simulation conditions are the same as above. The corresponding DOA estimation results are given in Figure 4B for an SNR of 10 dB.

Compared to Figure 3C, these results show that the sidelobe levels increase with position errors. The STMV method has the greatest performance degradation, and the DCV method has false peaks which may significantly affect the estimation accuracy. The GWRT method still has the lowest sidelobe levels and so has good robustness even with position errors.

4.3 Multiple sources

To further evaluate the proposed method, the performance with three targets is now obtained. The three broadband targets are located at the far field of the array with directions $\cos\theta = -0.05$, $\cos\theta = 0.05$, and $\cos\theta = 0.2$. The SNRs of these targets are -5dB, 10dB and -5dB, respectively. The other simulation conditions are the same as above. Figure 5A presents $\Psi(k|f)$ and the DOA estimation results are shown in Figure 5B.

These results indicate the CBF and STMV methods only identify the second and third targets. This is because the directions of the first and second targets are close and the energy difference is large. The DCV and GWRT methods are able to distinguish all three targets. Although these methods have similar mainlobe widths, the former method produces false peaks and has higher sidelobe levels, making it easy for weak targets to be missed. The peak-energy lines in Figure 5A corresponding to the first and third targets are barely distinguishable due to the strong interference from the second target. However, the proposed

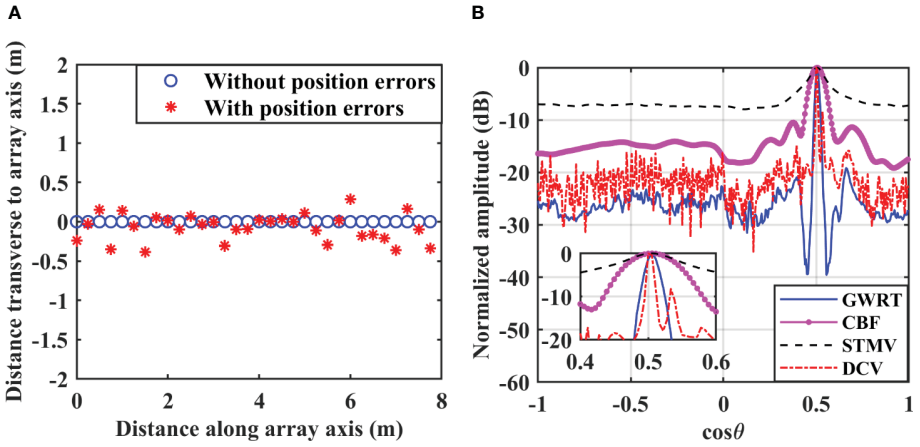


FIGURE 4
DOA estimation results for four methods with position errors. **(A)** Top view of the array configuration. **(B)** SNR=10 dB.

method uses information in the azimuth-energy domain which confirms the advantages of the GWRT. Therefore, the proposed method herein has excellent anti-jamming capability and high-resolution performance even with multiple targets having different SNRs.

4.4 DOA estimation versus SNR and array size

The performance of the proposed method is now evaluated for different SNRs and numbers of array elements. The other simulation conditions are the same as in Section 4.1.

4.4.1 Effect of SNR

An increase in noise and/or interference affects the sidelobe levels and so can degrade performance as noise suppression and interference discrimination are determined by these levels (Ma et al., 2021). In this section, the highest sidelobe level in the azimuth spectrum and the root mean square error (RMSE) of the estimated azimuth are considered as the SNR varies from -10 dB to 10 dB. The RMSE of the estimated azimuth is calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\hat{\theta}_j - \theta)^2} \tag{12}$$

where J is the number of Monte Carlo trials, and $\hat{\theta}_j$ and θ are the estimated azimuth for the jth independent experiment and the true orientation of the target, respectively. The scanning interval is 1/18000 rad herein. The average results for 100 Monte Carlo trials are given in Figure 6. Figure 6A shows that the highest sidelobe level decreases with increasing SNR for all four methods. For SNR>5dB, the highest sidelobe level with the CBF method is around -13 dB, and the performance of the STMV method is slightly worse than with the DCV method. The sidelobe levels with the GWRT method are lower than the other methods for all SNR values, and at least 4 dB less than with the DCV method which is the second best. The SNR in underwater acoustic applications is often low so the proposed method is preferable. Additionally, the Cramér-Rao bound (CRB) (Feng and Huang, 2007) is included as a reference for DOA estimation performance, as shown in Figure 6B. The

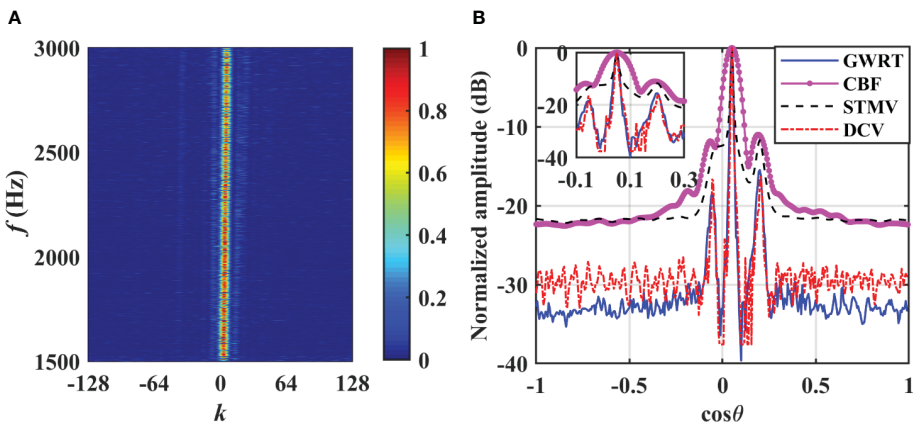


FIGURE 5
DOA estimation results for multiple sources. **(A)** $\Psi(k|f)$. **(B)** Azimuth spectrum.

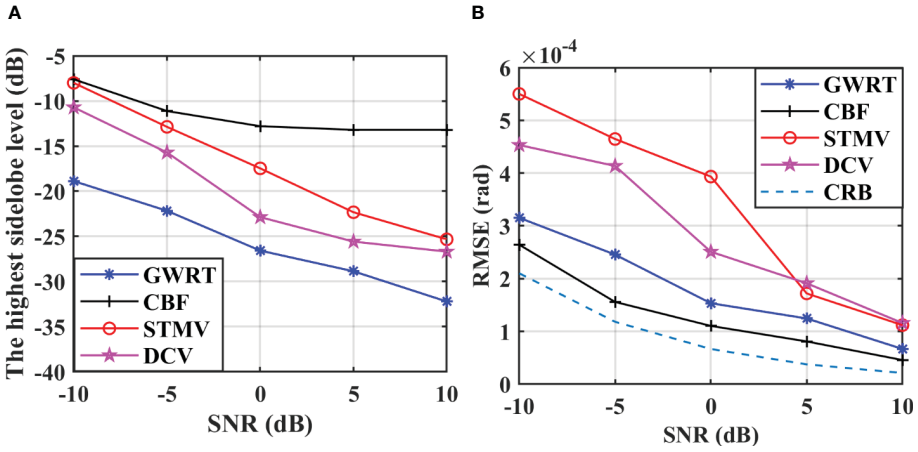


FIGURE 6
The performance versus SNR for four methods. **(A)** The highest sidelobe level versus SNR for four methods. **(B)** The RMSE versus SNR for four methods.

RMSE of the GWRT method is smaller in comparison to the DCV and STMV methods, with only a slight increase relative to the CBF method.

4.4.2 Effect of the number of array elements

Angle resolution is the smallest angle difference between the directions of two targets and is an important criterion in evaluating DOA estimation methods. The angle resolution $\Delta\theta$ of the four algorithms was evaluated for different numbers of array elements and frequency bands. The number of elements varies from 12 to 36, and the frequency bands are 2500 Hz to 3000 Hz and 1500 Hz to 3000 Hz, respectively. The average angle resolution for 100 Monte Carlo trials for an SNR of 10 dB is given in Figure 7. This shows that the angle resolution improves with an increase in the number of elements with all four methods. The DCV method exhibits the highest resolution, followed by the GWRT method and the STMV, all of which outperform the CBF. Comparing Figures 7A, B indicates that the performance of the STMV method is severely degraded with a wider frequency band. The main reason is that the covariance matrix is obtained by averaging the covariance matrices for each frequency point, and increasing the number of frequency points decreases the accuracy of this matrix and thus the angle resolution. However, the GWRT method has better robustness with broadband signals. Furthermore, as discussed in Section 3, the resolution of the GWRT can be less than $\frac{2c}{Mf_{\max}d}$, which is consistent with these results.

5 Sea-trial results

To evaluate the performance of the proposed method in practical applications, DOA estimation results were obtained for a towed line array with 32 elements uniformly spaced at 2 m. The experiments were conducted in Huanghai Sea, China in the summer of 2014. The water depth was approximately 40 m and the towed array was about 20 m above the sea floor. The recorded data suggests that the received signals include signals from passing vessels, experimental vessels, and the broadband pulses

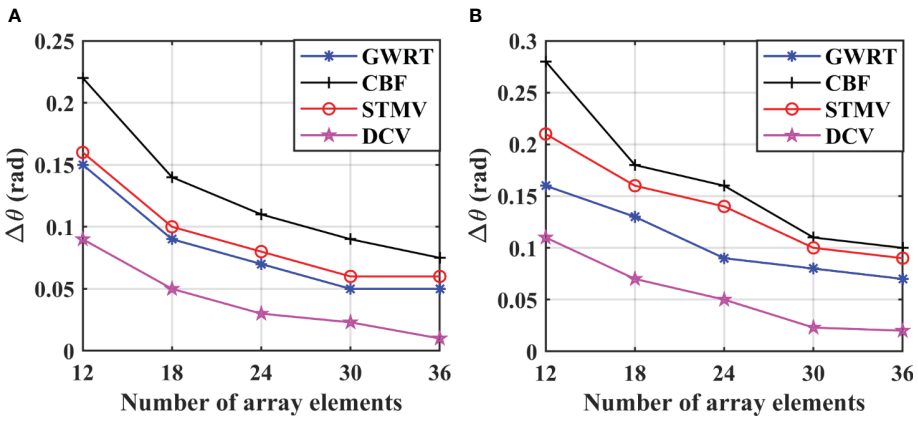


FIGURE 7
The angle resolution versus the number of array elements for two frequency bands and four methods. **(A)** 2500~3000Hz. **(B)** 1500~3000Hz.

transmitted by the experimental vessels. The sampling frequency was equal to 8 kHz, with a total of 120 data frames, each comprising 4096 samples. The data within the frequency range of 1500 to 3000 Hz was processed using the CBF, STMV, DCV, and GWRT methods and the bearing time records (BTRs) are given in Figure 8. Additionally, the low-frequency analysis and recording (LOFAR) for the data from one element is shown in Figure 9.

The BTRs for the CBF method in Figure 8A show that due to the Rayleigh limit and the ambient ocean noise, the mainlobe width for each target is relatively wide so it is not possible to distinguish the targets located in the directions around $\cos\theta = 0.78$ and $\cos\theta = 0.88$. Figure 8B indicates that the STMV method has a narrower mainwidth but still fails to distinguish the two targets. The BTRs for the DCV and GWRT methods in Figures 8C, D, respectively, have much clearer backgrounds than with the CBF and STMV methods. The target trajectories are clearly distinguishable with a much narrower mainlobe width for each target. Figure 8 also shows a set of broadband pulse signals in the direction of around $\cos\theta = 0.30$. There are two clear focused points in the red circle in Figure 8D which are not as well distinguished by other methods. This indicates that the GWRT method has a lower background noise level and thus better weak target detection and anti-interference capability.

Figure 10 gives the BTRs for the four methods at 60 s. These results indicate that the target located in the direction of around $\cos\theta = 0.78$ cannot be distinguished by the CBF and STMV methods due to the strong inference from the target located in the direction of around $\cos\theta = 0.88$. Conversely, both the DCV and GWRT methods clearly distinguish these targets. The GWRT method has a mainlobe width similar to that of the DCV method but the sidelobe levels are lower. Thus, it is better able to suppress the interference due to strong targets and noise which makes it easier to detect weak targets. Therefore, the proposed method provides better high-resolution performance in multiple target environments.

6 Conclusion

A generalized weighted Radon transform to estimate the DOA for broadband targets was proposed. The GWRT was used on the conditional wavenumber spectrum density to convert image

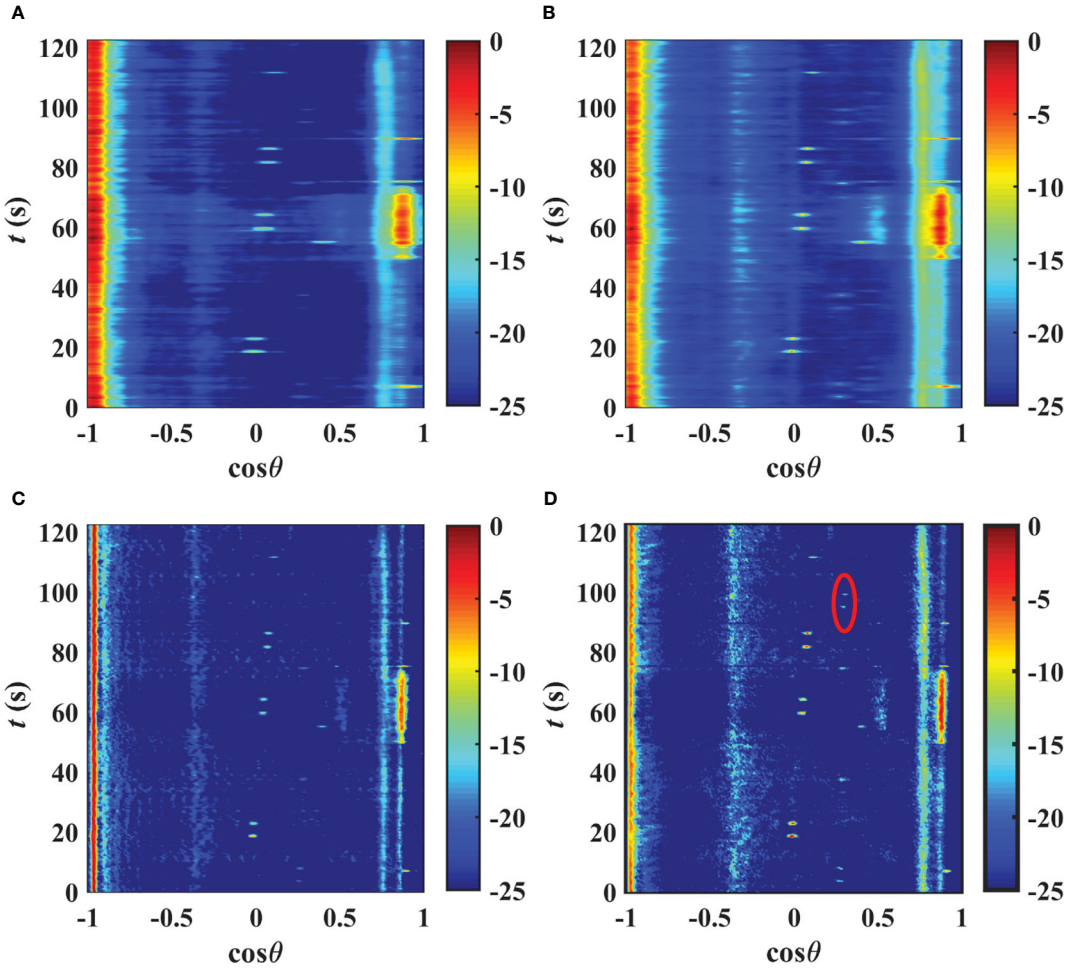


FIGURE 8
BTRs for four methods. **(A)** CBF. **(B)** STMV. **(C)** DCV. **(D)** GWRT.

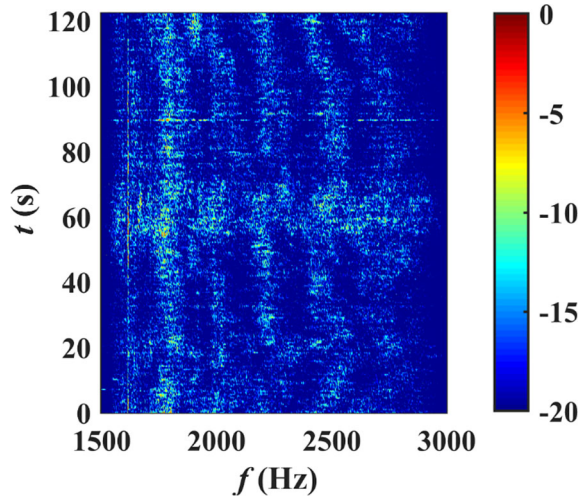


FIGURE 9
LOFAR for the data from one element.

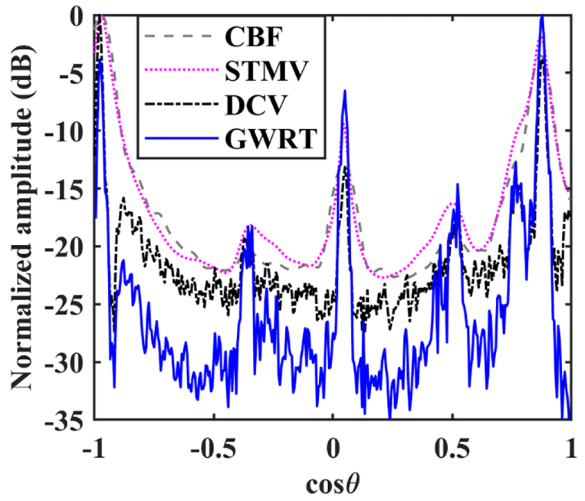


FIGURE 10
BTRs for four methods at 60 s.

information in the *f-k* domain to the azimuth-energy domain for high-resolution DOA estimation. Simulation and sea-trial results were presented which show that the proposed method is simple and effective and does not require a priori information. It is not sensitive to the outliers and thus provides good robustness even with position errors. Furthermore, it produces a narrow mainlobe with low sidelobe levels which results in good performance when there are multiple targets and the target SNR is low. However, the proposed method is only applicable for broadband signals and it is not suitable for real-time applications. Therefore, a short-time model for DOA estimation with narrow-band signals will be considered as future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ML: Conceptualization, Validation, Writing – original draft, Methodology. DS: Supervision, Writing – review & editing, Funding acquisition, Project administration. TG: Supervision, Writing – review & editing, Validation. YL: Supervision, Writing – review & editing, Data curation, Validation. JM: Data curation, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China Grant No. 61871144.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alpatov, B. A., Babayan, P. V., and Shubin, N. Y. (2015). Weighted Radon transform for line detection in noisy images. *J. Electronic Imaging.* 24 (2), 23023. doi: 10.1117/1.JEI.24.2.023023

Baggeroer, A. B., and Cox, H. (1999). "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," in *Conference record of the thirty-third asilomar conference on signals, systems, and computers* (USA: IEEE), 103–108. doi: 10.1109/ACSSC.1999.832304

Beall, J. M., Kim, Y. C., and Powers, E. J. (1982). Estimation of wavenumber and frequency spectra using fixed probe pairs. *J. Appl. Physics.* 53 (6), 3933–3940. doi: 10.1063/1.331279

Bengtsson, M., and Ottersten, B. (2001). A generalization of weighted subspace fitting to full-rank models. *IEEE Trans. Signal Processing.* 49 (5), 1002–1012. doi: 10.1109/78.917804

Burg, J. P. (1975). *Maximum entropy spectral analysis* (California, United States: Stanford University).

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE.* 57 (8), 1408–1418. doi: 10.1109/PROC.1969.7278

Chen, X., Zhang, H., Gao, Y., and Wang, Z. (2023). DOA estimation of underwater acoustic co-frequency sources for the coprime vector sensor array. *Front. Mar. Science.* 10. doi: 10.3389/fmars.2023.1211234

Cheng, F., Xia, J., Xu, Z., Hu, Y., and Mi, B. (2018). Frequency–wavenumber (FK)-based data selection in high-frequency passive surface wave survey. *Surveys Geophysics.* 39, 661–682. doi: 10.1007/s10712-018-9473-3

Feng, X., and Huang, J. (2007). Cramer-Rao bound of DOA estimation for underwater wide band sources. *Acta Armamentarii.* 28 (3), 291–295. doi: 10.3321/j.issn:1000-1093.2007.03.008

Hansen, K. V., and Toft, P. A. (1996). Fast curve estimation using preconditioned generalized Radon transform. *IEEE Trans. Image Process.* 5(12), 1651–1661. doi: 10.1109/83.544572

Krolik, J., and Swingler, D. (1989). Multiple broad-band source location using steered covariance matrices. *IEEE Trans. Acoustics Speech Signal Processing.* 37 (10), 1481–1494. doi: 10.1109/29.35386

Li, X., Huang, H., Li, Y., Ye, Q., and Zhang, Y. (2019). A broadband high resolution direction of arrival estimation algorithm based on conditional wavenumber spectral density. *Chin. J. Acoustics.* 44 (4), 585–594. doi: 10.15949/j.cnki.0371-0025.2019.04.019

Luo, X., and Shen, Z. (2021). A space-frequency joint detection and tracking method for line-spectrum components of underwater acoustic signals. *Appl. Acoustics.* 172, 107609. doi: 10.1016/j.apacoust.2020.107609

Ma, C., Sun, D., Mei, J., and Teng, T. (2021). Spatiotemporal two-dimensional deconvolution beam imaging technology. *Appl. Acoustics.* 183, 108310. doi: 10.1016/j.apacoust.2021.108310

Radon, J. (1986). On the determination of functions from their integral values along certain manifolds. *IEEE Trans. Med. Imaging.* 5 (4), 170–176. doi: 10.1109/TMI.1986.4307775

Ramm, A. G., and Katsevich, A. I. (2020). *The Radon transform and local tomography* (Florida: CRC press).

Rao, B. D., and Hari, K. S. (1989). Performance analysis of root-MUSIC. *IEEE Trans. Acoustics Speech Signal Processing.* 37 (12), 1939–1949. doi: 10.1109/29.45540

Roy, R., and Kailath, T. (1989). ESPRIT-estimation of signal parameters *via* rotational invariance techniques. *IEEE Trans. Acoustics Speech Signal Processing.* 37 (7), 984–995. doi: 10.1109/29.32276

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagation.* 34 (3), 276–280. doi: 10.1109/TAP.1986.1143830

Somasundaram, S. D. (2012). Wideband robust capon beamforming for passive sonar. *IEEE J. Oceanic Engineering.* 38 (2), 308–322. doi: 10.1109/JOE.2012.2223560

Somasundaram, S. D., Butt, N. R., Jakobsson, A., and Hart, L. (2015). Low-complexity uncertainty-set-based robust adaptive beamforming for passive sonar. *IEEE J. Oceanic Engineering.* 99, 1–17. doi: 10.1109/JOE.2015.2474495

Stoica, P., and Nehorai, A. (1989). MUSIC, maximum likelihood, and Cramer-Rao bound. *IEEE Trans. Acoustics Speech Signal processing.* 37 (5), 720–741. doi: 10.1109/29.17564

Sun, D., Lu, M., Mei, J., Wang, S., and Pei, Y. (2021). Generalized Radon transform approach to target motion parameter estimation using a stationary underwater vector hydrophone. *J. Acoustical Soc. America.* 150 (2), 952–968. doi: 10.1121/10.0005813

Sun, D., Ma, C., Mei, J., and Shi, W. (2020). Improving the resolution of underwater acoustic image measurement by deconvolution. *Appl. Acoustics.* 165, 107292. doi: 10.1016/j.apacoust.2020.107292

Sun, D., Ma, C., Yang, T. C., Mei, J., and Shi, W. (2019). Improving the performance of a vector sensor line array by deconvolution. *IEEE J. Oceanic Engineering.* 45 (3), 1063–1077. doi: 10.1109/JOE.2019.2912586

Vincent, L. (1992). "Morphological grayscale reconstruction: definition, efficient algorithm and applications in image analysis," in *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (USA: IEEE), 633–635.

Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Trans. Image Processing.* 2 (2), 176–201. doi: 10.1109/83.217222

Xie, Z., Lin, R., Wang, L., Zhang, A., Lin, J., and Tang, X. (2023). Data augmentation and deep neural network classification based on ship radiated noise. *Front. Mar. Science.* 10. doi: 10.3389/fmars.2023.1113224

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Applications.*, 1–17. doi: 10.1007/s11042-023-16757-0

Yang, T. C. (2017). Deconvolved conventional beamforming for a horizontal line array. *IEEE J. Oceanic Engineering.* 43 (1), 160–172. doi: 10.1109/JOE.2017.2680818

Yang, T. C. (2018). Performance analysis of superdirectivity of circular arrays and implications for sonar systems. *IEEE J. Oceanic Engineering.* 44 (1), 156–166. doi: 10.1109/JOE.2018.2801144

Yang, T. C. (2019). Superdirective beamforming applied to SWellEx96 horizontal arrays data for source localization. *J. Acoustical Soc. America.* 145 (3), 179–184. doi: 10.1121/1.5092580

Ye, K., Cai, Y., Hong, S., and Sun, H. (2023). Direction-of-arrival estimation based on difference-sum co-array of a special coprime array. *Electron. Letters.* 59 (2), e12701. doi: 10.1049/ell2.12701

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sensing.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhao, M., Wang, W., Ren, Q., Ni, H., Xiao, X., and Ma, L. (2023). Modified you-only-look-once model for joint source detection and azimuth estimation in a multi-interfering underwater acoustic environment. *J. Acoustical Soc. America.* 153 (4), 2393–2393. doi: 10.1121/10.0017828

Zhou, L., Ye, K., Qi, J., and Sun, H. (2022). DOA estimation based on pseudo-noise subspace for relocating enhanced nested array. *IEEE Signal Process. Letters.* 29, 1858–1862. doi: 10.1109/LSP.2022.3199149

Zywicki, D. J., and Rix, G. J. (1999). Frequency-wavenumber analysis of passive surface waves. *Symposium on the Application of Geophysics to Engineering and Environmental Problems 1999. Soc. Explor. Geophysicists*, 75–84. doi: 10.3997/2214-4609-pdb.202.1999_009

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 January 2024
DOI 10.3389/fmars.2023.1331635

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Xin Qing,
Harbin Engineering University, China
Jiahua Zhu,
National University of Defense Technology, China

*CORRESPONDENCE
Tingting Lv
tingtinglu@ouc.edu.cn

†These authors have contributed equally to this work

RECEIVED 01 November 2023
ACCEPTED 14 December 2023
PUBLISHED 11 January 2024

CITATION
Zhou M, Zhang H, Lv T, Huang W, Duan Y and Gao Y (2024) A simplified decision feedback Chebyshev function link neural network with intelligent initialization for underwater acoustic channel equalization.
Front. Mar. Sci. 10:1331635.
doi: 10.3389/fmars.2023.1331635

A simplified decision feedback Chebyshev function link neural network with intelligent initialization for underwater acoustic channel equalization

Manli Zhou[1†], Hao Zhang[1,2†], Tingting Lv[1*], Wei Huang[1], Yingying Duan[1] and Yong Gao[1]

[1]Department of Electronic Engineering, Ocean University of China, Qingdao, China, [2]Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada

Introduction: In shallow-water environments, the reliability of underwater communication links is often compromised by significant multipath effects. Some equalization techniques such as decision feedback equalizer, and deep neural network equalizer suffer from slow convergence and high computational complexity.

Methods: To address this challenge, this paper proposes a simplified decision feedback Chebyshev function link neural network equalizer (SDF-CFLNNE). The structure of the SDF-CFLNNE employs Chebyshev polynomial function expansion modules to directly and non-linearly transform the input signals into the output layer, without the inclusion of hidden layers. Additionally, it feeds the decision signal back to the input layer rather than the function expansion module, which significantly reduces computational complexity. Considering that, in the training phase of neural networks, the random initialization of weights and biases can substantially impact the training process and the ultimate performance of the network, this paper proposes a chaotic sparrow search algorithm combining the osprey optimization algorithm and Cauchy mutation (OCCSSA) to optimize the initial weights and thresholds of the proposed equalizer. The OCCSSA utilizes the Piecewise chaotic population initialization and combines the exploration strategy of the ospreywith the Cauchy mutation strategy to enhance both global and local search capabilities.

Rseults: Simulations were conducted using underwater multipath signals generated by the Bellhop Acoustic Toolbox. The results demonstrate that the performance of the SDFCFLNNE initialized by OCCSSA surpasses that of CFLNN-based and traditional nonlinear equalizers, with a notable improvement of 2-6 dB in terms of signal-to-noise ratio at a bit error rate (BER) of 10–4 and a reduced mean square error (MSE). Furthermore, the effectiveness of the proposed equalizer was validated using the lake experimental data, demonstrating lower BER and MSE with improved stability.

Discussion: This underscores thepromise of employing the SDFCFLNNE initialized by OCCSSA as a promising solution to enhance the robustness of underwater communication in challenging environments.

KEYWORDS

decision feedback equalizer, Chebyshev function link artificial neural network, sparrow search algorithm, osprey optimization algorithm, chaotic mapping, Cauchy mutation

1 Introduction

The shallow water acoustic environment is complex and changeable, which often exhibits intricate signal multipath effects and Doppler frequency shifts (Stojanovic and Preisig, 2009; Huang et al., 2018). The multipath propagation of underwater acoustic (UWA) signals originates from the effects of acoustic boundaries (such as reflection from the water surface and seabed), refraction caused by the non-uniformly distributed dissound speed in the water, as well as scattering from particles. The complex multipath results in significant signal time spreading, thereby causing severe intersymbol interference. In typical shallow water acoustic communications, intersymbol interference may span over hundreds of symbols. Consequently, at the receiver end, it is essential for the channel equalization to possess strong adaptive channel tracking capabilities (Song et al., 2006; Wang et al., 2021). This poses a significant challenge for reliable and efficient UWA communication (Zhang et al., 2018).

To combat intersymbol interference caused by time-varying multipath propagation, extensive research has been conducted on various channel equalization techniques. Single-carrier schemes and time-domain equalization techniques offer high spectral efficiency and robustness, albeit at the cost of high receiver complexity (Stojanovic and Preisig, 2009; Zhang et al., 2018). The proposed adaptive step-size least mean square performs well for many channel types, but for certain complex non-stationary UWA channels, the rapid tracking capability of recursive least square is essential (Freitag et al., 1997). To achieve reliable coherent communication over UWA channels, a receiver was designed which combines the recursive least square algorithm with a second-order digital phase-locked loop for carrier synchronization and performs fractionally spaced decision feedback equalization of the received signals. The parameters of this receiver are adaptively adjusted (Stojanovic et al., 1994). An adaptive nonlinearity (piecewise linear) was introduced into the channel equalization algorithm and its effectiveness was demonstrated through highly realistic experiments conducted on real-field data as well as accurate simulations of UWA channels (Kari et al., 2017). In recent years, in order to alleviate propagation errors, expedite convergence speed, and further enhance receiver performance, there has been growing research on adaptive turbo equalization (He et al., 2019; Xi et al., 2019; Qin et al., 2020). Considering the sparsity inherent in UWA channels, sparse matrices have been utilized to construct sparse equalizers, aiming to achieve faster convergence and lower error rates (Xi et al., 2020; Wang et al., 2021; Wang et al., 2021). Additionally, the equalization challenges in an impulsive interference single-carrier modulation system based on a parameterized model are addressed, and a two-step equalization algorithm is proposed (Ge et al., 2022). The robust equalization for single-carrier underwater acoustic communication in sparse impulsive interference environment was proposed (Wei et al., 2023). This algorithm is based on the framework of variational Bayesian inference and possesses the unique capability of simultaneously accounting for the sparsity inherent in the channel and impulse interference. At the same time, several waveform design (Zhu et al., 2023) and enhanced receiver schemes (Zhang et al., 2021; Liu et al., 2023) were proposed to further address inter-symbol interference and multipath propagation issues. However, the complex multipath effect of the UWA channels contributes to the slow convergence rate and extensive computational requirements of traditional equalization algorithms. As a result, there is substantial room for improvement in UWA communication systems.

In recent years, machine learning techniques have garnered attention across various fields. Particularly, deep learning (DL) technology holds tremendous potential for addressing non-parametric problems such as object detection and recognition (Tsai et al., 2013), speech recognition (Zhang and Wang, 2016), target tracking (Milan et al., 2017), wireless communication (Wang et al., 2017; Ma et al., 2018; van Heteren, 2022; Mishra et al., 2023). In order to reduce the computational costs of traditional equalizers, machine learning-based equalizers have been introduced to mitigate intersymbol interference. Channel equalization can be viewed as a classification problem, where the equalizer is designed as a decision device with the motivation to classify the transmitted signals as accurately as possible (Zhang and Yang, 2020). Gibson et al. introduced an adaptive equalizer employing a neural network architecture based on multilayer perceptrons (MLP) to counter intersymbol interference on linear channels with Gaussian white noise (Gibson et al., 1989). Chang et al. proposed a neural network-based decision feedback equalizer (DFE) that obviates the need for time-consuming complex-valued backpropagation training algorithms (Chang and Wang, 1995). Gao et al. demonstrated

that in underwater digital communication scenarios, their proposed blind equalizer achieves faster convergence speed and smaller mean square error (MSE) compared to original MLP-based equalizers that require training data (Gao et al., 2009). Zhang et al. proposed a DL-based time-varying UWA channel single-carrier communication receiver to adapt to the dynamic characteristics of UWA channels. The receiver operates in an alternating mode between online training and testing (Zhang et al., 2019b). Zhang et al. introduced a DL-based UWA communication orthogonal frequency-division multiplexing receiver. A stack of convolutional layers with skip connections effectively extracts meaningful features from the received signal and reconstructs the original transmitted symbols (Zhang et al., 2022). Radial basis function (RBF) neural networks have garnered the attention of many researchers due to their simple structure and high learning efficiency, and have been utilized for addressing channel equalization issues (Lee and Sankar, 2007; Guha and Patra, 2009; Ning et al., 2009).

However, with a higher channel order, a greater number of RBF centers are required, ultimately resulting in an excessive computational burden. To overcome these drawbacks of MLP and RBF, another novel single-layer neural network, known as the Functional Link Neural Network (FLNN), was proposed by Paul. Due to the non-linear processing of signals in the FLNN, it can generate arbitrarily complex decision regions (Patra et al., 1999). This network features a simple structure with only input and output layers, and the hidden layer is entirely replaced by non-linear mappings. These mappings are introduced through the expansion of input patterns using trigonometric polynomials and other basis functions like Gaussian polynomials, orthogonal polynomials, Legendre polynomials, and Chebyshev polynomials (Burse et al., 2010). The FLNN increases the dimensionality of the input signal space by a set of linearly independent non-linear functions, thus reducing computational load and allowing for straightforward hardware implementation (Patra et al., 2008; Zhang and Yang, 2020). Moreover, research indicates that non-linear equalizers based on FLNN outperform MLP, RBF, and PPN equalizers in terms of MSE, convergence rate, bit error rate (BER), and computational complexity (Patra et al., 1999). Lee et al. introduced a Chebyshev Neural Network for static function approximation, which is more computationally efficient than trigonometric polynomials when expanding the input space for extended static function approximation and non-linear dynamic system identification (Lee and Jeng, 1998). Patra et al. have employed Chebyshev Functional Link Neural Networks (CFLNN) for channel equalization of four quadrature amplitude modulation signals (Patra and Kot, 2002; Patra et al., 2005). Hussain combined traditional DFE with FLNN, proposing a Decision Feedback Functional Link Neural Network Equalizer (DFFLNN) (Hussain et al., 1997). Building upon this, they introduced a Chebyshev orthogonal polynomial cascaded FLNN for non-linear channel equalization (Zhao and Zhang, 2008) and an adaptive DFE based on the combination of the FIR and FLNN (Zhao et al., 2011). Moreover, Convolutional Neural Network (He et al., 2023), Recurrent Neural Networks (Kechriotis et al., 1994; Chagra et al., 2005; Xiao et al., 2008; Zhao et al., 2010; Li et al., 2021; Qiao et al., 2022), Fuzzy Neural Networks (Heng et al., 2006; Chang and Ho, 2009; Chang and Ho, 2011), Extreme Learning Machines (Yang et al., 2018; Liu et al., 2019), Wavelet Neural Networks (Xiao and Dong, 2015), Support Vector Machines (Zhang et al., 2019a), other neural network models and Deep Reinforcement Learning (He and Tao, 2023) have been employed for channel equalization.

Swarm intelligence optimization algorithms are a class of bio-inspired algorithms inspired by the behavioral patterns of certain social organisms in the natural world. The central idea is to conduct both global and local searches within a solution space to find optimal solutions. These algorithms provide a new approach to solving complex problems without centralized control or a global model. In recent years, new swarm intelligence optimization algorithms have continuously emerged. Scholars have drawn inspiration from the behavior of various animals such as ants, wolves, birds, moths, whales, sparrows, and more to propose a series of swarm intelligence optimization algorithms, including the Particle Swarm Optimization (PSO) algorithm (Kennedy and Eberhart, 1995), the Grey Wolf Optimization (GWO) algorithm (Mirjalili et al., 2014), the Whale Optimization Algorithm (WOA) (Mirjalili and Lewis, 2016), the Bald Eagle Search (BES) algorithm (Alsattar et al., 2020), the Sparrow Search Algorithm (SSA) (Xue and Shen, 2020), the Cooperation Search Algorithm (CSA) (Feng et al., 2021), artificial gorilla troops optimizer(GTO) (Abdollahzadeh et al., 2021), white shark optimizer(WSO)(Braik et al., 2022), dung beetle optimizer(DBO)(Xue and Shen, 2023) and Osprey Optimization Algorithm(OOA)(Dehghani and Trojovskỳ, 2023). The Sparrow Search Algorithm (SSA) was first introduced by Xue et al. in 2020 (Xue and Shen, 2020). In comparison to other algorithms, SSA offers several advantages, including fast convergence, strong optimization capabilities, and a wider range of application scenarios. As a result, SSA has garnered the attention of researchers from various fields. However, SSA does have limitations in terms of initial population quality, search capabilities, and population diversity. To address these issues, the Improved Sparrow Search Algorithm (ISSA) was proposed (Song et al., 2020). ISSA introduces non-linear decay in the position updates of producers, which facilitates the exploration and utilization of the search space. ISSA incorporates a mutation strategy to update the positions of scavengers with lower energy, combining chaotic search with local development by higher-energy scavengers. This enhances diversity and prevents falling into local optima. At the same time, the Tent mapping is used to initialize the population. Then, for the producers, an adaptive weight strategy is combined with the Levy flight mechanism, making the fusion search approach more comprehensive and flexible. Finally, in the scavenger stage, a variable spiral search strategy is employed to provide a more detailed search scope (Ouyang et al., 2021).

Traditional network equalizers suffer from problems such as large steady-state errors, slow convergence, susceptibility to local minima during the search process, and the curse of dimensionality. Moreover, in the training phase of neural networks, the random initialization of weights and biases can substantially impact the

training process and the ultimate performance of the network. In contrast, swarm intelligence optimization algorithms exhibit strong convergence and high precision advantages in the optimization process of practical problems. Therefore, they have become popular research topics in the field of equalizer optimization methods. A modified constant modulus algorithm digital channel equalizer learning algorithm based on PSO is proposed by Sahu (Sahu and Majumder, 2021). The particle swarm algorithm is employed as the training algorithm, resulting in a shorter convergence time and better performance compared to traditional LMS algorithms. This equalizer avoids introducing any phase ambiguity and does not get trapped in local optima. A novel training strategy using the Fuzzy Firefly Algorithm is proposed for channel equalization (Mohapatra et al., 2022). By employing an appropriate network topology and parameters, the suggested training system exhibits enhanced exploration and exploitation capabilities, as well as the ability to address local minima issues. An enhanced Grasshopper Optimization Algorithm (GOA) is proposed for nonlinear wireless communication channel equalization (Ingle and Jatoth, 2023). By combining Levy flights and greedy selection operators with the basic GOA, the diversity of the swarm is increased. Simulation results on four nonlinear channels demonstrate the exploration and exploitation capabilities of the improved Grasshopper Optimization Algorithm in terms of MSE and BER performance. An effective equalizer based on artificial neural networks is proposed by Shwetha (Shwetha et al., 2023). The Battle Royale Optimization method, as introduced, is utilized to train the weights of the neural network. The effectiveness of this approach is demonstrated through the evaluation of performance metrics such as MSE, mean squared residual error, and BER.

In shallow water acoustic propagation, there often exists severe multipath effects. Traditional equalization techniques may require hundreds of taps, greatly increasing system complexity. While the DFFLNNE (Hussain et al., 1997) outperforms FLNNE and traditional DFE, but it increases the dimensionality of the input layer, raising the complexity of the network structure. Simultaneously, during the training phase of the network, random initialization of weights and biases can affect the neural network's training process and final performance. Improper initialization can lead to problems such as gradient vanishing or exploding, causing training to be infeasible or overly slow. To enhance communication reliability without increasing system complexity, this paper proposes a simplified decision feedback Chebyshev functional link neural network equalizer (SDF-CFLNNE) initialized with swarm intelligence optimization algorithms. The papers contributions can be summarized as follows.

1. To address the issue of unreliability in underwater communication links caused by significant strong multipath effects in shallow-water environments, we propose a simplified decision feedback Chebyshev function link neural network equalizer.
2. To optimize the initial weights and thresholds of the proposed equalizer, We propose a Chaotic Sparrow Search Algorithm combining osprey optimization algorithm and Cauchy mutation. This approach mitigates the instability resulting from random weight initialization in the network equalizer.

The rest of this paper is organized as follows. In Section 2, a novel simplified decision feedback Chebyshev functional link neural network equalizer is proposed to address the unreliability of communication due to multipath effects. In Section 3, a chaotic sparrow search algorithm combining osprey optimization algorithm and Cauchy mutation is proposed for intelligent optimization of network weight and bias initialization. We validate the method through simulation and lake experimental data processing in Section 4. Finally, conclusions are given in Section 5.

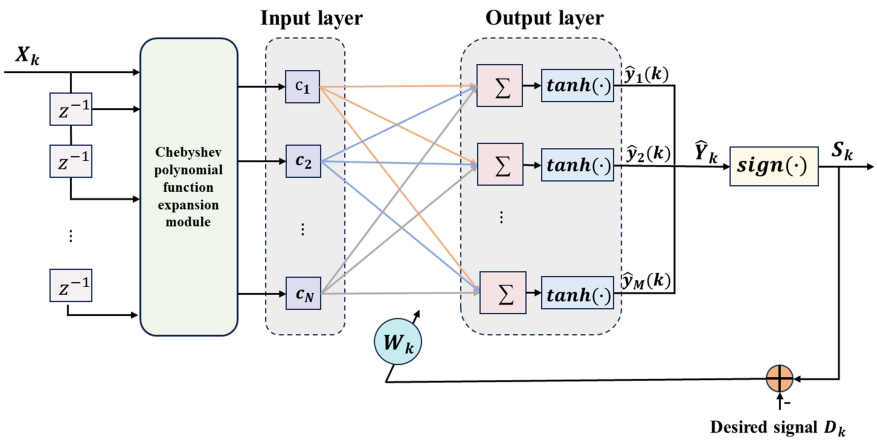


FIGURE 1
The structural diagram of the CFLNNE.

2 A novel simplified decision feedback Chebyshev functional link neural network equalizer

2.1 Decision feedback Chebyshev functional link neural network equalizer

To overcome intersymbol interference caused by multipath effects, a channel equalizer is embedded in the receiver to restore the transmitted signal. The FLNNE (Patra et al., 1999) has no hidden layers and is composed solely of a function extension module and a single-layer perceptron. This composition enables the generation of complex decision regions through the creation of nonlinear decision boundaries. In contrast to the linear weighting of input patterns generated by linear connections in MLP, the function expansion module enhances the dimensionality of input patterns by applying a set of linearly independent functions to elements or the entire pattern itself, thus enhancing its representation in high-dimensional space. Moreover, due to its single-layer structure, this FLNN structure exhibits lower computational complexity and faster convergence speed compared to other traditional neural networks. As widely recognized, utilizing the optimal approximation theory, Chebyshev orthogonal polynomials possess a robust capability for nonlinear approximation (Patra et al., 2005). The function expansion module in this context is composed of Chebyshev polynomials and their outer products, serving to simulate nonlinear channels, to construct the Chebyshev Functional Link Neural Network Equalizer (CFLNNE). The Figure 1 illustrates the structure of the CFLNNE.

Chebyshev polynomials are a set of orthogonal polynomials defined as solutions to the Chebyshev differential equation, denoted as $T_n(x)$. Chebyshev polynomials are computationally more tractable compared to trigonometric polynomials. The first several Chebyshev polynomials are given by $T_0(x) = 1$, $T_1(x) = x$, and $T_2(x) = 2x^2 - 1$. When the input signal is $\mathbf{X}_k = [x_1(k), x_2(k), \ldots, x_M(k)]^T$, the higher-order Chebyshev polynomials for $-1 < x < 1$ can be generated using the following recursion formula Equation 1:

$$\begin{aligned} c_1(\mathbf{X}_k) &= T_0(\mathbf{X}_k) = 1, \\ c_2(\mathbf{X}_k) &= T_1(\mathbf{X}_k) = \mathbf{X}_k, \\ &\vdots \\ c_{n+2}(\mathbf{X}_k) &= T_{n+1}(\mathbf{X}_k) = 2\mathbf{X}_k T_n(\mathbf{X}_k) - T_{n-1}(\mathbf{X}_k) \end{aligned} \tag{1}$$

In CFLNNE, the input signal denoted as $\mathbf{X}_k$, is expanded into N linearly independent functions using Chebyshev polynomials, and can be represented as $\mathbf{C}_k = [c_1(\mathbf{X}_k) c_2(\mathbf{X}_k) \ldots c_N(\mathbf{X}_k)]^T$.

Through forward propagation, the j-th neuron of the output layer can be represented as Equations 2 and 3:

$$u_j(k) = \sum_{i=1}^{N} w_{ji}(k) c_i(X_k) + b_j(k) \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots M\,. \tag{2}$$

$$\hat{y}_j(k) = f(u_j(k)) = f\left(\sum_{i=1}^{N} w_{ji}(k) c_i(X_k) + b_j(k)\right) \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots M\,. \tag{3}$$

where w represents the weight coefficients from the input layer to the output layer, and b represents the bias of the output layer. The nonlinear activation function here is $f(\cdot) = tanh(\cdot)$, and its derivative is denoted as $f'(\cdot)$.

The output signal after decision device can be represented as Equation 4:

$$s(k) = sign(\hat{y}(k)) = \begin{cases} -1 & \text{if } \hat{y}(k) < 0 \\ 1 & \text{if } \hat{y}(k) \geq 0 \end{cases} \tag{4}$$

Taking advantage of the traditional decision feedback equalizer's ability to mitigate inter-symbol interference introduced by the preceding information symbol, Hussain et al. integrated the DFE with a FLNN,

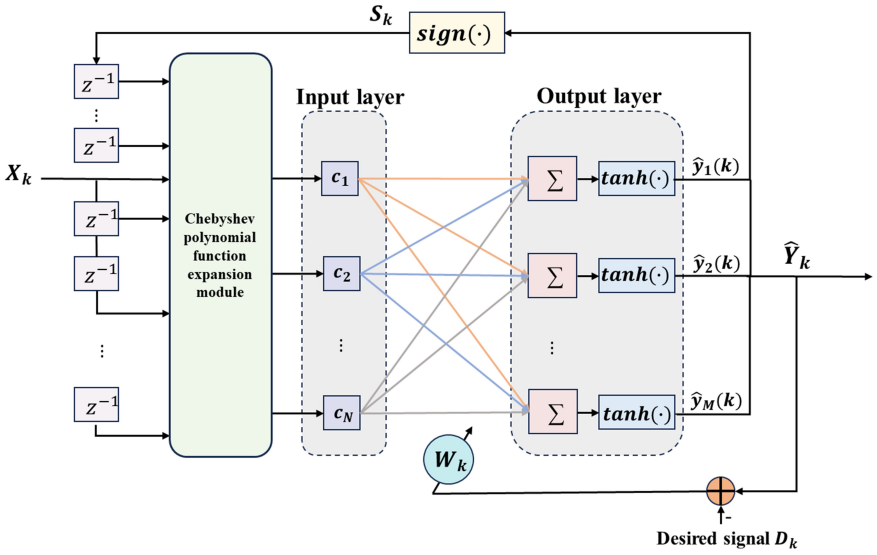


FIGURE 2
The structural diagram of the DF-CFLNNE.

creating the decision feedback functional link neural network (Hussain et al., 1997). To further enhance the nonlinear approximation capabilities of the function link module, the DFE is combined with CFLNNE to form a decision feedback Chebyshev functional link neural network (DF-CFLNNE), as illustrated in Figure 2.

The input signal $\mathbf{X}_k = [x_1(k), x_2(k), \ldots, x_M(k)]^T$ and feedback signal $\mathbf{S}_k = [s_1(k), \ldots, s_{N_2}(k)]^T$ from the decision device are jointly used as the input signal $\mathbf{Z}_k = [\mathbf{X}_k, \mathbf{S}_k]$ for the DF-CFLNNE, where N_2 represents the order of the feedback delay path. The key distinction from CFLNNE is that CFLNNE takes only $\mathbf{X}_k$ as its input signal, without the feedback signals from the decision device. Subsequently, the input signal $\mathbf{Z}_k$ of DF-CFLNNE is expanded into N linearly independent functions using Chebyshev polynomials, denoted as $\mathbf{C}_k = [c_1(\mathbf{Z}_k), c_2(\mathbf{Z}_k), \ldots, c_N(\mathbf{Z}_k)]^T$, where $\mathbf{C}_k$ serves as the input to the network's input layer.

Through forward propagation, the j-th neuron of the output layer can be represented as Equations 5 and 6:

$$u_j(k) = \sum_{i=1}^{N} w_{ji}(k) c_i(Z_k) + b_j(k) \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots M. \tag{5}$$

$$\hat{y}_j(k) = f(u_j(k)) = f\left(\sum_{i=1}^{N} w_{ji}(k) c_i(Z_k) + b_j(k)\right) \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots M. \tag{6}$$

For convenience, these values of functions can be represented in matrix form as Equation 7:

$$\hat{\mathbf{Y}}_k = f(\mathbf{W}_k \mathbf{C}_k + \mathbf{B}_k) \tag{7}$$

where $\mathbf{W}_k$ is an $M \times N$ dimensional matrix, i.e., $\mathbf{W}_k = [w_{j1}, w_{j2}, \ldots, w_{jN}]$. $\mathbf{B}_k$ is an $M \times 1$ dimensional matrix, i.e., $\mathbf{B}_k = [b_1, b_2, \ldots, b_M]$. The output of the entire network can be represented in matrix form as $\hat{\mathbf{Y}}_k = [\hat{y}_1, \hat{y}_2, \ldots \hat{y}_M]^T$.

We use the MSE as the loss function, which can be represented as Equation 8:

$$\mathbf{J}_k = \sum (\mathbf{D}_k - \hat{\mathbf{Y}}_k)^2 \tag{8}$$

where $\mathbf{D}_k$ represents the desired output sequence at time instant k.

The backpropagation algorithm is employed here to train the DF-CFLNN. The training process is expressed as follows Equations 9 and 10:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{W}_k} &= \frac{\partial \mathbf{J}_k}{\partial \hat{\mathbf{Y}}_k} \cdot \frac{\partial \hat{\mathbf{Y}}_k}{\partial \mathbf{U}_k} \cdot \frac{\partial \mathbf{U}_k}{\partial \mathbf{W}_k} \\ &= 2(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot f'(\mathbf{U}_k) \cdot \mathbf{C}_k \\ &= 2(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot (1 - \hat{\mathbf{Y}}_k^2) \cdot \mathbf{C}_k \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{B}_k} &= \frac{\partial \mathbf{J}_k}{\partial \hat{\mathbf{Y}}_k} \cdot \frac{\partial \hat{\mathbf{Y}}_k}{\partial \mathbf{U}_k} \cdot \frac{\partial \mathbf{U}_k}{\partial \mathbf{B}_k} \\ &= 2(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot f'(\mathbf{U}_k) \\ &= 2(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot (1 - \hat{\mathbf{Y}}_k^2) \end{aligned} \tag{10}$$

According to the gradient descent algorithm, there will be Equations 11 and 12:

$$\mathbf{W}_k = \mathbf{W}_k - \mu \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{W}_k} \tag{11}$$

$$\mathbf{B}_k = \mathbf{B}_k - \mu \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{B}_k} \tag{12}$$

where the parameter μ denotes the learning factor.

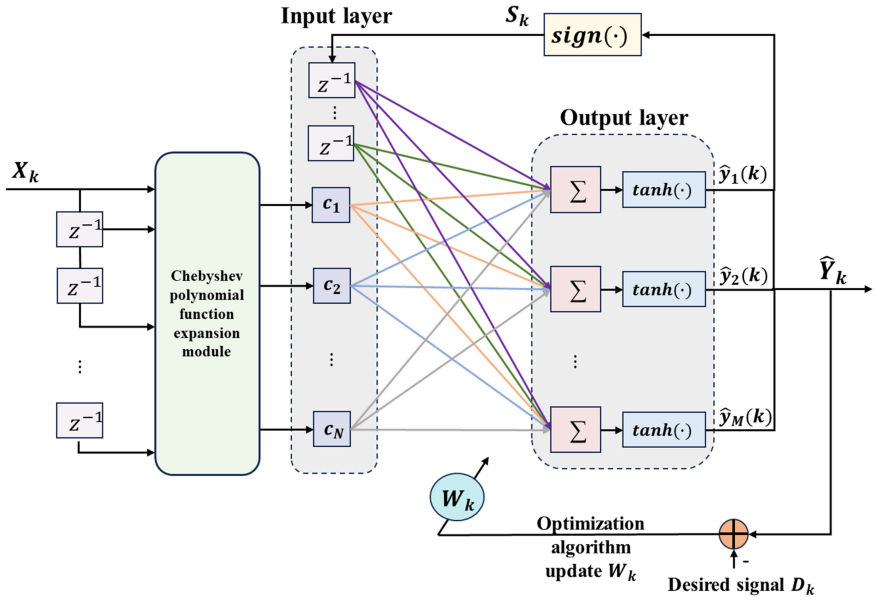


FIGURE 3
The structural diagram of the SDF-CFLNNE.

2.2 Simplified decision feedback Chebyshev functional link neural network equalizer

DF-CFLNNE increases the system's performance at the cost of increased complexity. In order to reduce hardware costs without compromising system performance, a simplified DF-CFLNNE (SDF-CFLNNE) structure is proposed as illustrated in Figure 3.

In the SDF-CFLNNE structure, the post-decision output signal is directly fed back to the input layer of the neural network, rather than being used as an input signal to the network, and it no longer goes through the function expansion module. Namely, the input layer signal is composed of the Chebyshev polynomial function expansion of the received signal, denoted as $\mathbf{C}_k = [c_1(\mathbf{X}_k), c_2(\mathbf{X}_k), \ldots, c_N(\mathbf{X}_k)]^T$, and the feedback signal fafter decision, denoted as $\mathbf{S}_k = [s_1(k), \ldots, s_{N_2}(k)]^T$, which can be represented as $\mathbf{G}_k = [c_1(\mathbf{X}_k), c_2(\mathbf{X}_k), \ldots, c_N(\mathbf{X}_k), s_1(k), \ldots, s_{N_2}(k)]^T = [g_1(k), \ldots, g_P(k)]^T, P = N + N_2$. It can be observed from Figures 2, 3 that the number of input signals in SDF-CFLNNE is fewer compared to DF-CFLNNE. Consequently, the number of signals after the function expansion module for SDF-CFLNNE is significantly reduced compared to DF-CFLNNE. This streamlined system structure enhances computational efficiency.

Through forward propagation, the j-th neuron of the output layer can be represented as Equations 13 and 14:

$$u_j(k) = \sum_{i=1}^{N} w_{ji}(k) g_i(k) + b_j(k) \quad i = 1, 2, \ldots, P; j = 1, 2, \ldots M. \tag{13}$$

$$\hat{y}_j(k) = f(u_j(k)) = f\left(\sum_{i=1}^{N} w_{ji}(k) g_i(k) + b_j(k)\right) \quad i = 1, 2, \ldots, P; j = 1, 2, \ldots M. \tag{14}$$

where w represents the weight coefficients from the input layer to the output layer, and b represents the bias of the output layer. The nonlinear activation function here is $f(\cdot) = tanh(\cdot)$.

For convenience, these values of functions can be represented in matrix form as Equation 15:

$$\hat{\mathbf{Y}}_k = f(\mathbf{W}_k \mathbf{G}_k + \mathbf{B}_k) \tag{15}$$

where $\mathbf{W}_k$ is an $M \times P$ dimensional matrix, i.e., $\mathbf{W}_k = [w_{j1}, w_{j2}, \ldots, w_{jP}]$. $\mathbf{B}_k$ is an $M \times 1$ dimensional matrix, i.e., $\mathbf{B}_k = [b_1, b_2, \ldots, b_M]$. The output of the entire network can be represented in matrix form as $\hat{\mathbf{Y}}_k = [\hat{y}_1, \hat{y}_2, \hat{y}_M]^T$. We still adopt the MSE, as given in Equation 8, as the loss function.

The BP algorithm is employed here to train the SDF-CFLNN. According to Equations 9 and 10, the training process is expressed as follows Equations 16 and 17:

$$\begin{aligned} \mathbf{W}_{k+1} &= \mathbf{W}_k - \mu \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{W}_k} \\ &= \mathbf{W}_k - \mu \frac{\partial \mathbf{J}_k}{\partial \hat{\mathbf{Y}}_k} \cdot \frac{\partial \hat{\mathbf{Y}}_k}{\partial \mathbf{U}_k} \cdot \frac{\partial \mathbf{U}_k}{\partial \mathbf{W}_k} \\ &= \mathbf{W}_k - 2\mu(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot f'(\mathbf{U}_k) \cdot \mathbf{G}_k \\ &= \mathbf{W}_k - 2\mu(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot (1 - \hat{\mathbf{Y}}_k^2) \cdot \mathbf{G}_k \end{aligned} \tag{16}$$

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{B}_k - \mu \frac{\mathrm{d}\mathbf{J}_k}{\mathrm{d}\mathbf{B}_k} \\ &= \mathbf{B}_k - \mu \frac{\partial \mathbf{J}_k}{\partial \hat{\mathbf{Y}}_k} \cdot \frac{\partial \hat{\mathbf{Y}}_k}{\partial \mathbf{U}_k} \cdot \frac{\partial \mathbf{U}_k}{\partial \mathbf{B}_k} \\ &= \mathbf{B}_k - 2\mu(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot f'(\mathbf{U}_k) \\ &= \mathbf{B}_k - 2\mu(\mathbf{D}_k - \hat{\mathbf{Y}}_k) \cdot (1 - \hat{\mathbf{Y}}_k^2) \end{aligned} \tag{17}$$

where the parameter μ denotes the learning factor.

SDF-CFLNNE directly inputs the decision feedback signal into the network's input layer instead of the function expansion module, reducing the number of neurons in the input layer. In this way, we can obtain the improvement of system performance from the feedback signal without increasing the number of neurons in the input layer of the network. It reduces system complexity, enhances computational efficiency, and accelerates convergence speed.

3 A novel chaotic sparrow search algorithm combining osprey optimization algorithm and Cauchy mutation

In this section, a OCCSSA is proposed to solve the impact of the random initialization of network weights on the convergence of the training process and network performance. This algorithm utilizes chaotic mapping for random population initialization and combines the osprey optimization algorithm with the Cauchy mutation criterion to update the positions in the SSA. The use of the osprey optimization algorithm.

(OOA) in the initial phase provides a global exploration strategy, where a random attack on one of the food sources helps mitigate the SSA's over-reliance on the previous generation's sparrow positions for updates. In the second phase, Cauchy mutation is applied to perturb individuals in the sparrow positions, thereby expanding the search scope of the SSA and enhancing its ability to escape local optima.

3.1 Preliminaries

We briefly introduce the basic framework of SSA, chaotic mapping, OOA, Cauchy mutation, and some basic concepts.

3.1.1 Sparrow search algorithm

Sparrows are typically gregarious birds. Captive house sparrows come in two different types, referred to as "producers" and "scroungers" (Barnard and Sibly, 1981). The producers actively search for sources of food, while the scroungers obtain food through the producers. Additionally, these birds are typically capable of flexibly employing behavioral strategies and switching between producing and scrounging (Liker and Barta, 2002). It can be said that, in order to find food, the sparrows often utilize both producer and scrounger strategies simultaneously (Barnard and Sibly, 1981; Xue and Shen, 2020).

Assuming there are N sparrows in a d-dimensional search space, the position of each sparrow can be represented by the following matrix Equation 18:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & \cdots & x_{N,d} \end{bmatrix} \tag{18}$$

The positions of sparrows in the search space are randomly initialized using Equation 19.

$$\mathrm{x_{i,j}} = lb_j + r_{i,j} \cdot (ub_j - lb_j) \tag{19}$$

where $x_{i,j}$ represents the position of the i-th sparrow in the j-th dimension. $r_{i,j}$ is a random number in the interval [0,1]. lb_j and ub_j are the lower and upper bounds of the j-th dimension of the problem variables, respectively.

The fitness values of all sparrows can be represented by the following vector Equation 20:

$$\mathbf{F_X} = \begin{bmatrix} f([x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,d}]) \\ f([x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,d}]) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f([x_{N,1} & x_{N,2} & \cdots & \cdots & x_{N,d}]) \end{bmatrix} \tag{20}$$

The first stage is the exploration phase. The producers with better fitness values are given preference when it comes to acquiring food during the search process. Additionally, because producers take on the responsibility of food searching and guiding the entire population's movement, they have a wider search area compared to the scroungers. Moreover, when a sparrow detects a predator, it initiates an alarm by chirping. If the alarm value surpasses a predefined safety threshold, producers must lead all the scroungers to a safe zone. Throughout each iteration, the positions of producers are updated as follows Equation 21:

$$\mathrm{X_{i,j}^{t+1}} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\frac{-i}{\alpha \cdot \mathrm{iter}_{\max}}\right) & \text{if } R_2 < ST \\ X_{i,j}^{t} + Q \cdot L & \text{if } R_2 \geq ST \end{cases} \tag{21}$$

where t represents the current iteration number, $j = 1, 2, \ldots, d$. $X_{i,j}^{t+1}$ represents the value of the j-th dimension for the i-th sparrow at the t-th iteration. $\mathrm{iter}_{\max}$ signifies the maximum number of iterations. $\alpha \in (0,1]$ is a random number. Q is a random number following a normal distribution. L is a matrix of size $1 \times d$ in which every element is equal to 1. R_2 (where $R_2 \in [0,1]$) represents the alarm value. ST (where $ST \in [0.5,1]$) stands for the safety threshold.

If $R_2 < ST$, it signifies an absence of predators in the vicinity, prompting the producers to transition into an expansive search mode. However, when $R_2 \geq ST$, it signifies that certain sparrows have detected predators, necessitating a swift relocation of all sparrows to alternative safe areas.

The second phase is the development phase. Scroungers follow the producers who can offer the best food to search for nourishment. Meanwhile, some scroungers may continuously monitor the producers, and if they notice a producer has found good food, they immediately leave their current location to compete for the food. If they succeed, they can acquire the producer's food immediately. The position update formula for the scroungers is as follows Equation 22:

$$\mathbf{X_{i,j}^{t+1}} = \begin{cases} Q \cdot \exp\left(\frac{X_{\mathrm{worst}}^{t} - X_{i,j}^{t}}{i^2}\right) & \text{if } i > N/2 \\ X_P^{t+1} + \left|X_{i,j}^{t} - X_P^{t+1}\right| \cdot A^{+} \cdot L & \text{otherwise} \end{cases} \tag{22}$$

where X_P represents the optimal position occupied by the producer. X_{worst} represents the current global worst location. A is a $1 \times d$ matrix where each element is randomly set as 1 or −1, and $A^{+} = A^{T}(AA^{T})^{-1}$. If $i > N/2$, it indicates that the i-th scrounger, with the worst fitness value, is highly likely to be in a starved state.

When they sense danger, sparrows located at the edge of the flock quickly move towards a safe area, while those in the middle of the flock move randomly to get closer to others. We assume that the sparrows aware of the danger constitute between 10% and 20% of the total population. The initial positions of these sparrows are randomly generated within the entire population and can be expressed using the following formula Equation 23:

$$\mathrm{X_{i,j}^{t+1}} = \begin{cases} X_{\mathrm{best}}^{t} + \beta \cdot \left|X_{i,j}^{t} - X_{\mathrm{best}}^{t}\right| & \text{if } f_i > f_g \\ X_{i,j}^{t} + K \cdot \left(\frac{\left|X_{i,j}^{t} - X_{worst}^{t}\right|}{(f_i - f_w) + \epsilon}\right) & \text{if } f_i = f_g \end{cases} \tag{23}$$

where X_{best} is the current global optimal location. β, as a step size control parameter, is a random number following a normal distribution with a mean of 0 and a variance of 1. $K \in [-1,1]$ is a random number and denotes the direction in which the sparrow moves and is also the step size control coefficient. ϵ is a small constant to avoid division by zero errors. Here, f_i represents the fitness value of the current sparrow, and f_g and f_w are the current global best and worst fitness values, respectively.

If $f_i > f_g$, this signifies that the sparrow is positioned at the group's periphery. X_{best} denotes the location of the population center and is considered safe. If $f_i = f_g$, it implies that sparrows in the middle of the group have sensed danger and must approach the others.

3.1.2 Chaotic mapping

A chaotic matrix is a typical source of "ordered chaos," exhibiting unique characteristics of randomness and state transitivity. Under certain "rules," chaotic sequences traverse all different states within a defined range. Chaotic sequences generally possess several key features, including nonlinearity, sensitivity to initial conditions, transitivity, randomness, strange attractors (chaotic attractors), global stability and local instability, and long-term unpredictability.

In the context of intelligent optimization algorithms, random initialization of the population is often achieved using a uniform distribution. Compared to standard random search based on conventional probability distributions, the use of chaotic mappings in intelligent optimization algorithms can help populations escape local minima and enable faster iterative searches.

3.1.3 Osprey optimization algorithm

The OOA was proposed by Mohammad Dehghani and Pavel Trojovský in 2023(Dehghani and Trojovskỳ, 2023), simulating the hunting behavior of ospreys.

The first phase is the exploration phase, involving the locating and capturing of fish. Ospreys, with their powerful vision, are formidable predators capable of spotting fish beneath the water's surface. Once they've pinpointed a fish's location, they dive underwater to attack and capture it. The initial stage of population update in OOA draws inspiration from this natural osprey behavior. Modeling how ospreys hunt fish results in substantial alterations to the ospreys' positions within the search space. This, in turn, enhances OOA's ability to explore and locate optimal regions while avoiding local optima.

Let's assume there are N ospreys in a d-dimensional search space. For each osprey, the positions of other ospreys in the search space that have a better objective function value are considered underwater fish. The set of fish for each osprey is specified using Equation 24.

$$FP_i = \{X_k \mid k \in \{1, 2, \ldots, N\} \wedge F_k < F_i\} \cup \{X_{\text{best}}\} \tag{24}$$

where FP_i is the set of fish positions for the i-th osprey and X_{best} is the best candidate solution.

The osprey employs a random process to detect the location of one of these fishes, and it initiates an attack. Through modeling the osprey's movement as it approaches the fish, a new position is computed for the osprey by Equation 25 and Equation 26.

$$x_{i,j}^{P1} = x_{i,j} + r_{i,j} \cdot (SF_{i,j} - I_{i,j} \cdot x_{i,j}) \tag{25}$$

$$x_{i,j}^{P1} = \begin{cases} x_{i,j}^{P_1}, lb_j \le x_{i,j}^{P_1} \le ub_j \\ lb_j, x_{i,j}^{P_1} < lb_j \\ ub_j, x_{i,j}^{P_1} > ub_j \end{cases} \tag{26}$$

This new position, if it results in a better objective function value, replaces the osprey's previous position by Equation 27.

$$X_i = \begin{cases} X_i^{P1}, F_i^{P1} < F_i \\ X_i, \text{ else} \end{cases} \tag{27}$$

where $x_{i,j}^{P1}$ is the new position of the i-th osprey in the j-th dimension in the first phase, F_i^{P1} is its fitness value, and $SF_{i,j}$ is the fish chosen by the i-th osprey in the j-th dimension. $r_{i,j}$ is a random number within the range [1,2], and $I_{i,j}$ is a random number chosen from the set {1,2}.

The second phase is known as the development stage. After successfully capturing a fish, the osprey relocates it to a secure and suitable spot for consumption. This modeling, involving the relocation of the fish, introduces minor adjustments to the osprey's positions within the search space. Consequently, it enhances OOA's capability for exploiting the local search and converging towards improved solutions around the identified ones.

In the OOA design, the emulation of osprey behavior involves initially determining a new random position for each individual in the population, akin to a "fish-eating spot." This calculation is based on Equation 28. Subsequently, if this new position results in an improved objective function value, it is employed to replace the previous position of the respective osprey according to Equation 29.

$$x_{i,j}^{P2} = x_{i,j} + \frac{lb_j + r \cdot (ub_j - lb_j)}{t} \tag{28}$$

$$x_{i,j}^{P2} = \begin{cases} x_{i,j}^{P_2}, lb_j \le x_{i,j}^{P_2} \le ub_j \\ lb_j, x_{i,j}^{P_1} < lb_j \\ ub_j, x_{i,j}^{P_1} > ub_j \end{cases} \tag{29}$$

This new position, if it results in a better objective function value, replaces the osprey's previous position by Equation 30.

$$X_i = \begin{cases} X_i^{P2}, F_i^{P2} < F_i \\ X_i, \text{ else} \end{cases} \tag{30}$$

where $x_{i,j}^{P2}$ is the new position of the i-th osprey in the j-th dimension in the second phase, F_i^{P2} is its fitness value, and $SF_{i,j}$ is the fish chosen by the i-th osprey in the j-th dimension. r is a random number within the range of [1,2], t represents the current iteration count, and T is the maximum number of iterations.

3.1.4 Cauchy mutation

The Cauchy mutation is derived from the Cauchy distribution. The probability density function of the one-dimensional Cauchy distribution is given by Equation 31:

$$f(x) = \frac{1}{\pi} \cdot \frac{a}{a + x^2} \tag{31}$$

here, when $a = 1$, it is the standard Cauchy distribution.

The Cauchy distribution is similar to the standard normal distribution. It is a continuous probability distribution that has smaller values near the origin, is more elongated towards the ends, and approaches zero at a slower rate. Therefore, compared to the normal distribution, it can introduce larger disturbances. By utilizing Cauchy mutation for perturbing individuals in the sparrow position updates, the SSA's search scope is expanded, leading to an improved ability to escape local optima.

3.2 Chaotic sparrow search algorithm combining osprey optimization algorithm and Cauchy mutation

Traditional SSA employs a random initialization method for the population, which can lead to premature convergence and slower convergence speed. To address this, this paper adopts a chaotic population initialization approach. This ensures randomness in the population while enhancing the algorithm's convergence performance and diversifying the population. This helps prevent algorithm stagnation caused by a homogenous population.

The positions of sparrows in the search space are initialized using Piecewise chaotic mapping, as shown in Equation 32.

$$\mathbf{x_{i,j}} = lb_j + chaos_{i,j} \cdot (ub_j - lb_j) \tag{32}$$

Where $chaos_{i,j}$ represents the chaotic mapping.

The first phase is the exploration phase. For producers, the Equation 25 of OOA's global exploration strategy in the first phase replaces the original producer position update Equation 21 of the SSA. OOA aims to address the SSA's overreliance on the update method based on the positions of the previous generation of sparrows.

The update method for the positions of producers in the sparrow algorithm is determined based on the simulation of the osprey's movement toward fish. For each sparrow, the locations of other sparrows in the search space with superior fitness values are considered as food. Equation 33 is utilized to determine the set of superior food chosen by each sparrow.

$$FP_i = \{X_k \mid k \in \{1, 2, \ldots, N\} \wedge F_k < F_i\} \cup \{X_{\text{best}}\} \tag{33}$$

where FP_i represents the food collection for the i-th sparrow, and X_{best} is the position of the best sparrow.

The sparrows randomly detect the position of one of the foods and go hunting. During each iteration, the positions of the producers are updated according to Equation 34. If the updated position is better, the sparrow's previous position is replaced.

$$\mathbf{X_{i,j}^{t+1}} = \begin{cases} X_{i,j}^t + r_{i,j} \cdot (SF_{i,j} - I_{i,j} \cdot X_{i,j}^t) & \text{if } R_2 < ST \\ X_{i,j}^t + Q \cdot L & \text{if } R_2 \geq ST \end{cases} \tag{34}$$

where $SF_{i,j}$ represents the food chosen by the i-th sparrow in the j-th dimension.

The second phase is the development stage. Scroungers often focus their search around the best discoverers. Food competition can also occur during this period, where a scrounger tries to become the producer. To prevent the algorithm from getting trapped in local optima, a Cauchy mutation strategy is introduced into the equation for updating the scroungers. The updated scrounger position equation, replacing the original SSA's scrounger position update equation, is as follows:

$$X_{i,j}^{t+1} = X_{best}(t) + cauchy(0,1) \oplus X_{best}(t) \tag{35}$$

where X_{best} represents the current global best position. *cauchy* (0,1) is the standard Cauchy distribution function, and $\oplus$ denotes multiplication.

The sparrows that sense danger still undergo updates according to Equation 23.

The pseudocode for OCCSSA, which we have proposed, is presented in **Algorithm 1**.

TABLE 1 Bellhop simulation parameters setup.

Parameter	CH1	CH2	CH3
Modulation type	QPSK	QPSK	QPSK
Sound source frequency	10 *kHz*	10 *kHz*	10 *kHz*
Sound pressure level	195 *dB*	195 *dB*	195 *dB*
Angle of sound wave emitted by transmitter point	−20° ~ + 20°	−20° ~ + 20°	−20° ~ + 20°
Sea water depth	200 *m*	200 *m*	200 *m*
Transmitter depth	80 *m*	80 *m*	150 *m*
Receiver depth	40 *m*	40 *m*	40 *m*
Distance from receiving point to transmitter point	3 *km*	3 *km*	3 *km*
P-wave speed of sound in bottom	1511.96 *m/s*	1511.96 *m/s*	1511.96 *m/s*
Density at the cutoff depth of the seabed medium	1.421 g/cm^3	2.034 g/cm^3	1.421 g/cm^3
Attenuation coefficient of the seabed medium	0.078 *dB/ wavelength*	0.479 *dB/ wavelength*	0.078 *dB/ wavelength*

4 Simulation and lake experiments results

4.1 Simulation validation and comparison

4.1.1 Comparison of the neural network equalizers

We used Bellhop Acoustic Toolbox to generate a time-varying UWA channel model to evaluate equalizers (Zhou et al., 2022). The parameters for the time-varying Bellhop channel simulator are listed in Table 1. We conducted simulations in three different underwater acoustic channel environments, with the different transmitter depth and seabed medium. The sound speed profile of 200 m is shown in Figure 4. The acoustic transmission loss obtained using the acoustic toolbox is shown in the Figures 5A, C, E. The maximum transmission loss is approximately 70 dB. The channel impulse response plot from the sound source to the receiving point is shown in the Figures 5B, D, F.

Input: *Max_iter*: maximum iteration; *PD*: the number of producers; *SD*: the number of sparrows who perceive the danger; R_2: the alarm value; *n*: the population size.
Output: X_{best}: the current global best position.
Initialize: Using Equation 32 to initialize the population
while *iter*< *Max_iter* **do**
Sort the fitness values to find the current best and worst individuals.
$R_2 = rand(1)$
for *i* =1 : *PD* **do**
Find a better food location using Equation 33.
Producers randomly select a food and update their positions using Equation 34.
end for
Find the optimal population and record as X_{best}.
for *i* = (*PD* + 1) : *n* **do**
Scroungers use the Cauchy mutation strategy to update their positions using Equation 35.

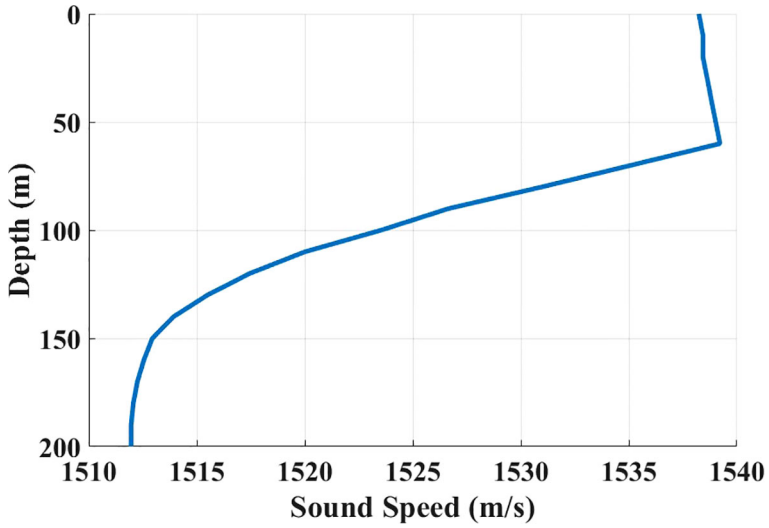


FIGURE 4
The sound speed profile of 200 m.

```
end for
for i = 1 : SD do
Sparrows partially aware of danger update their positions using Equation 23.
end for
Obtain the global best value X_best.
iter = iter + 1
end while
return X_best
```

Algorithm 1. The pseudocode of OCCSSA.

We will compare equalizers based on MLP (Gibson et al., 1989), CFLNN (Patra et al., 2005), DFCFLNN (Zhao and Zhang, 2008)and traditional DFE with phase-locked loop (PLL) (Stojanovic et al., 1994).

CFLNN has 10 feedforward taps, and its input signal is represented as $\mathbf{X}_k = [r_{k,i}, r_{k,q}, r_{k+1,i}, r_{k+1,q}, \ldots, r_{k+9,i}, r_{k+9,q}]^T$. Here, $r_{k,i}$represents the I-path of the k-th signal, and $r_{k,q}$ represents the Qpath of the k-th signal. The input signal is transformed into 121 dimensions through a six-order Chebyshev transformation. Therefore, the input layer has 121 nodes, and the output layer has 2 nodes. In DF-CFLNN, the tap coefficient of feedback is 2, and the input signal consists of $\mathbf{X}_k = [r_{k,i}, r_{k,q}, r_{k+1}, i, r_{k+1,q}, \ldots, r_{k+9}, i, r_{k+9,q}]^T$ and the feedback signal $\mathbf{S}_k = [s_{k,i}, s_{k,q}, s_{k-1,i}, s_{k-1,q}]$ from the decision device. The input signal is transformed into 97 dimensions through a four order Chebyshev transformation. Therefore, the input layer has 97 nodes, and the output layer has 2 nodes. In SDF-CFLNN, the tap coefficient of feedback is 2, and the input signal $\mathbf{X}_k = [r_{k,i}, r_{k,q}, r_{k+1,i}, r_{k+1,q}, \ldots, r_{k+9,i}, r_{k+9,q}]^T$ is transformed into 41 dimensions through a two-order Chebyshev transformation. The input layer has 45 nodes, including 41 nodes for input signals and 4 nodes for feedback signals. The parameters for each equalizer are as shown in Table 2.

To enhance the reliability, we conducted 30 independent trials on each equalizer. In our setup, we use 10,000 QPSK signals as input data, with 80% serving as training data and 20% as testing data. In each trial, the maximum iteration count was set to 1000. Since this paper focuses on the learning and equalization capabilities of neural networks, in our simulations, we assumed perfect time sequence recovery.

The simulated BER graph is shown in the Figures 6A, C, E. It can be observed that our proposed SDFCFLNNE exhibits the best BER performance in two different underwater acoustic environments, followed by DF-CFLNNE, traditional DFE-PLL, and CFLNNE in descending order. In complex underwater environments, the MLP equalizer performs poorly and is the least effective. The MSE iteration curve at SNR=10 dB is shown in the Figures 6B, D, F. SDF-CFLNNE converges the fastest, with minimal initial oscillations, and exhibits smooth and stable convergence. It also has the smallest MSE value when reaching a steady state. In CH1, SDF-CFLNNE reaches convergence in about 20 iterations, while DFCFLNNE and CFLNNE reach convergence around 300 iterations with minor oscillations. MLP achieves basic convergence in approximately 70 iterations but experiences significant oscillations. In CH2, SDFCFLNNE reaches convergence in about 40 iterations, while DF-CFLNNE, CFLNNE, and MLP all converge around 300 iterations with minor oscillations.

4.1.2 Comparison of the swarm intelligence optimization algorithms

4.1.2.1 Benchmark test functions

Benchmark test functions are typically utilized to evaluate the performance of optimization algorithms. We utilized the CEC2005 benchmark test functions as provided in Table 3 (Suganthan et al., 2005) to assess the applicability and effectiveness of the proposed OCSSA algorithm.

4.1.2.2 Comparison of chaotic mapping methods

In order to select the more effective chaotic mapping, we initialized the population of OCCSSA using ten different chaotic

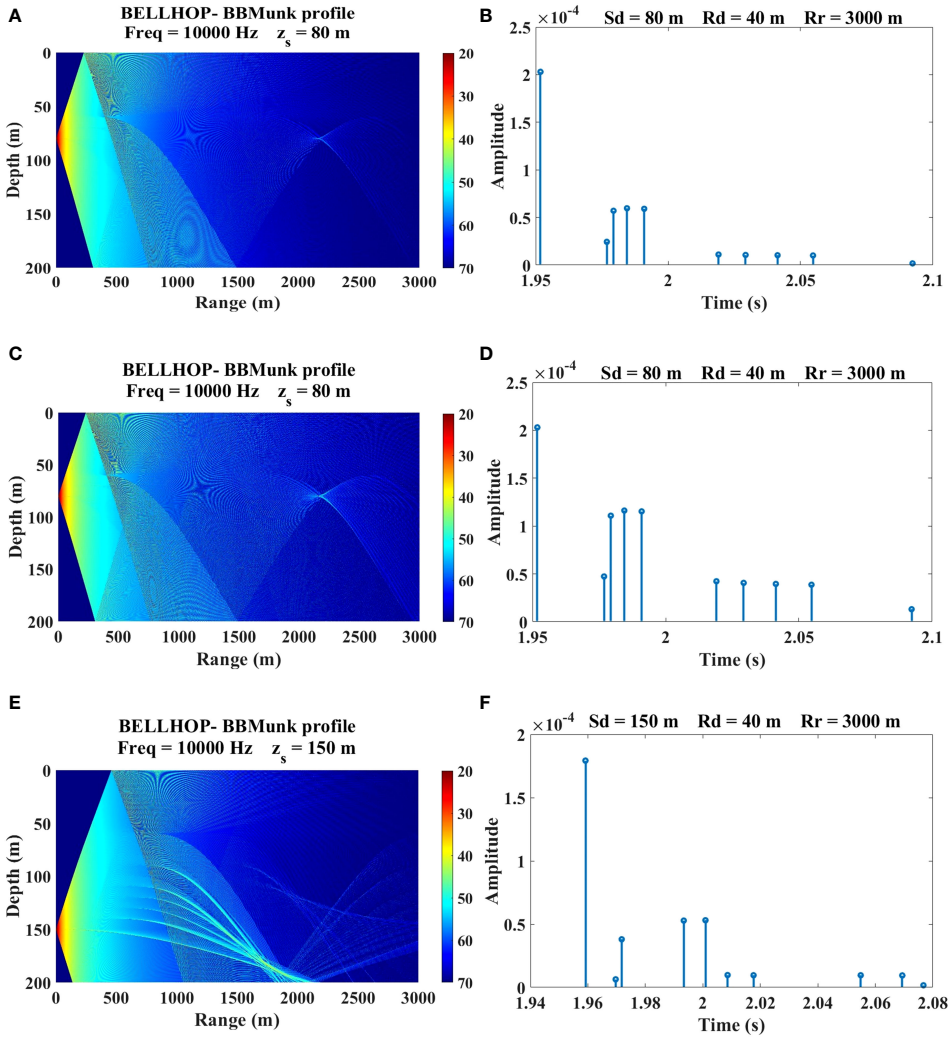


FIGURE 5
The channel information. **(A)** The transmission loss at CH1. **(B)** the impulse response at CH1. **(C)** The transmission loss at CH2. **(D)** the impulse response at CH2. **(E)** The transmission loss at CH3. **(F)** the impulse response at CH3.

mappings, namely Tent, Logistic, Cubic, Chebyshev, Piecewise, Sinusoidal, Sine, ICMIC, Circle, and Bernoulli.

The experimental test environment is as follows: 12th Gen Intel (R) Core(TM) i7-12700H CPU with a base frequency of 2.70 GHz and 16.0 GB of RAM. The operating system used is Windows 11, and the integrated development environment (IDE) is Matlab 2021a.

To increase the credibility of the algorithm, we conducted 30 independent trials on each test function. The maximum iteration count was set to 500, and the population size was 50. The

TABLE 2 Network parameter configuration.

Parameter	DFE-PLL	MLP	CFLNN	DF-CFLNN	SDF-CFLNN
Tap coefficient of feedforward	100	10	10	10	10
Tap coefficient of feedback	50	–	2	2	2
Order of Chebyshev polynomials	–	–	6	4	2
Number of nodes in the input layer	–	20	121	97	45
Number of nodes in the hidden layer	–	41	–	–	–
Number of nodes in the output layer	–	4	4	4	4
Step size	–	0.005	0.025	0.025	0.025

TABLE 3 Details of the CEC2005 benchmark suite.

Classification	Function	Dim	[Xmin, Xmax]	Fmin
Unimodal benchmark functions	$F_1(x) = \sum_{i=1}^{n} x_i^2$	30	[100,100]	0
	$F_2(x) = \sum_{i=1}^{n} \lvert x_i^2 \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	30	[100,100]	0
	$F_3(x) = \sum_{i=1}^{n} \left(\sum_{j-1}^{i} x_j \right)^2$	30	[100,100]	0
	$F_4(x) = \max_i \{ \lvert x_i \rvert \}, 1 \le i \le n$	30	[100,100]	0
	$F_5(x) = \sum_{i}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right]$	30	[30,30]	0
	$F_6(x) = \sum_{i=1}^{n} (\lvert x_i + 0.5 \rvert)^2$	30	[100,100]	0
	$F_7(x) = \sum_{i=1}^{n} i x_i^4 + random[0,1)$	30	[1.28,1.28]	0
Multimodal benchmark functions	$F_8(x) = \sum_{i=1}^{n} - x_i \sin\ (\sqrt{\lvert x_i \rvert})$	30	[500,500]	418.9829 × dim
	$F_9(x) = \sum_{i=1}^{n} [x_i^2 - 10 \cos\ (2\pi x_i) + 10]$	30	[5.12,5.12]	0
	$F_{10}(x) = -20 \exp\ \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \right)) - \exp\ \left(\frac{1}{2} \sum_{i=1}^{2} \cos\ (2\pi x_i) \right) + 20 + e$	30	[32,32]	0
	$F_{11}(x) = \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\ \left(\frac{x_i}{\sqrt{i}} \right) + 1$	30	[600,600]	0
	$F_{12}(x) = \frac{\pi}{n} \left\{ 10 \sin\ (\pi y_1) + \sum_{i=1}^{n} (y_i - 1)^2 [1 + 10 \sin^2 (\pi y_i)] + (y_n - 1)^2 \right\}$ $y_i = 1 + \frac{x_i + 1}{4} u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, x_i > a \\ 0, -a < x_i < a \\ k(-x_i - a)^m, x_i < -a \end{cases}$	30	[50,50]	0
	$F_{13} = 0.1 \left\{ \sin^2 (3\pi x_1) + \sum_{i=1}^{n} (x_i - 1)^2 [1 + \sin^2 (3\pi x_1 + 1)] + (x_n - 1)^2 [1 + \sin^2 (2\pi x_n)] \right\}$ $+ \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	30	[50,50]	0
Fixed-Dimension multimodal benchmark functions	$F_{14}(x) = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^{2} (x_i - a_{ij})^6} \right)^{-1}$	2	[65,65]	1
	$F_{15}(x) = \sum_{i=1}^{11} \left[a_i - \frac{x_1 (b_1^2 + b_1 x_2)}{b_1^2 + b_1 x_3 + x_4} \right]^2$	4	[5,5]	0.00030
	$F_{16}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3} x_1^6 + x_1 x_2 - 4x_2^2 + 4x_2^4$	2	[5,5]	1.0316
	$F_{17}(x) = \left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos\ x_1 + 10$	2	[5,5]	0.398
	$F_{18}(x) = [1 + (x_1 + x_2 + x_3)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2)]$ $\times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_2 + 48x_2 - 36x_1 x_2 + 27x_2^2)]$	2	[−2,2]	3
	$F_{19}(x) = -\sum_{i=1}^{4} c_i \exp\ \left(-\sum_{j=1}^{4} a_{ij} (x_j - p_{ij})^2 \right)$	3	[0,1]	3.86

(Continued)

TABLE 3 Continued

Classification	Function	Dim	[Xmin, Xmax]	Fmin
	$F_{20}(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2\right)$	6	[0,1]	3.32
	$F_{21}(x) = -\sum_{i=1}^{5} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	[0,10]	10.1532

experimental results indicate that overall, the Piecewise chaotic mapping exhibited superior convergence speed and accuracy. We selected the results for F15 and F21, where the convergence effects are more pronounced, for illustration, as shown in Figure 7. Therefore, we chose the Piecewise mapping as the method for random population initialization. The expression for the Piecewise mapping is as shown in Equation 36.

$$x(t+1) = \begin{cases} \frac{x(t)}{p}, & 0 \le x(t) < p \\ \frac{x(t)-p}{0.5-p}, & p \le x(t) < 0.5 \\ \frac{1-p-x(t)}{0.5-p}, & 0.5 \le x(t) < 1-p \\ \frac{1-x(t)}{p}, & 1-p \le x(t) < 1 \end{cases} \tag{36}$$

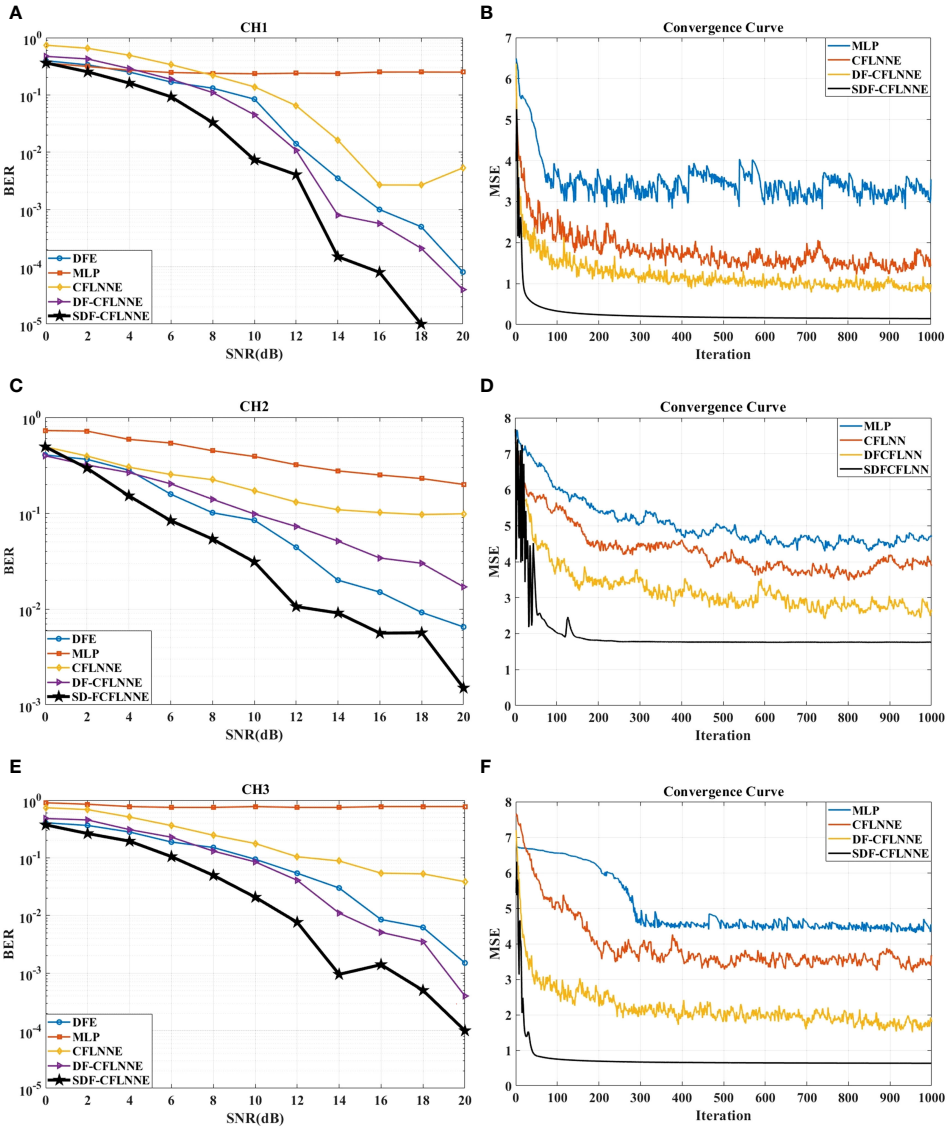


FIGURE 6
The comparison of the four CFLNN-based and DFE-PLL-based equalizers for CH1, CH2 and CH3. **(A)** The BER performance for CH1. **(B)** The MSE performance for CH1 at SNR=10 dB. **(C)** The BER performance for CH2. **(D)** The MSE performance for CH2 at SNR=10 dB. **(E)** The BER performance for CH3. **(F)** The MSE performance for CH3 at SNR=10 dB.

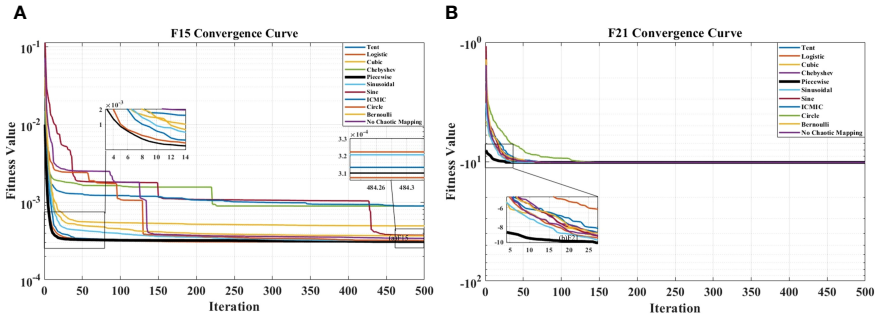


FIGURE 7
The convergence curves of initializing the OCCSSA population with ten different chaotic mappings: **(A)** F15. **(B)** F21.

where $p = 0.4$, $x(1) = rand$.

4.1.2.3 Comparison with the improved SSA-type optimization algorithms

Precision and convergence speed are important indicators for measuring the quality of an algorithm. In order to better validate the effectiveness of the proposed algorithm, this section assesses performance metrics, including convergence precision, using CEC2005 benchmark test functions. This section reproduces SSA (Xue and Shen, 2020), ISSA (Song et al., 2020), and the Adaptive Spiral Flying Sparrow Search Algorithm (ASFSSA) (Ouyang et al., 2021) to compare their performance against the proposed optimization algorithm.

To enhance the algorithm's reliability, we conducted 30 independent trials on each test function. In each trial, the maximum iteration count was set to 500, and the population size was 50. The safety threshold was set to 0.8, and the number of producers and sparrows sensing danger was both set to 20% of the population size. Tables 4–6 provide the optimization indicators for each algorithm during the optimization process, including the average objective function value, standard deviation, median, best value, average runtime, and algorithm ranking. Under the same standard test functions, the average represents the convergence accuracy of the algorithm, while the standard deviation reflects its stability. The algorithm ranking in this paper is based on both the average and standard deviation, where smaller values indicate better

TABLE 4 The comparative data for the improved SSA-type optimization algorithms on the CEC2005 multi-modal functions.

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	SSA	-3952.9522	-2788.2072	-3355.182	300.0432	-3338.7384	0.28386	4
	ASFSSA	-4189.8289	-3081.3567	-3601.1772	276.0404	-3594.6031	0.65211	2
	ISSA	-4189.8289	-2641.0282	-3190.4933	322.1179	-3151.9709	0.56114	3
F8	OCCSSA	**-4189.8289**	**-4189.8289**	**-4189.8289**	**1.8807e-12**	**-4189.8289**	**0.36753**	**1**
	SSA	0	0	0	0	0	0.2154	4
	ASFSSA	**0**	0	0	**0**	0	0.57664	**1**
	ISSA	0	**0**	**0**	0	**0**	0.54976	3
F9	OCCSSA	0	0	0	0	0	0.36489	2
	SSA	4.4409e-16	4.4409e-16	4.4409e-16	0	4.4409e-16	**0.23346**	4
	ASFSSA	**4.4409e-16**	**4.4409e-16**	**4.4409e-16**	0	**4.4409e-16**	0.61848	**1**
	ISSA	4.4409e-16	4.4409e-16	4.4409e-16	0	4.4409e-16	0.54772	3
F10	OCCSSA	**4.4409e-16**	**4.4409e-16**	**4.4409e-16**	0	**4.4409e-16**	0.37073	2
	SSA	0	0	0	0	0	0.36387	4
	ASFSSA	**0**	0	0	**0**	0	0.66729	**1**
	ISSA	0	**0**	**0**	0	**0**	0.6481	3
F11	OCCSSA	**0**	**0**	**0**	**0**	**0**	0.453	2

(Continued)

10.3389/fmars.2023.1331635

TABLE 4 Continued

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	SSA	4.7116e-32	1.2762e-28	7.7739e-30	2.3715e-29	3.8534e-31	0.43226	2
	ASFSSA	4.7864e-31	9.1663e-27	7.9713e-28	2.0578e-27	4.4766e-29	1.1384	4
	ISSA	4.7116e-32	**4.9052e-32**	**4.7649e-32**	**6.1095e-34**	**4.7237e-32**	1.0566	**1**
F12	OCCSSA	**5.2199e-32**	2.0845e-27	1.102e-28	3.8796e-28	3.7765e-30	0.76728	3
	SSA	1.473e-32	1.4984e-26	9.0349e-28	2.7677e-27	3.5795e-29	**0.66066**	3
	ASFSSA	2.0893e-32	8.6921e-27	7.6682e-28	2.13e-27	2.6357e-29	1.1247	2
	ISSA	1.3128e-30	0.010987	0.0018312	0.0041648	4.867e-28	1.0698	4
F13	OCCSSA	**1.3498e-32**	**9.19e-28**	**3.9444e-29**	**1.6717e-28**	**1.3786e-30**	0.79072	**1**
	SSA	0.998	12.6705	5.868	5.5657	2.9821	1.8161	4
	ASFSSA	0.998	10.7632	3.9379	3.9286	2.9821	2.4586	3
	ISSA	0.998	12.6705	1.7179	2.1848	0.998	1.9249	2
F14	OCCSSA	**0.998**	**0.998**	**0.998**	**2.9156e-16**	**0.998**	**1.5888**	**1**
	SSA	3.0749e-4	3.5402e-4	3.0904e-4	8.4967e-06	3.0749e-4	0.079858	2
	ASFSSA	3.0749e-4	0.0012232	4.3684e-4	2.5835e-4	3.0749e-4	0.19687	3
	ISSA	3.0749e-4	0.0015941	6.61e-4	4.7977e-4	3.764e-4	0.13209	4
F15	OCCSSA	**3.0749e-4**	**3.075e-4**	**3.0749e-4**	**3.599e-09**	**3.0749e-4**	**0.1641**	**1**
	SSA	-1.0316	-1.0316	-1.0316	6.3208e-16	-1.0316	**0.083546**	3
	ASFSSA	-1.0316	-1.0316	-1.0316	5.2964e-16	-1.0316	0.5866	2
	ISSA	-1.0316	-1.0316	-1.0316	6.5843e-16	-1.0316	0.31525	4
F16	OCCSSA	**-1.0316**	**-1.0316**	**-1.0316**	**4.9651e-16**	**-1.0316**	0.38597	**1**
	SSA	0.39789	0.39789	0.39789	0	0.39789	**0.2027**	2
	ASFSSA	0.39789	0.39789	0.39789	0	0.39789	0.53944	3
	ISSA	**0.39789**	**0.39789**	**0.39789**	**0**	**0.39789**	0.29137	**1**
F17	OCCSSA	0.39789	0.39789	0.39789	0	0.39789	0.41833	4
	SSA	3	30	3.9	4.9295	3	**0.22676**	2
	ASFSSA	**3**	**3**	**3**	**1.2176e-15**	**3**	0.61142	**1**
	ISSA	3	30	3.9	4.9295	3	0.32566	3
F18	OCCSSA	3	30	7.5	10.2343	3	0.41765	4
	SSA	-3.8628	-3.8628	-3.8628	2.5391e-15	-3.8628	0.28393	3
	ASFSSA	-3.8628	-3.8628	-3.8628	2.3557e-15	-3.8628	0.70925	2
	ISSA	-3.8628	-3.8549	-3.8625	0.001439	-3.8628	0.45951	4
F19	OCCSSA	**-3.8628**	**-3.8628**	**-3.8628**	**2.1787e-15**	**-3.8628**	0.504	**1**
	SSA	-3.322	-3.2031	-3.2507	0.059241	-3.2031	0.27128	3
	ASFSSA	-3.322	-3.2031	-3.2744	0.059241	-3.322	0.68092	2
	ISSA	-3.322	-3.2031	-3.2784	0.058273	-3.322	0.48302	4
F20	OCCSSA	**-3.322**	**-3.2031**	-3.3141	**0.030164**	-3.322	0.47212	**1**
	SSA	-10.1532	-5.0552	-9.6434	1.5555	-10.1532	**0.26862**	2
F21	ASFSSA	-10.1532	-5.0552	-8.4682	2.3059	**-10.0281**	0.55033	4

(Continued)

TABLE 4 Continued

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	ISSA	-10.1532	-5.0552	-8.6174	2.372	-10.1511	0.33206	3
	OCCSSA	**-10.1532**	**-10.1532**	**-10.1532**	**5.1842e-15**	-10.1532	0.48869	1

TABLE 5 The comparative data for the swarm intelligence optimization algorithms on the CEC2005 multimodal functions.

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
F8	PSO	-8.7193e+3	-3.9441e+3	-6.6761e+3	1.0476e+3	-6.7351e+3	0.097544	8
	GWO	-7.9899e+3	-5.0501e+3 -	-6.4564e+3 -	6.2055e+3	-6.5501e+3	0.21649	9
	WOA	-1.2568e+4	8.5968e+3	1.1133e+4	1.4966e+3	-1.1671e+4	0.085122	4
	BES	-9.8392e+3	-4.8897e+3	-6.8080e+3	1.5443e+3	-6.3557e+3	1.7086	7
	CSA	-1.2567e+4	-5.7760e+3	-8.0776e+3	2.6827e+3	-6.3615e+3	0.082586	6
	GTO	**-1.2569e+4**	**-1.2569e+4**	**-1.2569e+4**	**6.8563e-07**	**-1.2569e+4**	0.35746	**1**
	WSO	-4.4925e+3	-3.0767e+3	-3.6360e+3	3.6998e+3	-3.5819e+3	**0.068351**	10
	DBO	-1.2352 + 4	-6.8907e+3	-9.3852e+3	1.3641e+3	-9.3650e+3	0.076335	5
	OOA	-1.2569e+4	-9.0163e+3	-1.1266e+4	1.7415e+3	-1.2569e+4	0.17756	3
	OCCSSA	**-1.2569e+4**	-1.2451e+4	**-1.2569e+4**	21.6229	**-1.2569e+4**	0.49263	2
F9	PSO	101.3281	218.3835	152.4089	26.5922	151.7649	0.075587	9
	GWO	0	8.9196	1.8651	2.6026	4.2633e-13	0.18007	6
	WOA	0	0	0	0	0	0.051678	4
	BES	0	0	0	0	0	1.3967	5
	CSA	7.9697	47.8554	27.9259	11.4001	25.4951	0.054526	8
	GTO	0	0	0	0	0	0.29075	3
	WSO	15.7193	146.5208	39.0134	27.0612	29.6164	0.068292	10
	DBO	0	25.8689	2.1348	5.4652	0	0.061137	7
	OOA	0	0	0	0	0	**0.093629**	2
	OCCSSA	**0**	**0**	**0**	**0**	**0**	0.41125	1
F10	PSO	1.318	3.306	2.3647	0.46286	2.3992	0.077702	9
	GWO	3.9524e-14	6.4393e-14	4.4142e-14	5.8377e-15	4.3077e-14	0.17897	7
	WOA	4.4409e-16	7.5495e-15	3.8784e-15	2.5523e-15	3.9968e-15	**0.055302**	6
	BES	**4.4409e-16**	**4.4409e-16**	**4.4409e-16**	**0**	**4.4409e-16**	1.4385	**1**
	CSA	19.9668	19.9668	19.9668	4.9591e-12	19.9668	0.057118	8
	GTO	4.4409e-16	4.4409e-16	4.4409e-16	0	4.4409e-16	0.29324	4
	WSO	2.9679	5.4777	4.0622	0.65712	4.0067	0.072419	10
	OOA	4.4409e-16	4.4409e-16	4.4409e-16	0	4.4409e-16	0.10506	3
	DBO	4.4409e-16	3.9968e-15	5.6251e-16	6.4863e-16	4.4409e-16	0.06724	5
	OCCSSA	**4.4409e-16**	**4.4409e-16**	**4.4409e-16**	**0**	**4.4409e-16**	0.42693	2
F11	PSO	0.033408	0.16374	0.093846	0.033652	0.091152	0.13796	8
	GWO	0	0.027649	0.0036764	0.0076973	0	0.23287	6

(Continued)

TABLE 5 Continued

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	WOA	0	0.13137	0.0043788	0.023984	0	0.10538	7
	BES	0	0	0	0	0	1.0244	4
	CSA	1.0139	1.7255	1.1649	0.14593	1.1152	0.10955	9
	GTO	0	0	0	0	0	0.3944	3
	WSO	1.1382	2.3146	1.5244	0.25651	1.4776	0.12313	10
	DBO	0	0	0	0	0	**0.072711**	5
	OOA	0	0	0	0	0	0.19774	2
	OCCSSA	**0**	**0**	**0**	**0**	**0**	0.51511	**1**
F12	PSO	0.0030888	0.12305	0.037425	0.030208	0.028614	0.4018	8
	GWO	0.013497	0.048772	0.02788	0.0086555	0.02709	0.52616	7
	WOA	0.0013778	0.11719	0.0092898	0.020855	0.0041805	0.40458	6
	BES	**1.5705e-32**	**3.0953e-32**	**1.669e-32**	**3.7477e-33**	**1.5705e-32**	2.6525	**1**
	CSA	1.8485	11.4071	3.9982	1.784	3.6541	0.39943	10
	GTO	2.024e-13	7.5911e-09	7.1513e-10	1.5397e-09	1.1799e-10	0.98566	4
	WSO	0.19453	2.8473	1.1124	0.63026	0.94623	0.40955	9
	DBO	1.0548e-11	1.2499e-07	4.8707e-09	2.2709e-08	2.9129e-10	**0.26447**	5
	OOA	1.5705e-32	1.0317e-11	4.11e-13	1.8928e-12	1.5705e-32	0.84577	3
	OCCSSA	**1.5705e-32**	1.9101e-27	1.1224e-28	3.9124e-28	4.7087e-32	1.0582	2
F13	PSO	0.11726	0.69786	0.35562	0.17123	0.33052	0.41634	7
	GWO	0.098496	0.73036	0.38326	0.17023	0.38972	0.5469	6
	WOA	0.047162	0.69122	0.21665	0.17243	0.16514	0.40542	5
	BES	0.097371	2.9661	2.6729	0.79559	2.9661	2.6043	8
	CSA	2.3803	71.6642	33.0108	23.2431	37.6155	0.41566	10
	GTO	1.1212e-12	0.010987	0.0011045	0.0033508	4.046e-09	0.99464	3
	WSO	3.5379	35.3039	15.231	8.2623	15.3338	**0.41094**	9
	DBO	1.7766e-09	0.29615	0.076464	0.078188	0.092316	0.26475	4
	OOA	1.3498e-32	7.8895e-10	2.6368e-11	1.4403e-10	1.733e-29	0.79109	2
	OCCSSA	**1.3498e-32**	**1.6778e-26**	**1.047e-27**	**3.4861e-27**	**1.2296e-30**	0.9622	**1**

algorithm performance and higher rankings. In cases where the average and standard deviation are the same, the comparison is based on the convergence speed in the convergence curves. When the convergence speeds are similar, the average runtime is considered. The optimal values, algorithms with a ranking of 1, and the shortest average running time among all compared algorithms are highlighted in bold.

The data results for the CEC2005 tests are presented in Table 4. Since the optimization of neural network weights involves multi-modal functions, we utilized the second-class functions (F8-F13) and third-class functions (F14-F21) from the CEC2005 benchmark test functions. In the case of CEC2005 multi-modal functions, despite these functions having multiple local optima, the proposed algorithm was able to successfully solve the optimization problems. OCCSSA demonstrated the overall best performance, especially in F8, F13-F15, and F19-F21, where it achieved the best values for each indicator, ranking first. For F9-F11, ASFSSA performed the best, with OCCSSA ranking second. The runtime falls within a moderate range. From the convergence curve plots in Figure 8, it is evident that OCCSSA exhibited superior convergence speed and accuracy in F8, F14-F17, and F19-F21 compared to other algorithms. However, in the case of F18, OCCSSA performed poorly, with lower convergence accuracy than other algorithms. Overall, OCCSSA demonstrated good convergence speed and accuracy, as well as strong resistance to local optima in multi-modal functions. The introduction of multiple strategies significantly improved the algorithm's stability and search capabilities.

10.3389/fmars.2023.1331635

TABLE 6 The comparative data for the swarm intelligence optimization algorithms on the CEC2005 fixed dimension multi-modal functions.

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
F14	PSO	0.998	6.9033	1.8561	1.363	1.992	0.92824	7
	GWO	0.998	12.6705	4.4237	4.2869	2.9821	1.0034	9
	WOA	0.998	10.7632	2.2114	2.471	0.998	0.95681	8
	BES	0.998	12.6705	3.208	4.3568	0.998	4.1981	10
	CSA	0.998	1.992	1.0311	0.18148	0.998	0.96765	4
	GTO	0.998	0.998	0.998	0	0.998	2.0796	2
	WSO	0.998	0.998	0.998	2.4089e-13	0.998	**0.91439**	3
	DBO	0.998	2.9821	1.0641	0.36225	0.998	0.39187	5
	OOA	0.998	2.0263	1.1648	0.37943	0.998	1.982	6
	OCCSSA	**0.998**	**0.998**	**0.998**	**2.9733e-16**	**0.998**	1.7201	**1**
F15	PSO	5.5339e-4	1.0887e-3	8.9695e-4	1.466e-4	8.9653e-4	1.2332e-2	4
	GWO	3.075e-4	2.0363e-2	5.0635e-3	8.5883e-3	3.1016e-4	0.03277	10
	WOA	3.0888e-4	1.5269e-3	6.5179e-4	3.8321e-4	5.3985e-4	**0.01804**	9
	BES	3.0749e-4	1.2232e-3	3.3801e-4	1.671e-4	3.0749e-4	0.95058	5
	CSA	3.0749e-4	1.2232e-3	5.1769 e-4	3.5264e-4	3.0749e-4	0.035304	7
	GTO	3.0749e-4	1.2232e-3	3.6853e-4	2.3232e-4	3.0749e-4	0.2191	6
	WSO	3.0749e-4	3.0749e-4	3.0749 e-4	**2.4691e-8**	3.0749	0.03571	2
	DBO	3.0749e-4	1.4887e-3	7.3946e-4	3.5861e-4	7.8266e-4	0.060859	8
	OOA	3.0987e-4	8.2964e-4	4.4708e-4	1.3792e-4	3.870e-4	0.1038	3
	OCCSSA	**3.0749e-4**	**3.0749e-4**	**3.0749e-4**	2.8651e-08	**3.0749e-4**	0.38096	**1**
F16	PSO	-1.0316	-1.0316	-1.0316	5.1334e-16	-1.0316	0.020588	2
	GWO	-1.0316	-1.0316	-1.0316	1.5011e-08	-1.0316	0.055229	8
	WOA	-1.0316	-1.0316	-1.0316	5.1596e-10	-1.0316	0.045389	7
	BES	-1.0316	-1.0316	-1.0316	4.1946e-11	-1.0316	1.3678	6
	CSA	-1.0316	-1.0316	-1.0316	6.5195e-16	-1.0316	0.038116	5
	GTO	-1.0316	-1.0316	-1.0316	6.4539e-16	-1.0316	0.21813	4
	WSO	-1.0316	-1.0316	-1.0316	1.4226e-06	-1.0316	0.036043	9
	DBO	-1.0316	-1.0316	-1.0316	6.3877e-16	-1.0316	0.056909	3
	OOA	-1.0316	-1.0312	-1.0316	8.2782e-05	-1.0316	0.10267	10
	OCCSSA	**-1.0316**	**-1.0316**	**-1.0316**	**4.8787e-16**	**-1.0316**	0.40119	**1**
F17	PSO	0.39789	0.39789	0.39789	0	0.39789	0.0087385	4
	GWO	0.39789	0.3979	0.39789	1.8026e-06	0.39789	0.039837	6
	WOA	0.39789	0.3979	0.39789	3.0404e-06	0.39789	0.034851	7
	BES	0.39789	0.3979	0.39789	3.2532e-06	0.39789	1.3276	8
	CSA	0.39789	0.39789	0.39789	0	0.39789	0.027288	5
	GTO	0.39789	0.39789	0.39789	0	0.39789	0.20295	2
	WSO	0.39789	0.39793	0.39789	7.3998e-06	0.39789	0.027323	9
	DBO	0.39789	0.39789	0.39789	0	0.39789	0.049667	**1**

(Continued)

10.3389/fmars.2023.1331635

TABLE 6 Continued

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	OOA	0.39789	0.40128	0.39801	6.1782e-04	0.39789	0.086509	10
	OCCSSA	**0.39789**	**0.39789**	**0.39789**	**0**	**0.39789**	0.39572	3
F18	PSO	3	3	3	4.7063e-15	3	**0.0080774**	2
	GWO	3	3.0001	3	1.6003e-05	3	0.038497	5
	WOA	**0.0013778**	**0.11719**	**0.0092898**	0.020855	**0.0041805**	0.40458	10
	BES	3	3	3	1.3143e-15	3	1.3119	6
	CSA	3	3	3	1.2315e-15	3	0.027287	7
	GTO	3	3	3	**7.8233e-16**	3	0.19681	**1**
	WSO	3	3	3	1.7992e-15	3	0.024549	4
	DBO	3	3	3	1.1186e-15	3	0.046369	8
	OOA	3	3.0674	3.0023	0.012296	3	0.079621	9
	OCCSSA	3	3	3	3.4586e-15	3	0.37656	3
F19	PSO	-3.8628	-3.8628	-3.8628	2.3646e-15	-3.8628	0.024082	2
	GWO	-3.8628	-3.8552	-3.8618	0.0022799	-3.8627	0.060494	8
	WOA	-3.8628	-3.8504	-3.8609	0.0032281	-3.8622	0.052052	9
	BES	-3.8628	-3.8628	-3.8628	2.7101e-15	-3.8628	1.4188	6
	CSA	-3.8628	-3.8628	-3.8628	2.6543e-15	-3.8628	0.043274	4
	GTO	-3.8628	-3.8628	-3.8628	2.6823e-15	-3.8628	0.22995	5
	WSO	-3.8628	-3.8628	-3.8628	2.7101e-15	-3.8628	0.040928	7
	DBO	-3.8628	-3.8628	-3.8628	2.6543e-15	-3.8628	0.059415	3
	OOA	-3.8628	-3.679	-3.8038	0.046934	-3.8076	0.11594	10
	OCCSSA	**-3.8628**	**-3.8628**	**-3.8628**	**2.2629e-15**	**-3.8628**	0.42646	**1**
F20	PSO	-3.322	-3.2031	-3.2705	0.059923	-3.322 -	**0.030683**	3
	GWO	-3.322	-3.1345	-3.2654	0.072981	3.322	0.073937	5
	WOA	-3.322 -	-2.8401	-3.2149	0.10919	-3.1884	0.047003	9
	BES	3.322	-3.2031	-3.2586	0.060328	-3.2031	1.443	6
	CSA	-3.322	-3.2014	-3.2585	0.060388	-3.2031	0.039499	7
	GTO	-3.322	-3.2031	-3.2705	0.059923	-3.322	0.24303	4
	WSO	-3.322	-3.2031	-3.318	0.021707	-3.322	0.044449	2
	DBO	-3.322	-3.0839	-3.2459	0.08448	-3.2625	0.061138	8
	OOA	-3.2121	-1.8054	-2.6624	0.29428	-2.6389	0.11226	10
	OCCSSA	-3.322	-3.322	**-3.322**	**1.6739e-15**	**-3.322**	0.4264	**1**
F21	PSO	-10.1532	-2.6305	-7.6382	3.0198	-10.1532	0.031031	10
	GWO	-10.153	-5.0552	-9.3055	1.9253	-10.1518	0.077393	5
	WOA	-10.1531	-5.0551	-8.9603	2.1912	-10.147	0.061021	7
	BES	-10.1532	-5.0552	-7.6028	2.5911	-7.5825	1.4532	8
	CSA	-10.1532	-5.1008	-9.1427	2.0555	-10.1532	0.050115	6
	GTO	-10.1532	-10.1532	-10.1532	6.3278e-15	-10.1532	0.24091	3

(Continued)

TABLE 6 Continued

Function Name	Algorithm Name	Best Value	Worst Value	Average Value	StandDP	Median Value	Average Running Time	Rank
	WSO	**-10.1532**	**-10.1532**	**-10.1532**	**5.0459e-15**	**-10.1532**	0.049681	**1**
	DBO	-10.1532	-2.6305	-6.7838	2.675	-5.1008	0.065969	9
	OOA	-10.1532	-10.1532	-10.1532	3.0155e-07	-10.1532	0.12924	4
	OCCSSA	**-10.1532**	**-10.1532**	**-10.1532**	5.3086e-15	**-10.1532**	0.41797	2

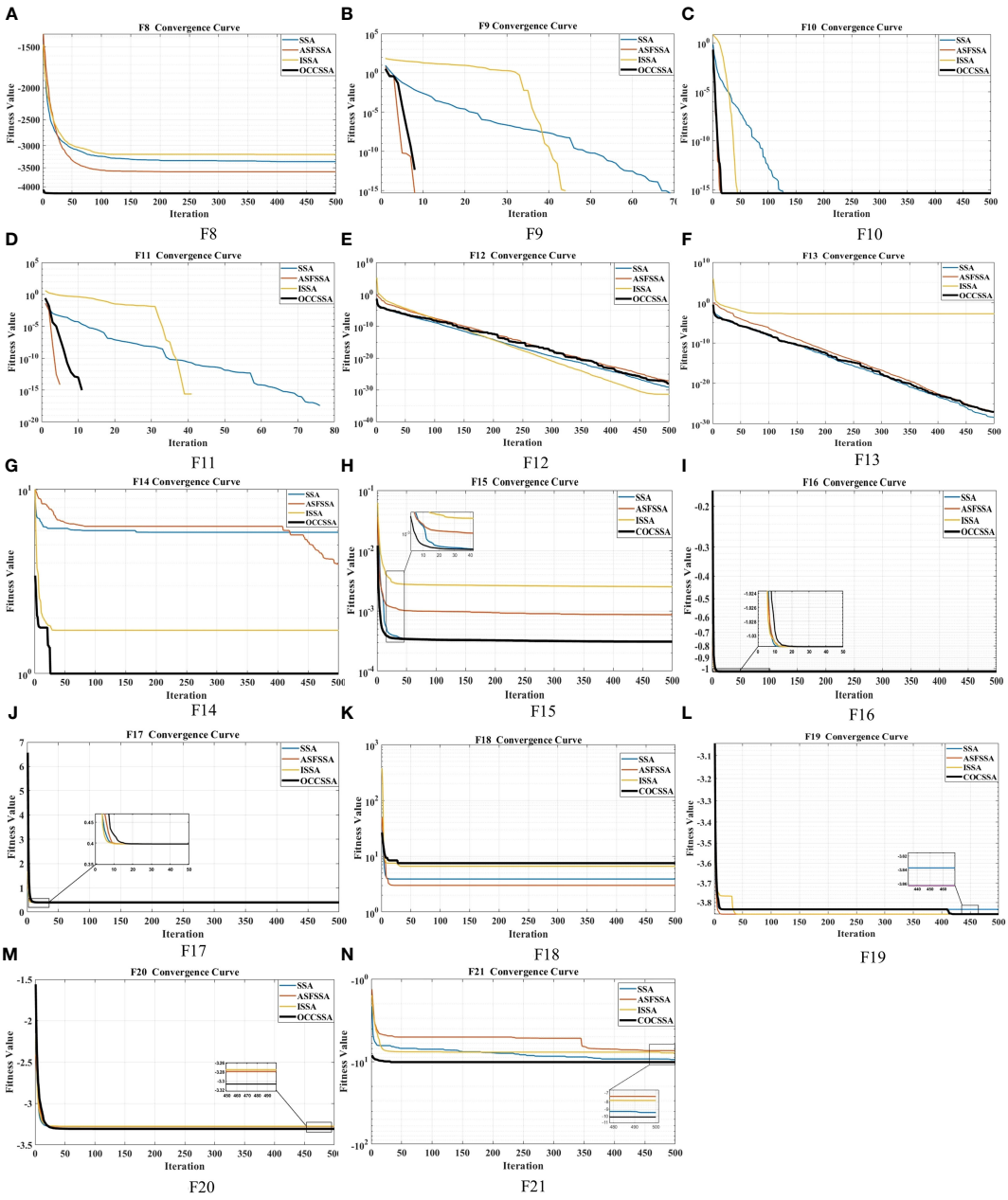


FIGURE 8
The convergence curve comparison graph of the improved SSA-type optimization algorithms on CEC2005 Multi-Modal functions: **(A)** F8. **(B)** F9. **(C)** F10. **(D)** F11. **(E)** F12. **(F)** F13. **(G)** F14. **(H)** F15. **(I)** F16. **(J)** F17. **(K)** F18. **(L)** F19. **(M)** F20. **(N)** F21.

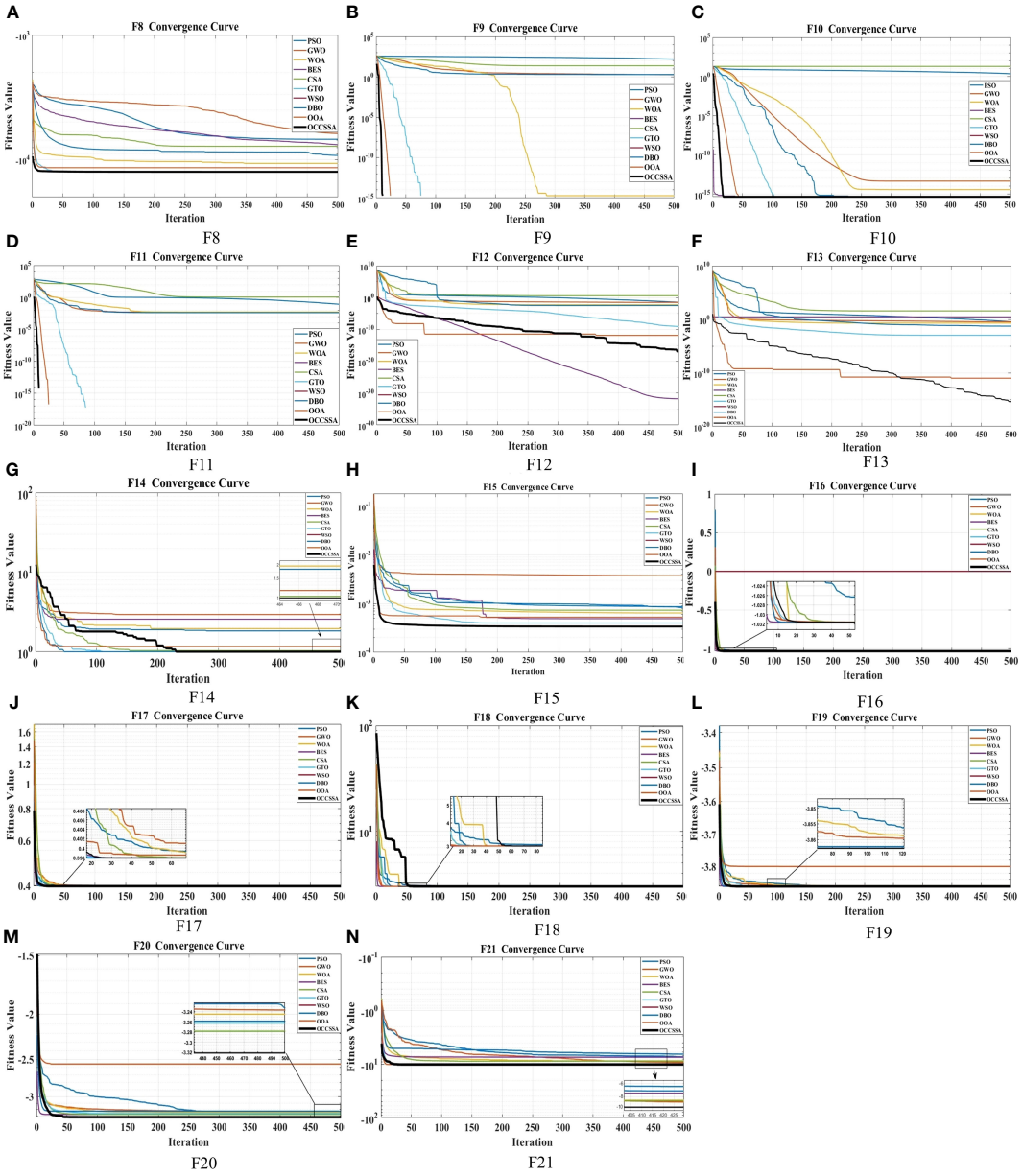


FIGURE 9
The convergence curve comparison graph of the swarm intelligence optimization algorithms on CEC2005 Multi-Modal functions: **(A)** F8. **(B)** F9. **(C)** F10. **(D)** F11. **(E)** F12. **(F)** F13. **(G)** F14. **(H)** F15. **(I)** F16. **(J)** F17. **(K)** F18. **(L)** F19. **(M)** F20. **(N)** F21.

4.1.2.4 Comparison with other swarm intelligence optimization algorithms

To further validate the effectiveness of the proposed algorithm, we compared OCCSSA with other recent swarm intelligence optimization algorithms. These include PSO (Kennedy and Eberhart, 1995), GWO (Mirjalili et al., 2014), WOA (Mirjalili and Lewis, 2016), BES (Alsattar et al., 2020), CSA (Feng et al., 2021), GTO (Abdollahzadeh et al., 2021), DBO (Braik et al., 2022), WSO (Xue and Shen, 2023), and OOA (Dehghani and Trojovskỳ, 2023).

To enhance the credibility of the algorithm, we conducted 30 independent trials on each test function. In each trial, the maximum iteration count was set to 500, and the population size was 50.

The data results for CEC2005 tests are presented in Tables 5, 6. In CEC2005 multi-modal functions, despite the presence of multiple local optima, the proposed algorithm was able to successfully solve the optimization problems. OCCSSA demonstrated the best overall performance, especially in F8-F11, F13-F16, and F20, where it achieved the best values for each indicator, ranking first. For F12 and F21, OCCSSA ranked second, just behind BES and WSO, respectively. In the case of F17 and F18, OCCSSA ranked third. Due to its higher algorithm complexity, the runtime was in the middle to lower range. From the convergence curve plots in Figure 9, it is evident that OCCSSA exhibited overall better convergence speed and accuracy compared

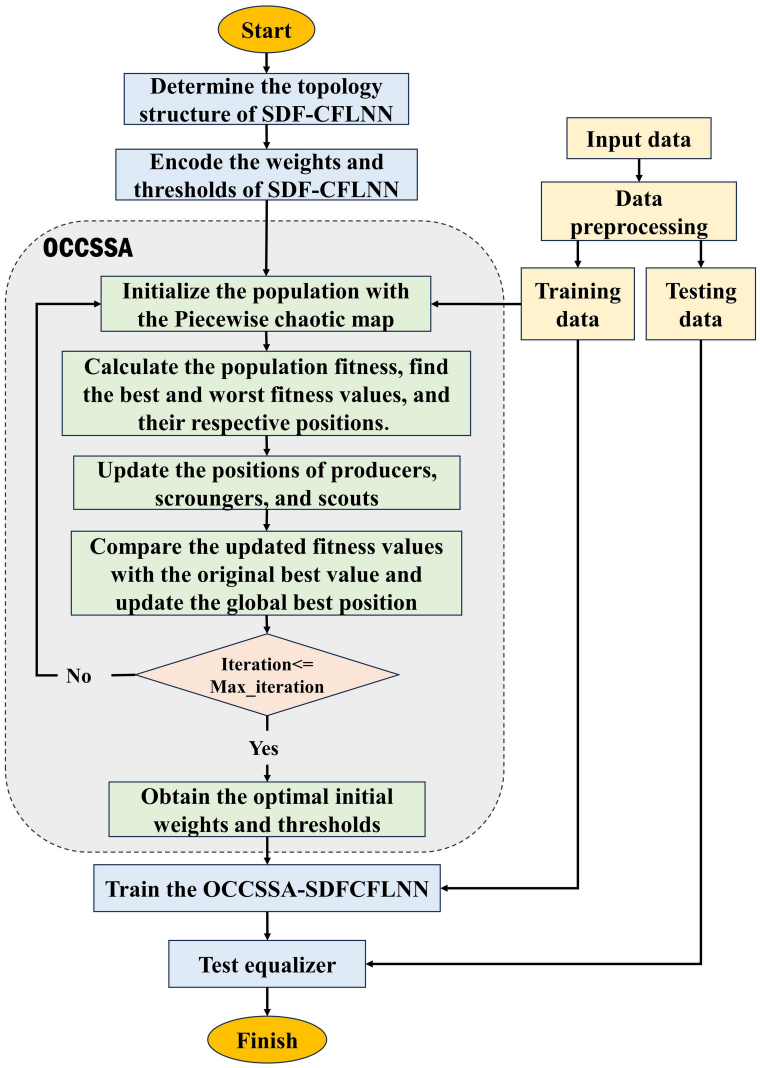


FIGURE 10
The flowchart of the OCCSSA-SDFCFLNNE.

to other algorithms. Overall, OCCSSA displayed strong resistance to local optima in multi-modal functions, and the introduction of multiple strategies significantly improved the algorithm's stability and search capabilities, with good convergence speed and accuracy.

4.1.3 Comparison of the random initialization and OCCSSA initialization of SDF-CFLNNE

During the training phase of a neural network, random initialization of weights and biases can significantly impact the training process and the final performance of the network. To address the issues related to improper initialization, such as gradient vanishing or exploding, and infeasible or slow convergence of the training process, the SDF-CFLNNE initialized intelligently with OCCSSA is proposed. The flowchart is as shown in the Figure 10.

First, determine the topology of SDF-CFLNN and encode its weights and thresholds. Then, input the encoded population into OCCSSA for initialization using the Piecewise chaotic map. Next, calculate the fitness of the initial population and identify the best and worst population members. OCCSSA updates the positions of producers, scroungers, and scouts. The updated fitness is compared to the original best value, and the global best position is updated. When the maximum iteration is reached, obtain the best population as the initial weights and thresholds for training and testing the network.

We compared the SDFCFLNN equalizers with random initialization (R-SDFCFLNNE) and OCCSSA initialization (OCCSSA-SDFCFLNNE) in both CH1 and CH2 channel environments. To enhance the reliability of the algorithm, we conducted 10 independent experiments. In each experiment, the maximum iteration limit was set to 200, and SDF-CFLNNE still used the parameters from Table 2. For OCCSSA, we used a population size of 50, a safety threshold of 0.8, and the number of producers and the number of sparrows sensing danger were both set to 20% of the population size.

The BER performance of the R-SDFCFLNNE and OCCSSA-SDFCFLNNE are shown in Figures 11A, C, E. OCCSSA-

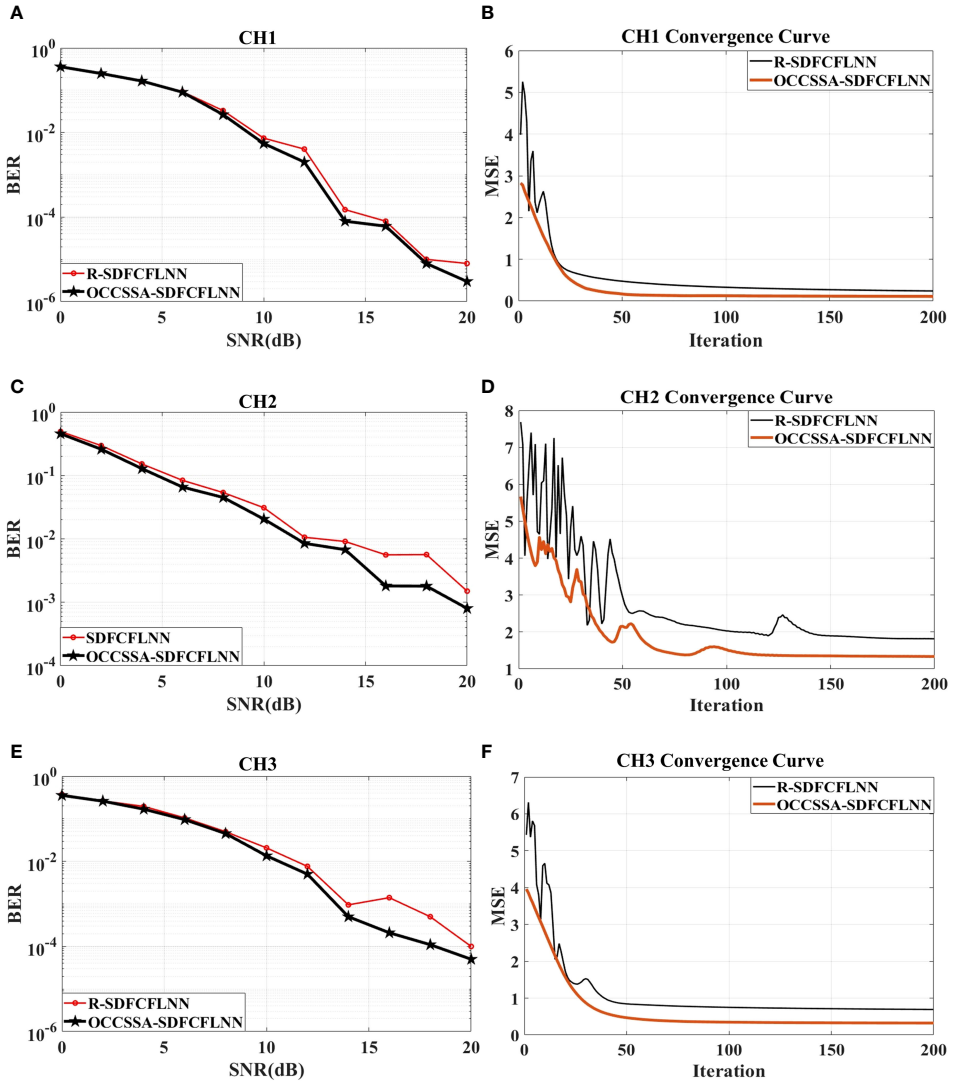


FIGURE 11
The comparison of the R-SDFCFLNNE and OCCSSA-SDFCFLNNE for CH1, CH2 and CH3. **(A)** The BER performance for CH1. **(B)** The MSE performance for CH1 at SNR=10 dB. **(C)** The BER performance for CH2. **(D)** The MSE performance for CH2 at SNR=10 dB. **(E)** The BER performance for CH3. **(F)** The MSE performance for CH3 at SNR=10 dB.

SDFCFLNNE exhibits a slightly lower BER compared to R-SDFCFLNNE. Using CH1 as an example, as observed from Figures 6A, 11A, it is evident that at a BER of 10^{-3}, CFLNNE requires the SNR exceeding 20 dB, DF-CFLNNE requires the SNR of 16 dB, DFE requires the SNR of 13.8 dB, SDF-CFLNNE requires the SNR of 12.8 dB, and OCCSSA-SDFCFLNNE requires the SNR of 12.5 dB. OCCSSA-SDFCFLNNE demonstrates an improvement SNR of 0.2-8 dB compared to CFLNN-based and traditional equalizers. At a BER of 10^{-4}, CFLNNE requires the SNR exceeding 20 dB, DFE requires the SNR of 19.7 dB, DF-CFLNNE requires the SNR of 18.8 dB, SDF-CFLNNE requires the SNR of 15.5 dB, and OCCSSA-SDFCFLNNE requires the SNR of 13.5 dB. OCCSSA-SDFCFLNNE outperforms CFLNN-based and traditional equalizers, demonstrating an improvement SNR of 2-6 dB.

The MSE performance of the R-SDFCFLNNE and OCCSSA-SDFCFLNNE are shown in Figures 11B, D, F. Both of them converge at almost the same rate. However, R-SDFCFLNNE exhibits minor oscillations in the early iterations, and the curve becomes smooth after convergence. In contrast, OCCSSA-SDFCFLNNE has an extremely smooth convergence curve, which is more stable. The MSE value of OCCSSA-SDFCFLNNE is smaller when it reaches a steady state, indicating more accurate signal recovery. When using OCCSSA initialization, it takes into account the specific characteristics and constraints of the communication channel to provide an optimal set of weight values for network initialization. This can lead to better initial conditions for the network, resulting in improved convergence and signal recovery.

4.2 Lake experiments and results

The analysis of lake experimental data has been presented in this part. On the day of the experiment, there was a slight surface fluctuation on the lake. Before conducting the lake experiments, the hydrophones

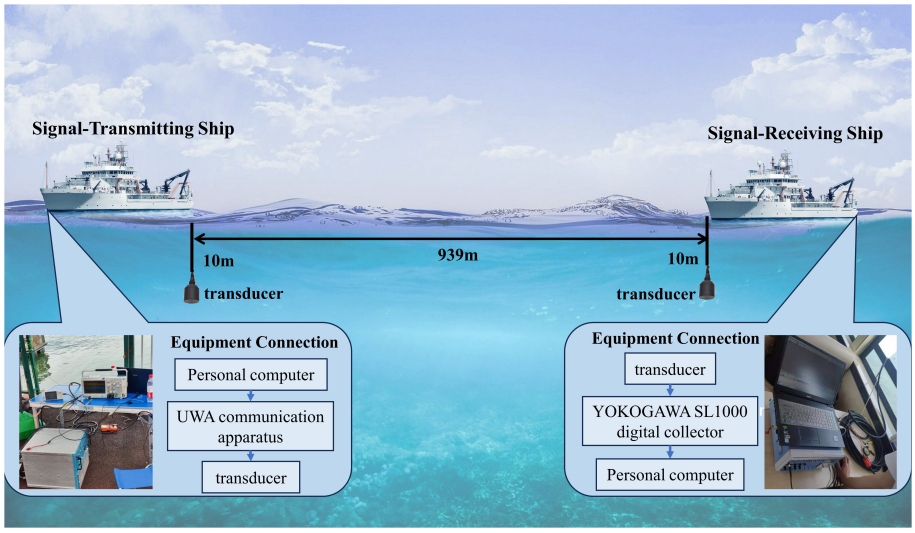


FIGURE 12
The layout diagram of transmitter and receiver.

	Doppler Estimation and Frame Synchronization Head	Training Sequences	Data Information	Frame Interval
Data I	1023 chips balanced Gold sequences	1500 chips	1000 chips	217 chips of data information
Data Q	1023 chips balanced Gold sequences	1500 chips	1000 chips	

Data Frame 1 | Data Frame 2 | ...

FIGURE 13
The frame format of the transmitting QPSK signal.

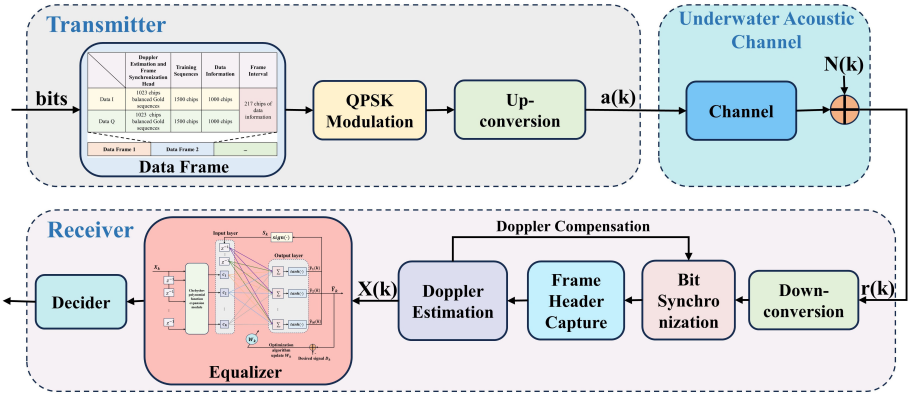


FIGURE 14
The signal processing flow.

TABLE 7 The BER comparison of every frame.

Frame	DFE-PLL	R-SDFCFLNNE	OCCSSA-SDFCFLNNE
1	0.1641	0.1222	0.0941
2	0.0962	0.0587	0.0425
3	0.1377	0.0959	0.0877
4	0.0854	0.0311	0.0195
5	0.1176	0.0734	0.0567

and other experimental equipment underwent meticulous calibration performed. Additionally, we carried out tasks such as assessing electrical connections and confirming the reliability of communication links within a controlled water tank environment. The equipment connection and layout for the transmitting ship and the receiving ship are as shown in the Figure 12. The distance between the transmitting ship and the receiving ship is either 939m. Both the transmitting and receiving transducers are positioned 10 meters underwater.

The system employs QPSK modulation to facilitate data transmission. The signal frame format, as illustrated in Figure 13, has been specifically tailored for complex underwater acoustic conditions. Each data frame incorporates key elements such as Doppler estimation, frame synchronization headers, training sequences, data content, and frame intervals. This frame format is engineered to offer robust anti-Doppler capabilities and effectively mitigate cumulative timing errors. The signal undergoes amplification by a power amplifier and is then transmitted via a transducer. Simultaneously, multiple cycles of underwater acoustic signals are collected by the receivers using a digital collector linked to the transducer. The collected data are processed by DFE-PLL, R-SDFCFLNNE and OCCSSA-SDFCFLNNE. The signal processing flow at the receiving end is depicted in the Figure 14.

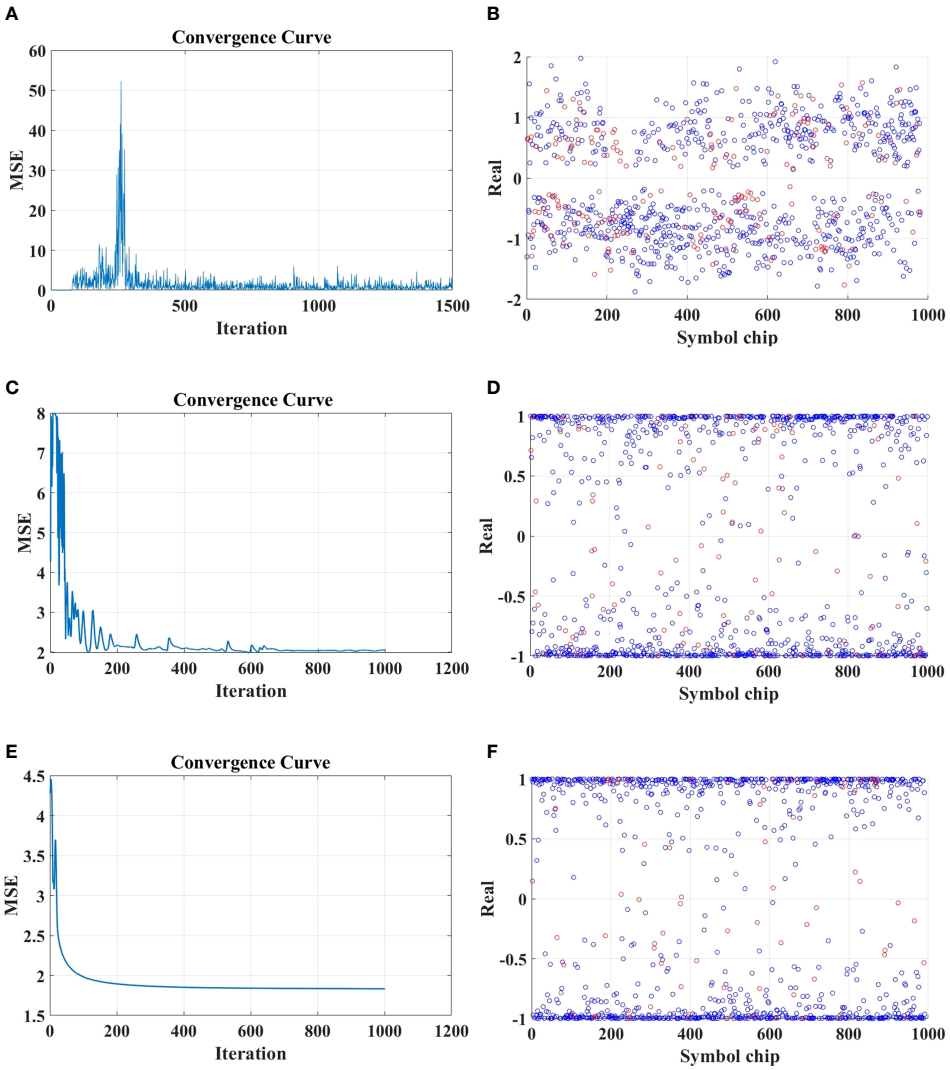


FIGURE 15
The lake experimental results. **(A)** The MSE performance for DFE-PLL. **(B)** The Demodulation correct symbol (blue) and transmission error symbol (red) for DFE-PLL. **(C)** The MSE performance for R-SDFCFLNNE. **(D)** The Demodulation correct symbol (blue) and transmission error symbol (red) for R-SDFCFLNNE. **(E)** The MSE performance for OCCSSA-SDFCFLNNE. **(F)** The Demodulation correct symbol (blue) and transmission error symbol (red) for OCCSSA-SDFCFLNNE.

The experimental results are shown in Table 7. We can observe that in the data processing of each frame, OCCSSA-SDFCFLNNE exhibits the lowest BER, followed by R-SDFCFLNNE, and DFE-PLL performing the least favorably. We've displayed the MSE convergence curves and the correct and incorrect symbol plots for the second frame of data. Figures 15A, C, E clearly demonstrate that the MSE curve in OCCSSA-SDFCFLNNE converges smoothly and reaches the lowest value, indicating a superior resistance to interference. While R-SDFCFLNNE exhibits minor initial fluctuations, DFE-PLL's MSE values fluctuate significantly during the initial phase. In Figures 15B, D, F, where blue represents correctly demodulated symbols and red represents incorrectly demodulated symbols, it's evident that OCCSSA-SDFCFLNNE has the fewest incorrect symbols and the best overall performance.

5 Conclusion

In the shallow-water environments, underwater communication links are susceptible to significant multipath effects. To address issues such as slow convergence and high system complexity in traditional channel equalizers, this paper proposes a simplified decision feedback Chebyshev function link neural network equalizer (SDF-CFLNNE). The SDF-CFLNNE structure's innovative approach employs Chebyshev polynomial function expansion modules, eliminating the need for hidden layers and enabling a direct, nonlinear transformation of input signals into the output layer. Additionally, it incorporates the feedback of decision signals into the input layer of the SDF-CFLNN directly, instead of the function expansion module, which significantly reduces computational complexity. However, the effectiveness of neural networks crucially depends on the initial weights and biases, and random initialization can profoundly impact both the training process and the eventual performance of the network. To address this challenge, a novel chaotic sparrow search algorithm combining osprey optimization algorithm and Cauchy mutation (OCCSSA) is proposed. OCCSSA leverages a Piecewise chaotic population initialization strategy, combining the osprey's exploration tactics with the Cauchy mutation strategy to bolster global and local search capabilities. Simulation experiments, utilizing underwater multipath signals generated by the Bellhop Acoustic Toolbox, unequivocally demonstrate that the SDFCFLNNE initialized by OCCSSA outperforms both CFLNN-based and traditional nonlinear equalizers. Notably, it achieves an impressive 2-6 dB improvement in SNR at a BER of 10^{-4} and exhibits a significantly reduced MSE. Furthermore, lake experimental data was employed to validate the effectiveness of the proposed equalizer. These results underscore the remarkable potential of the SDFCFLNNE initialized by OCCSSA as a compelling solution for significantly enhancing the reliability of underwater communication, particularly in the face of the challenges posed by complex underwater environments. This research paves the way for more robust and efficient underwater communication systems, promising increased performance and greater accuracy in signal recovery.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

MZ: Conceptualization, Data curation, Methodology, Validation, Writing – original draft, Writing – review & editing. HZ: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. TL: Writing – review & editing, Funding acquisition, Project administration, Resources, Supervision. WH: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. YD: Writing – review & editing, Data curation, Formal analysis, Investigation. YG: Data curation, Formal analysis, Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Natural Science Foundation of China (Grant No. 91938204 and 62271459), the Marine S and T fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2018SDKJ0210).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abdollahzadeh, B., Soleimanian Gharehchopogh, F., and Mirjalili, S. (2021). Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems. *Int. J. Intelligent Syst.* 36, 5887–5958. doi: 10.1002/int.22535

Alsattar, H. A., Zaidan, A., and Zaidan, B. (2020). Novel meta-heuristic bald eagle search optimisation algorithm. *Artif. Intell. Rev.* 53, 2237–2264. doi: 10.1007/s10462-019-09732-5

Barnard, C. J., and Sibly, R. M. (1981). Producers and scroungers: a general model and its application to captive flocks of house sparrows. *Anim. Behav.* 29, 543–550. doi: 10.1016/S0003-3472(81)80117-0

Braik, M., Hammouri, A., Atwan, J., Al-Betar, M. A., and Awadallah, M. A. (2022). White shark optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems. *KnowledgeBased Syst.* 243, 108457. doi: 10.1016/j.knosys.2022.108457

Burse, K., Yadav, R. N., and Shrivastava, S. (2010). Channel equalization using neural networks: A review. *IEEE Trans. systems man cybernetics Part C (Applications Reviews)* 40, 352–357. doi: 10.1109/TSMCC.2009.2038279

Chagra, W., Bouani, F., Abdennour, R. B., Ksouri, M., and Favier, G. (2005). Equalization with decision delay estimation using recurrent neural networks. *Adv. Eng. Software* 36, 442–447. doi: 10.1016/j.advengsoft.2005.01.011

Chang, Y.-J., and Ho, C.-L. (2009). "Decision feedback equalizers using self-constructing fuzzy neural networks," in *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*. 1483–1486 (IEEE).

Chang, Y.-J., and Ho, C.-L. (2011). Scfnn-based decision feedback equalization robust to frequency offset and phase noise. *Circuits Systems Signal Process.* 30, 929–940. doi: 10.1007/s00034-010-9258-5

Chang, P.-R., and Wang, B.-C. (1995). Adaptive decision feedback equalization for digital satellite channels using multilayer neural networks. *IEEE J. selected areas Commun.* 13, 316–324. doi: 10.1109/49.345876

Dehghani, M., and Trojovskỳ, P. (2023). Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems. *Front. Mechanical Eng.* 8, 1126450. doi: 10.3389/fmech.2022.1126450

Feng, Z.-k., Niu, W.-j., and Liu, S. (2021). Cooperation search algorithm: A novel metaheuristic evolutionary intelligence algorithm for numerical optimization and engineering optimization problems. *Appl. Soft Computing* 98, 106734. doi: 10.1016/j.asoc.2020.106734

Freitag, L., Johnson, M., and Stojanovic, M. (1997). "Efficient equalizer update algorithms for acoustic communication channels of varying complexity," in *Oceans' 97. MTS/IEEE Conference Proceedings* (Halifax, NS, Canada: IEEE), Vol. 1. 580–585.

Gao, M., Guo, Y.-c., Liu, Z.-x., and Zhang, Y.-p. (2009). "Feed-forward neural network blind equalization algorithm based on super-exponential iterative," in *2009 International Conference on Intelligent HumanMachine Systems and Cybernetics* (Hangzhou, China: IEEE), Vol. 1. 335–338.

Ge, W., Wang, Z., Yin, J., and Han, X. (2022). Robust equalization for single-carrier underwater acoustic communications based on parameterized interference model. *IEEE Wireless Commun. Lett.* doi: 10.1109/LWC.2022.3223533

Gibson, G. J., Siu, S., and Cowen, C. (1989). "Multilayer perceptron structures applied to adaptive equalisers for data communications," in *International Conference on Acoustics, Speech, and Signal Processing* (Glasgow, UK: IEEE), 1183–1186.

Guha, D. R., and Patra, S. K. (2009). "Isi and burst noise interference minimization using wilcoxon generalized radial basis function equalizer," in *2009 Fifth International Conference on MEMS NANO, and Smart Systems* (Dubai, United Arab Emirates: IEEE), 89–92.

He, C., Jing, L., Xi, R., Wang, H., Hua, F., Dang, Q., et al. (2019). Time-frequency domain turbo equalization for single-carrier underwater acoustic communications. *IEEE Access* 7, 73324–73335. doi: 10.1109/ACCESS.2019.2919757

He, Y., and Tao, Y. (2023). "Deep reinforcement learning based cognitive equalization algorithm research in underwater communication," in *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)* (Taiyuan, China: IEEE), 348–352.

He, Q., Tao, J., Kong, X., Zhuang, Y., Jiang, M., and Qiao, Y. (2023). "Channel replay aided neural network equalizer for underwater acoustic communications," in *OCEANS 2023-Limerick* (Limerick, Ireland: IEEE), 1–5).

Heng, S., Chen, W., Yin, H., Shiping, M., and Jizhang, Z. (2006). Decision feedback equalizer based on non-singleton fuzzy regular neural networks. *J. Syst. Eng. Electron.* 17, 896–900. doi: 10.1016/S1004-4132(07)60034-6

Huang, J.-g., Wang, H., He, C.-b., Zhang, Q.-f., and Jing, L.-y. (2018). Underwater acoustic communication and the general performance evaluation criteria. *Front. Inf. Technol. Electronic Eng.* 19, 951–971. doi: 10.1631/FITEE.1700775

Hussain, A., Soraghan, J. J., and Durrani, T. S. (1997). A new adaptive functional-link neural-networkbased dfe for overcoming co-channel interference. *IEEE Trans. Commun.* 45, 1358– 1362. doi: 10.1109/26.649741

Ingle, K. K., and Jatoth, R. K. (2023). Non-linear channel equalization using modified grasshopper optimization algorithm. *Appl. Soft Computing* 110091. doi: 10.1016/j.asoc.2023.110091

Kari, D., Vanli, N. D., and Kozat, S. S. (2017). Adaptive and efficient nonlinear channel equalization for underwater acoustic communication. *Phys. Communication* 24, 83–93. doi: 10.1016/j.phycom.2017.06.001

Kechriotis, G., Zervas, E., and Manolakos, E. S. (1994). Using recurrent neural networks for adaptive communication channel equalization. *IEEE Trans. Neural Networks* 5, 267–278. doi: 10.1109/72.279190

Kennedy, J., and Eberhart, R. (1995). "Particle swarm optimization," in *Proceedings of ICNN'95international conference on neural networks* (WA, Australia: IEEE), Vol. 4, 1942–1948.

Lee, T.-T., and Jeng, J.-T. (1998). The chebyshev-polynomials-based unified model neural networks for function approximation. *IEEE Trans. Systems Man Cybernetics Part B (Cybernetics)* 28, 925–935.

Lee, J., and Sankar, R. (2007). Theoretical derivation of minimum mean square error of rbf based equalizer. *Signal Process.* 87, 1613–1625. doi: 10.1016/j.sigpro.2007.01.008

Li, Y., Geng, T., Tian, R., and Gao, S. (2021). Machine-learning based equalizers for mitigating the interference in asynchronous mimo owc systems. *J. Lightwave Technol.* 39, 2800–2808. doi: 10.1109/JLT.2021.3057396

Liker, A., and Barta, Z. (2002). The effects of dominance on social foraging tactic use in house sparrows. *Behaviour*, 1061–1076. doi: 10.1163/15685390260337903

Liu, W., Liu, J., Liu, T., Chen, H., and Wang, Y.-L. (2023). Detector design and performance analysis for target detection in subspace interference. *IEEE Signal Process. Lett.* doi: 10.1109/LSP.2023.3270080

Liu, J., Mei, K., Zhang, X., Ma, D., and Wei, J. (2019). Online extreme learning machine-based channel estimation and equalization for ofdm systems. *IEEE Commun. Lett.* 23, 1276–1279. doi: 10.1109/LCOMM.2019.2916797

Ma, X., Ye, H., and Li, Y. (2018). "Learning assisted estimation for time-varying channels," in *2018 15th international symposium on wireless communication systems (ISWCS)* (Lisbon, Portugal: IEEE), 1–5.

Milan, A., Rezatofighi, S. H., Dick, A., Reid, I., and Schindler, K. (2017). "Online multi-target tracking using recurrent neural networks," in *Proceedings of the AAAI conference on Artificial Intelligence*, Vol. 31.

Mirjalili, S., and Lewis, A. (2016). The whale optimization algorithm. *Adv. Eng. software* 95, 51–67. doi: 10.1016/j.advengsoft.2016.01.008

Mirjalili, S., Mirjalili, S. M., and Lewis, A. (2014). Grey wolf optimizer. *Adv. Eng. software* 69, 46–61. doi: 10.1016/j.advengsoft.2013.12.007

Mishra, J. P., Singh, K., and Chaudhary, H. (2023). "Recent advancement of ai technology for underwater acoustic communication," in *AIP Conference Proceedings* (AIP Publishing). (Jaipur, India: IEEE), Vol. 2752.

Mohapatra, P. K., Rout, S. K., Bisoy, S. K., and Sain, M. (2022). Training strategy of fuzzy-firefly based ann in non-linear channel equalization. *IEEE Access* 10, 51229–51241. doi: 10.1109/ACCESS.2022.3174369

Ning, X., Liu, Z., and Luo, Y. (2009). "Research on variable step-size blind equalization algorithm based on normalized rbf neural network in underwater acoustic communication," in *Advances in Neural Networks–ISNN 2009: 6th International Symposium on Neural Networks, ISNN 2009 Wuhan, China, May 26-29, 2009 Proceedings, Part III 6* (Berlin Heidelberg: Springer), 1063–1070.

Ouyang, C., Qiu, Y., and Zhu, D. (2021). Adaptive spiral flying sparrow search algorithm. *Sci. Programming* 2021, 1–16. doi: 10.1155/2021/6505253

Patra, J. C., Chin, W. C., Meher, P. K., and Chakraborty, G. (2008). "Legendre-flann-based nonlinear channel equalization in wireless communication system," in *2008 IEEE international conference on systems, man and cybernetics* (Singapore: IEEE), 1826–1831.

Patra, J. C., and Kot, A. C. (2002). Nonlinear dynamic system identification using chebyshev functional link artificial neural networks. *IEEE Trans. Systems Man Cybernetics Part B (Cybernetics)* 32, 505–511. doi: 10.1109/TSMCB.2002.1018769

Patra, J. C., Pal, R. N., Baliarsingh, R., and Panda, G. (1999). Nonlinear channel equalization for qam signal constellation using artificial neural networks. *IEEE Trans. Systems Man Cybernetics Part B (Cybernetics)* 29, 262–271. doi: 10.1109/3477.752798

Patra, J. C., Poh, W. B., Chaudhari, N. S., and Das, A. (2005). "Nonlinear channel equalization with qam signal using chebyshev artificial neural network," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005* (Montreal, QC, Canada: IEEE), Vol. 5. 3214–3219.

Qiao, G., Liu, Y., Zhou, F., Zhao, Y., Mazhar, S., and Yang, G. (2022). Deep learning-based m-ary spread spectrum communication system in shallow water acoustic channel. *Appl. Acoustics* 192, 108742. doi: 10.1016/j.apacoust.2022.108742

Qin, X., Qu, F., and Zheng, Y. R. (2020). Bayesian iterative channel estimation and turbo equalization for multiple-input–multiple-output underwater acoustic communications. *IEEE J. Oceanic Eng.* 46, 326–337. doi: 10.1109/JOE.2019.2956299

Sahu, J., and Majumder, S. (2021). "A particle swarm optimization based training algorithm for mcma blind adaptive equalizer," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)* (Pune, India: IEEE), 462–465.

Shwetha, N., Priyatham, M., and Gangadhar, N. (2023). Artificial neural network based channel equalization using battle royale optimization algorithm with different initialization strategies. *Multimedia Tools Appl.* 1–26. doi: 10.1007/s11042-023-16161-8

Song, H. C., Hodgkiss, W., Kuperman, W., Stevenson, M., and Akal, T. (2006). Improvement of timereversal communications using adaptive channel equalizers. *IEEE J. Oceanic Eng.* 31, 487–496. doi: 10.1109/JOE.2006.876139

Song, W., Liu, S., Wang, X., and Wu, W. (2020). "An improved sparrow search algorithm," in *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/ SustainCom)*. 537–543 (IEEE).

Stojanovic, M., Catipovic, J. A., and Proakis, J. G. (1994). Phase-coherent digital communications for underwater acoustic channels. *IEEE J. oceanic Eng.* 19, 100–111. doi: 10.1109/48.289455

Stojanovic, M., and Preisig, J. (2009). Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Commun. magazine* 47, 84–89. doi: 10.1109/MCOM.2009.4752682

Suganthan, P. N., Hansen, N., Liang, J. J., Deb, K., Chen, Y.-P., Auger, A., et al. (2005). Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization. *KanGAL Rep. 2005005* 2005.

Tsai, C.-W., Lai, C.-F., Chiang, M.-C., and Yang, L. T. (2013). Data mining for internet of things: A survey. *IEEE Commun. Surveys Tutorials* 16, 77–97. doi: 10.1109/SURV.2013.103013.00206

van Heteren, M. (2022). Link adaptation and equalization for underwater acoustic communication using machine learning.

Wang, Z., Chen, F., Yu, H., and Shan, Z. (2021). Sparse decision feedback equalization for underwater acoustic channel based on minimum symbol error rate. *Int. J. Naval Architecture Ocean Eng.* 13, 617–627. doi: 10.1016/j.ijnaoe.2021.07.004

Wang, T., Wen, C.-K., Wang, H., Gao, F., Jiang, T., and Jin, S. (2017). Deep learning for wireless physical layer: Opportunities and challenges. *China Commun.* 14, 92–111. doi: 10.1109/CC.2017.8233654

Wei, G., Yizhen, J., Xiao, H., Xiao, Z., and Wentao, T. (2023). Robust equalization for single-carrier underwater acoustic communication in sparse impulsive interference environment. *Appl. Acoustics* 214, 109706. doi: 10.1016/j.apacoust.2023.109706

Xi, J., Yan, S., Xu, L., and Hou, C. (2020). Sparsity-aware adaptive turbo equalization for underwater acoustic communications in the mariana trench. *IEEE J. Oceanic Eng.* 46, 338–351. doi: 10.1109/JOE.2020.2982808

Xi, J., Yan, S., Xu, L., Zhang, Z., and Zeng, D. (2019). Frequency–time domain turbo equalization for underwater acoustic communications. *IEEE J. Oceanic Eng.* 45, 665–679. doi: 10.1109/JOE.2019.2891171

Xiao, Y., and Dong, Y. (2015). Instantaneous gradient based dual mode wavelet neural network blind equalization for underwater acoustic channel. *Appl. Mathematics Inf. Sci.* 9, 1467. doi: 10.12785/amis/090341

Xiao, Y., Dong, Y., and Li, Z. (2008). "Blind equalization in underwater acoustic communication by recurrent neural network with bias unit," in *2008 7th World Congress on Intelligent Control and Automation* (Chongqing, China: IEEE), 2407–2410.

Xue, J., and Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst. Sci. control Eng.* 8, 22–34. doi: 10.1080/21642583.2019.1708830

Xue, J., and Shen, B. (2023). Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. *J. Supercomputing* 79, 7305–7336. doi: 10.1007/s11227-022-04959-6

Yang, R., Yang, L., Zhang, J., Sun, C., Cong, W., and Zhu, S. (2018). "Blind equalization of qam signals via extreme learning machine," in *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)* (Xiamen, China: IEEE), 34–39.

Zhang, Y., Li, C., Wang, H., Wang, J., Yang, F., and Meriaudeau, F. (2022). Deep learning aided ofdm receiver for underwater acoustic communications. *Appl. Acoustics* 187, 108515. doi: 10.1016/j.apacoust.2021.108515

Zhang, Y., Li, J., Zakharov, Y. V., Li, J., Li, Y., Lin, C., et al. (2019b). Deep learning based single carrier communications over time-varying underwater acoustic channel. *IEEE Access* 7, 38420–38430. doi: 10.1109/ACCESS.2019.2906424

Zhang, X.-L., and Wang, D. (2016). A deep ensemble learning method for monaural speech separation. *IEEE/ACM Trans. audio speech Lang. Process.* 24, 967–977. doi: 10.1109/TASLP.2016.2536478

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver sas imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, L., and Yang, L.-L. (2020). Machine learning for joint channel equalization and signal detection. *Mach. Learn. Future Wireless Commun.*, 213–241. doi: 10.1002/9781119562306.ch12

Zhang, G., Yang, L., Chen, L., Zhao, B., Li, Y., and Wei, W. (2019a). "Blind equalization algorithm for underwater acoustic channel based on support vector regression," in *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)* (IHMSC) (Hangzhou, China: IEEE), Vol. 2. 163–166.

Zhang, Y., Zakharov, Y. V., and Li, J. (2018). Soft-decision-driven sparse channel estimation and turbo equalization for mimo underwater acoustic communications. *IEEE Access* 6, 4955–4973. doi: 10.1109/ACCESS.2018.2794455

Zhao, H., Zeng, X., Zhang, J., and Li, T. (2010). Nonlinear adaptive equalizer using a pipelined decision feedback recurrent neural network in communication systems. *IEEE Trans. Commun.* 58, 2193–2198. doi: 10.1109/TCOMM.2010.08.080612

Zhao, H., Zeng, X., Zhang, X., Zhang, J., Liu, Y., and Wei, T. (2011). An adaptive decision feedback equalizer based on the combination of the fir and flnn. *Digital Signal Process.* 21, 679–689. doi: 10.1016/j.dsp.2011.05.004

Zhao, H., and Zhang, J. (2008). Functional link neural network cascaded with chebyshev orthogonal polynomial for nonlinear channel equalization. *Signal Process.* 88, 1946–1957. doi: 10.1016/j.sigpro.2008.01.029

Zhou, M., Zhang, H., Lv, T., Li, H., Xiang, D., Huang, S., et al. (2022). "Underwater acoustic channel modeling under different shallow seabed topography and sediment environment," in *OCEANS 2022Chennai* (Chennai, India: IEEE), 1–7.

Zhu, J., Song, Y., Jiang, N., Xie, Z., Fan, C., and Huang, X. (2023). Enhanced doppler resolution and sidelobe suppression performance for golay complementary waveforms. *Remote Sens.* 15, 2452. doi: 10.3390/rs15092452

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 January 2024
DOI 10.3389/fmars.2023.1341198

OPEN ACCESS

EDITED BY
Arata Kaneko,
Hiroshima University, Japan

REVIEWED BY
Xiao-Hua Zhu,
Ministry of Natural Resources, China
Minmo Chen,
Harbin Engineering University, China

*CORRESPONDENCE
Juan Liu
liujuan22@nudt.edu.cn

RECEIVED 20 November 2023
ACCEPTED 29 December 2023
PUBLISHED 17 January 2024

CITATION
Guo W, Liu J, Xu G, Li G and Xu P (2024) Long-term statistics and wind dependence of near-bottom and deep-sea ambient noise in the northwest South China Sea. *Front. Mar. Sci.* 10:1341198. doi: 10.3389/fmars.2023.1341198

Long-term statistics and wind dependence of near-bottom and deep-sea ambient noise in the northwest South China Sea

Wei Guo[1], Juan Liu[1]*, Guojun Xu[1], Guangming Li[2] and Pan Xu[1]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China, [2]Cutting-edge Technology Research Center, National Innovation Institute of Defense Technology, Beijing, China

Research on ocean ambient noise is highly important for environment monitoring, marine mammal protection, underwater communication and navigation. In this paper, we present the long-term statistics and wind dependence of near-bottom and deep-sea ambient noise in the northwest South China Sea, at a depth of 1240 m. The data were collected from 11th July 2022 to 31st December 2022 together with local wind speeds ranging from 1 to 58 knots (two typhoons involved), and the processing frequency band is between 20 and 2000 Hz. The long-term mean noise level is calculated along with its skewness, kurtosis and percentile distributions. Diurnal and monthly average of noise levels are analyzed, and the large fluctuations in lower ($\leq$100 Hz) and higher ($\geq$400 Hz) frequencies are respectively caused by the variation of the number of nearby and distant ships and the diverse distributions of the windspeeds in individual months. We find that the noise level in winter (Dec.) is 10~11 dB higher than that in summer (Jul.) at higher frequencies. The probability densities of noise levels in the situation of a fixed wind speed are likely to obey the Burr distributions in low frequencies (50 and 100 Hz) and the Weibull distributions in high frequencies (400 and 1000 Hz). In addition, the mean noise levels for different Beaufort scales match well with the 5-dB-addtion Wenz curves, and a mathematic relationship is acquired between the noise level and wind speed in the experimental site. The results are of great representativeness, and are significant to data-driven noise modelling, evaluation and improvement of sonar performance in the region of South China Sea with an incomplete deep-water sound channel.

KEYWORDS

Northwest South China Sea, near-bottom and deep-sea ambient noise observation, long-term statistics, monthly variations, wind dependence

1 Introduction

Ocean ambient noise has been studied and measured for several decades, and a considerable amount of research has been performed to analyze the statistics and characteristics of ambient noise. Most notable is the approximate empirical law proposed by (Wenz, 1962) that ambient noise is a composite of multiple components, generated by distinct types of sources. These noise sources vary in origin, frequency and time period (Hildebrand, 2009). In general, earthquake activity and surface wave interactions dominate ocean ambient noise at the frequencies less than 10 Hz, anthropogenic noise from ships and natural noise driven by wind compete in the frequency band from 50 to 500 Hz. Wind-driven noise dominates between 500 Hz to 25 kHz, and its generation is due to the interaction of the wind with the sea surface (Deane and Stokes, 2002). Earlier studies have pointed out that the level of wind-driven noise is locally dependent on the wind speed in the vicinity of the receiver (Wenz, 1962; Urick, 1983). Additionally, marine animals (Širović et al., 2013; Romagosa et al., 2017) and rainfall events (Medwin et al., 1992) are also contributors of ocean ambient noise, and the generated noise possesses seasonal and occurrent features.

Long-term statistics of ocean ambient noise have been investigated at plenty of locations in the global ocean, ranging from tropical/subtropical Pacific region (Farrokhrooz et al., 2017; Niu et al., 2021; Yang et al., 2023) and South China Sea (Da et al., 2014; Jiang et al., 2017; Shi et al., 2019), and temperate North Pacific region (McDonald et al., 2006; Seger et al., 2015; Schwock and Abadi, 2021), to polar region (Chen and Schmidt, 2017; Bonnel et al., 2021; Mo et al., 2023). The motivation of these previous works is to acquire the diel, monthly or seasonal variations of ambient noise in different areas, and describe the relationship between noise and meteorological data at the recording locations. It has been proven that the environmental conditions and regions in the global ocean can both affect the results of ambient noise level and corresponding spectral characteristics, e.g., diverse wind speeds in regions with different bathymetry profiles and sound speed distributions lead to deviation of wind-driven noise (Barclay and Lin, 2019). As a consequence, it becomes necessary to choose a representative region to analyze the characteristics of ambient noise. The northwest South China Sea is a typical region in which there are significant shipping lanes, incomplete sound channels in deep water (usually indicating a deep-water environment having a sound channel axis and meanwhile the sound speed at water-seabed interface is lower than that at water-air interface), dramatically varying bathymetry profiles and frequent oceanic/atmospheric dynamic processes. But up to the present, there are few studies on long-term statistics and characteristics of ambient noise in this region. The existing long-term observations in South China Sea concerns mostly the noise recorded by hydrophones deployed at the depths above or near the sound channel axis (Jiang et al., 2017; Shi et al., 2019). The near-bottom noise is studied yet in the shallow water region (water depth<200 m) (Da et al., 2014). It is thus valuable to acquire the data and analyze the characteristics of near-bottom ambient noise in the deep-water region of South China Sea having an incomplete deep-water sound channel, since it could provide extra valuable information for subaqueous activities that rely on sound for detection, communication, navigation and echolocation.

In this paper, we investigate the long-term statistics and wind-dependent characteristics of ambient noise using data collected from a noise measurement experiment conducted at a near-bottom deep-water site in the northwest South China Sea. The autonomous, bottom-mounted acoustic recording system was deployed at the deep-water site in the northwest South China Sea at 17°08′N 112°02′E with water depth of 1240 m (marked as the red solid circle in Figure 1). The recorder (hydrophone) was moored near sea bottom with sampling frequency at 4000 Hz and effective bandwidth from

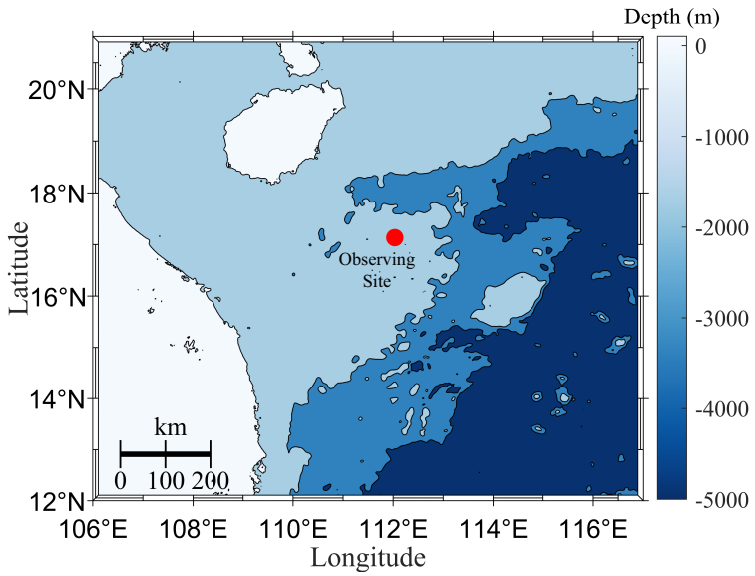


FIGURE 1
The experimental site in the northwest South China Sea. The red solid circle gives the observing site.

20 to 2000 Hz. The ambient noise data was collected for 5 min every one hour from 11th July 2022 to 31st December 2022. With anomalous, incomplete and pile-driving noise data excluded, the selected dataset consists of 19090 time samples, each of 1 min. The synchronous 10-m wind speed data was provided by Ventusky[1] at 7 km resolution in the observing area and at 3h intervals throughout the experimental time period. Monthly variations of noise levels are analyzed. Additionally, distributions of noise levels at a fixed wind speed are discussed, and average noise levels for different wind speed groups (Beaufort scales) are presented, along with the comparison with Wenz curves. The mathematic expression of wind dependence on noise level is acquired, which can be used to predict wind-driven noise levels near the experimental area in South China Sea. The paper's main contributions are twofold. First, we take the lead in measuring and analyzing the statistics of near-bottom ambient noise in the region of South China Sea with an incomplete deep-water sound channel, giving the detailed results about the spectrogram, diurnal and monthly averages of noise levels. Further, we provide the expression of wind-driven noise levels in the experimental area through regression. These results have representativeness for other incomplete deep-water areas in South China Sea, and can provide guidance on analysis of seasonal and yearly variations of ocean ambient noise in South China Sea, data-driven noise modelling, communications, sonar designing and ocean tomography.

2 Long-term statistics of ambient noise

Figure 2 is the spectrogram of the dataset. The processed frequency band is from 20 to 2000 Hz. The raw data for any one hour (5 min) is processed using the Welch's method with a 10s-Hamming window, and a 50%-overlapping FFT algorithm is introduced, leading to 59 periodograms with frequency resolution given by 0.1 Hz. These resulting periodograms are then averaged and converted to decibels (dB) in the 1/3-octave bandwidth. Here the 1/3-octave frequency band from 20 to 2000 Hz is divided into 21 1/3-octave frequency bands with center frequencies of 20, 25, 31.5, 40, 50, 63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600 and 2000 Hz. Figure 2 is acquired when repeating the process mentioned above for every 1h data segment until the entire dataset is processed. The 10-m wind speeds (in red solid line, knots) during the experimental period, ranging from 1 to 58 knots, are also overlaid to show the correlation with noise levels. In addition, Figure 2 indicates clearly the effect of the typhoon 'Nesat' and 'Nalgae' occurred from 16th Oct. to 19th Oct. and from 29th Oct. to 2nd Nov., resp., together with that of a series of strong and rapidly varied wind speeds occurred from 14th Dec. to 18th Dec. It can be roughly seen in Figure 2 that wind speeds and ambient noise levels are very well correlated in the higher frequency band $f \in [400, 2000]$ Hz, i.e., higher wind speed correlates with higher noise intensity. Figure 3A shows the detailed spectrogram during the typhoons 'Nesat' and 'Nalgae', and Figures 3B, C give respectively the relevant variations of ambient noise levels at 100 Hz, 400 Hz and 1000 Hz versus wind speeds. We can see that except for some coincidently abnormal peaks for all the three frequencies, trends of the noise levels at 400 Hz and 1000 Hz basically coincide with wind speeds. Figures 2, 3 indicate there is no obvious correlation between wind speeds and ambient noise levels in the lower frequency band $f \in [20, 100]$ Hz. The specific correlation coefficient between ambient noise and wind speeds will be demonstrated in Sec. 3. By the way, it is noticed that the ambient noise levels during November are much higher than other months in the frequency band from 20 to 200 Hz. It is possibly due to higher shipping noise during that period. Considering that shipping noise depends mainly on the kinds, number, distribution, speeds and tracks of nearby ships and sound propagation, analysis on shipping noise is quite challenging and complicated. Since we focus on the characteristics and wind-dependence of ocean ambient noise, we will not discuss the effect of shipping noise in this paper. More details about shipping noise to interpret the features of low-frequency noise will be studied in future work.

The long-term statistics of the ambient noise levels are shown in Figure 4. The mean, standard deviation (STD), 5%, 50% (median) and 95% percentiles, which describe the average value and heterogeneity of the data, are shown in Figure 4A. The maximum and minimum of the mean ambient noise level are 88.5 and 57.9 dB re 1 μPa^2/Hz occurring at 31.5 Hz and 1600 Hz, resp. The mean value is larger than the median value in the frequency band $f \in [20, 125]$ Hz, while the median value prevails from 160 to 2000 Hz. The absolute differences between the mean values and median values are not greater than 1.3 dB re 1 μPa^2/Hz. The 5% and 95% percentiles represent the lower- and upper-limits of the overwhelming majority of the data. Figure 4B shows skewness, kurtosis and standard deviation for the entire noise levels. Skewness represents tailedness of the distribution and kurtosis represents peakedness giving information about outliers. Distant shipping leads to positive skewness ($f \in [20, 200]$ Hz) while wind-generated noise leads to negative skewness ($f \in [400, 2000]$ Hz). In general, it shows that the higher mean values are accompanied by higher skewness and kurtosis. But abnormities emerge in the frequency band $f \in [63.5, 125]$ Hz, in which the mean values decrease steadily while the values of skewness and kurtosis show abrupt peaks. The abnormal higher skewness and kurtosis suggest that much of the variance is the result of the competition between wind-generated noise and shiping noise with a wide fluctuation range of windspeeds and number of ships. Considering that the STD exhibits local minima (3.3 to 3.6 dB re 1 μPa^2/Hz) in this frequency band, we deduce that the probability density of the noise levels obeys a centralized and non-normal distribution.

In certain area, ocean ambient noise has significant diurnal variability (Klusek and Lisimenka, 2016). Here we focus on the diurnal variation of ambient noise in the experimental area. Figure 5 shows the day and night comparison of ambient noise level. There is only a small difference between the mean day and mean night values in the frequency band from 20 to 2000 Hz. The maximum deviation is 0.72 dB re 1 μPa^2/Hz (at 1600 Hz). At the frequency band between 20 and 500 Hz, the deviation is not more than 0.23 dB re 1 μPa^2/Hz. The STD during the day is larger from 80 to 2000 Hz, and is less in

1 Ventusky. Czech Republic. www.ventusky.com.

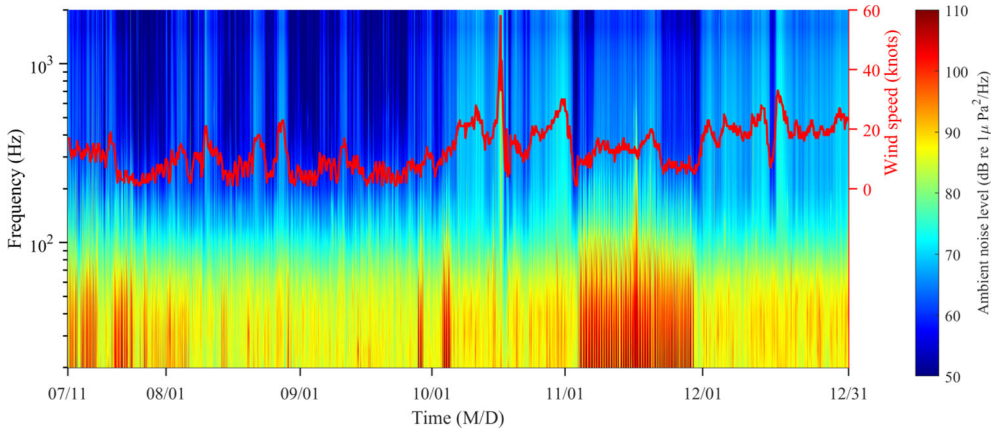


FIGURE 2
Spectrogram of the ambient noise from 11/07/2022 to 31/12/2022 with wind speed (red solid line, in knots). The frequencies range from 20 to 2000 Hz, and the color bar value represents the noise level in dB re 1 μPa²/Hz. M/D: Month/Day.

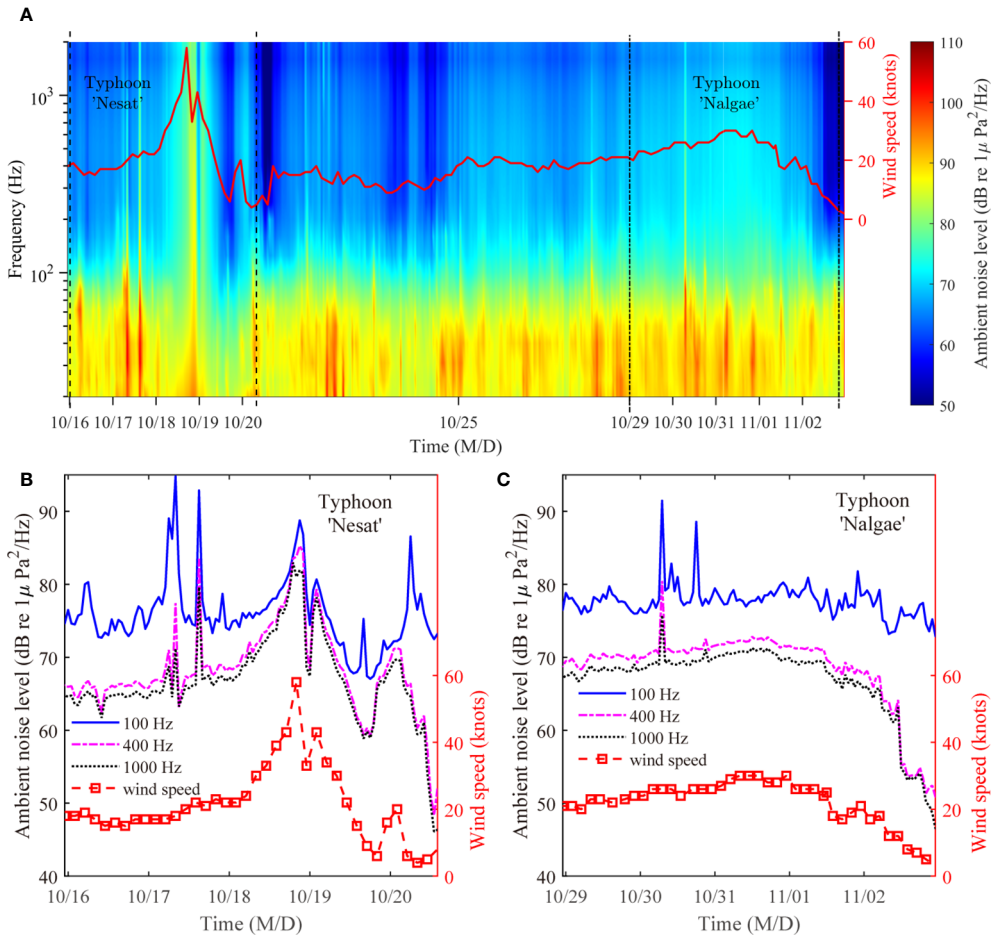


FIGURE 3
(A) Spectrogram of the ambient noise during the typhoons 'Nesat' and 'Nalgae' with wind speed (red solid line, in knots). Ambient noise levels at 100 Hz (blue solid line), 400 Hz (magenta dash-dotted line) and 1000 Hz (black dotted line) during the typhoons: **(B)** 'Nesat' and **(C)** 'Nalgae' with wind speed (red squares). M/D: Month/Day.

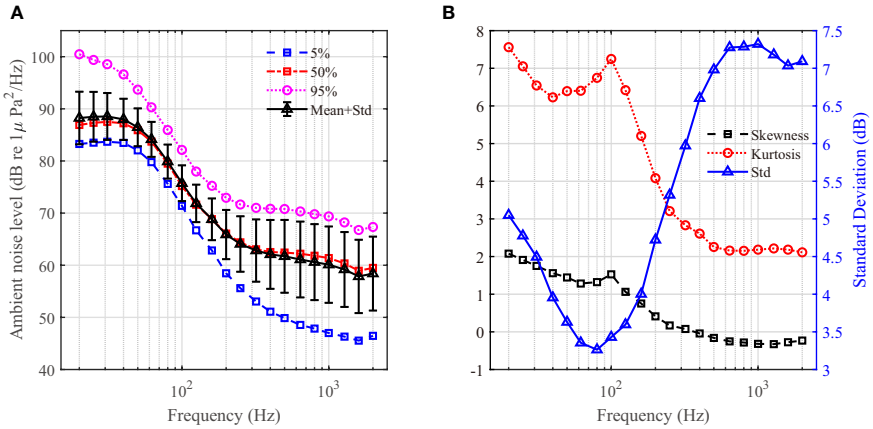


FIGURE 4
The long-term statistics of ambient noise levels: **(A)** 5th, 50th and 95th percentiles, mean and standard deviation. **(B)** Skewness, kurtosis, and standard deviation.

the remaining frequency band. There is thus no remarkable difference between the mean day and mean night noise levels.

Figure 6 shows the monthly average ambient noise levels varied over time. The monthly variation is relatively large (approximately 6 dB re 1 μPa^2/Hz at 200 Hz and 12 dB re 1 μPa^2/Hz at 20 and 2000 Hz). Due to the noises of distant shipping, the levels at frequencies below 100 Hz are higher in all months, and meanwhile, the average level for November is 7~12 dB re 1 μPa^2/Hz greater than those for other months, which is shown in Figure 2 as well. In this case nearby ships play a nonnegligible role in the noise level. Noise levels at frequencies above 200 Hz are relatively lower (<70 dB re 1 μPa^2/Hz), which are dominated mainly by wind and waves. The spread of noise levels in this frequency band is nearly 11 dB re 1 μPa^2/Hz, which is principally caused by the diverse distributions of the wind speeds in individual months. The monthly average wind speeds are given in Figure 6 as well, which indicates clearly that the higher monthly average wind speed corresponds to the higher monthly average noise level at frequencies above 400 Hz. In addition, it can be seen that at higher frequencies (≥400 Hz), the noise level during December (in winter) is highest, and is 10~11 dB higher than that during July (in summer). This phenomenon coincides with the conclusion drawn in (Niu et al., 2021), and the reason is that wind speeds are higher in winter, which is consistent with the variation in wind speeds in the experimental area (see Figure 2).

3 Wind-dependent characteristics of ambient noise

In this section, we investigate the dependence of the ambient noise in the experimental site on wind speeds. Figure 7 shows the

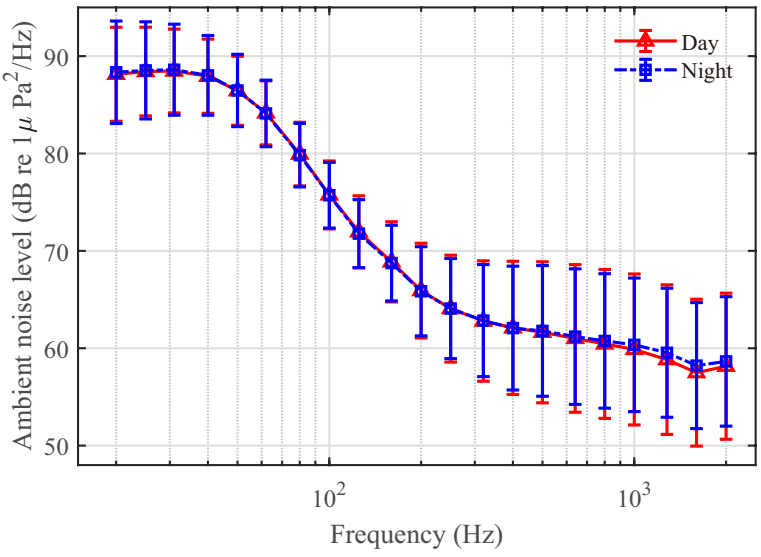


FIGURE 5
Day and night comparison of mean value and standard deviation of ambient noise levels.

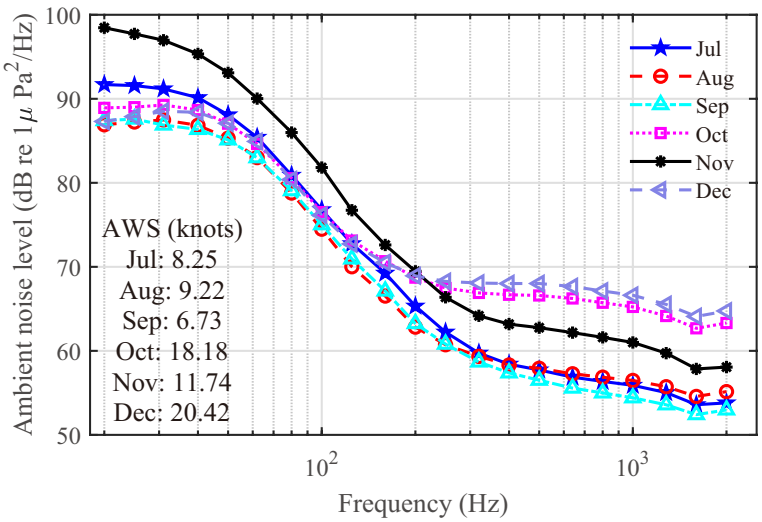


FIGURE 6
Average monthly ambient noise levels from 20 Hz to 2000 Hz together with the monthly average wind speeds (AWS) in knots. Different colors indicate individual months.

probability histogram of 10-m wind speeds during the entire period of acoustic recordings. The wind speeds are given in units of knots and with 3h intervals. It can be seen that over half of the wind speeds are within the interval from 7 to 21 knots, corresponding to Beaufort scales from 3 to 5. We then analyze the distribution of ambient noise levels when the wind speed is fixed. Figure 8 show the probability histograms of noise levels at 50Hz, 100 Hz, 400 Hz and 1000 Hz, reps., when the wind speed is 12 knots. The red solid lines and black dashed lines represent respectively the Burr distributions and Weibull distributions fitted to the noise levels at four frequencies. For the cases of 50 Hz and 100 Hz, the noise levels are dominated by ships rather than the wind speed. It is shown in Figures 8A, B that the spreads of the noise levels are approximately 23~24 dB re 1 μPa^2/Hz. The large spreads of the noise levels result from the variation of the number of nearby and distant ships. Besides, the probability density distributions appear to match the Burr distribution better. While for the cases of 400 Hz and 1000 Hz, wind speeds play a dominant role in noise levels, Figures 8C, D show that the spreads of the noise levels are approximately 12~14 dB re 1 μPa^2/Hz, and the probability density distribution is likely to match the Weibull distribution better. The distributions of noise levels at different frequencies can be applied to simulating non-Gaussian and non-white spectra and time series of ambient noise, and provide effective assistance for accurately predicting the

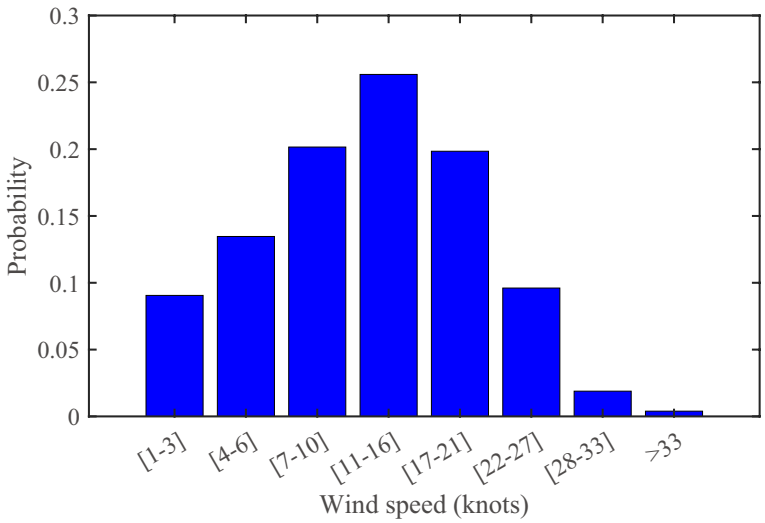


FIGURE 7
Probability distribution of wind speeds during the entire period of acoustic recordings.

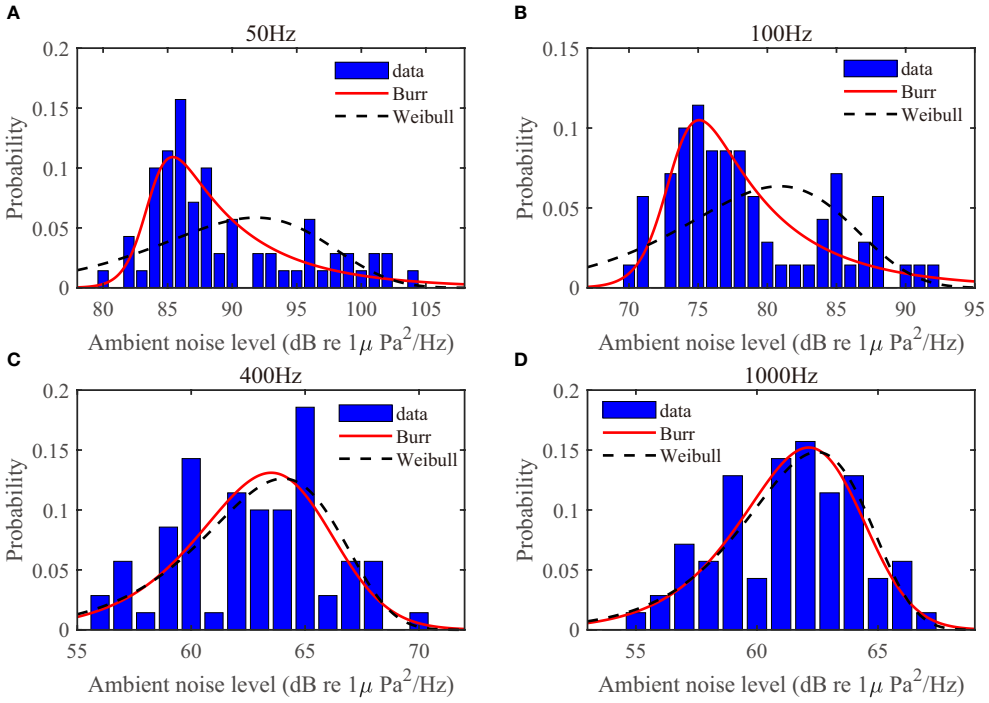


FIGURE 8
Probability distribution of ambient noise levels at **(A)** 50Hz, **(B)** 100 Hz, **(C)** 400 Hz and **(D)** 1000 Hz with fixed wind speed at 12 knots.

capacity of underwater communications or performance of detection.

Figure 9 shows the mean ambient noise levels for multiple Beaufort scales from 2 to 8. It can be seen that the wind dependence of the noise level is weak in the lower frequency region (≤100 Hz), and tends to be noticeable at frequencies larger than 200 Hz, where the noise level increases with the Beaufort scale. The increments of noise levels between adjacent Beaufort scales are

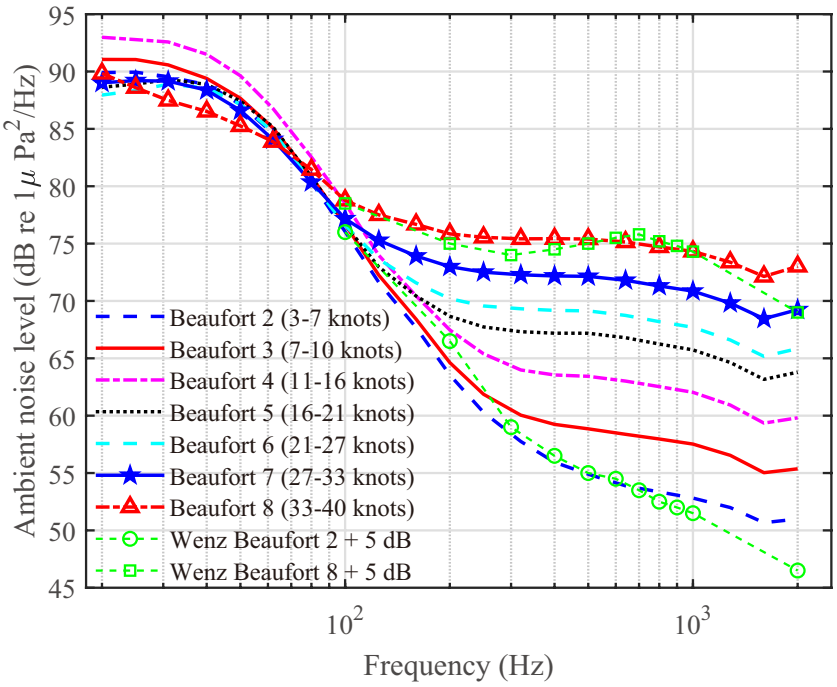


FIGURE 9
Average noise levels from 20 Hz to 2000 Hz for the Beaufort scales ranging from 2 to 8 compared with the 5-dB-addition Wenz curves (green marked lines).

nearly 2 dB re 1 μPa^2/Hz at 200 Hz and 3.5 dB re 1 μPa^2/Hz at 1000 Hz. A comparison of data from (Wenz, 1962) and this work is also presented. The case of comparison here is specified as in Figure 2B of (Wenz, 1962), representing noise levels in deep ocean. The Wenz curves are converted to the units used throughout this paper, and are exhibited as the green dashed lines with circles (Beaufort 2) and with squares (Beaufort 8) with an extra 5 dB re 1 μPa^2/Hz added on the original noise levels. It is known that the received level of wind-driven noise is determined by two factors. One is the noise source spectral level which should be only dependent on wind speeds. The other is the environmental conditions, i.e., sound speed profiles, bathymetry, etc., which influence the propagation of wave generated by noise sources. Thus, in different observing regions there should be diverse differences between the observing noise levels and Wenz curve. Recent work has revealed that the mean noise levels acquired at 6 locations of the Pacific Ocean at different Beaufort scales are 7-dB higher than Wenz curve(Yang et al., 2023). Here the additional 5 dB re 1 μPa^2/Hz can be regarded as the correction caused by the difference of environmental conditions between two observing areas. It is apparent that the presented data and the 5-dB-added Wenz curves are in excellent agreement.

It has been found by many authors (Crouch and Burt, 2005; Klusek and Lisimenka, 2016) that the noise level and wind speed can be linked by the following expression

$$NL(f) = 10 \cdot n(f)\log_{10}U + M(f), \quad (1)$$

where NL is the measured noise level at the central frequency of the 1/3-octave frequency band in dB re 1 μPa^2/Hz, U is the 10-m wind speed in knots, and f represents the central frequency in the 1/3-octave frequency band. n and M are the wind-dependence coefficient and offset, resp., and they are both dependent on the frequency.

The correlation coefficients of the noise levels and 10-m wind speeds of the presented dataset are analyzed here to inspect the dependence. The Pearson correlation coefficients $r(f)$ are written as

$$r(f) = \frac{\mathrm{cov}(NL,\ U)}{\sigma_{NL} \cdot \sigma_U} = \frac{\sum_{i=1}^{N}(NL_i - \overline{NL})(\log_{10}(U_i) - \overline{\log_{10} U})}{\sqrt{\sum_{i=1}^{N}(NL_i - \overline{NL})^2} \cdot \sqrt{\sum_{i=1}^{N}(\log_{10}(U_i) - \overline{\log_{10} U})^2}} \quad (2)$$

where cov (,) and σ represent the covariance and STD, resp. $\overline{(\cdot)}$ represents the mean value. Figure 10 shows the errorbar of $r(f)$ in the frequency band from 20 to 2000 Hz, which is computed from Equation (2). For the frequency between 20 and 160 Hz, ambient noise is dominated by shiping noise, the correlation coefficient is less than 0.4. The noise level is weakly correlated with the (logarithm of) wind speed. In the frequency band from 200 to 315 Hz, the correlation coefficient increases gradually, $r(f) \in (0.4,\ 0.8)$, indicating that the impact of 10-m wind speed on the noise level are non-negligible. In the frequency band from 400 to 2000 Hz, the correlation coefficient is above 0.8, indicating a strong correlation between the noise level and wind speed. It means that the noise level is dominantly determined by the wind speed in this frequency band, which can be verified in Figure 2 as well. The STD ranges from 0.020 to 0.033, which is quite small and corroborates the previous conclusion that wind dominates the ocean ambient noise between 400 and 2000 Hz. The wind-dependence coefficient n and the offset M are then computed in the frequency band from 400 to 2000 Hz using a LMSE regression. These results are listed in Table 1 together with the correlation coefficients. These correlation coefficients $r(f)$ are identical with the mean values in Figure 10. Since the correlation coefficients are relatively small from 20 to

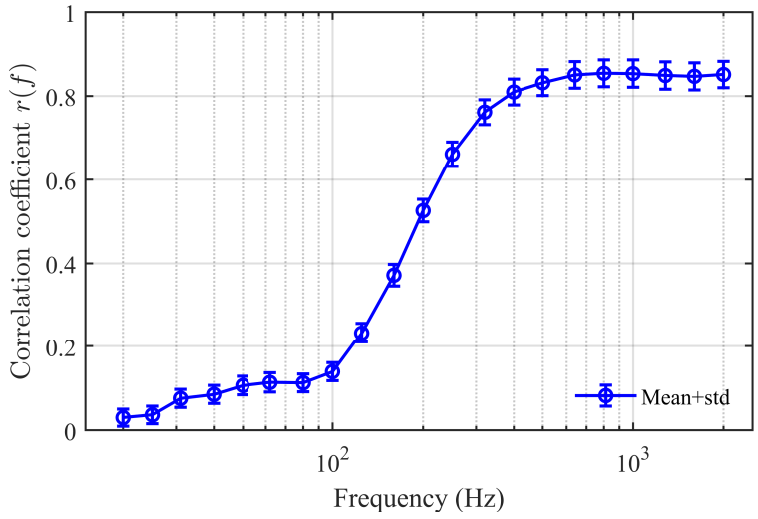


FIGURE 10
Errorbar (mean and standard deviation) of correlation coefficients between noise levels and wind speeds. The frequency band ranges from 20 Hz to 2000 Hz.

TABLE 1 Noise-wind speed dependence coefficients and correlation coefficients as a function of the center frequencies in 1/3-octave frequency from 20 to 2000 Hz.

Frequency (Hz)	$n(f)$	$M(f)$	$r(f)$	Frequency (Hz)	$n(f)$	$M(f)$	$r(f)$
20	/	/	0.03	250	/	/	0.66
25	/	/	0.04	315	/	/	0.76
31.5	/	/	0.08	400	1.64	45.53	0.81
40	/	/	0.09	500	1.80	43.64	0.83
50	/	/	0.11	630	1.89	41.97	0.85
63	/	/	0.11	800	1.90	41.40	0.85
80	/	/	0.11	1000	1.91	40.85	0.85
100	/	/	0.14	1250	1.86	40.32	0.85
125	/	/	0.23	1600	1.82	39.27	0.85
160	/	/	0.37	2000	1.86	39.43	0.85
200	/	/	0.53				

315 Hz, it is not necessary to analyze the wind-dependence coefficient and the offset. We observe an approximately quadratic wind dependence in the frequency band from 500 to 2000 Hz. These results can be adopted to predict the noise level in the area surrounding the experimental site employing Equation (1).

4 Conclusions

This paper reports on the ambient noise measured by a near-bottom acoustic recording system deployed at the deep-water site in the northwest South China Sea, together with the synchronous 10-m wind speeds ranging from 1 to 58 knots, covering the effect of two typhoons passing by the experiment site, during the latter half of 2022. To our knowledge, this study was the first to report on the near-bottom ambient noise properties in the region of South China Sea with an incomplete deep-sea sound channel. The collected noise data provides significant information about the long-term statistics of ambient noise in this area. In the absence of transient events, the noise levels at low frequencies range between 58~100 dB re 1 μPa^2/Hz (20 to 200 Hz) and 51~72 dB re 1 μPa^2/Hz (200 to 400 Hz). In the high frequency range (400 to 2000 Hz), which is typically dominated by wind, the noise levels range from 45 to 70 dB re 1 μPa^2/Hz. No significant difference is observed between mean day and mean night noise levels, but there are great monthly variations (7~12 dB re 1 μPa^2/Hz) in the mean levels from 20 to 200 Hz, revealing a wide varying range of the number of nearby and distant ships. At the frequencies larger than 400 Hz, the monthly variations in the mean levels are mainly caused by the differences among the average wind speeds in individual months, and the noise is 10~11 dB higher in winter (December) than in summer (July). Besides, in the situation of fixed wind speed (12 knots), we have observed that the probability densities of noise levels obey the Burr distribution at lower frequencies (50 and 100 Hz), and are likely to obey Weibull distribution at higher frequencies (400 and 1000 Hz). These distributions of noise levels can be employed to simulate the spectra and time series of the non-Gaussian and non-white ambient noise. Ambient noise levels exhibit continual increase with increasing wind speed group or Beaufort scale for frequencies between 200 to 2000 Hz, and the noise levels for different Beaufort scales are in excellent agreement with the Wenz curves with a 5-dB addition. We analyze the wind-dependence coefficient and offset of noise levels, and acquire a mathematic expression of wind dependence from 400 to 2000 Hz. The present observing site is at a depth of over 1200 m, corresponding to the deep ocean with an incomplete sound channel. In the northwest South China Sea, most of areas except coastal regions are at the depth between 1000 and 3000 m, thus the present data can at least cover most of the northwest part of South China Sea, and the results can be employed to predict noise levels in the experimental area. We hope the presented results could be utilized to discuss the trends in the ambient noise statistics and characteristics in relation to changes in ocean traffic and meteorological condition in South China Sea.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WG: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft. JL: Formal analysis, Funding acquisition, Supervision, Validation, Writing – review & editing. GX: Data curation, Formal analysis, Funding

acquisition, Supervision, Validation, Writing – review & editing. GL: Formal Analysis, Writing – review & editing. PX: Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the China Postdoctoral Science Foundation (Grant No. 2023M734311); and National Key Research and Development Program of China (Grant No. 2022YFC3103404).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Barclay, D. R., and Lin, Y.-T. (2019). Three-dimensional ambient noise modeling in a submarine canyon. *J. Acoust. Soc Am.* 146 (3), 1956–1967. doi: 10.1121/1.5125589

Bonnel, J., Kinda, G. B., and Zitterbart, D. P. (2021). Low-frequency ocean ambient noise on the Chukchi Shelf in the changing Arctica. *J. Acoust. Soc Am.* 149 (6), 4061–4072. doi: 10.1121/10.0005135

Chen, R., and Schmidt, H. (2017). Characterization of arctic ambient noise environment. *J. Acoust. Soc Am.* 141, 3532–3532. doi: 10.1121/1.4987457

Crouch, W. W., and Burt, P. J. (2005). The logarithmic dependence of surface-generated ambient-sea-noise spectrum level on wind speed. *J. Acoust. Soc Am.* 51 (3B), 1066–1072. doi: 10.1121/1.1912926

Da, L., Wang, C., Han, M., and Zhang, L. (2014). Ambient noise spectral properties in the north area of Xisha. *Acta Oceanol. Sin.* 33 (12), 206–211. doi: 10.1007/s13131-014-0569-4

Deane, G. B., and Stokes, M. D. (2002). Scale dependence of bubble creation mechanisms in breaking waves. *Nature* 418 (6900), 839–844. doi: 10.1038/nature00967

Farrokhrooz, M., Wage, K. E., Dzieciuch, M. A., and Worcester, P. F. (2017). Vertical line array measurements of ambient noise in the North Pacific. *J. Acoust. Soc Am.* 141 (3), 1571–1581. doi: 10.1121/1.4976706

Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. *Mar. Ecol. Prog. Ser.* 395, 5–20. doi: 10.3354/meps08353

Jiang, D., Li, Z., Qin, J., Peng, Z., and Shao, H. (2017). Characterization and modeling of wind-dominated ambient noise in South China Sea. *Sci. China Phys. Mech. Astron.* 60 (12), 124321. doi: 10.1007/s11433-017-9088-5

Klusek, Z., and Lisimenka, A. (2016). Seasonal and diel variability of the underwater noise in the Baltic Sea. *J. Acoust. Soc Am.* 139 (4), 1537–1547. doi: 10.1121/1.4944875

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (2006). Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California. *J. Acoust. Soc Am.* 120 (2), 711–718. doi: 10.1121/1.2216565

Medwin, H., Nystuen, J. A., Jacobus, P. W., Ostwald, L. H., and Snyder, D. E. (1992). The anatomy of underwater rain noise. *J. Acoust. Soc Am.* 92 (3), 1613–1623. doi: 10.1121/1.403902

Mo, X., Wen, H., Yang, Y., Zhou, H., Yin, J., Han, X., et al. (2023). Ocean ambient noise on the Chukchi Plateau and its environmental correlates. *Mar. Environ. Res.* 188, 106024. doi: 10.1016/j.marenvres.2023.106024

Niu, F., Xue, R., Yang, Y., Chen, B., Ruan, H., and Luo. (2021). Baseline assessment of ocean ambient noise in the western Clarion Clipperton Zone, Pacific Ocean. *Mar. Pollut. Bull.* 173, 113057. doi: 10.1016/j.marpolbul.2021.113057

Romagosa, M., Cascão, I., Merchant, N. D., Lammers, M. O., Giacomello, E., et al. (2017). Underwater ambient noise in a baleen whale migratory habitat off the azores. *Front. Mar. Sci.* 4. doi: 10.3389/fmars.2017.00109

Schwock, F., and Abadi, S. (2021). Statistical analysis and modeling of underwater wind noise at the northeast pacific continental margin. *J. Acoust. Soc Am.* 150 (6), 4166–4177. doi: 10.1121/10.0007463

Seger, K. D., Thode, A. M., Swartz, S. L., and Urbán R., J. (2015). The ambient acoustic environment in Laguna San Ignacio, Baja California Sur, Mexico. *J. Acoust. Soc Am.* 138 (5), 3397–3410. doi: 10.1121/1.4935397

Shi, Y., Yang, Y., Tian, J., Sun, C., Zhao, W., et al. (2019). Long-term ambient noise statistics in the northeast South China Sea. *J. Acoust. Soc Am.* 145 (6), EL501–EL507. doi: 10.1121/1.5110740

Širović, A., Wiggins, S. M., and Oleson, E. M. (2013). Ocean noise in the tropical and subtropical Pacific Ocean. *J. Acoust. Soc Am.* 134 (4), 2681–2689. doi: 10.1121/1.4820884

Urick, R. J. (1983). *Principles of underwater sound* (New York: McGraw-Hill).

Wenz, G. M. (1962). Acoustic ambient noise in the ocean: spectra and sources. *J. Acoust. Soc Am.* 34, 1936–1956. doi: 10.1121/1.1909155

Yang, J., Nystuen, J. A., Riser, S. C., and Thorsos, E. I. (2023). Open ocean ambient noise data in the frequency band of 100 Hz–50 kHz from the Pacific Ocean. *JASA Express Lett.* 3 (3), 036001. doi: 10.1121/10.0017349

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 19 January 2024
DOI 10.3389/fmars.2023.1341989

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Guangming Li,
PLA Academy of Military Sciences, China
Xiao-Hua Zhu,
Ministry of Natural Resources, China

*CORRESPONDENCE
Won-Ki Kim
konekee@chungbuk.ac.kr

RECEIVED 21 November 2023
ACCEPTED 29 December 2023
PUBLISHED 19 January 2024

CITATION
Bae HS and Kim W-K (2024) Measurement of backscattering strength of artificial bubbles in the Southern Sea of the Korean Peninsula. *Front. Mar. Sci.* 10:1341989. doi: 10.3389/fmars.2023.1341989

Measurement of backscattering strength of artificial bubbles in the Southern Sea of the Korean Peninsula

Ho Seuk Bae[1] and Won-Ki Kim[2*]

[1]Maritime Technology Research Institute, Agency for Defense Development, Changwon, Republic of Korea, [2]Department of Earth and Environmental Sciences, Chungbuk National University, Chungju, Republic of Korea

Bubbles greatly affect the propagation of sound waves as the acoustic characteristics of a bubble cluster in water are distinct from those of undisturbed water. Therefore, bubble clusters affect the ability of sensors to detect underwater targets by blocking or reflecting sound waves. Additionally, the bubble wakes produced by ships in movement can be detected by wake-homing torpedoes, thus greatly threatening the safety of the ship. Thus, research on bubble dynamics in water is crucial for the development of military technology. Recently, we conducted a quantitative estimation of the acoustic characteristics of artificial bubbles, including their backscattering strength, existence time, population density spectrum level and void fraction in an ideal water-tank environment. Based on our previous findings, the present study sought to measure the acoustic characteristics of artificial bubbles using an acoustic Doppler current profiler (ADCP) in the southern sea of the Korean Peninsula, which is a real marine environment. Additionally, we validated the ADCP measurements by comparing them to those obtained using a scattering strength measurement system (SSMS) developed by our team. Collectively, our findings provide a basis for the development of military technology, as well as for the study of bubble in water.

KEYWORDS

backscattering strength, artificial bubble, acoustic Doppler current profiler, sea experiment, military technology

1 Introduction

Understanding the characteristics of bubbles naturally generated in real marine environments is a very important topic in the fields of oceanography and underwater acoustics. Therefore, many studies have explored the characteristics of naturally occurring bubbles. Keiffer et al. (1995) measured the surface backscattering and reverberation strengths induced by a bubble layer below the water surface and compared them with

sound wave propagation and scattering models in the low to moderate frequency range (Keiffer et al., 1995). Caruthers et al. (1999) studied bubble distribution by measuring the acoustic attenuation caused by the bubbles generated in rip currents of very shallow waters (Caruthers et al., 1999). Trevorrow (2003) determined the volume scattering strength of a bubble layer below the water surface in the high-frequency band. To compare the acoustic scattering of bubbles caused by natural wind currents to that induced by the sea bottom roughness, Dahl (2003) measured the acoustic backscattering strength of the sea surface. The gas columns that naturally leak from the sea bottom are also often studied because they have a significant impact on the ocean environment (Schneider von Deimling et al., 2007; Gulf of Mexico Expedition, 2011; Schneider von Deimling and Weinrebe, 2014; Urban et al., 2017).

The study of bubbles in water is very important not only because it provides insights into natural phenomena but also for military purposes. Bubbles in water rise to the surface due to their buoyancy and dissipate gradually. However, smaller bubbles remain in the water for several tens of minutes (United States Office of Scientific Research and Development, 1946). Therefore, these small bubbles could potentially be used in military applications. Bubble layers can interfere with acoustic wave propagation due to their strong attenuation properties. Therefore, bubble clusters can hinder underwater target detection by affecting the detection performance of sonar, thus protecting underwater facilities and equipment from potential attacks. Similarly, the noise emitted by a moving ship can be detected by an enemy sonar, which highlights the vital importance of noise reduction. Noise can be attenuated by generating artificial bubbles around the ship. Additionally, given that wake-homing torpedoes, which is a strong threat to the ship, acoustically track the wake generated by moving ships, the small bubbles in the wake that persist in the water for long periods are very unfavorable for the ship's defense. Therefore, the characteristics of artificial bubbles in water have attracted the attention of many military experts and researchers.

The ship wakes generated by propeller rotation and hull resistance are dominant factors that determine the formation of artificial bubbles in the ocean. Therefore, many studies have sought to measure ship wakes using a wide variety of equipment to extract the signatures of the ship. Culver and Trujillo (2007) measured the density and size distribution of the ship wake using an upward single-beam sonar mounted on an autonomous underwater vehicle. Stanic et al. (2009) derived the void fraction and the sound speed of the wake using bi-static sonar. Ji et al. (2009) characterized the spatial distribution of bubbles by measuring their backscattering intensity at several positions under the ship wake using a self-manufactured sonar instrument, and Leighton et al. (2011) acquired the backscattering signal of the wake of a 3,900-ton ship using a wake penetration sonar developed by the research team. Li et al. (2014) acquired spatial scattering images of ship wakes with a multi-beam bathymetric sonar, and Karoui et al. (2015) measured ship wakes using forward-looking sonar, which is often equipped on unmanned vehicles.

In addition to the wake caused by ships, several studies have measured the effects of artificially generated bubble clusters. Rustemeier et al. (2012) collaborated with Atlas Elektronik in Germany to create a bubble curtain by injecting compressed air into lake water, after which they measured the changes in acoustic attenuation by the bubble curtain. To reduce the noise generated by ships, compressed air was sprayed to generate a bubble column in a large cavitation tunnel, after which the insertion loss was measured by Park et al. (2021). Except for the aforementioned studies, very few studies have assessed the use of artificial bubbles as a defense mechanism against wake-homing torpedoes for military purposes (Guelou et al., 2002; Guelou et al., 2005).

The quantitative characterization of the acoustic properties of bubble clusters is known to be quite challenging. Unlike in natural bubble studies, there are physical restrictions that affect the generation of artificial bubbles, thus limiting their study in natural environments. Our team recently estimated the acoustic properties (backscattering strength, existence time, population density spectrum level, and void fractions) of artificial bubble clusters using a bubble-generating material (BGM) designed by our team using an acoustic Doppler current profiler (ADCP) (Bae et al., 2022). This study was conducted in a water-tank (i.e., in stable experimental conditions) to quantify the acoustic characteristics of the generated bubbles, and the results were compared to those obtained with highly reliable commercial equipment. However, the salinity, temperature, density, flow, and concentration of organic and inorganic compounds in seawater are different from those of fresh waters in the tank experiments. Moreover, marine environments exhibit strong spatiotemporal variation compared to that in the water-tank experiments. Therefore, the acoustic characteristics of artificial bubbles in marine environments will likely be different.

The present study sought to expand upon our previous work by characterizing the acoustic characteristics of artificial bubbles in a real marine environment (the southern sea of the Korean Peninsula) using ADCP. To validate the backscattering strength derived from the ADCP results, we simultaneously acquired another dataset using a scattering strength measurement system (SSMS) developed by our research team. Section 2 briefly introduces the measurement devices and methods. Section 3 describes our experimental design in the southern sea of the Korean Peninsula. Section 4 describes our quantitative results and Section 5 summarizes our findings and conclusions.

2 Measurement devices and approach

We measured the backscattering strength of artificial bubble clusters generated in the water-tank using an ADCP in our previous study (Bae et al., 2022). The ADCP can measure the echo intensities reflected from bubble clusters, and received intensities can be converted into backscattering strength by the equation presented by Bae et al. (2022). Here, the acoustic characteristics of artificial bubble clusters were estimated using the same procedures. To validate the ADCP-based estimations, an SSMS (i.e., a stand-alone sonar device) was developed to record the acoustic signals with a high resolution.

Figure 1A shows the ADCP used in our experiment. We selected the Workhorse Sentinel 300 model produced by Teledyne

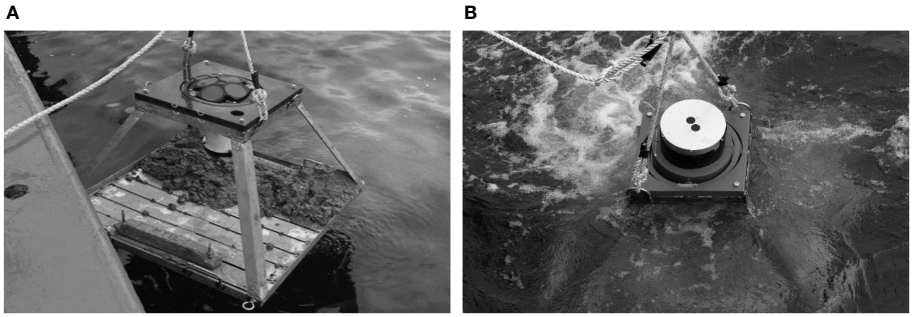


FIGURE 1
Acoustic measurement devices: **(A)** ADCP; **(B)** SSMS.

RD Instruments) because it is widely used in Korea and our group is highly experienced in its use and deployment. Additionally, this equipment was used in our previous water-tank experiment (Bae et al., 2022). Figure 1B shows the SSMS developed by our team, which records acoustic signals to overcome limitations when measuring underwater bubbles with an ADCP. Basically, the method of receiving acoustic signals is the same as ADCP. However, it has a higher resolution than the commercial ADCP. Moreover, it features a vertical direction beam instead of a tilted beam such as that in the ADCP system, which enables it to identify the exact target position without additional slant correction. The acoustic hydrophone and projector were physically separated from each other to reduce electrical noise. This device was designed to use an arbitrary operating frequency range from 450 to 550 kHz to avoid acoustic interferences with other acoustic sonar devices. Afterward, acoustic signals at a 64 kHz sampling frequency were recorded via frequency modulation (FM). This SSMS can transmit the pulse type continuous wave (CW), FM, or arbitrary signals with a length of 0.5–10.0 ms. The acoustic source level varies depending on the operating frequency, but in the frequency used in this study, the magnitude of the source level was 197.0 dB.

The scattering strength induced by the bubble cluster can be easily derived using both sonar instruments. In the case of the ADCP instrument, the manufacturer provides a transform equation to determine the backscattering strength from the echo intensity as a function of depth (Deines, 1999; Mullison, 2017). Scattering strength can also be directly obtained through the sonar equation from the volume scattering theory (Urick, 1983). In this study, scattering strength was estimated using the sonar equation because it can account for the precise transmission loss in the ocean environment. Additionally, it allows for the direct application of the parameters from our SSMS device. From the sonar equation, backscattering strength can be determined as described by Equation (1) (Urick, 1983; Bae et al., 2022):

$$S_v = RL + 2TL - SL - 10logV \quad (1)$$

where the RL is the received level reflected from the bubble cluster and the TL is the transmission loss. We applied Jurng's approach to obtain the TL (Jurng, 1996). SL is the transmitted acoustic pressure level and V is the volume of the acoustic beam.

3 Sea experiment design

3.1 Overview of experiment

To obtain the backscattering strength of the artificial bubbles in a real marine environment, various factors were considered when designing our experimental plan. Figure 2 shows a schematic of our sea experiment. First, the most suitable experimental area to stably deploy the measurement equipment was identified, after which data were acquired in the calmest conditions possible. Moreover, we identified the most appropriate measuring period and time to obtain better quality data by accounting for environmental factors including the weather and a tide. Artificial bubbles were then generated after deploying the measurement devices. A drone was used to avoid interference from other acoustic signals. The drone can perfectly drop our designed BGMs for generating artificial bubbles in a wide area without any interferences. The BGMs were dropped near the sea surface to artificially generate a bubble cluster, after which the backscattered acoustic signals reflected and scattered from the bubbles were recorded with the two measurement devices.

The main purpose of this measurement was to obtain acoustic characteristics related exclusively to the artificial bubbles. Therefore, mock-up models with the same shape, size, and density as the BGMs were first dropped to obtain reference data in the absence of bubbles. Afterward, the acoustic characteristics of the bubbles were measured by dropping the BGMs with the drone.

3.2 Experiment area

Various factors affect the conditions of marine environments, including water depth, water current and tide, fisheries and harbors, and the presence of restricted zones. Therefore, to identify the optimal conditions for our experiments, potential interference factors must be identified and addressed in advance. Among the aforementioned factors, water current and tide are the most influential. Therefore, potential areas were identified using numerical methods and an electrical marine chart. Water depth and bottom properties were examined using the marine database,

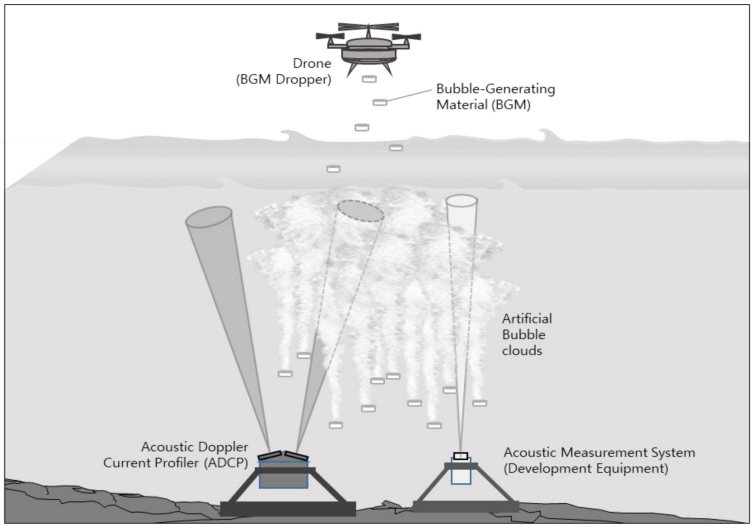


FIGURE 2
Schematic of the sea experiment.

and information on fisheries and harbor and water current speeds were obtained from the Korea Hydrographic and Oceanographic Agency (KHOA). Additionally, other limitations such as drone flight restriction zones and no-flight zones were verified using the spatial information open platform provided by the Ministry of Land, Infrastructure, and Transport of Korea (MLIT). After visually identifying the interference factors such as actual ship traffic and fishing nets, the experimental area was finally decided (Figure 3). The selected area was located in the southern part of the Korean Peninsula, which is characterized by its complex coastline and the presence of several small islands that act as a natural barrier against wind currents. Additionally, there is a wave-absorbing revetment in the area and the water current during neap tide is relatively slow compared to other candidate sites that were considered. These environmental conditions were favorable to successfully conduct our experiments.

3.3 Artificial bubble generation

The bubbles to be measured were generated with a BGM, which generates artificial bubbles through a chemical reaction with water while sinking slowly from the sea surface. The BGM was shaped as pellets to ensure stable sinking to the desired depth. Additionally, precautions were taken to ensure that the formulation of the BGM did not adversely affect the ecosystem or human health.

3.4 BGM dropping method

The main concern in determining how to drop the BGMs was to avoid physical interference with other objects such as a boat. Therefore, a drone was adapted for this purpose. As shown in Figure 4, a medium-size drone was selected to minimize the effects

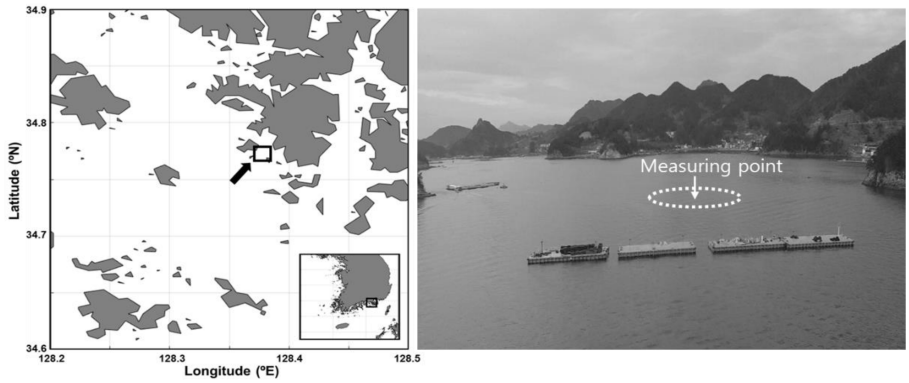


FIGURE 3
Location and photograph of the sea experiment.

of meteorological factors such as the wind. To evenly distribute the BGM pellets on the sea surface, an additional rotation plate was mounted under the drone body so that when the pellets passed through it when they were ejected. The rotation rate of the rotation plate and open angle of the gate were adjusted by conducting land experiments in advance. Another consideration of drone operation is related to the operation environment. For example, landing coordinates change because the drone is operated on the ship deck.

Figure 5 shows the BGM pellets distributed on the sea surface using the drone. Figure 5A is the top view at the moment of dropping the BGM and Figure 5B shows the artificial bubble cluster on the sea surface after dropping the pellets.

3.5 Measurement equipment setting

To measure the backscattering strength of the artificial bubbles, the ADCP was settled on the sea bottom, after which water current speed data were acquired using its operating software. The acquisition bin size was set to 1 m, which is directly related to its spatial resolution. The distribution of bubbles in the water varies in time and space, and therefore an acoustic ping was transmitted and received at 1-second intervals to characterize the variation of backscattering strength in more detail. Finally, the depth-dependent echo intensities of the four transmitted pings toward different directions were obtained for every second during the experiment.

Measurement parameters of the SSMS can set up more suitable for the experiemnt compared to ADCP. The frequency band was set to 470.0 kHz to avoid acoustic interference and to enhance the resolution relative to that of ADCP (307.2 kHz). To increase the time resolution, the transmission interval was set to 0.1 seconds with a 1.0 ms pulse length to acquire 10 times more pings than the ADCP system. To improve the spatial resolution, the sample rate was set to 64 kHz, which is sampling numer per second, which is the sampling unit. Both measurement devices were mounted horizontally on the sea bottom using a gimbal (Figure 1).

4 Results

4.1 Marine environment analysis

The main measurement was conducted for 90 minutes from 10:00 a.m. to 11:30 a.m., and the CTD (conductivity, temperature, and depth) sensor was operated for 30 minutes before the main measurement to obtain marine environmental changes. Figure 6 shows the observed water temperature, salinity, and sound speed. From 09:35 to 11:35, the profiles of the water temperature and salinity remained largely constant at 15.3°C and 33.6 PSU (Figures 6A, B). In Figure 6C, the sound speed profiles derived from the observed CTD data were also within a relatively constant range of 1,506.2 to 1,506.7 m/s. These constant features are characteristic of the areas surrounding the Korean Peninsula in winter, which is attributed to the coverage of the entire surface layer by the water mass of the Tsushima Warm Current (Lim, 1976). These conditions are very favorable for the acquisition of acoustic data for quantitative analysis.

Figure 7 shows the water currents in the experiment site using the Aanderaa RCM Blue 5450 (gray arrow) and Aanderaa Seaguard II DCP (black arrow) instruments, both of which are water current measurement sonars. The RCM was deployed at 3.4 m and 8.4 m depths with 1 minutes intervals, and the DCP system was moored at depths of 6.0 to 24.5 m with a depth interval of 0.5 m and a time interval of 10 minutes. In Figure 7, the lengths of the arrows represent the current speed and the arrow directions represent the direction of the current. Comparing the data of the RCM system installed at a depth of 8.4 m with the data of the DCP system moored at a depth of 8.5 m, the overall speed and direction of the water current were largely equal.

From 10:00 a.m. to 11:30 a.m., the direction of the water current according to the RCM changed from a southwest direction (240°) to a west direction (280°) at a depth of 3.4 m with an average speed of 4.5 cm/s. At a depth of 8.4 m, the direction of the water current changed from a southwest direction (240°) to a northwest direction (300°) with an average speed of 5.2 cm/s. The water current

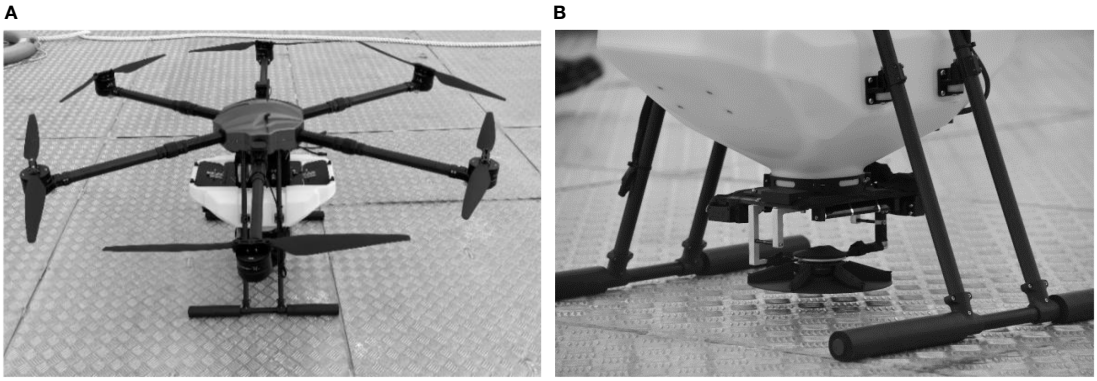


FIGURE 4
(A) Drone and **(B)** mounted rotation plate.

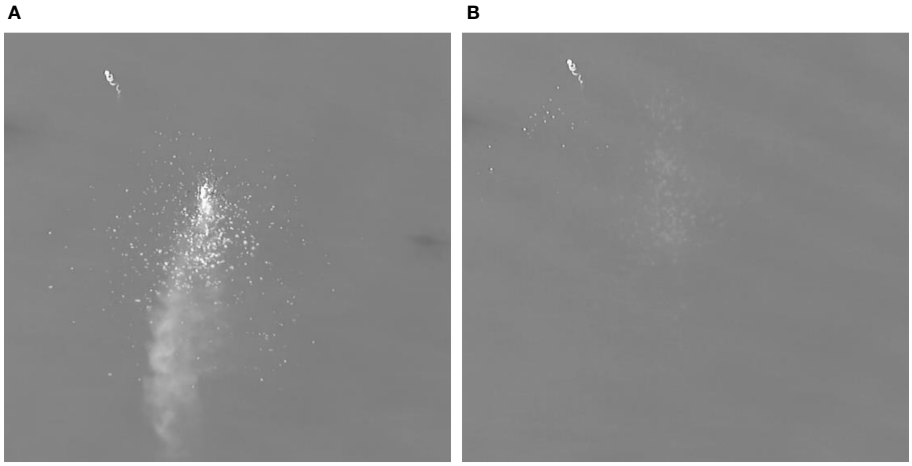


FIGURE 5
Top views of distributed results of the BGM: **(A)** at the moment of dropping the BGM; **(B)** distributed bubble cluster after dropping.

measured by DCP showed a directional change from 250° to 320° with an average speed of 4.9 cm/s. Therefore, the observed water current changes determined using the two instruments were almost identical to each other. Although the depth-dependent speed changes of the water current measured using the DCP system were not significant, the waters appeared to rotate clockwise as depth increased. Particularly, the rotation of the water current increased with depth at 11:00. The rotation angles at 6 m and 23.5 m depths were 274° and 329°, respectively.

Upon analyzing the variability of the water current with depth and time, the current speed during the measurement was approximately 4.9 cm/s, meaning that the BGM pellets did not move substantially from the experiment site. As time goes by and the depth increases, the water current rotated from southwest to northwest (i.e., clockwise) but the difference in the direction of the current was small (within 60°). These findings suggested that the BGM pellets may have moved in an almost constant direction while they sank into the water column. Therefore, the measured results were highly reliable because the marine environment was ideal, similar to the indoor watertank.

4.2 Backscattering strength measurements using ADCP

In this sea experiment, the backscattering stength of the bubble cluster was obtained using ADCP. The measured acoustic signal from ADCP was converted to backscattering stength based on the equation described in Section 2.

Figure 8 shows two-dimensional (2-D) sections for the backscattering strength converted from the echo intensity recorded using the ADCP for 4,500 seconds. In the figure, the x-axis represents time (or ping number) and the y-axis indicates the distance from the sea bottom. Figures 8A-D correspond to beam 1

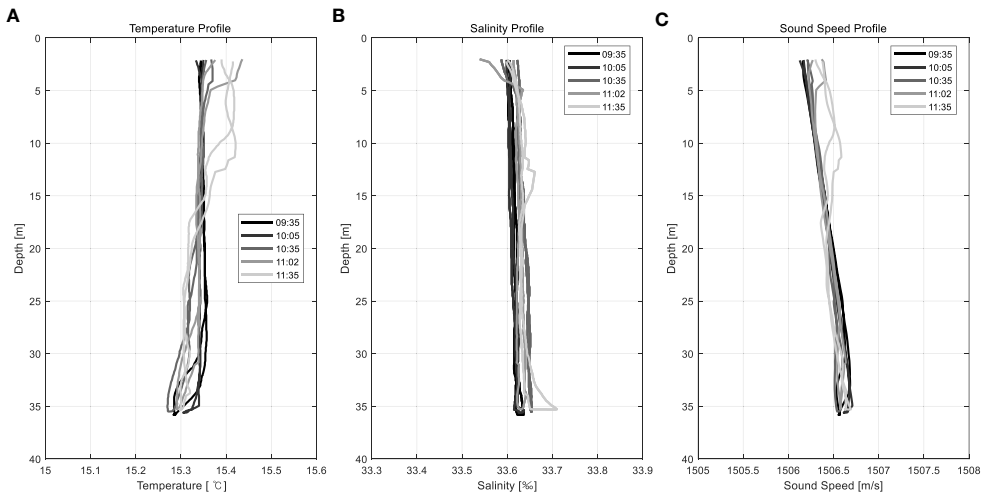


FIGURE 6
CTD measurement results: **(A)** water temperature, **(B)** salinity, and **(C)** sound speed.

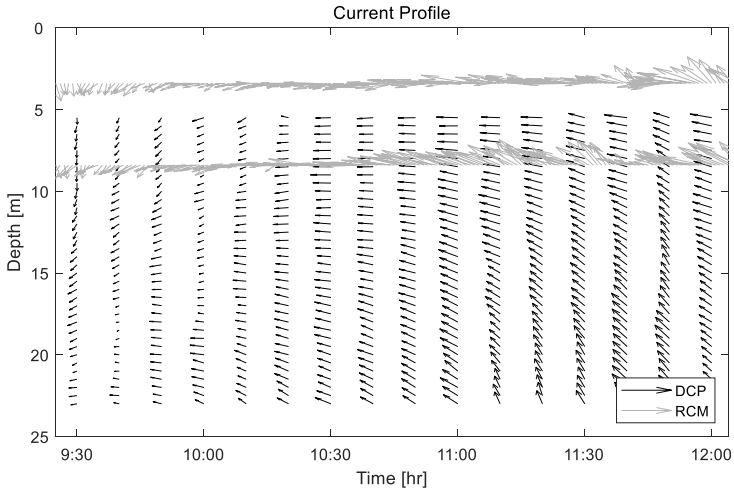


FIGURE 7
Measurement results of the water/tidal current.

and beam 4, respectively. The backscattering strength shown in Figure 8C was the strongest, meaning that the BGMs were mainly distributed near the direction of beam 3. From these results, it is possible to identify the distribution characteristics of BGM that are difficult to directly confirm in sea experiments. Based on the results of this analysis, we mainly used data from beam #3.

Artificial bubbles are continuously created as the BGM sinks. Therefore, to understand the distribution characteristics of bubble clusters, it is necessary to understand the sinking characteristics of the BGM. In this experiment, we attempted to understand the characteristics of the bubble cluster created as the BGM sinks by comparing it with the sea experimental results using the BGM Mock-up model. Figure 9 shows 2-D sections of the backscattering strength of the reference experiment (mock-up model) and the artificial bubble (BGM) experiments derived using beam #3 of the ADCP, which is the direction in which the recorded backscattering strength was the strongest, as shown in Figure 8C. Figure 9A shows the result when mock-up models that do not generate artificial bubbles were dropped as a reference. Figures 9B-D show the backscattering strengths of the artificial bubbles of the different BGM dropping experiments with a white contour line indicating –50 dB. When the mock-up model was dropped, only the reflected signal by the models could be identified, and it was confirmed that the sinking speed was approximately 13 cm/s based on the change in water depth over time, which was in good agreement with our desired value. In the figure, the backscattering signals by the mock-up models were identified only to a depth of approximately 15 m from the sea surface. This indicates that the available observation

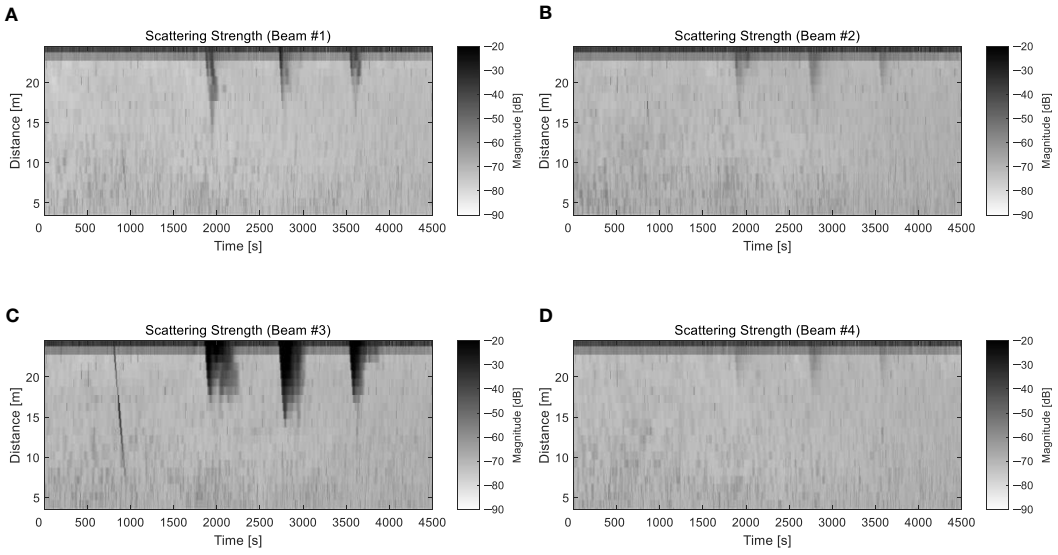


FIGURE 8
Backscattering strength derived from ADCP data: beams **(A)** #1, **(B)** #2, **(C)** #3, and **(D)** #4 of the ADCP.

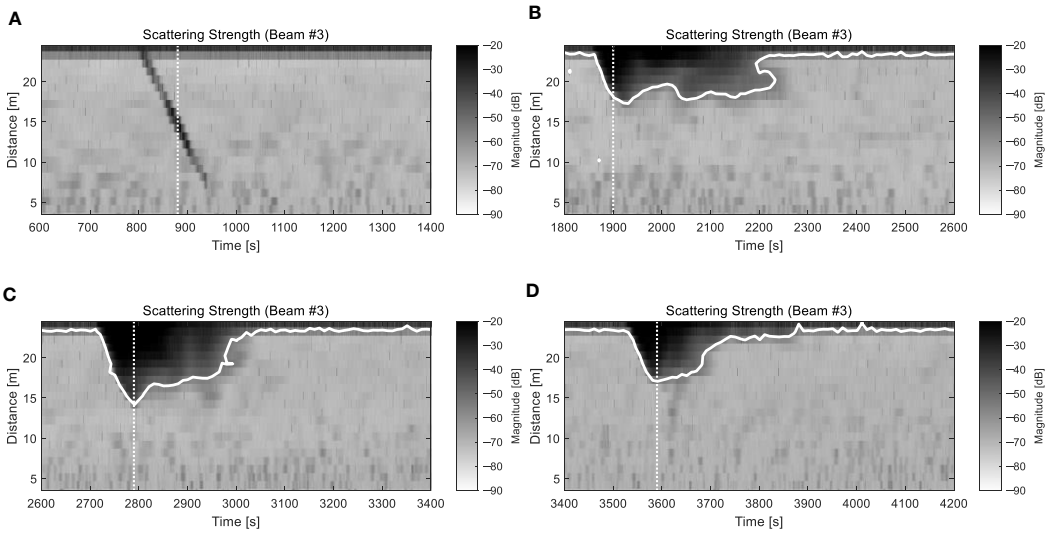


FIGURE 9
(A) Reference backscattering strength and **(B-D)** backscattering strengths of the bubble cluster for different BGM dropping experiments.

depth in the steering direction using ADCP was approximately 15 m from the surface. Our experiments also confirmed the generation of artificial bubbles as the BGM pellets sank (Figures 9B-D). The backscattering signal was weakened due to a gradual dissipation of bubbles in the lower part of the bubble cluster. Therefore, our results not only allowed for the estimation of the sinking speed of the BGM pellets but also demonstrated that the artificial bubbles occurred at depths of up to 8–10 m from the surface. Additionally, our observations confirmed that the generated bubbles persisted for more than a few minutes. Figure 10 shows the depth-dependent backscattering strength of the reference data at the time indicated by the vertical white dashed line in Figure 9A. Figure 10A shows the individual backscattering strength of the four steering beams. In beam #3, the backscattering strength by the mock-up models reached –25.0 to –30.0 dB, whereas the scattering strength at other depths (except near the sea surface) was approximately –65.0 and –70.0 dB and remained mostly constant. In general, the backscattering strength increases as it reaches the sea surface due to the presence of floating particles moved by the wind, waves, and tidal currents, and these characteristics are shown in Figure 10A. Our results were generally 15.0 dB higher than the backscattering strength of –80.0 and –85.0 dB observed in our previous watertank study (Bae et al., 2022). Figure 10B compares the backscattering strength derived from the sonar equation in Equation (1) with the result obtained using the method described in the previous study, which was provided by the manufacturer. Both methods exhibited

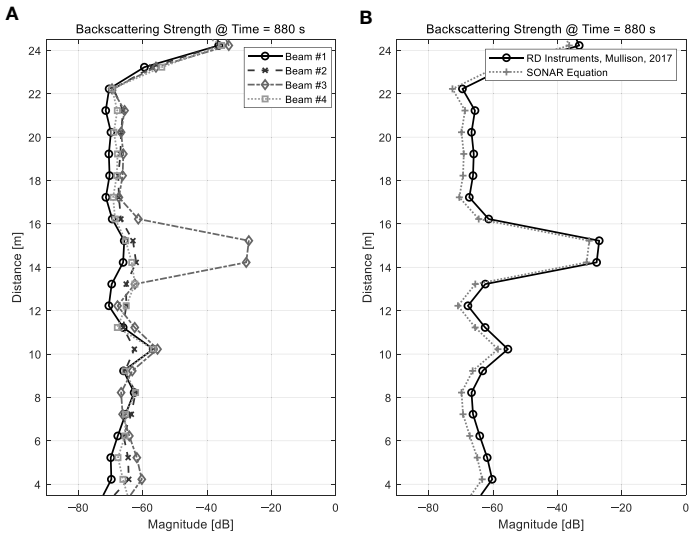


FIGURE 10
Backscattering strength profiles for the mock-up models: **(A)** each beam responses and **(B)** converted results with different equations.

similar trends, but some quantitative bias errors were observed. We estimated that these biases derived from some differences between the approaches. Specifically, the transmission loss for the manufacturer's method does not account for the characteristics of the area including the absorption. Furthermore, parameter correction for water temperature is also required for the manufacturer's approach, but this was not taken into account.

Figure 11 shows the magnitude of the backscattering strength for each depth at the time indicated by the vertical white dashed line in the presence of the artificial bubbles depicted in Figures 10B-D. The results for the four steering beams are shown together. Although there were slight differences in the results of the three experiments due to the spatiotemporal variations of the marine environment, the magnitude of the backscattering strength by the bubble cluster increased up to 50.0 dB compared to when there were no bubbles. These results are also consistent with those of our previous watertank experiments (Bae et al., 2022). In the vicinity of the sea surface, the maximum backscattering strength was calculated at approximately −10.0 dB, and assuming −50.0 dB as the bubble boundary, the distribution depth of the artificial bubbles used in the experiment was estimated to range from 8.0 to 10.0 m. These results indicate that the backscattering strength of artificial bubbles can be effectively measured through sea experiments using ADCP.

4.3 Backscattering strength using the SSMS

Using ADCP, we were able to effectively measure the backscattering strength of the artificial bubbles. However, ADCP is a device that steers a beam with a specific frequency in a specific direction. Therefore, some verifications are needed because unwanted distortions may occur. For this reason, we developed SSMS with high resolution and vertical beam steering, and attempted to compare the results of maritime experiments using SSMS with the results obtained through ADCP for verification.

Figure 12A shows the calculated backscattering strength of the mock-up model (i.e., without artificial bubbles) and corresponds to the case illustrated in Figure 9A. The constructed SSMS allowed for the acquisition of data with higher temporal and spatial resolutions than the ADCP. Moreover, the ADCP results could not be observed at a depth of ≥15 m from the sea surface due to the tilted steering beam angle. In contrast, the beam of the SSMS was toward the sea surface directly and therefore the acquired data could be observed for all depths. Figure 12B shows the results of the backscattering strength acquired by the SSMS for the same experiment illustrated in Figure 9C, which corresponds to the ADCP data obtained in the presence of artificial bubbles. The calculated backscattering strength was very similar to that of our ADCP experiments. Furthermore, the identified characteristics in the ADCP data were also observed with very high resolution.

Figure 13A shows the backscattering strength profiles at 180 seconds, when the bubbles reached their maximum depth in the water column (Figure 12B). In the figure, the dashed lines represent the raw data, whereas the solid line represents the results after applying a moving average. The processed results exhibited similar trends and magnitudes compared to the ADCP #3 beam data in Figure 11B, and therefore our results validated the ADCP data. Figure 13B shows the backscattering strengths of the artificial bubbles at different times with 60-second intervals (vertical dotted white lines in Figure 12B). According to our results, the backscattering strength increases from the vicinity of the sea surface to the bottom as the bubble cluster extended. Additionally, the backscattering strength at a depth of less than 10 m from surface increased gradually from its original magnitude at 60 seconds to the magnitude at 180 seconds, which values is the strongest magnitude, and decreased gradually.

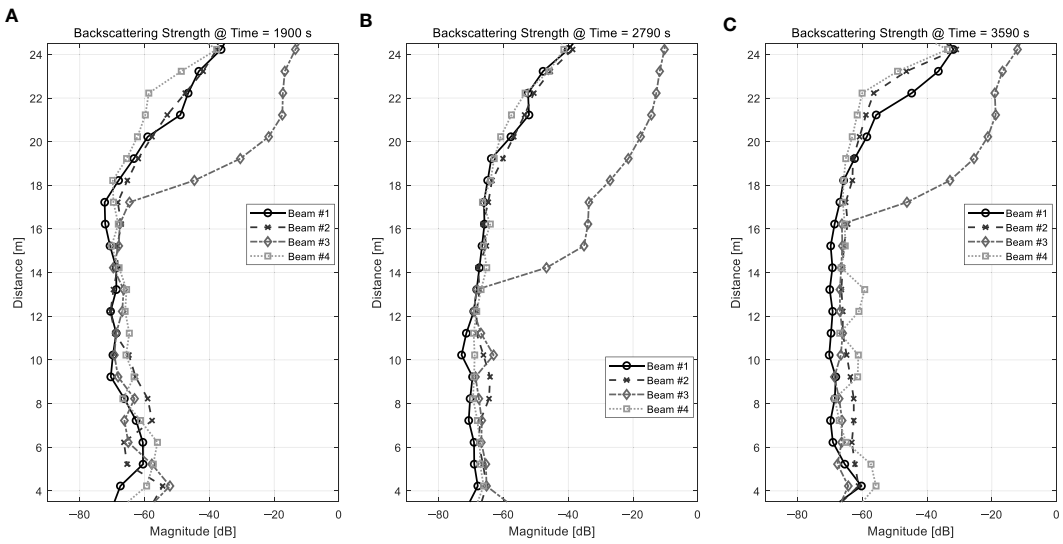


FIGURE 11
Backscattering strength profiles for the bubble cluster obtained by ADCP: **(A)** 1,900 s, **(B)** 2,790 s, and **(C)** 3,590 s.

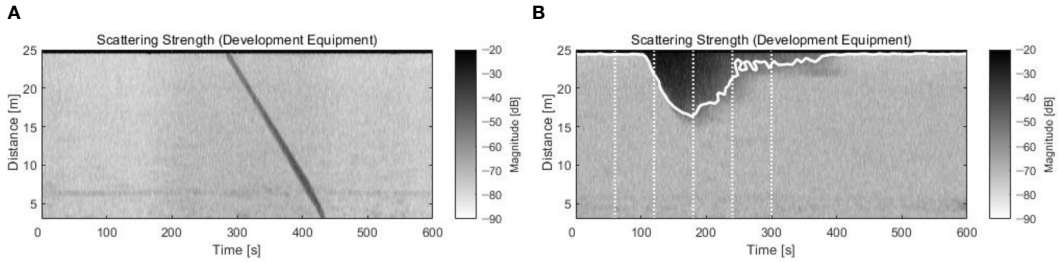


FIGURE 12
Backscattering strength obtained from the SSMS: **(A)** mock-up model and **(B)** bubble cluster data.

5 Conclusions

The study of bubble characteristics in water is not only important in the field of natural sciences but also provides a basis for the development of military technology. Particularly, measuring the acoustic characteristics of artificial bubbles is the first step toward their military application. In our previous study, we measured the acoustic characteristics of artificial bubble clusters generated by BGM pellets in an ideal watertank environment. In this study, we measured the acoustic characteristics of artificial bubbles in a real marine environment where, unlike in controlled laboratory conditions, the water temperature, salinity, and concentrations of organic and inorganic compounds change in a time- and space-dependent manner.

To quantitatively measure the acoustic properties of artificial bubbles in a real ocean environment, a sea site suitable for our experimental purposes was selected in the southern sea of the Korean Peninsula. Additionally, to avoid any interference between the bubbles generated by the BGM pellets and other materials during the dropping, the BGMs were remotely dropped using a drone.

During the measurements, water temperature and salinity were largely constant throughout the entire layer, thus mimicking the ideal conditions of a watertank. Because the measurement was carried out during a neap tide, the observed current speed was approximately 4.9 cm/s, thus providing ideal conditions to conduct our experiments.

To obtain the backscattering strength of the artificial bubbles, echo intensities were acquired using an ADCP. Moreover, the ADCP approach was validated by comparing its measurement results with those obtained with the SSMS. The measured results for each BGM dropping experiment showed similar trends. Specifically, artificial bubbles were generated as the BGM sank and backscattering strength weakened as bubbles began to dissipate from the lower part of the cluster. In addition to observing the acoustic characteristics of the artificial bubbles, physical properties were also evaluated, such as the sinking speed of the BGM pellets.

Our experimental design had some inherent limitations because real marine environments change considerably in a time- and space-dependent manner. However, the measurement method proposed herein is expected to be highly applicable because it provides a simple way to measure the acoustic signatures of bubbles in water. Collectively, our findings demonstrated that our proposed approach could be applied not only for the exploration of basic science concepts but also for the development of military applications.

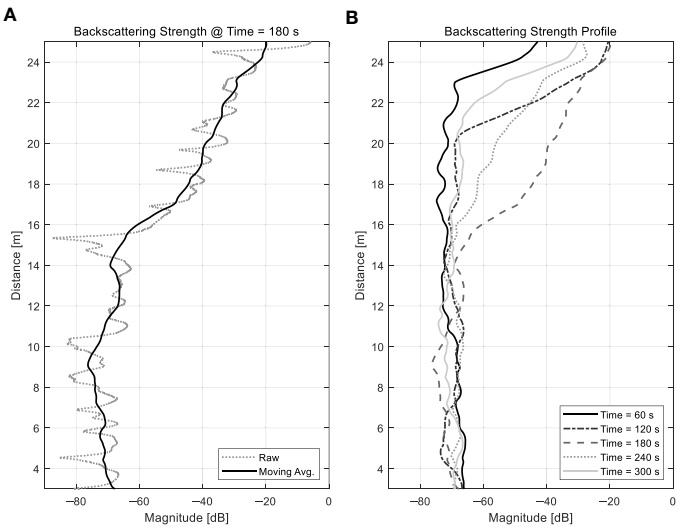


FIGURE 13
Backscattering strength profiles for the bubble cluster obtained by the SSMS: **(A)** raw and moving average example and **(B)** result profiles with different time represented in Figure 12B.

Data availability statement

The datasets presented in this article are not readily available because of national security issues. Requests to access the datasets should be directed to corresponding author Won-Ki Kim, konekee@chungbuk.ac.kr.

Author contributions

HB: Investigation, Resources, Software, Visualization, Writing – original draft. W-KK: Conceptualization, Investigation, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Agency for Defense Development by the Korean Government (912763101).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bae, H. S., Kim, W.-K., Son, S.-U., Kim, W.-S., and Park, J.-S. (2022). An estimation of the backscattering strength of artificial bubbles using an acoustic Doppler current profiler. *Sensors* 22, 1812. doi: 10.3390/s22051812

Caruthers, J. W., Stanic, S. J., Elmore, P. A., and Goodman, R. R. (1999). Acoustic attenuation in very shallow water due to the presence of bubbles in rip currents. *J. Acoust. Soc Am.* 106, 617–625. doi: 10.1121/1.427030

Culver, R. L., and Trujillo, M. F. (2007). "Measuring and modeling bubbles in ship wakes, and their effect on acoustic propagation," in *Proceedings of the 2nd International Conference on Underwater Acoustic Measurements: Technologies and results*. (Greece: FORTH, Crete), 25–29 June.

Dahl, P. H. (2003). The contribution of bubbles to high-frequency sea surface backscatter: A 24-h time series of field measurements. *J. Acoust. Soc Am.* 113, 769–780. doi: 10.1121/1.1532029

Deines, K. L. (1999). "Backscatter estimation using broadband acoustic Doppler current profilers," in *Proceedings of the IEEE Sixth Working Conference on Current Measurement (Cat. No. 99CH36331), San Diego, CA, USA, 13 March 1999* (Piscataway, NJ, USA: IEEE), 249–253.

Guelou, Y., Mourry, P., and Gillot, C. (2002). EP Patent 00942865. Muret (FR): Etienne Lacroix Tous Artifices S.A.

Guelou, Y., Mourry, P., and Gillot, C. (2005). US Patent 06889590. Muret (FR): Etienne Lacroix Tous Artifices S.A.

Gulf of Mexico Expedition (2011). Available at: https://oceanexplorer.noaa.gov/okeanos/explorations/ex1105/welcome.html (Accessed 17 March 2022).

Ji, Y.-H., Lee, J.-H., Kim, J.-S., Kim, J.-H., Kim, W.-S., and Choi, S.-M. (2009). Modeling of scattered signal from ship wake and experimental verification. *J. Acoust. Soc Korea.* 28, 10–18.

Jurng, M. S. (1996). *Estimation of zooplankton biomass by volume scattering strength in coastal waters* (Seoul, Korea: Ph.D. Thesis, Seoul National University).

Karoui, I., Quidu, I., and Legris, M. (2015). Automatic sea-surface obstacle detection and tracking in forward-looking sonar image sequences. *IEEE Trans. Geosci. Remote Sens.* 53, 4661–4669. doi: 10.1109/TGRS.2015.2405672

Keiffer, R. S., Novarini, J. C., and Norton, G. V. (1995). The impact of the background bubble layer on reverberation-derived scattering strengths in the low to moderate frequency range. *J. Acoust. Soc Am.* 97, 227–234. doi: 10.1121/1.412990

Leighton, T. G., Finfer, D. C., Chua, G. H., and White, P. R. (2011). Clutter suppression and classification using twin inverted pulse sonar in ship wakes. *J. Acoust. Soc Am.* 130, 3431–3437. doi: 10.1121/1.3626131

Li, H., Li, S., Chen, B., Xu, C., Zhu, J., and Du, W. (2014). *Research on ship wake acoustic imaging based on multi-beam sonar* (NL, Canada: 2014 Oceans – St. John's).

Lim, D. B. (1976). The movements of the waters off the south coast of Korea. *J. Oceanol. Soc Korea.* 11, 77–88.

Mullison, J. (2017). "Backscatter estimation using broadband acoustic Doppler current profilers-updated," in *Proceedings of the ASCE Hydraulic Measurements & Experimental Methods Conference*, (Durham, NH, USA: ASCE), 9–12 July.

Park, C., Jeong, S. W., Kim, G. D., Park, Y., Moon, I., and Yim, G. (2021). An empirical model of air bubble size for the application to air masker. *J. Acoust. Soc Korea.* 40, 320–329. doi: 10.7776/ASK.2021.40.4.320

Rustemeier, J., Griesmann, T., and Rolfes, R. (2012). *Underwater sound mitigation of bubble curtains with different bubble size distributions. In Proceedings of Meetings on Acoustics, 17, 070055* (Edinburgh, Scotland).

Schneider von Deimling, J., Brockhoff, J., and Greinert, J. (2007). Flare imaging with multibeam systems: Data processing for bubble detection at seeps. *Geochem. Geophys. Geosyst.* 8, Q06004. doi: 10.1029/2007GC001577

Schneider von Deimling, J., and Weinrebe, W. (2014). Beyond bathymetry: Water column imaging with multibeam echo sounder systems. *Hydrographische. Nachrichten.* 31, 6–10.

Stanic, S., Caruthers, J. W., Goodman, R. R., Kennedy, E., and Brown, R. A. (2009). Attenuation measurements across surface-ship wakes and computed bubble distributions and void fractions. *IEEE J. Ocean. Eng.* 34, 83–92. doi: 10.1109/JOE.2008.2008411

Trevorrow, M. V. (2003). Measurements of near-surface bubble plumes in the open ocean with implications for high-frequency sonar performance. *J. Acoust. Soc Am.* 114, 2672–2684. doi: 10.1121/1.1621008

United States Office of Scientific Research and Development. (1946). *National Defense Research Committee, Issuing Body* (The Physics of Sound in the Sea, Washington D.C.: Office of Scientific Research and Development, National Defense Research Committee, Division 6).

Urban, P., Koser, K., and Greinert, J. (2017). Processing of multibeam water column image data for automated bubble/seep detection and repeated mapping. *Limnol. Oceanogr.: Methods* 15, 1–21. doi: 10.1002/lom3.10138

Urick, R. J. (1983). *Principles of underwater sound. 3rd ed* (New York, NY, USA: McGaw-Hill), 240–244.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 22 February 2024
DOI 10.3389/fmars.2024.1350337

OPEN ACCESS

EDITED BY
Arata Kaneko,
Hiroshima University, Japan

REVIEWED BY
Chanhyung Jeon,
Pusan National University, Republic of Korea
Haocai Huang,
Zhejiang University, China

*CORRESPONDENCE
Xiao-Hua Zhu
xhzhu@sio.org.cn

RECEIVED 05 December 2023
ACCEPTED 05 February 2024
PUBLISHED 22 February 2024

CITATION
Zhang C, Zhu Z-N, Xiao C, Zhu X-H and Liu Z-J (2024) Acoustic tomographic inversion of 3D temperature fields with mesoscale anomaly in the South China Sea. *Front. Mar. Sci.* 11:1350337. doi: 10.3389/fmars.2024.1350337

Acoustic tomographic inversion of 3D temperature fields with mesoscale anomaly in the South China Sea

Chuanzheng Zhang[1,2], Ze-Nan Zhu[1], Cong Xiao[1], Xiao-Hua Zhu[1,2,3]* and Zhao-Jun Liu[1,2]

[1]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, [2]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China, [3]School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

Acoustic tomographic inversion is based on travel times measured along the transmission paths between all station pairs to reconstruct three-dimensional temperature structures with mesoscale anomalies. In this study, tomographic simulation experiments were designed based on the Hybrid Coordinate Ocean Model (HYCOM) reanalysis data to reconstruct mesoscale phenomena from travel time data obtained from five, seven, and nine stations in the South China Sea over a domain of 100 × 100 km. The travel times for each station pair were calculated in the vertical section using the Bellhop acoustic ray simulation method. Six Empirical orthogonal function (EOF) modes of sound speed along the sound transmission paths in a vertical slice were used to formulate the inversion equations. The horizontal-slice distributions of temperature in the tomography domain were reconstructed using the grid-segmented method for each depth layer. For station-to-station distances greater than 100 km, the performance of inversion was best for the seven-station case rather than for the nine-station case, with the highest horizontal resolution of the three cases. This case study concluded that the seven-station case rather than the nine-station case provided an optimal station number for reconstructing the three-dimensional temperature fields.

KEYWORDS

ocean acoustic tomography, inversion of three-dimensional temperature fields, mesoscale phenomena, HYCOM data, South China Sea

1 Introduction

Ocean mesoscale eddies are globally widespread and play important roles in ocean heat transport and energy dissipation. Mesoscale phenomena are the best targets for ocean acoustic tomography because of the movement and variability of eddies (Munk et al., 1995). In the northern part of the South China Sea (SCS), mesoscale eddies are generated by Kuroshio intrusion through the Luzon Strait (Liu et al., 2008).

Ocean acoustic tomography (OAT) is an innovative method that is widely used in oceanography (Munk and Wunsch, 1979; Munk et al., 1995; Kaneko et al., 2020). The OAT was proposed as an advanced underwater remote sensing technology, which was applicable to reconstruct the three-dimensional structures of ocean dynamic parameters. Acoustic stations are located at the periphery of an observation area, tomography domain is measured by sound traveling among the acoustic stations (Zheng et al., 1997; Zhu et al., 2013, Zhu et al., 2017; Syamsudin et al., 2019) to realize synchronous observation of rapidly varying mesoscale temperature fields, which are difficult to achieve using conventional shipboard and point mooring observations (Zhang et al., 2015). Several different vertical temperature distributions were assumed to be suitable for performing vertical inversion in coastal seas (Park et al., 2021). However, due to the complexity of the deep-sea environment, the sound speed profile is difficult to represent using a simple function. The Empirical orthogonal function (EOF) decomposes the vertical structure of sound speed into multiple principal components, and the characteristic feature of sound speed profiles can be accurately reconstructed by inverting the coefficients of the major EOF modes (LeBlanc and Middleton, 1980; Fukumori and Wunsch, 1991). The propagation time information observed by acoustic tomography can also be used to invert the coefficients of individual EOF modes. Tomographic mapping of three-dimensional mesoscale temperature fields has frequently been studied, with stochastic inversion (the Gauss–Markov method) being applied in most studies. However, the solution provides less flexibility because the covariance of the expected solution is required prior to inversion (Cornuelle et al., 1985; Howe et al., 1987; Yuan et al., 1999). Consequently, more flexible inversion methods are preferred.

The SCS is the largest marginal sea adjacent to the northwestern Pacific Ocean. Recent observations have shown that the SCS exhibits frequent mesoscale eddies (Wang et al., 2003, Wang et al., 2008; Chen et al., 2011; Nan et al., 2011; Chu et al., 2020). In this study, tomographic inversion of mesoscale eddies was performed for a model domain of 100 × 100 km in the northern SCS. The acoustic tomography network strategy was designed using temperature and salinity outputs from the Hybrid Coordinate Ocean Model (HYCOM) data. This study aimed to reconstruct three-dimensional mesoscale sound speed fields in the northern SCS using tomographic inversion under different station configurations.

A new method was proposed by combining the EOF method in the vertical slice and the grid-segmented method in the horizontal slice. Sections 2 and 3 describe the model and the forward formulation, respectively. The process and method of inversion are described in Section 4. Section 5 presents the simulation results under different station configurations. The discussion is presented in Section 6. Finally, Section 7 concludes the study.

2 Model

2.1 Ray simulation in a vertical slice

The process of sound propagation between acoustic stations is affected by many factors, the most important of which are the sound speed profile (SSP), sound frequency, and bottom topography. The BELLHOP ray tracing method was used to simulate the sound propagation process between each of the transmitter-receiver station pairs. This was achieved using the time- and domain-averaged sound speed profile (the reference sound speed), calculated from the temperature and salinity data within the tomography region (MacKenzie, 1981). In the process of determining the ray patterns along the transmission paths between the station pairs, the sound speed distribution used is independent of range, while the bottom topography used is related to range. Figure 1 shows a typical ray pattern along each transmission path, along with the reference sound speed profile. Surface-bottom reflected rays were constructed in all sections (Figures 1A, B) that traversed the entire vertical section of the water body, indicating that the tomographic technique could measure the sound speed field over the entire section. All the parameters of multipath travel time, ray path, and ray length were required to execute the vertical section inversion. To distinguish the multipath travel time, this study used a 480 Hz sound source as an example, and the travel time difference in the selected typical sound rays was greater than the time resolution of acoustic tomography (2.1 ms), defined as the one-digit length of the M sequence for modulation number=1 (Q=1). The M sequence is a pseudo-random signal that has no correlation with ambient noises. It is a powerful tool to delete the effect of ambient noises in received signals and increases remarkably signal-to-noise ratio (SNR). For station distances varying between different station pairs, the number of typical sound rays was 38 at a maximum of 105 km distance (Figure 1A) and 12 at a minimum of 33 km distance (Figure 1B), and the average number was 25 for all station distances at 30–140 km. Note that apart from the bottom topography, the distance between the acoustic stations was a major factor affecting the density of acoustic rays within the vertical section (Figure 1C). A shadow zone is a space that cannot be covered with acoustic rays. The shorter the distance, the smaller the density of the acoustic rays and the larger the shadow zones. As long as a sufficient number of acoustic rays are obtained, the influence of shadow zones on inversion results is weak.

2.2 Empirical orthogonal function method

From the HYCOM reanalysis data from 1993 to 2011, we accumulated the vertical profiles of the sound speed to take seasonal averages using the daily temperature, salinity, and depth data, interpolated to 1 m interval data, and then subtracted the climatologically averaged sound speed from the daily data to obtain the deviated sound speed profiles (Figure 2A). The deviated SSP varied greatly with time and showed different vertical structures in the upper ocean in the different seasons. In the upper 200 m, the deviated sound speeds notably varied owing to the main thermocline, with a variation range of approximately ±15 m/s. The range of variation in the deviated sound speeds decreased with increasing depth. At a 1000 m depth, the variation in deviated sound speeds was only within ±2 m/s. Different seasonal profiles showed that the deviated sound speed varied within ±5 m/s in the subsurface layer, where the seasonal thermocline of the SSC was present (Wang et al., 2022). EOF decomposition was applied to the deviated sound-speed profile

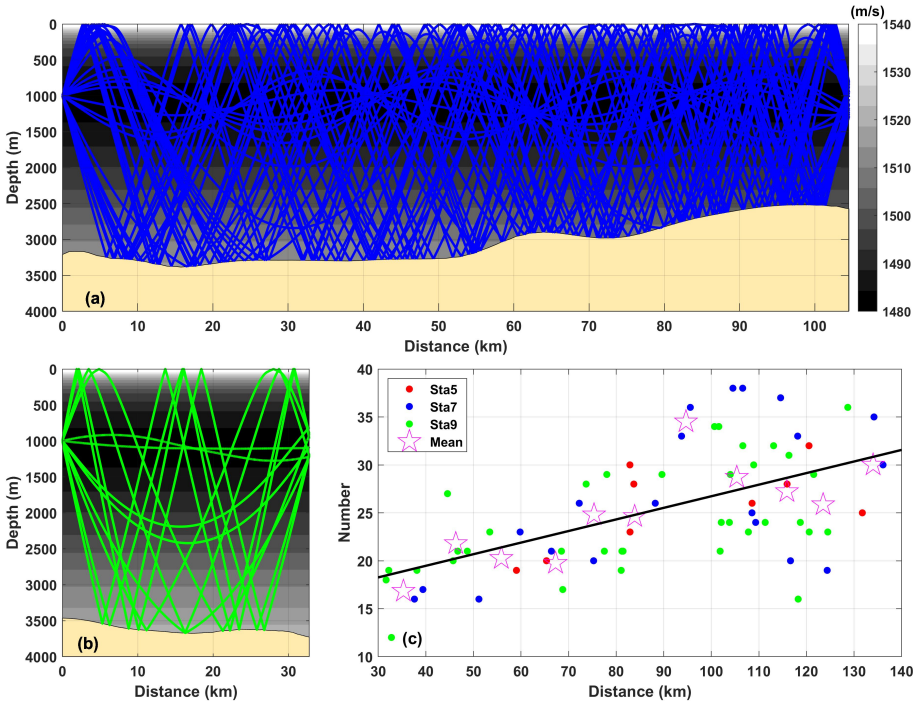


FIGURE 1
The typical ray pattern simulated using the Bellhop ray tracing scheme together with the reference sound speed profile. **(A, B)** show the typical ray patterns for L=105 km and L=33 km, respectively. The background color indicates the time- and domain-averaged reference sound speed profiles. **(C)** shows the number of rays plotted against the station distances. The open star marks indicated the mean of the ray number, calculated for every horizontal grid box of 10 km.

data, and the first six EOF modes were calculated (Figure 2B). The contribution rates of the six EOF modes were 81.7%, 12.4%, 3.5%, 1.0%, 0.9%, 0.5%, respectively, and the first three modes were the major modes accounting for the contribution rate of 97.6%. The deviated sound speed showed a large variability in the upper 400 m, and the variability diminished rapidly with depth. For the first several modes of the EOF, near-surface-intensified phenomena were also prominent in the upper 150 m, where the deviated sound speed had a large value. At depths greater than 600 m, the speed of the deviating sound decreased rapidly.

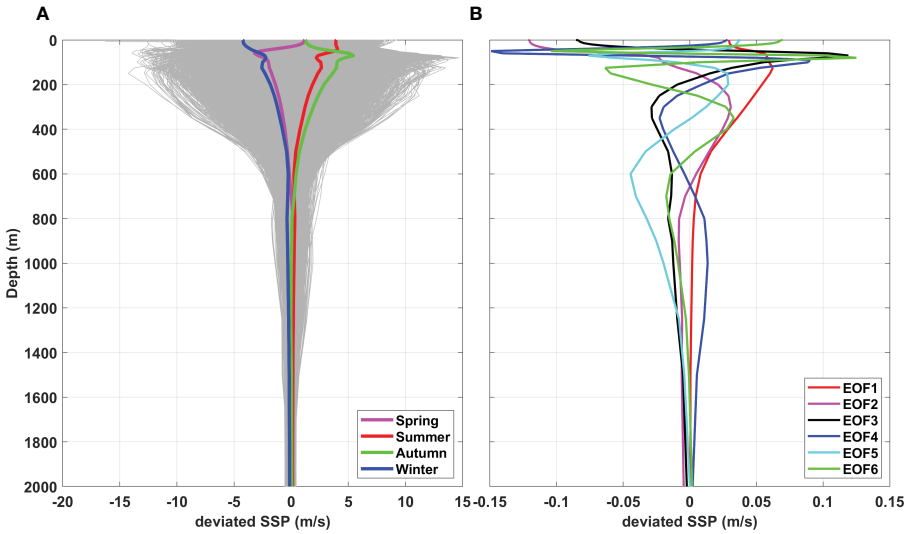


FIGURE 2
EOF decomposition for the deviated sound speed profiles. **(A)** The vertical profiles of deviated sound speed data accumulated from 1993 to 2011; **(B)** The vertical profiles of deviated sound speeds for the first six EOF modes.

2.3 Configuration of the projected rays in a horizontal slice

The HYCOM reanalysis product was developed earlier and is currently the longest-time-scale high-resolution dataset with eddy-resolving resolution (Adams et al., 2011). Based on the hydrographic data (temperature and salinity) with a time resolution of 1 d and a horizontal resolution of 1/12° in the HYCOM dataset, we selected a 1° × 1° latitude-longitude model domain (approximately 104 × 111 km) in the southwest of the Luzon Strait, where the Kuroshio frequently intrudes. Mesoscale eddies that frequently appear in this area (Chen et al., 2011), where are the targets of acoustic tomography. In the analysis, the effects of station number on the accuracy of the inversion and the standard configuration of stations with asymmetry were considered. Figure 3 shows the standard configurations for the five-, seven-, and nine-station cases. The total number of rectangular grids was M=7 × 7 = 49. With an increase in the number of acoustic stations, the number of acoustic transmission lines increased exponentially. The five-, seven-, and nine-station cases constructed ten, twenty-one, and thirty-six transmission paths in the horizontal domain, respectively. The observational information, acquired from sound propagation among tomographic stations, increased with exponential growth rather than the linear growth obtained by conventional mooring stations (Zhang et al., 2017).

2.4 Warm eddy

The temperature distribution on January 6, 2011, was selected with a focus on the warm eddy in the simulated domain. To facilitate comparison, the horizontal-slice distributions of the depth-averaged temperatures are shown at every 50-m depth with a 7 × 7 grid, as shown in Figure 4. In the upper layer (100–600 m), the core of temperature anomaly due to warm eddies intensified with a maximum temperature anomaly of approximately 2°C at a depth of 100–150 m. The eddies penetrated deeply, extending from 100 m to 600 m. The core of the temperature anomaly gradually weakened as depth increased, constructing an incline toward the southeast.

3 Forward formulation

3.1 Vertical slice

The propagation of acoustic signals in the ocean can be approximated by using refracted acoustic rays. The propagation of ray paths in the vertical section depends on the vertical distribution of sound speed and velocity (Munk et al., 1995). This study focused on the influence of sound speed on propagation time without considering the velocity. The travel time deviation ($\delta\tau_i$) for the i-th ray path traveling between the acoustic station pair is expressed as follows (Zhang et al., 2015):

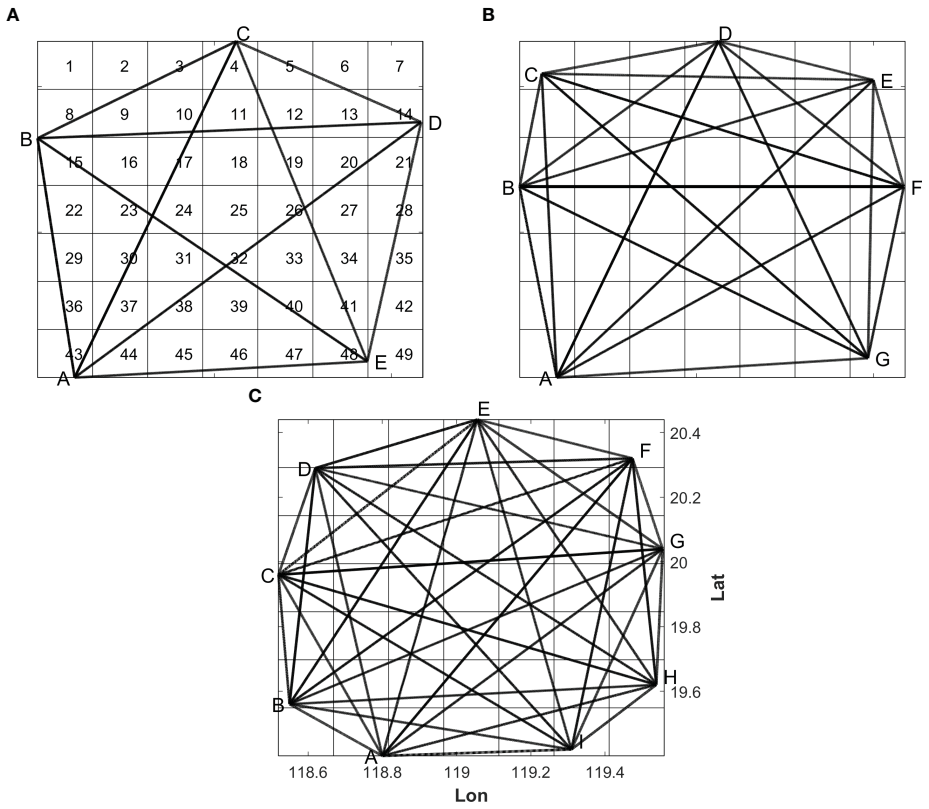


FIGURE 3
The standard configuration for **(A)** five, **(B)** seven, and **(C)** nine stations superimposed on the horizontal-slice inversion grid. The grids are numbered as 1–49; A–I are the names of the stations.

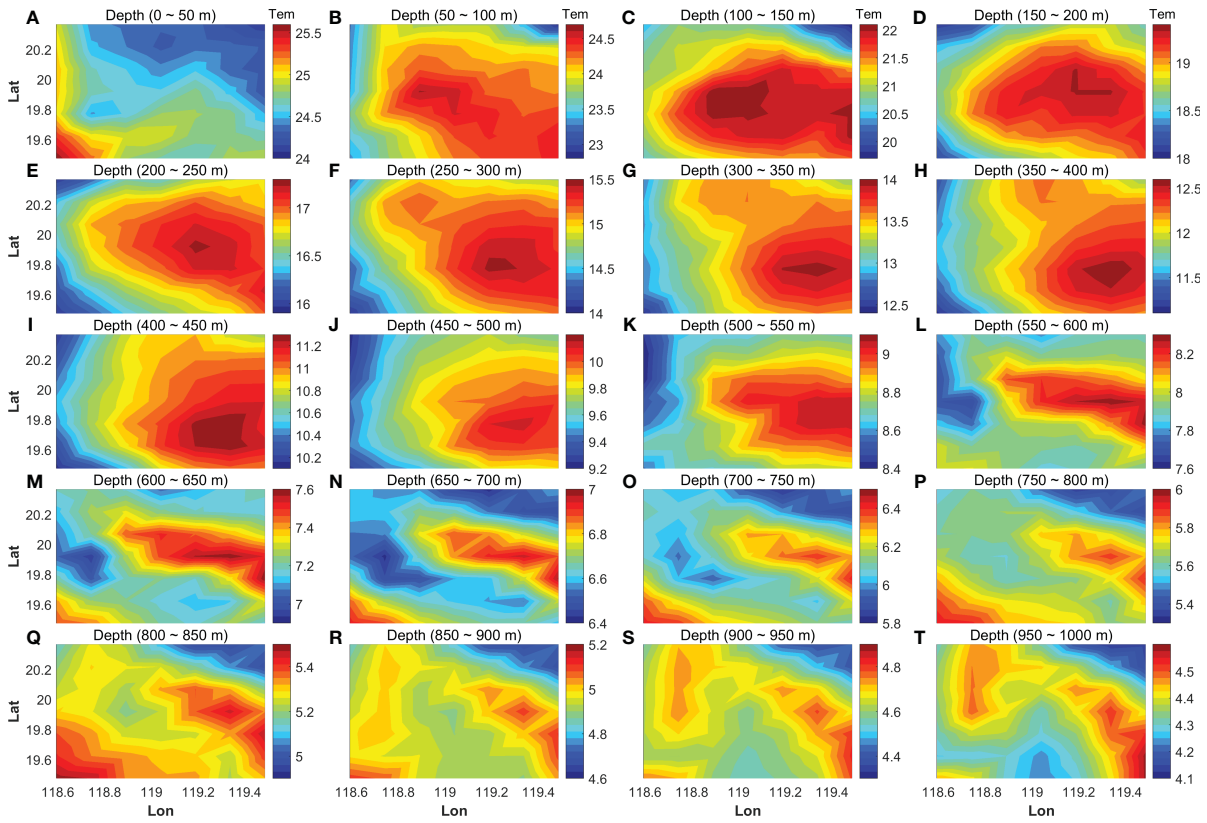


FIGURE 4
The horizontal-slice depth-average temperature distribution every 50 m with the 7 × 7 grid. The depth information is indicated at the top of each figure **(A–T)**. The color bar of temperature is also indicated at the right of each figure.

$$\delta\tau_i = \tau_i - \tau_{0i} \approx \int_{\Gamma_i} \frac{\delta C(x,z)}{C_0(z)^2} dL \tag{1}$$

When sound is transmitted from a source, it propagates in a stratified ocean; refracted rays passing through different depths are constructed, and various travel times are obtained from a receiver. For the vertical-slice inversion, the tomographic domain is decomposed into N depth layers such that the travel time deviation (Equation 1) is reduced to a discrete form as follows:

$$\delta\tau_i \approx -\sum_{z=1}^{N} \frac{l_{iz}\delta\overline{C_z(x)}}{C_{0z}^2} \tag{2}$$

where l_{iz} represents the actual length of the i-th ray through the z-th layer and C_{0z} and $\delta\overline{C_z(x)}$ represent the reference sound speed and the average sound speed deviation of the z-th layer, respectively.

The purpose of vertical-slice inversion was to reconstruct the vertical distribution of the layer-average sound speed deviations ($\delta\overline{C_z(x)}$) using the travel time deviation as known variables. In this study, we considered a method in which the depth layer is decomposed into 4000 sublayers with an interval of 1 m. Because of the limited travel time information, only the sound speed deviation for typical layers was calculated (Taniguchi et al., 2013; Dai et al., 2023).

The EOF decomposition method was introduced to reduce the number of unknown inversion variables. In this study, the first six EOF modes (M=6) for the sound speed deviation were introduced. Subsequently, Equation (2) was transformed into

$$\delta\tau_i \approx -\sum_{z=1}^{N}\sum_{j=1}^{J} \frac{l_{iz}M_z(j)}{C_{0z}^2} a_j \tag{3}$$

where $M_z(j)$ is the value of the j-th EOF mode crossing the z-th layer, and a_j is the coefficient for the j-th EOF mode. The goal of formulation is to determine the unknown variables (a_j) at every observation time when travel time deviation ($\delta\tau_i$) is acquired. Finally, the vertical slice distribution of the average sound speed deviation was reconstructed using the following formula:

$$\delta\overline{C_z(x)} = M_z(j)a_j \quad (\text{for j} = 1, 2 \ldots 6) \tag{4}$$

As a result of applying the EOF mode method, the number of unknown variables (the coefficient number of the EOF modes) was considerably reduced.

3.2 Horizontal slice

The tomography domain was surrounded by five, seven, and nine acoustic stations, typical cases for constructing ten, twenty-one, and thirty-six transmission paths, respectively. The horizontal-layered distribution of the sound speed deviation obtained in the vertical-slice inversion and the lengths of the transmission paths crossing individual grids were used in the horizontal-slice inversion as known variables. The formulation for the horizontal-slice inversion is represented as follows:

$$\overline{\delta C_z(x)} = \sum_{x=1}^{M} \frac{l_{iz}(x)\delta C_z(x)}{L_{iz}} \quad (5)$$

where $\overline{\delta C_z(x)}$ is the layer-average sound speed deviation for the x-th grid at the z-th layer obtained from Equation (4), $l_{iz}(x)$ is the length of the i-th ray projected onto the z-th horizontal layer crossing the x-th grid, $\delta C_z(x)$ is the sound speed deviation of the x-th grid at the z-th layer, and L_{iz} is the length of the i-th projected ray at the z-th layer.

Three-dimensional mesoscale sound speed fields were reconstructed by combining the vertical and horizontal slice inversions. The simulation region was divided into a depth range of 50 m from the surface to 1000 m. To better describe mesoscale eddies, the three-dimensional distribution of sound speed was converted into a three-dimensional distribution of temperature according to the correlation formula using the sound speed deviation and taking depth as a variable. From the surface to a depth of 1000 m at an interval of 50 m, a change of 1 ms^{-1} in sound speed was equivalent to temperature changes of 0.5–0.2 °C.

4 Inversion

Equations (3) and (5), which correspond to the vertical- and horizontal-slice inversions, respectively, are expressed in matrix form as follows:

$$y = Ex + n \quad (6)$$

where y is the simulated data set vector, x is the unknown variable vector, E is the transform matrix, and n is the noise vector.

In this study, the tapered least-squares method was adopted to obtain the optimal solution. In the tapered least squares method, we can express the objective function as follows:

$$J = (y - Ex)^T(y - Ex) + \alpha^2 x^T x \quad (7)$$

where α is the damping parameter. The expected solution $\tilde{x}$ is obtained, so that to minimize the objective function as follows:

$$\tilde{x} = (E^T E + \alpha^2 I)^{-1} E^T y \quad (8)$$

The L-curve method developed by Hansen and O'Leary (1993) was used to determine the optimal value of α. Consequently, the optimal solution of the sound speed deviation fields was obtained more flexibly than with stochastic inversion, which requires the covariance of the expected solution prior to inversion.

5 Results

Figure 5 shows the horizontal-slice contour maps of temperature fields with warm eddy anomaly, reconstructed by inversion for the five-station number. The results of 20 depth

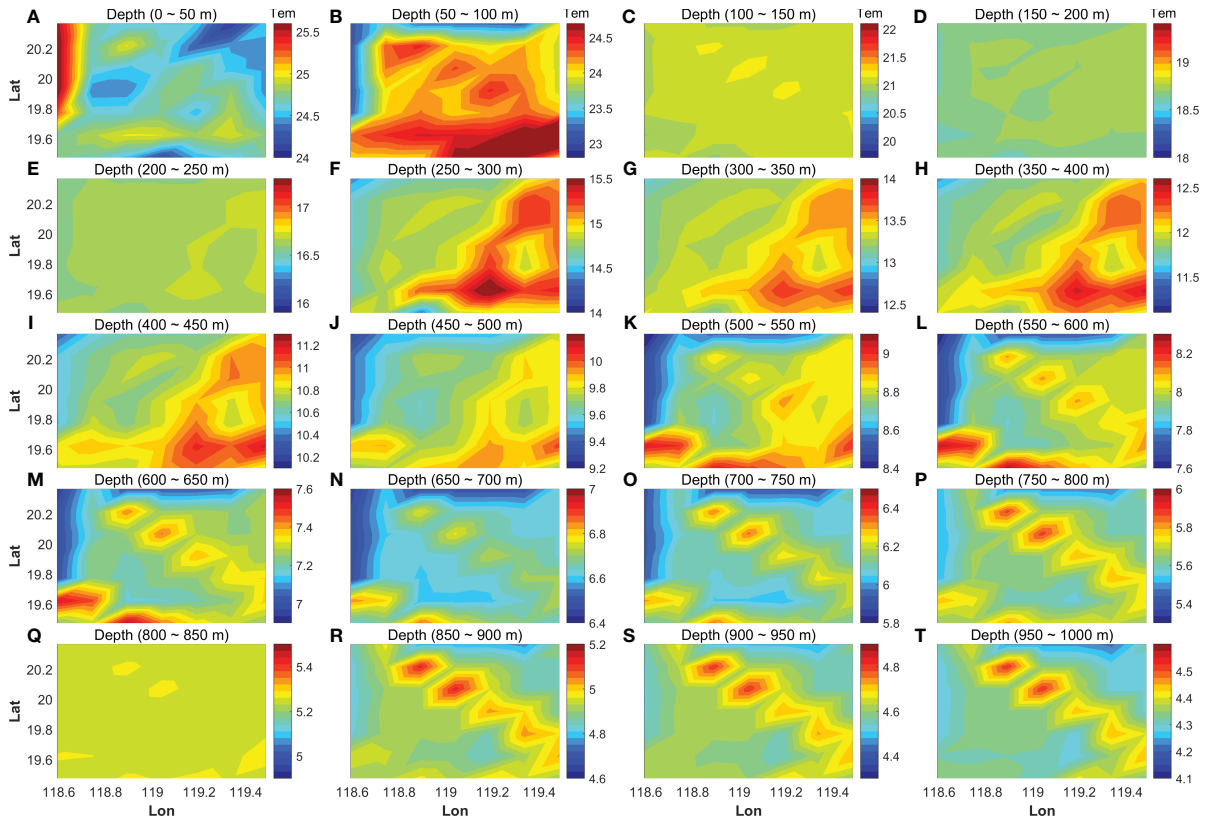


FIGURE 5
Contour maps of the horizontal-slice temperature fields, reconstructed for five acoustic stations. The depth range is indicated at the top of each figure **(A–T)**. The color bar of temperature is also indicated at the right of each figure.

averaged temperature fields at a 50 m interval from the surface to a depth of 1000 m are shown in Figures 5A–T, 6A–T, 7A–T show the contour maps of temperature for seven- and nine-station cases, respectively. The mesoscale anomaly was intensified in the depth layers of 100–600 m, as visible in the 50 m-interval depth-average temperature field (Figure 4), and rapidly diminished with depth in the layers deeper than 600 m. The core of the anomaly, which corresponded to the mesoscale eddies existing in the depth range of 50–600 m, was not clear in Figure 5 (five-station case). The cores were reconstructed with almost the same horizontal positions as Figure 4 in the depth range of 100–550 m for Figure 6 (seven-station case) and in the depth range of 150–500 m for Figure 7 (nine-station case).

6 Discussion

6.1 Correlation coefficient and root mean squares difference

To evaluate the performance of the tomographic inversion, the inverted temperature at the m-th grid point of each layer was compared with the HYCOM data using two indices, CCOE and RMSD, as follows:

$$CCOE = \frac{\sum_{i=1}^{t}(T_{hycom} - \overline{T_{hycom}})(T_{inv} - \overline{T_{inv}})}{\sqrt{\sum_{i=1}^{t}(T_{hycom} - \overline{T_{hycom}})^2 \sum_{i=1}^{t}(T_{inv} - \overline{T_{inv}})^2}} \quad (9)$$

$$RMSD = \sqrt{\frac{1}{t}\sum_{t=1}^{t}(T_{inv} - T_{hycom})_2} \quad (10)$$

where t is the time and T_{hycom} and T_{inv} represent the HYCOM and inverted temperatures at each layer, respectively. $\overline{T_{hycom}}$ and $\overline{T_{inv}}$ represent the average values of HYCOM and the inverted temperatures over all grid points in each layer, respectively.

6.2 Temporal change of horizontal-slice inversion results

The CCOE (Equation 9) and RMSD (Equations 10) for the HYCOM data and the inversion results averaged over the simulation domain were calculated for the three acoustic stations. Figure 8 shows the time plot of the comparison results for the entire year of 2011 in the second layer, where the sound speed (temperature) anomaly had the largest value. Within one year, the temperature varied from 20.0 °C to 26.5 °C. In general, all three sets of temporal data varied with similar tendencies. The CCOE and RMSD were 0.980 and 0.297 °C for the five-station case, 0.991 and

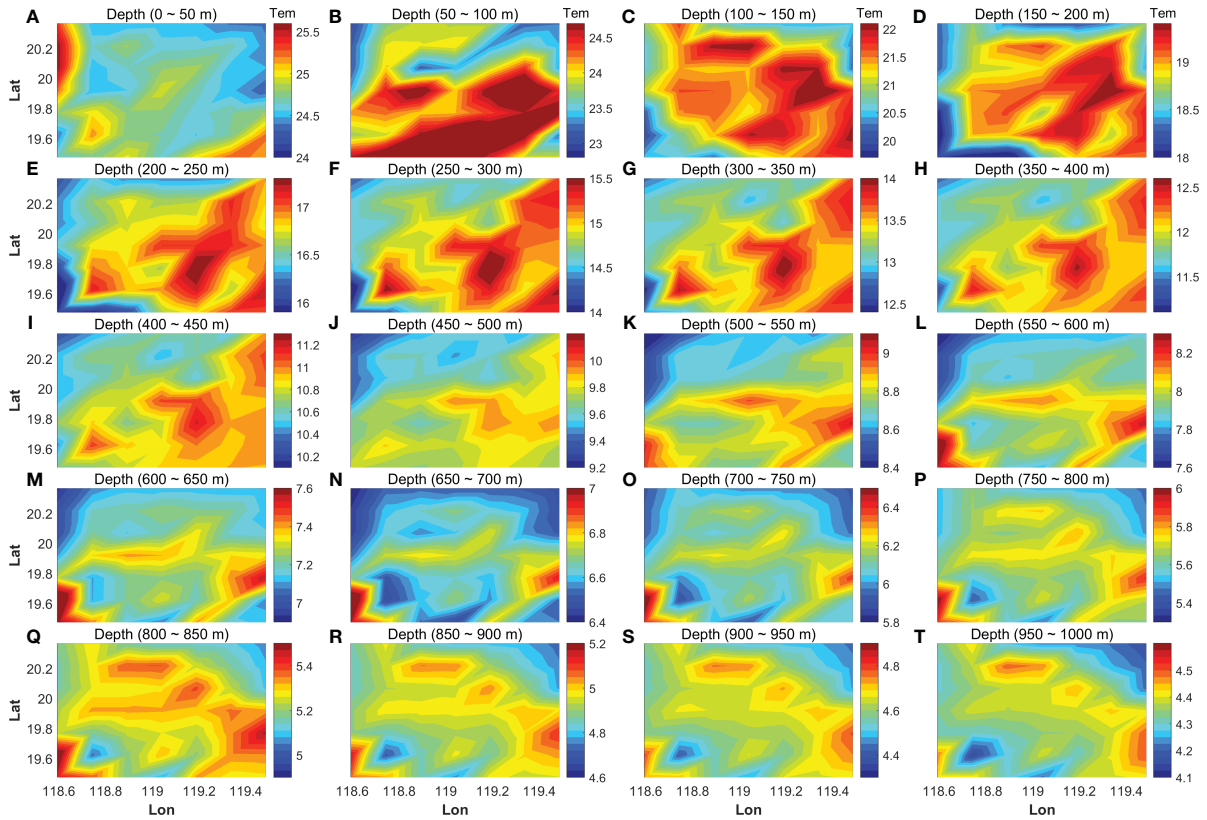


FIGURE 6
Contour maps of the horizontal-slice temperature fields, reconstructed for seven acoustic stations. Others are similar to Figure 5.

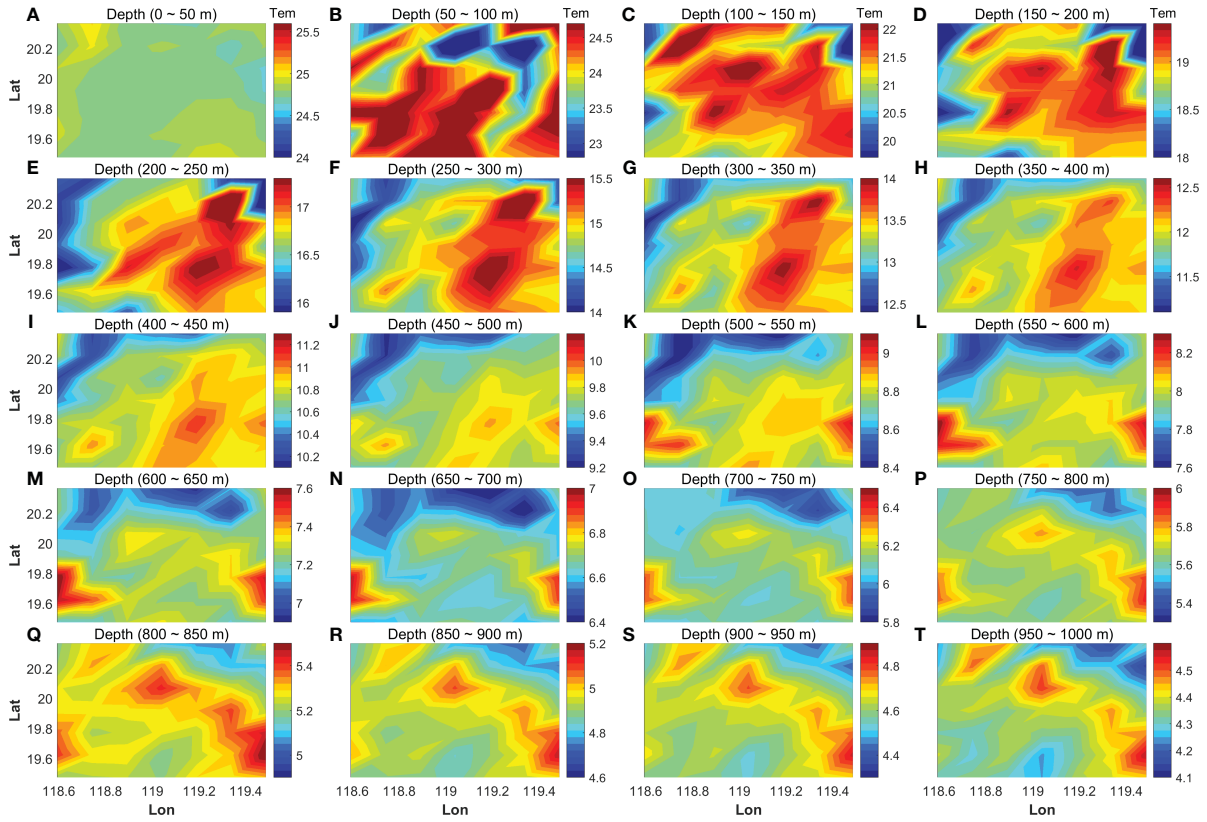


FIGURE 7
Contour maps of the horizontal-slice temperature fields, reconstructed for nine acoustic stations. Others are similar to Figure 5.

FIGURE 8
Time series of the 2011 HYCOM data (red line) and inversion results for five-station (black line), seven-station (blue line), and nine-station (green line) cases in the second layer. CCOE and RMSD are indicated in the lower right of the figure.

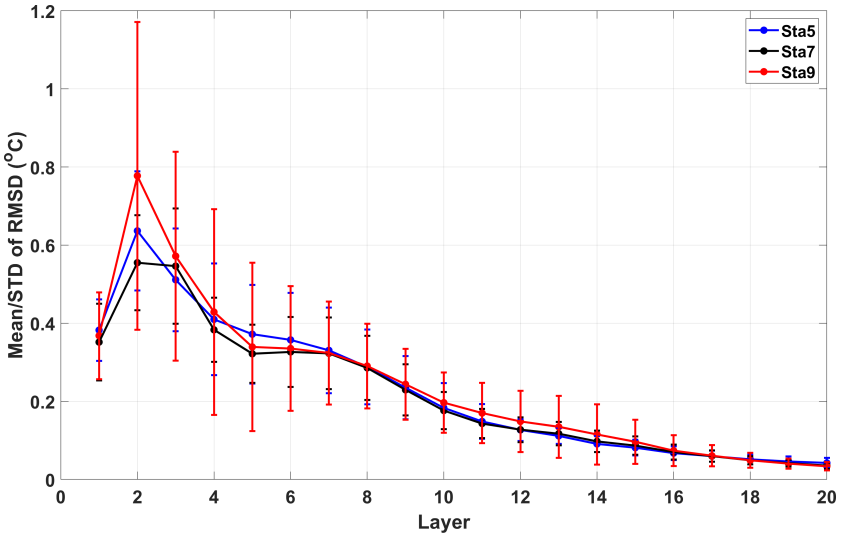


FIGURE 9
Layer dependence of the mean and STD of RMSD for the HYCOM and inverted data at each depth layer. The dots and vertical bars indicate the mean and STD, respectively. The results for five-, seven-, and nine-station cases are indicated with the blue, black, and red colors as simulated in the legend at the upper right of the figure.

0.197 °C for the seven-station case, and 0.989 and 0.220 °C for the nine-station case, respectively. The RMSD was the smallest for the seven-station case, whereas the nine-station case had the highest horizontal resolution of the three cases.

6.3 Performance of inversion

Statistical analyses of the RMSD at each layer provided an important indicator of the inversion performance. In the following analysis, the entire depth layer, from the surface to a depth of 1000 m, was divided into 20 sublayers every 50 m.

The mean and standard deviation (STD) of the RMSD were calculated for every grid of each depth layer, implying that the inversion accuracy varied with depth (Figure 9). The mean and STD of the RMSD were maximal in the second layer around the main thermocline. Its value was 0.64 °C for the five-station case, 0.55 °C for the seven-station case, and 0.78 °C for the nine-station case, showing that the seven-station case had a minimum value of mean. The results also showed that the nine-station case had the maximum value of the STD around the second layer. The mean and STD of the RMSD showed a decreasing trend with the number of depth layers. Therefore, the performance of inversion was best for the seven-station case rather than for the nine-station case, with the highest resolution of the three cases.

TABLE 1 Number of stations, number of transmission paths, and horizontal-slice spatial resolutions calculated for the three cases of station configurations.

Number of stations	Number of transmission paths	Horizontal Resolution (km)
5	10	34.0
7	21	23.4
9	36	17.9

The accuracy of horizontal-slice inversion depends on the spatial resolution (Park and Kaneko, 2001; Zhang et al., 2017). The number of stations determines the number of transmission paths required. The spatial resolution was formulated using the area (A) of the simulation domain and the number (N_r) of acoustic transmission paths, as follows:

$$L_{Hr} = \sqrt{\frac{A}{N_r}} \tag{11}$$

where A = 104 × 111 km =11,544 km^2. For the three station configurations on a horizontal slice, the spatial resolutions were calculated using Equation (11) and are listed in Table 1.

In this study, the results for the mesoscale anomaly inversion showed that the inversion performance for the seven-station case was better than that for the nine-station case, with a higher spatial resolution. This means that the accuracy of three-dimensional inversion depends on the horizontal-slice spatial resolution as well as the vertical-slice spatial resolution. To illustrate the vertical-slice inversion performance more directly, the RMSD for the vertical-slice inverted temperature of the full water depth and HYCOM data were plotted against station distance and ray number (Figure 10). When the distance between stations increased from 30 km to 100 km, the RMSD for the vertical-slice inversion significantly decreased with increasing distances. For distances greater than 100 km, the RMSD was approximately constant (Figure 10A). Similarly, the RMSD for the vertical-slice inversion decreased as the number of acoustic rays increased (Figure 10B). With increasing distance, oceanic signals with smaller length scales were filtered out, increasing the smoothness of the inversion. In addition, the larger the number of refracted rays, the smaller the shadow zone, resulting in a higher vertical-slice resolution.

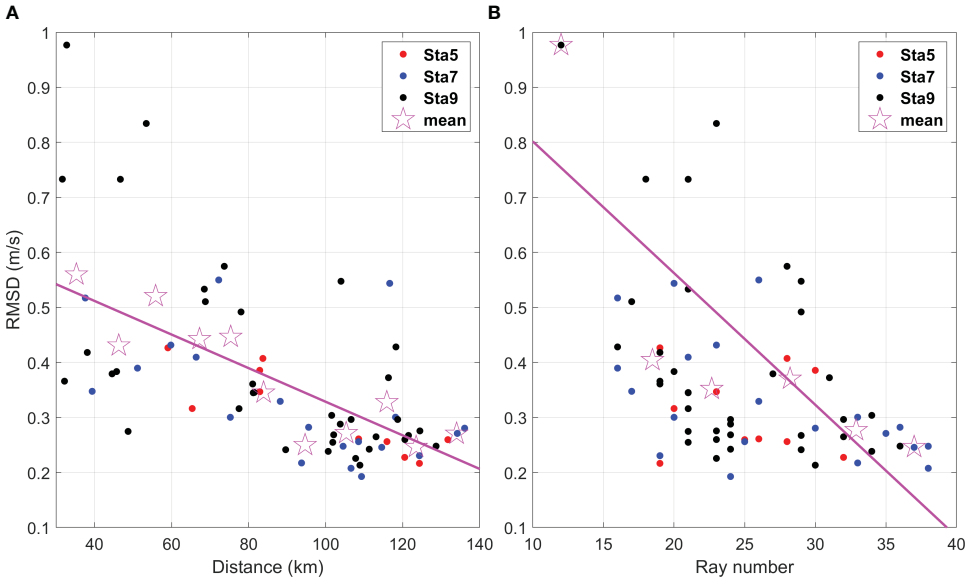


FIGURE 10
RMSDs for the inverted and HYCOM temperatures, plotted against **(A)** the station distance and **(B)** the ray number. The pink straight line is the regression line in each figure. The open star marks indicate the mean value of RMSD calculated for every horizontal grid box of 10 km and five rays, respectively.

The RMSD was close to a constant value for ray numbers greater than 30. The RMSD of the vertical slice was larger due to the small distance between station pairs of the nine-station case. However, the RMSD was more balanced due to the more appropriate distances in the seven-station case. Significantly, the comparison of temperature profiles from the vertical-slice inversion and HYCOM results for different station configurations during mesoscale eddy period is presented in Supplementary Figure S1 of the Supplementary material.

7 Conclusion

Mesoscale eddies are difficult to measure synchronously using conventional shipboard methods because their horizontal scales are greater than 100 km. This can be achieved through acoustic tomography methods using a network of multiple acoustic stations located at the periphery of the observation domain. By simulating the OAT experiment using HYCOM reanalysis data, we tested three types of OAT station configurations: five, seven, and nine acoustic stations located at the periphery of a 104 × 111 km domain. The simulation fields were in the northern SCS, with mesoscale eddies generated due to the intrusion of the Kuroshio through the Luzon Strait.

A new inversion method was proposed by combining the EOF method in a vertical slice and the grid-segmented method in a horizontal slice. The tapered least-squares method combined with the L-curve method was used in the inversion process. The performance of the inversion was evaluated for the three cases of five, seven, and nine stations using the correlation coefficient (CCOE) and root mean square difference (RMSD) for the inverted and HYCOM reanalysis data. For a 100 × 100 km domain, the seven-station case provided an optimal number to reconstruct the mesoscale eddy phenomena rather than the nine-station case, with the highest horizontal resolution of the three cases. This means that the accuracy of the three-dimensional inversion depends on the horizontal-slice spatial resolution as well as the vertical-slice spatial resolution. Furthermore, the horizontal-slice spatial resolution, ray number, and station distance affect the vertical-slice inversion accuracy. The fewer the number of refracted rays, the greater the shadow zone. A longer station distance smoothens out oceanic phenomena at smaller scales.

In this study, we proposed ocean acoustic tomography as an underwater remote sensing technique that is fully capable of observing mesoscale eddies.

Data availability statement

The HYCOM reanalysis data can be downloaded online (https://hycom.org/dataserver/gofs-3pt0/reanalysis). The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

CZ: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. Z-NZ: Methodology, Writing – review & editing. CX: Methodology, Writing – review & editing. X-HZ: Conceptualization, Writing – review & editing. Z-JL: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study

was sponsored by the National Key Research and Development Program of China (2021YFC3101502), the National Natural Science Foundation of China (Grants 41920104006, 41976001 and 52071293), the Scientific Research Fund of Second Institute of Oceanography, MNR (Grant JZ2001), Zhejiang Provincial Natural Science Foundation of China (LY24D060002), and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311020004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1350337/full#supplementary-material

References

Adams, D. K., McGillicuddy, D. J., Zamudio, L., Thurnherr, A. M., Liang, X., Rouxel, O., et al. (2011). Surface-generated mesoscale eddies transport deep-sea products from hydrothermal vents. *Science* 332, 580–583. doi: 10.1126/science.1201066

Chen, G., Hou, Y., and Chu, X. (2011). Mesoscale eddies in the South China Sea: Mean properties, spatiotemporal variability, and impact on thermohaline structure. *J. Geophys. Res.* 116, 102–108. doi: 10.1029/2010JC006716

Chu, X., Chen, G., and Qi, Y. (2020). Periodic mesoscale eddies in the South China sea. *J. Geophys. Res. Ocean* 125, e2019JC015139. doi: 10.1029/2019JC015139

Cornuelle, B., Wunsch, C., Behringer, D., Birdsall, T., Brown, M., Heinmiller, R., et al. (1985). Tomographic maps of the ocean mesoscale. Part 1: pure acoustics. *J. Phys. Oceanogr.* 15, 133–152. doi: 10.1175/1520-0485(1985)015<0133:TMOTOM>2.0.CO;2

Dai, L., Xiao, C., Zhu, X.-H., Zhu, Z.-N., Zhang, C., Zheng, H., et al. (2023). Tomographic reconstruction of 3D sound speed fields to reveal internal tides on the continental slope of the South China Sea. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1107184

Fukumori, I., and Wunsch, C. (1991). Efficient representation of the North Atlantic hydrographic and chemical distributions. *Prog. Oceanogr* 27, 111–195. doi: 10.1016/0079-6611(91)90015-E

Hansen, P., and O'Leary, D. P. (1993). The use of the L-curve in the regularization of discrete ill-posed problems. *Siam J. Sci. Comput.* 14 (6), 1487–1503. doi: 10.1137/0914086

Howe, B. M., Worceser, P. F., and Spindel, R. C. (1987). Ocean acoustic tomography: mesoscale velocity. *J. Geophys. Res.* 92, 3785–3805. doi: 10.1029/JC092iC04p03785

Kaneko, A., Zhu, X.-H., and Ju Lin, J. (2020). *Coastal Acoustic Tomography* (Amsterdam, Netherlands: Elsevier). doi: 10.1016/B978-0-12-818507-0.00003-2

LeBlanc, L., and Middleton, F. H. (1980). An underwater acoustic sound velocity data model. *J. Acoust. Soc Am.* 67, 2055–2062. doi: 10.1121/1.384448

Liu, Q., Kaneko, A., and Su, J. (2008). Recent progress in studies of the South China Sea circulation. *J. Oceano.* 64, 753–762. doi: 10.1007/s10872-008-0063-8

MacKenzie, K. V. (1981). Nine-term equation for sound speed in the ocean. *J. Acoust. Soc Am.* 70, 807–812. doi: 10.1121/1.386920

Munk, W., Worceser, P. F., and Wunsch, C. (1995). *Ocean acoustic tomography* (Cambridge: Cambridge Univ. Press). 433. doi: 10.1017/CBO9780511666926

Munk, W., and Wunsch, C. (1979). Ocean acoustic tomography: a scheme for large scale monitoring. *Deep-Sea Res.* 26A, 123–161. doi: 10.1016/0198-0149(79)90073-6

Nan, F., He, Z., Zhou, H., and Wang, D. (2011). Three long-lived anticyclonic eddies in the northern South China Sea. *J. Geophys. Res.* 116, C05002. doi: 10.1029/2010JC006790

Park, Y., Jeon, C., Song, H., Choi, Y., Chae, J.-Y., Lee, E.-J., et al. (2021). Novel method for the estimation of vertical temperature profiles using a coastal acoustic tomography system. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.675456

Park, J.-H., and Kaneko, A. (2001). Computer simulation of the coastal acoustic tomography by a two-dimensional vortex mode. *J. Oceangr.* 57, 593–602. doi: 10.1023/A:1021211820885

Syamsudin, F., Taniguchi, N., Zhang, C., Hanifa, A. D., Li, G., Chen, M., et al. (2019). Observing internal solitary waves in the Lombok Strait by coastal acoustic tomography. *Geophys. Res. Lett.* 46, 10475–10483. doi: 10.1029/2019GL084595

Taniguchi, N., Huang, C., Kaneko, A., Liu, C., Howe, B. M., Wang, Y., et al. (2013). Measuring the Kuroshio Current with ocean acoustic tomography. *J. Acoust. Soc Am.* 134, 3272–3281. doi: 10.1121/1.4818842

Wang, X., Du, Y., Zhang, Y., Wang, T., Wang, Mi., and Jing, Z. (2022). Subsurface anticyclonic eddy transited from Kuroshio shedding eddy in the northern South China Sea. *J. Phys. Oceanogr.* 53 (3), 841–861. doi: 10.1175/JPO-D-22-0106.1

Wang, G., Su, J., and Chu, P. C. (2003). Mesoscale eddies in the South China Sea observed with altimeter data. *Geophys. Res. Lett.* 30, 2121–1226. doi: 10.1029/2003GL018532

Wang, D., Xu, H., Lin, J., and Hu, J. (2008). Anticyclonic eddies in the northeastern South China Sea during winter 2003/2004. *J. Oceangr.* 64 (6), 925–936. doi: 10.1007/s10872-008-0076-3

Yuan, G., Nakano, I., Fujimori, H., Nakamura, T., Kamoshida, T., and Kaya, A. (1999). Tomographic measurements of the Kuroshio Extension meander and its associated eddies. *Geophys. Res. Lett.* 26, 79–82. doi: 10.1029/1998GL900253

Zhang, C., Kaneko, A., Zhu, X.-H., and Gohda, N. (2015). Tomographic mapping of a coastal upwelling and the associated diurnal internal tides in Hiroshima Bay, Japan. *J. Geophys. Res.Oceans* 120, 4288–4305. doi: 10.1002/2014JC010676

Zhang, C., Zhu, X.-H., Zhu, Z.-N., Liu, W., Zhang, Z.-Z., Fan, X.-P., et al. (2017). High-precision measurement of tidal current structures using coastal acoustic tomography. *Estuar. Coast. Shelf Sci.* 193, 12–24. doi: 10.1016/j.ecss.2017.05.014

Zheng, H., Gohda, N., Noguchi, H., Ito, T., Yamaoko, H., Tamura, T., et al. (1997). Reciprocal sound transmission experiment for current measurement in the Seto Inland Sea, Japan. *J. Oceanogr.* 53, 117–127.

Zhu, X.-H., Kanko, A., Wu, Q.-S., Zhang, C., Taniguchi, N., and Gohda, N. (2013). Mapping tidal current structures in zhitouyang bay, China, using coastal acoustic tomography. *IEEE J. Oceanic Eng.* 38, 285–296. doi: 10.1109/JOE.2012.2223911

Zhu, Z.-N., Zhu, X.-H., Guo, X., Fan, X., and Zhang, C. (2017). Assimilation of coastal acoustic tomography data using an unstructured triangular grid ocean model for water with complex coastlines and islands. *J. Geophys. Res.Oceans* 122, 7013–7030. doi: 10.1002/2017JC012715

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 26 February 2024
DOI 10.3389/fmars.2024.1362335

OPEN ACCESS

EDITED BY
Arata Kaneko,
Hiroshima University, Japan

REVIEWED BY
Ze-Nan Zhu,
Ministry of Natural Resources, China
Satoshi Fujii,
University of the Ryukyus, Japan

*CORRESPONDENCE
Jae-Hun Park
jaehunpark@inha.ac.kr

RECEIVED 28 December 2023
ACCEPTED 08 February 2024
PUBLISHED 26 February 2024

CITATION
Hwang Y, Lee E-J, Song H, Kim B-N, Ha HK, Choi Y, Kwon J-I and Park J-H (2024) Estimating three-dimensional current fields in the Yeosu Bay using coastal acoustic tomography system.
Front. Mar. Sci. 11:1362335.
doi: 10.3389/fmars.2024.1362335

Estimating three-dimensional current fields in the Yeosu Bay using coastal acoustic tomography system

Yerin Hwang[1], Eun-Joo Lee[1], Hajin Song[1], Byoung-Nam Kim[2], Ho Kyung Ha[1], Yohan Choi[3], Jae-Il Kwon[4] and Jae-Hun Park[1]*

[1]Department of Ocean Sciences, Inha University, Incheon, Republic of Korea, [2]Marine Domain & Security Research Department, Korea Institute of Ocean Science and Technology, Busan, Republic of Korea, [3]Metocean 1team, BLUE Division, Underwater Survey Technology 21, Incheon, Republic of Korea, [4]Coastal Disaster and Safety Research Department, Korea Institute of Ocean Science and Technology, Busan, Republic of Korea

Observation of current speeds in coastal seas is crucial because it can provide useful information for ship operations, fishing activities, and rapid responses to marine disasters. Coastal acoustic tomography (CAT) is a technology that can continuously monitor environmental changes such as current velocity and water temperature using reciprocal acoustic signals between CAT stations in coastal seas. This technology is different from traditional pointwise or intermittent sectional observations in that it can produce time-varying two- or three-dimensional current fields. The results of previous studies using CAT systems have been limited to reproducing horizontal maps of depth-averaged two-dimensional current fields. Utilizing results from a high-resolution coastal ocean model, this study developed a novel technique for estimating three-dimensional (3-D) current fields by combining the inverse method with an artificial intelligence (AI) model. Following three steps are the procedure for the test of estimating the 3-D current fields. First, utilizing the ray tracing model 'Bellhop,' reciprocal travel times among five CAT stations using the coastal ocean model outputs are computed. These five stations correspond to the locations where *in-situ* CAT systems were established for continuous monitoring of current changes in Yeosu Bay, Korea. Subsequently, the range-averaged currents at the five layers were estimated by incorporating this travel time difference data into an AI model trained using the same coastal ocean model outputs. Finally, the inverse method is applied to each layer to estimate the 3-D current fields. The validation results revealed that the newly developed method performed well in both summer and winter. Time-varying two-layer-like current fields were reasonably produced, occasionally revealing an out-of-phase relationship between the upper and lower layers depending on the tidal phases. This method yielded average root-mean-squared errors of less than 4 cm/s on six simulation paths for acoustic signal propagation. Furthermore, when

the same method was applied to *in-situ* CAT observations, the average correlation coefficient (R) of the along-channel current of each layer was found to be approximately 0.9 or higher. These results suggest that this novel method can be effectively applied to the continuous monitoring of 3-D current fields in coastal seas using a CAT system.

KEYWORDS

coastal acoustic tomography, inverse method, three-dimensional current field estimation, empirical orthogonal function, artificial intelligence model

1 Introduction

1.1 Background

Coastal acoustic tomography (CAT) is an emerging technology designed to monitor coastal environments. This technology evolved from ocean acoustic tomography, which was originally developed by Munk and Wunsch (1979), and has been adapted for coastal applications. Unlike traditional *in-situ* current measurement methods such as stationary or intermittent sectional observations, CAT can estimate time-varying temperature and current fields using reciprocally transmitted acoustic signals between CAT stations. This approach is cost-effective and provides valuable observational results for many coastal regions (Kaneko et al., 2020).

Research estimating the current field using CAT has primarily focused on calculating the depth-averaged horizontal two-dimensional current field. Park and Kaneko (2001) presented an inverse method for estimating the current field by applying the L-curve method to CAT data. Subsequently, current field measurements were conducted by applying this inverse method to CAT observations among multiple stations (e.g., Yamoaka et al., 2002; Zhu et al., 2012; Zhang et al., 2017). Additionally, research has been conducted to estimate horizontal current fields considering coastal effects using coast-fitting tomographic inversion in semi-enclosed seas (Chen et al., 2020), as well as assimilating CAT observation data into numerical models to reproduce current fields (Park and Kaneko, 2000; Zhu et al., 2021). As the need for three-dimensional (3-D) current field observations in coastal areas has increased, 3-D current fields have been derived by assimilating CAT observation data into a numerical model with unstructured triangular grids (Zhu et al., 2017). However, 3-D current field estimation from inverse analysis rather than from the data assimilation method using a numerical ocean model with a large number of calculations and a complex calculation procedure has not been reported so far. Kaneko et al. (2020) proposed a method for 3-D mapping of the current field from the sound speed deviation data of CAT through a two-step inversion procedure from vertical to horizontal slices. This method is feasible when multi-ray identification of 2^{nd} or 3^{rd} rays which pass through multiple layers along the sound transmission path is possible. However, because this is almost impossible in coastal areas, where the distances between stations are short and the water depth is shallow, this method cannot be applied to CAT data.

In this study, we combine an artificial intelligence (AI) model and an empirical orthogonal function (EOF) with an existing inverse method to develop a new 3-D current field estimation method. This method has the advantage of fully reflecting the current pattern of the study area by applying EOF and simultaneously reducing the number of unknowns during inverse analysis, thereby enabling the effective estimation of currents in underdetermined systems with a minimal number of observations. Moreover, by employing a pre-trained AI model, this method allows the rapid estimation of the 2-D current field along the section between two CAT stations using single-ray acoustic observations of CAT. The newly developed method was applied to *in-situ* CAT data to demonstrate its applicability for continuous current field monitoring using the CAT system in coastal seas.

1.2 Study area

The study area was Yeosu Bay, located in the southern part of the Republic of Korea, as shown in Figure 1. Because Yeosu Bay is characterized by shallow depth, complex coastline, and active ship traffic, there has been a growing need for real-time monitoring of current fields in this region. This region is dominated by tidal currents and shows a typical two-layer structure with opposing flows; the upper-layer currents from the estuarine area flow southward, whereas the lower-layer currents flow northward (Pritchard, 1952; Lee and Kim, 2007).

In this study, the target area for current field estimation was selected as a channel with a high current velocity in Yeosu Bay (Figure 1C). Six transmission paths were established by designating two stations on the west (Yeosu side) and three stations on the east (Namhae side) and connecting the stations to the west and east, as presented in Figure 1C. The 'Transmission Path' S1 is between stations K1 and K3, S2 is between K1 and K2, S3 is between K1 and K5, S4 is between K4 and K3, S5 is between K4 and K2, and S6 is between K4 and K5. Hereinafter, the term 'path' follows the definition of 'Transmission Path.' The CAT stations obtained through the high-resolution ocean numerical

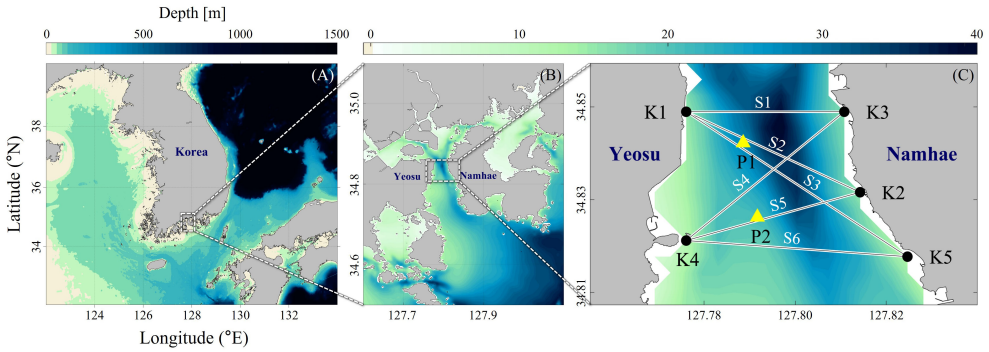


FIGURE 1
Study area in the Yeosu Bay, Korea. **(A)** Map of the Korean Peninsula. **(B)** Map of Yeosu and Namhae located in the South Sea of Korea, including the **(C)** Yeosu Bay. Black dots in **(C)** indicate the CAT stations and the lines connected between the CAT stations are paths for acoustic transmission simulation. Yellow triangles in **(C)** indicate ADCP mooring sites (P1, P2).

model results coincide with the actual locations where *in-situ* CAT observations are being conducted. The *in-situ* observation stations are the land station, except K4 which is located on the barge. This study utilized a numerical ray tracing model to validate the newly developed 3-D current field estimation method. The newly developed method was applied to *in-situ* CAT data. The locations of the five CAT stations and the two Acoustic Doppler Current Profiler (ADCP) mooring sites are listed in Table 1.

2 Data and methods

2.1 Data

The data utilized in this study were data assimilated real-time ocean prediction modeling results from the Korea Operational Oceanography System (KOOS), which was developed by the Korea Institute of Ocean Science and Technology (KIOST). The temporal and spatial resolutions of the data and their durations are summarized in Table 2. The depth, current, sea level, temperature, and salinity results from the KOOS model were used as input data for the Bellhop ray tracing model (Porter, 2011). The current data from the KOOS model were used for the EOF analysis, and the eigenvectors derived from the EOF analysis were used for inverse estimation. Additional information regarding the input data used to train the AI model is presented in Supplementary Figures 1, 2.

The *in-situ* application of the method was performed from September to October 2023 and validated against the ADCP mooring data at P1 and P2 (Figure 1C). The observation periods are presented in Table 2. The temporal resolutions were 20 min and 1 h, respectively, which are finer than that of the KOOS model; however, because the inverse method was built using the KOOS model, the *in-situ* application was also performed at 3-hour intervals.

2.2 Methods

In this study, the 3-D current field was estimated using the following three procedures. First, the reciprocal travel time difference (Δt) of each transmission path is computed through the ray tracing simulation. Second, $\Delta t'$ for each layer and path are computed using the AI model, which is trained with Δt. Finally, inverse analysis was applied to each layer, resulting in a 3-D current field consisting of five layers.

2.2.1 Ray tracing simulation

The first step, the ray tracing simulation, uses a numerical ray tracing model 'Bellhop' to calculate the reciprocal travel time difference (Δt). The Bellhop model requires the following files as inputs: '.bty', '.ati', '.env', and '.ssp'. The '.bty' and '.ati' files contain information on the topography and water level of the simulation

TABLE 1 Locations of CAT stations and ADCP mooring sites.

CAT stations	Latitude (°N)	Longitude (°E)	ADCP Mooring Sites	Latitude (°N)	Longitude (°E)
K1	34.8477	127.7755	P1	34.8423	127.7886
K2	34.8314	127.8144			
K3	34.8492	127.8113	P2	34.8262	127.7917
K4	34.8220	127.7778			
K5	34.8164	127.8278			

TABLE 2 Summary of the data used in this study.

Type		Period (MM/YYYY)	Resolution	
			Temporal	Spatial
Model	Ray tracing simulation	05/2019–12/2022	3 hours	300 m (horizontal) 14 m (vertical)
	EOF for Inverse estimation	01/2022–12/2022		
	Validation of Inverse estimation	01–02/2022 and 07–08/2022		
Observation	*In-situ* CAT data	09–10/2023	20 min	.
	ADCP mooring data	09–10/2023	1 hour	0.5 m (vertical)

path. The '.ssp' file contains the sound speed and current along the transmission path. The sound speed used here was calculated using temperature and salinity using the equation from Del Grosso (1974). The '.env' file contains parameters such as the depth of the source and receiver, number of beams, and launching degree. The Bellhop model uses these input data to output the travel time (t_i) between two stations from the current [$u(x,y,z,t)$] and sound speed [$C(x,y,z,t)$].

$$t_i = \oint_{L_i} \frac{ds}{\boldsymbol{C}(x,y,z,t) + \boldsymbol{u}(x,y,z,t) \cdot \boldsymbol{n}}$$

If the Bellhop model is performed for both directions, we get the reciprocal travel time (t^+, t^-) from '. ray' file. The final output of the ray tracing simulation is reciprocal travel time difference ($\Delta t = t^+ - t^-$). The ray tracing simulation is performed for each of the six transmission paths to obtain a time series of six reciprocal travel time differences ($\Delta t_1 - \Delta t_6$). Supplementary Figure 1 shows a detailed flowchart including the input and output data of the Bellhop model and the final output. Supplementary Figure 3 shows time series of ray tracing. 'Calculated v' is the current made from the ray tracing results, and 'model v' is the KOOS model current data. They show similar trends of mean velocity, implying that Δt is mainly affected by velocity along the path.

2.2.2 AI model

In the second process, the range-averaged currents in the five layers in the vertical section of the six transmission paths were obtained using the AI model. The input data for the AI model are presented in Supplementary Table 1, and the data locations are provided in Supplementary Figure 2. The data duration was approximately four years (May 2019 to April 2023), including the period used for training the AI model. The collected data were preprocessed to normalize and enhance their learning ability. The other designs used for training the AI model are presented in Supplementary Table 2. The design of the "training process" is as follows. The "test set" consisted of January–February and July–August 2022, the periods covered by the inverse analysis. During this period, no learning took place, and only real simulations were conducted. For all periods except the "test set," about 2/3 of the data is "training set," and the remaining 1/3 is "validation set." Independent training is conducted for each of the three non-overlapping "validation sets." The initial values of these models were randomized, and training was performed three times for each model, resulting in nine ensemble model sets. The final model results were obtained by averaging nine ensemble models. This ensemble process provided robust model results. Although the direction-based loss function typically uses the mean squared error (MSE), the model used in this study focuses on learning the upper modes by utilizing the EOF results as learning weights, resulting in the establishment of an optimized direction-based loss function.

The AI model was designed to estimate the principal component (PC) time series of EOF using appropriate AI model layers for input data with different structures and dimensions (see Supplementary Figure 4). To simulate the current caused by the difference in sea level between the southern and northern parts of Yeosu Bay, sea level data were only extracted at the southern and northern boundaries of the domain. Because the two boundaries have different lengths of data (11 and 8 nodes, respectively), a dense layer was utilized to have the same length of nodes, and then applied to a 1-D convolutional long short-term memory (ConvLSTM1D) filter to handle the spatial dimension. The calculation of the sea level difference was not entirely dependent on the neural network, and a subtraction layer was added to directly calculate the difference between the two lines. The dimension of sea level data passed through ConvLSTM1D is compressed from [time, space, feature] to [time, feature]. Ocean data, atmospheric data, and Bellhop model output data have dimensions of [time, feature], resulting in merging with compressed sea level data. The merged data pass through a dense layer and a hyperbolic tangent (nonlinear activation function), and then pass through an LSTM layer that handles the time series. In this procedure, the time dimension was removed, leaving the [feature] dimension. The tidal-current input field consisting of [latitude, longitude, depth, (U, V)] dimensions were passed through the 3-D convolutional layer and compressed into [feature] dimensions. Finally, the layer is merged with the layer that passes through the LSTM and is compressed into three dense layers to obtain the length of the PC time series (1^{st}–10^{th} modes) of the EOF. The PC time series estimated using this process was then dot-produced using the eigenvectors obtained from the EOF analysis to obtain the current fields along the six vertical sections.

The validation results are presented in Supplementary Table 3 and Supplementary Figures 5, 6. Supplementary Table 3 shows the average root-mean-squared error (RMSE) of the along-path velocity for the six vertical sections compared with the true value, and Supplementary Figure 5 shows the RMSE fields of the six vertical

sections. The bias fields are shown in the same format as in Supplementary Figure 6. The surface layer has a high variability in current and a relatively small number of modeling runs owing to sea level fluctuations, resulting in a high error. Finally, the output of the AI model is the range-averaged current at the five layers in the vertical section of the six transmission paths.

2.2.3 Inverse analysis

The third step is a new method of inverse analysis using EOF. Dividing the distance of each transmission path by the range-averaged current at five layers in the vertical section for the six transmission paths obtained from the AI model, a matrix Y_{ik} consisting of $\Delta t'$ for each layer and path is obtained as follows:

$$\boldsymbol{Y_{ik}} = \begin{bmatrix} \Delta t'_{11} & \cdots & \Delta t'_{1k} \\ \vdots & \ddots & \vdots \\ \Delta t'_{i1} & \cdots & \Delta t'_{ik} \end{bmatrix}$$

where i and k represent the six transmission paths and five layers, respectively. Using this matrix, inverse analysis yields the horizontal current fields for each layer.

The EOF analysis results using the KOOS 3-D current fields are shown in Figure 2. The KOOS model data for Yeosu Bay, which originally comprised 12 horizontal layers, were averaged into five layers. The layers were categorized based on the spatial pattern of the current value deviation over time. 1st and 2nd layers show the greatest variation across the domain; therefore, to reflect this, we averaged them and used them as Layer 1 of the inverse analysis. Layers 2, 3, 4, and 5 of the inverse analysis were used by averaging the 3rd–5th, 6th–7th, 8th–10th, and 11th–12th layers of the KOOS model, respectively.

Figure 2 shows the eigenvector field and PC time series for the first three modes with significant patterns obtained from EOF analysis. The results are presented for layers 1, 3, and 4, representing the upper, middle, and lower layers, respectively. This is characterized by the dominance of north-south reciprocating components in mode 1. The first five modes were used in the 'E matrix' of inverse analysis. The five modes explained 90.73% and 98.80% of the u- and v-component variance, respectively. The eigenvectors of the five modes are extracted for each transmission path and layer. Using the extracted vectors, 'E matrix' was defined as follows:

$$\boldsymbol{E_{ijk}} = \frac{2}{C_0^2}\int_0^{R_i} \boldsymbol{u_{jk}cos\theta_i + v_{jk}sin\theta_i}\ \ \boldsymbol{ds}$$

where, i, j and k are the paths, modes, and layers, respectively; R_i is the length of each path; and C_0 is the reference sound speed. θ_i is the angle between each transmission path and the x-axis. Then, E_{ijk}, Y_{ik}, the unknown matrix $\boldsymbol{X}$, and error $\boldsymbol{e}$ have the following relationships:

$$\boldsymbol{Y = EX + e}$$

Applying the L-curve method to this relationship yields the point at which error($\boldsymbol{e}$) and solution($\boldsymbol{X}$) are optimally balanced (Hansen and O'Leary, 1993; Park and Kaneko, 2001). This inverse analysis for layer 1 (k=1, skipped notation) can be expressed as the following matrix: When this is performed for all five layers, we obtain the $\boldsymbol{X}$ matrix (j*k), which is dot-produced with the eigenvectors to yield a 3-D current field (U_k, V_k) by summing over each mode, as follows:

$$\boldsymbol{X_{kj}} = \begin{bmatrix} a_{11}\cdots a_{1j}\ b_{11}\cdots b_{1j} \\ a_{21}\cdots a_{2j}\ b_{21}\cdots b_{2j} \\ \vdots\ \ \vdots\ \ \vdots\ \ \vdots \\ a_{k1}\cdots a_{kj}\ b_{k1}\cdots b_{kj} \end{bmatrix},$$

$$\boldsymbol{U_k = u_{k1}a_{k1} + u_{k2}a_{k2} + \cdots + u_{kj}a_{kj}}, \textbf{ and } \boldsymbol{V_k = v_{k1}b_{k1} + v_{k2}b_{k2} + \cdots + v_{kj}b_{kj}}.$$

3 Results

3.1 Validation of along-path current of KOOS

First, we validated the KOOS model output data used in this study. The shipboard ADCP data observed along the paths between two stations in the domain were used. The shipboard ADCP observation period and the number of transection observations for each path are listed in Table 3. Observations were conducted during both the spring and neap tidal periods. Comparisons of the along-path-averaged velocities between the observations and the KOOS outputs showed highly correlated features, as shown in Figure 3. The RMSE values were 0.09, 0.12, 0.07 m/s, correlation coefficient (R) values were 0.88, 0.78, and 0.95, and p-values were 0.00, 0.01, 0.05 for S2, S3, and S4, respectively. This confirmed that the current field reproduced by the KOOS model was suitable for this study.

3.2 Validation for three-dimensional current field estimation

3.2.1 Validation for the along-path current velocity

The method presented in Section 2.2 was applied to all five horizontal layers in the domain, resulting in a 3-D current field. The estimated current field was validated by comparison with KOOS outputs. Owing to the characteristics of CAT, sound waves propagating along a path are significantly affected by the along-path current velocity. For this reason, current velocity was converted to the along-path current velocity ($u\ \cos\theta + v\ \sin\theta$), which is used in validation. Therefore, S1 (K1–K3) and S6 (K4–K5), which are nearly zonal to the latitude line, were slightly influenced by the north-south components of the current.

Figure 4 shows a comparison between the KOOS and estimated current fields for the along-path current velocity on S4 (K4–K3). Each figure compares the true values with the inverse estimation results using scatter plots and time-series plots. The results are shown for layers 1, 3, and 4 to present the characteristics of the current fields in the upper, middle, and lower layers, respectively.

10.3389/fmars.2024.1362335

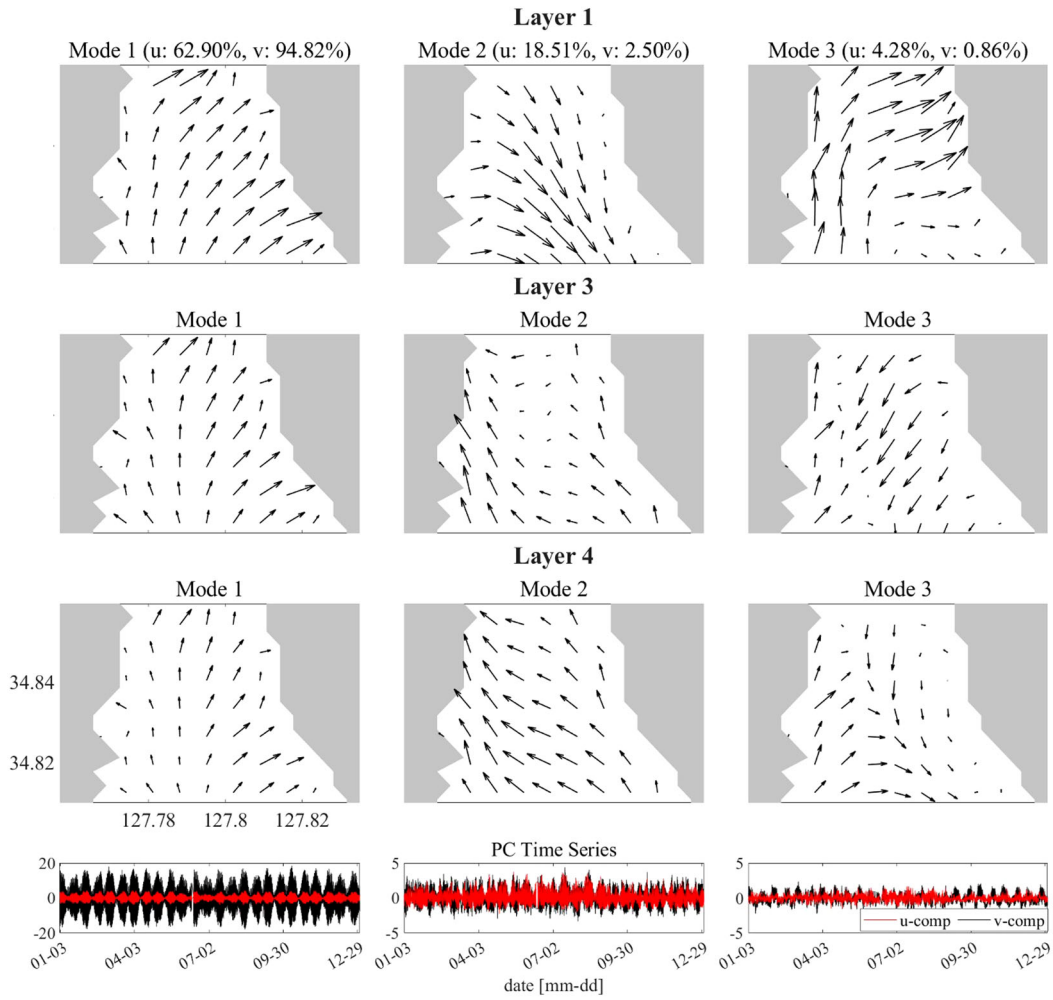


FIGURE 2
Three-dimensional EOF analysis results. (Rows) Eigenvectors for the first 3 modes (columns from left to right) at Layers 1, 3, and 4 (top three rows). Bottom panels represent time series of the principal components for the first 3 modes for u- (red) and v-components (black).

The estimated current field successfully reproduced the tidal variations, including the flood-ebb and spring-neap cycles, for the current velocity of the true values. Table 4 summarizes the average RMSE and R for all six paths and layers 1, 3, and 4. The estimated current field reproduced the KOOS current fields for all paths well, with average RMSE values less than 4 cm/s and average R values exceeding 0.9.

3.2.2 Vertical and horizontal current fields

Yeosu Bay is dominated by the v-component, and the upper and lower layers sometimes exhibit opposite phases, depending on the tidal phase. This can be observed from snapshots of v-component along the section, which are presented in Figure 5. When comparing the true values in the first column with the estimated current fields in the second column, the two-layer structure is reproduced similarly in the contours in the first row. In addition, the flood tidal period-averaged contours show a northward flow in all five layers, and the ebb tidal period-averaged contours show a southward flow, which is very similar between the estimated and true values. The difference (True - estimation) is higher than -3.8 cm/s and lower than 3.2 cm/s. For the other paths, snapshots are not presented because they show patterns like those of S1 (K1–K3).

Figure 6 shows snapshots of the vector representations of the horizontal current field in each layer. Both the southward and northward current periods produced current fields with low errors

TABLE 3 Summary of shipboard ADCP observations.

Period (MM/YYYY)	Path		Number of transection observation
05/2022	S2	K1–K2	6
	S3	K1–K5	4
07/2022–08/2022	S2	K1–K2	6
	S3	K1–K5	4
05/2023	S2	K1–K2	3
	S3	K1–K5	3
	S4	K4–K3	4

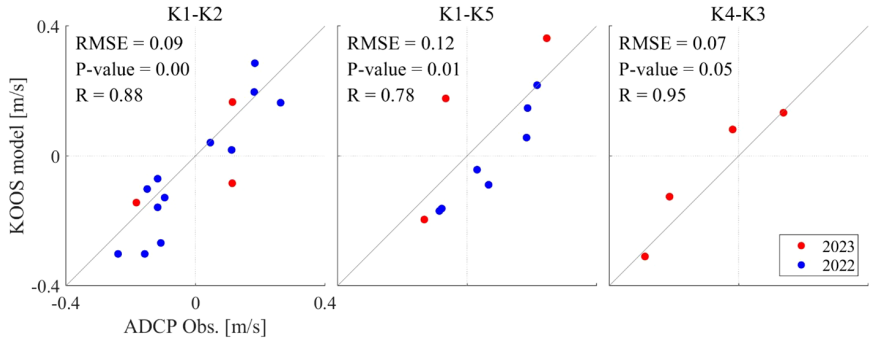


FIGURE 3
Comparison of along-path averaged velocity between ADCP observations and KOOS model outputs. Blue and red dots indicate ADCP observations carried out in 2022 and 2023, respectively. RMSE, R (correlation coefficient) values, and p-values are calculated for the entire period.

across the domain (Figure 6). In addition, the two-layer structure with a southward (northward) current in the upper (lower) layer was reproduced well (Figure 6). Figures 5, 6 present the validation results for the summer season (July and August 2022) when stratification is pronounced and the current structure is relatively complex.

4 Discussion

4.1 RMSE and PVE of three-dimensional current fields

The RMSE fields of the estimated current for each layer during the validation period are shown in Figure 7. The spatially averaged RMSE values for each layer are indicated in parentheses. The error of the u-component is slightly lower than that of the v-component. The errors were somewhat higher in the upper layer than in the middle and bottom layers and were larger in areas where the simulation path did not intersect or at the edges of the domain.

The percent of variance explained (PVE) for each layer indicates the degree to which the estimated current field reproduces the variability compared to the true value. It is calculated using the equation '$PVE = (1- \sigma_{err}^2/\sigma_{true}^2)*100$', where σ_{true}^2 and σ_{err}^2 represent the variance of the true value and the error (true value – calculated value), respectively. Figure 8 shows the PVE for each layer. The values in each title within the parentheses, expressed as percentages, represent the average PVE values within the domain enclosed by the six simulation paths. The average PVE for the u-component ranged from 49.8% to 68.9%, with approximately 10%

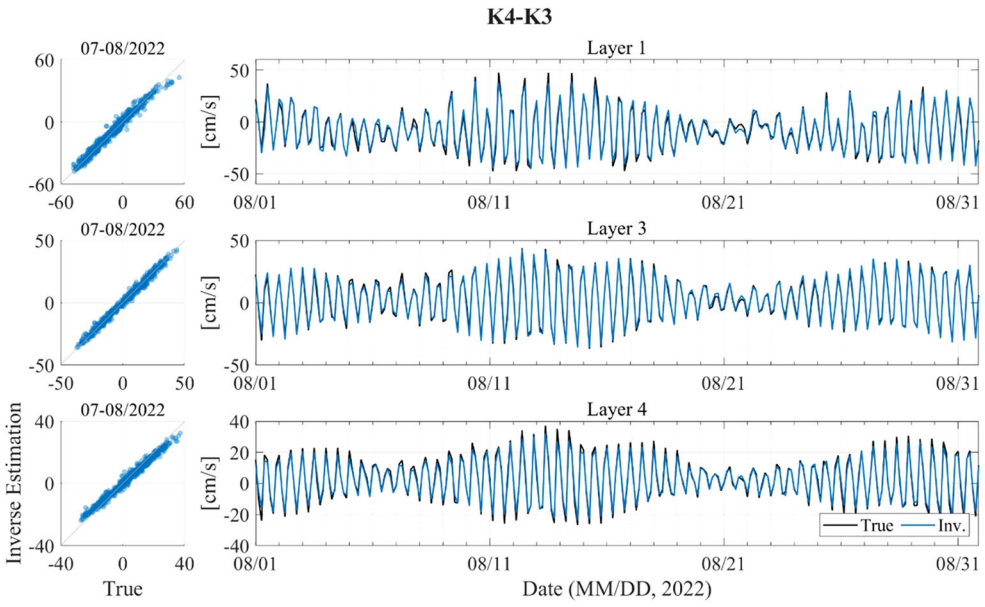


FIGURE 4
Comparison between the true value (KOOS model outputs, black) and results from the inverse estimation (blue) on S4 (K4–K3).

10.3389/fmars.2024.1362335

TABLE 4 RMSE (unit: cm/s) and R (correlation coefficient) values between the true current fields (KOOS model outputs) and inverse estimations.

Value	Layer	Transmission Path					
		S1	S2	S3	S4	S5	S6
RMSE [cm/s]	1	3.8	3.9	3.4	3.3	3.1	3.1
	3	2.0	2.1	2.1	2.1	1.6	1.7
	4	2.0	2.7	3.4	2.5	1.7	2.1
R	1	0.90	0.98	0.99	0.99	0.96	0.96
	3	0.91	0.99	0.99	0.99	0.96	0.98
	4	0.93	0.99	0.99	0.99	0.97	0.99

variation among the layers. For the v-component, the average PVE ranges from 93.8% to 96.0%, with a variation of approximately 1% among the layers. This result suggests that the higher RMSE in the upper layer compared to the middle and bottom layers can be attributed to the higher current velocity values and variability in the upper layer. The higher PVE for the v-component compared to the u-component is interpreted to be caused by the EOF related to the domain characteristics represented by the simple current pattern in the v-component.

4.2 Noise test using the AI model

In contrast to ray-tracing simulations utilizing the Bellhop model, *in-situ* data may encompass a diverse array of noise. To assess the impact of noise within *in-situ* data on inverse analysis, a noise test was conducted, artificially introducing noise to the results (Δt) of the ray-tracing simulation. The noise was configured to follow a random normal distribution, and 11 experiments were structured based on varying noise intensities. The application of noise in each experiment is governed by the following equation:

$$\Delta t_\sigma = \Delta t + N(0,1)\star\sigma,\ \ \sigma = \sigma' \star n$$

Here, n was established based on the maximum, minimum, and mean values derived from the observational results of CAT and the ray-tracing simulation results (Table 5). Table 6 lists the mean values of the RMSE for the along-path current calculated from the AI model noise test. The difference in RMSE between Case 1 (no noise) and Case 7 (noise with a magnitude of $10\times\sigma'$) was computed to be less than 1 cm/s. And RMSE difference between Case 1 and Case 11 (noise with a magnitude of $200\times\sigma'$) was calculated to be less

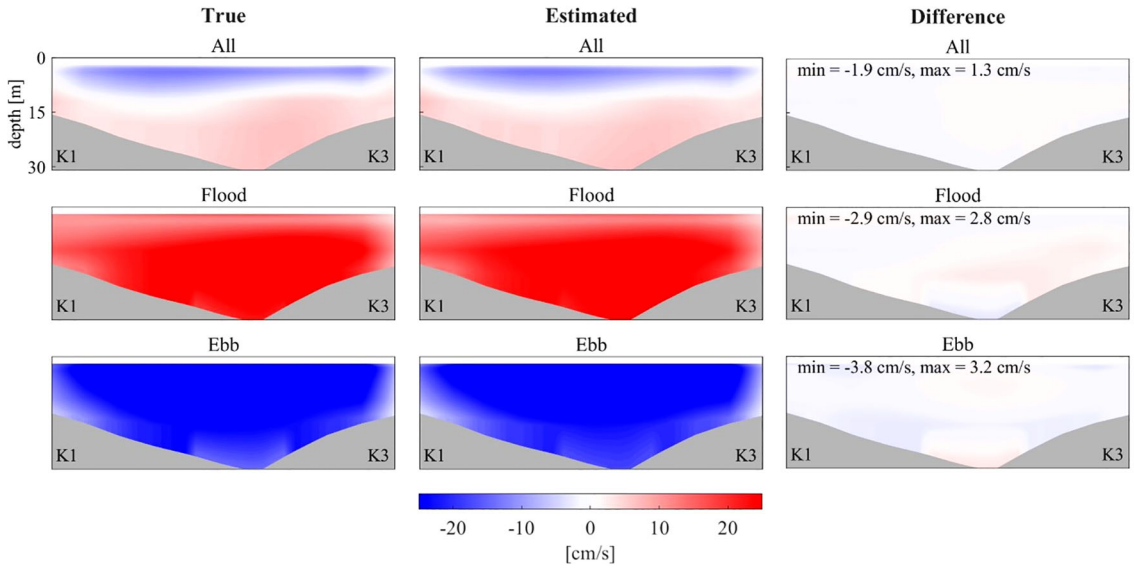


FIGURE 5
V-component along the section between K1 and K3 averaged (top panels) during July–August 2022, (middle panels) during the flood tidal period, and (bottom panels) during the ebb tidal period. First and second columns represent the true current field from the KOOS model and the results from the inverse estimation, respectively. Third column represents the difference between them (True - Estimation). Gray-shading indicates the sea floor.

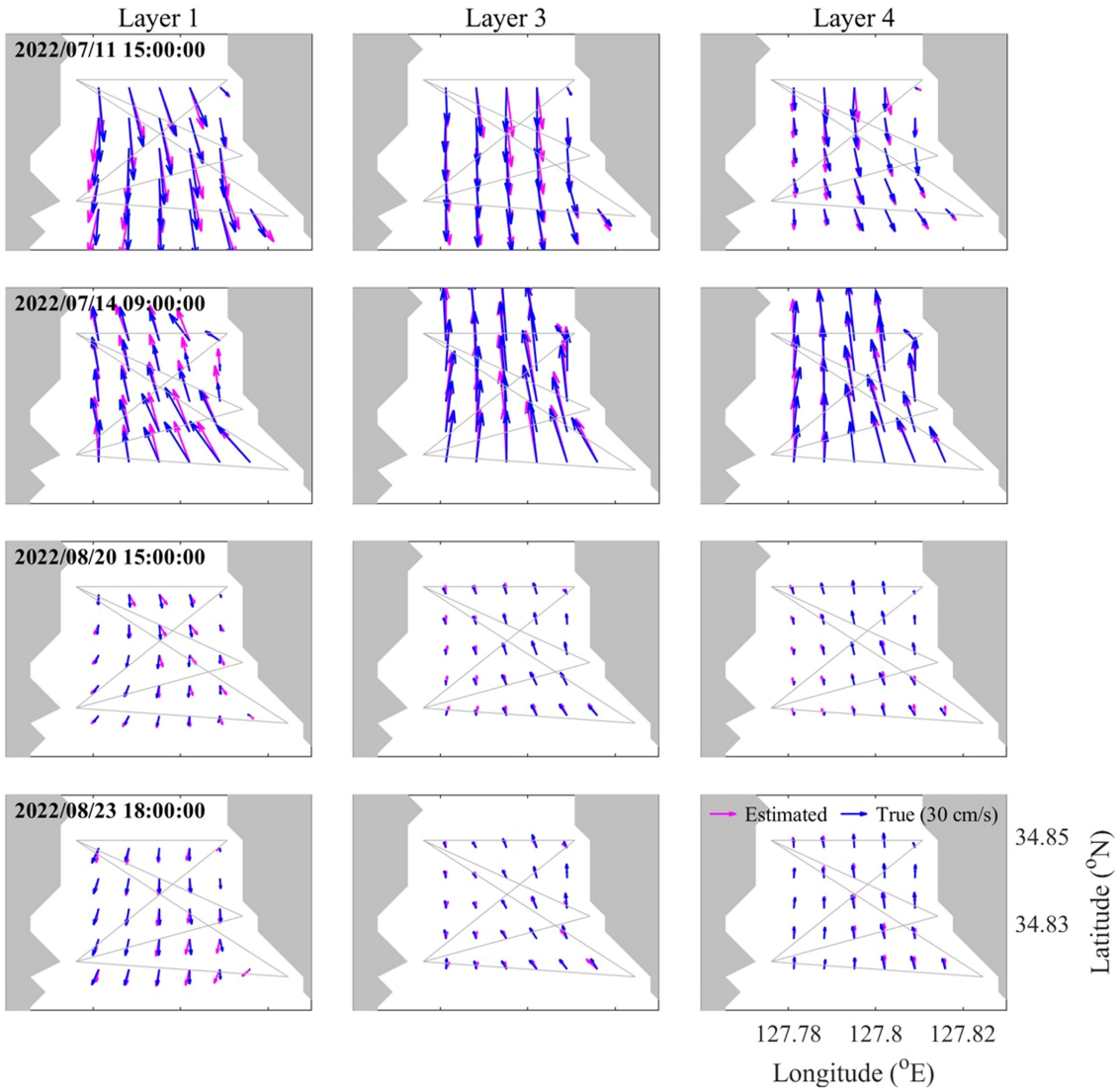


FIGURE 6
Snapshots of mapped current velocity at Layers 1, 3, and 4. Magenta and blue arrows indicate the results from the inverse estimation and the true value from KOOS model, respectively.

than 4.4 cm/s. This implies that even with a 20-fold increase in the magnitude of noise, the increase in error was less than 2.2 times. Consequently, the AI model demonstrated the capability of reducing the noise inherent in the observational results of CAT. Therefore, these experiments suggest the feasibility of conducting quality control of CAT data using an AI model.

4.3 Application to *in-situ* CAT observation

The method in Section 2.2 is applied to *in-situ* CAT observation data using same method and data except for Δt. Here, the *in-situ* data were utilized instead of Bellhop outputs. The validation of the estimated current field using *in-situ* CAT data was performed by comparison with the ADCP mooring data. Each subplot is a scatter plot for each layer and the ADCP mooring site. Note that the station K1 was moved southward (34.8397°N, 127.7748°E) to obtain stable and high quality *in-situ* observation data. The map in Figure 9 is provided to indicate the relocated K1 station.

The v-component exhibited a lower RMSE and higher correlation coefficient than the u-component, which is attributed to the alignment of the current direction of the v-component with the along-channel direction. Upon examining the KOOS model as of September and October 2022, it was observed that the v-component had a minimum of 3.8 times and a maximum of 9.9 times higher standard deviation in the five layers than the u-component at the nearest grid to P1 and P2, respectively. In contrast, the observed currents at P1 and P2 showed a standard deviation of at least

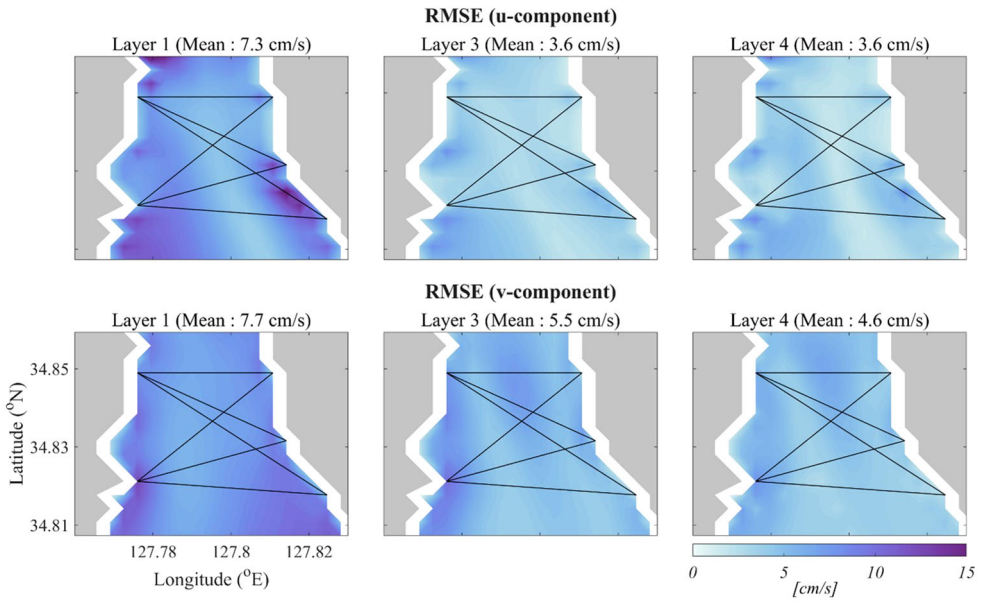


FIGURE 7
RMSE of u-component and v-component in January, February, July, and August 2022. Black lines indicate simulation paths, and gray-shaded parts are land.

2.3times and up to 6.1 times higher for the v-component than for the u-component in the five layers. This indicates a significant deviation between the two components of the current velocity in the KOOS model used to develop the proposed method in this study. These characteristics of the KOOS model output and its spatial resolution of 300 m appear to account for the differences between the components in the validation results. This issue may be addressed in a future study by improving our 3-D current field inverse method by utilizing a high-resolution coastal ocean model with spatial and temporal resolutions of 100 m and 30 min, respectively.

5 Conclusion

In this paper, we propose a new method for estimating the three-dimensional current field by combining AI and inverse methods.

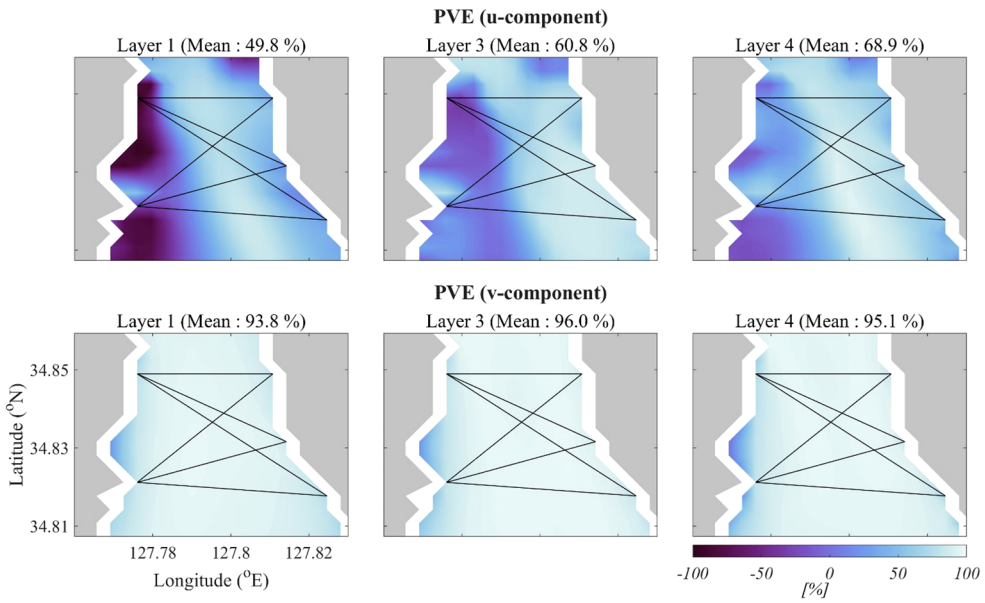


FIGURE 8
PVE of u-component and v-component in January, February, July, and August 2022. Black lines indicate simulation paths, and gray-shaded parts are land.

TABLE 5 CAT observation and Bellhop Model output data used for determining standard deviation (unit: sec, %).

Transmission Path	CAT observation (Δt, [sec])		Bellhop Model Output (Δt, [sec])	Ratio (CAT Observation/Bellhop Model, [%])	
	Before QC	After QC		Before QC	After QC
S1	4.11e-02	6.67e-04	2.92e-04	141.0	2.3
S2	8.04e-02	2.17e-02	6.62e-04	121.0	32.8
S3	7.36e-02	2.75e-02	1.10e-03	66.9	25.0
S4	2.78e-02	1.08e-03	6.30e-04	44.1	1.7
S5	6.64e-02	6.95e-03	3.30e-04	201.0	21.1
S6	4.57e-02	2.97e-03	4.30e-04	106.0	6.9
			min	44.1	1.7
			max	201.1	32.8
			mean	113.4	15.0

Ray-tracing simulations were performed using the Bellhop model, and the range-averaged currents at five layers and six simulation paths were obtained from the AI model. The inverse method is applied to each of the five horizontal layers, resulting a 3-D current fields. The significance of this study can be summarized as follows:

First, the 3-D current field was estimated for the first time by combining AI and inverse methods. The CAT *in-situ* observations are theoretically capable of identifying rays passing through all layers, but this is challenging in practice, making it difficult to estimate current fields in the vertical sections of the experimental paths. An AI model was employed to obtain the current fields in the vertical sections. Furthermore, applying the EOF of the current fields to the inverse method simplified the coastal boundary condition problem by incorporating the current-field characteristics of the domain through the first five EOF modes.

Second, the noise test of the AI model showed that it can handle the noise generated by the observations; therefore, it is applicable to CAT *in-situ* observations, which are expected to contain more noisy signals than ray tracing simulations. In fact, after applying the AI model to CAT *in-situ* observations taken in the domain over a one-month period starting on September 22, 2023, the estimated current fields showed that the along-channel velocity matched well with the ADCP mooring data at the two points inside the domain ($R > 0.85$). These results suggest that our novel 3-D current field estimation method is applicable to *in-situ*

TABLE 6 RMSE of the along-path-averaged velocity between true value (KOOS model output) and the results from AI model noise test (unit: cm/s).

Case	Std. ($\sigma = \sigma' * n$)	Path					
		S1	S2	S3	S4	S5	S6
1	$\sigma' * 0$	2.39	2.97	3.19	3.06	2.46	2.25
2	$\sigma' * 1$	2.40	2.99	3.20	3.07	2.48	2.26
3	$\sigma' * 2$	2.42	3.00	3.22	3.10	2.49	2.28
4	$\sigma' * 3$	2.47	3.04	3.25	3.12	2.52	2.32
5	$\sigma' * 4$	2.52	3.14	3.31	3.15	2.57	2.37
6	$\sigma' * 5$	2.53	3.15	3.38	3.18	2.64	2.42
7	$\sigma' * 10$	2.90	3.53	3.82	3.54	2.97	2.76
8	$\sigma' * 25$	3.85	4.61	4.93	4.53	3.72	3.66
9	$\sigma' * 50$	4.81	5.66	6.23	5.79	4.55	4.38
10	$\sigma' * 100$	5.36	6.40	7.17	6.62	5.03	4.86
11	$\sigma' * 200$	5.79	6.77	7.58	7.11	5.36	5.13

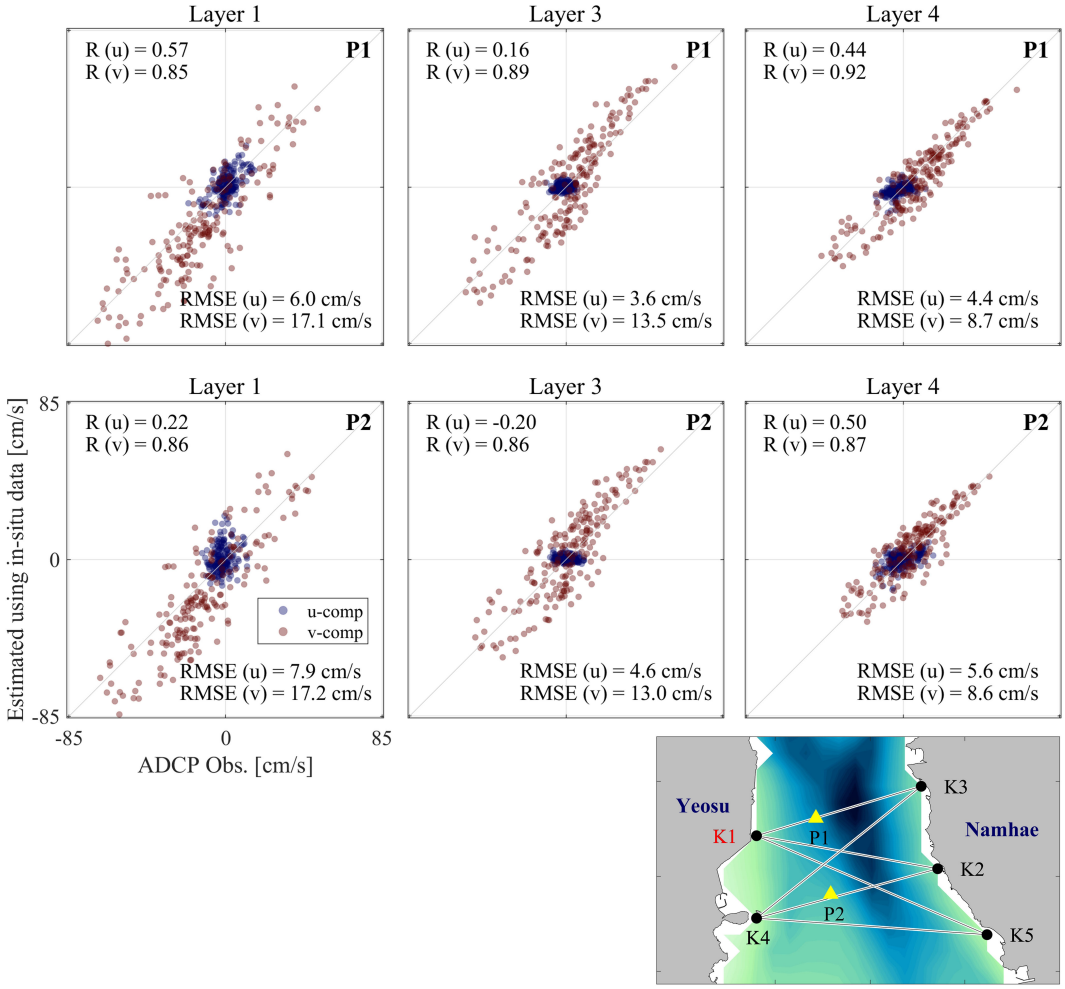


FIGURE 9
Scatter plots of currents between the observations from ADCP moorings and estimations from the inverse analysis using *in-situ* CAT data. Upper three panels are for mooring site P1 and lower three panels are for mooring site P2. Columns indicate Layers 1, 3, and 4 from left to right. Correlation coefficient (R) and RMSE values between the observed and estimated zonal and meridional (u and v) currents are shown in each panel. "Light blue and red dots indicate u- and v-components, respectively.".

CAT data in the Yeosu Bay. In addition, since the high-resolution KOOS model outputs are available all around the coastal seas of Korea, its application would be possible to other coastal areas where the CAT system is installed to continuously monitor 3-D current changes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YH: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization. EL: Data curation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. HS: Data curation, Formal analysis, Writing – review & editing. B-NK: Data curation, Investigation, Writing – review & editing. HH: Data curation, Investigation, Writing – review & editing. YC: Data curation, Investigation, Writing – review & editing. JK: Data curation, Writing – review & editing. JP: Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Methodology, Project administration, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Ministry of Oceans and Fisheries, Korea (grant number 20210642, "Development of 3-D Ocean Current Observation Technology for Efficient Response to Maritime Distress").

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1362335/full#supplementary-material

References

Chen, M., Zhu, Z. N., Zhang, C., Zhu, X. H., Wang, M., Fan, X., et al. (2020). Mapping current fields in a bay using a coast-fitting tomographic inversion. *Sensors* 20, 558. doi: 10.3390/s20020558

Del Grosso, V. A. (1974). New equation for the speed of sound in natural waters (with comparisons to other equations). *J. Acoustical Soc. America* 56, 1084–1091. doi: 10.1121/1.1903388

Hansen, P. C., and O'Leary, D. P. (1993). The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* 14, 1487–1503. doi: 10.1137/0914086

Kaneko, A., Zhu, X. H., and Lin, J. (2020). *Coastal acoustic tomography* (Amsterdam, Netherlands: Elsevier). doi: 10.1016/B978-0-12-818507-0.00014-7

Lee, J. C., and Kim, J. C. (2007). Current structure and variability in Gwangyang Bay in spring 2006. *The Sea: Journal of the korean society of oceanography*. 12 (3), 219–224. Available at: http://uci.or.kr/G704-000255.2007.12.3.016.

Munk, W., and Wunsch, C. (1979). Ocean acoustic tomography: A scheme for large scale monitoring. *Deep Sea Res. Part A. Oceanogr. Res. Papers* 26, 123–161. doi: 10.1016/0198-0149(79)90073-6

Park, J. H., and Kaneko, A. (2000). Assimilation of coastal acoustic tomography data into a barotropic ocean model. *Geophys. Res. Lett.* 27, 3373–3376. doi: 10.1029/2000GL011600

Park, J. H., and Kaneko, A. (2001). Computer simulation of coastal acoustic tomography by a two-dimensional vortex model. *J. Oceanogr.* 57, 593–602. doi: 10.1023/A:1021211820885

Porter, M. B. (2011). *The bellhop manual and user's guide: Preliminary draft* Vol. 260 (La Jolla, CA, USA: Heat, Light, and Sound Research, Inc.). Tech. Rep.

Pritchard, D. W. (1952). Salinity distribution and circulation in the Chesapeake Bay estuarine system. *J. Mar. Res.* 11 (2), 106–123. Available at: https://elischolar.library.yale.edu/journal_of_marine_research/763.

Yamoaka, H., Kaneko, A., Park, J. H., Zheng, H., Gohda, N., Takano, T., et al. (2002). Coastal acoustic tomography system and its field application. *IEEE J. Oceanic Eng.* 27, 283–295. doi: 10.1109/JOE.2002.1002483

Zhang, C., Zhu, X. H., Zhu, Z. N., Liu, W., Zhang, Z., Fan, X., et al. (2017). High-precision measurement of tidal current structures using coastal acoustic tomography. *Estuar. Coast. Shelf Sci.* 193, 12–24. doi: 10.1016/j.ecss.2017.05.014

Zhu, X. H., Kaneko, A., Wu, Q., Zhang, C., Taniguchi, N., and Gohda, N. (2012). Mapping tidal current structures in Zhitouyang Bay, China, using coastal acoustic tomography. *IEEE J. Oceanic Eng.* 38, 285–296. doi: 10.1109/JOE.2012.2223911

Zhu, Z. N., Zhu, X. H., Guo, X., Fan, X., and Zhang, C. (2017). Assimilation of coastal acoustic tomography data using an unstructured triangular grid ocean model for water with complex coastlines and islands. *J. Geophys. Res.: Oceans* 122, 7013–7030. doi: 10.1002/2017JC012715

Zhu, Z. N., Zhu, X. H., Zhang, C., Chen, M., Wang, M., Dong, M., et al. (2021). Dynamics of tidal and residual currents based on coastal acoustic tomography assimilated data obtained in Jiaozhou Bay, China. *J. Geophys. Res.: Oceans* 126, e2020JC017003. doi: 10.1029/2020JC017003

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 28 February 2024
DOI 10.3389/fmars.2024.1349478

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Hiroyuki Matsumoto,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Chuanxi Xing,
Yunnan Minzu University, China
Jia Liu,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Hanhao Zhu
zhuklmpg@163.com
Weihua Song
whsong@zjou.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 04 December 2023
ACCEPTED 26 January 2024
PUBLISHED 28 February 2024

CITATION
Wang J, Cui Z, Zhu H, Meng L, Song W and Liu X (2024) A back propagation neural network-based approach for inverting layered seabed acoustic parameters in shallow waters.
Front. Mar. Sci. 11:1349478.
doi: 10.3389/fmars.2024.1349478

A back propagation neural network-based approach for inverting layered seabed acoustic parameters in shallow waters

Jiahui Wang[1†], Zhiqiang Cui[2†], Hanhao Zhu[3*], Lei Meng[4], Weihua Song[5*] and Xu Liu[6]

[1]School of Information Engineering, Zhejiang Ocean University, Zhoushan, China, [2]Hydroacoustics Technology Co., Ltd., Zhoushan, China, [3]Fujian Provincial Key Laboratory of Marine Physical and Geological Processes, Xiamen, China, [4]Unit 92578 of the People's Liberation Army, Beijing, China, [5]National Engineering Research Center of Marine Facilities Aquaculture, Zhejiang Ocean University, Zhoushan, China, [6]School of Marine Science and Technology, Zhejiang Ocean University, Zhoushan, China

Introduction: Existing methods primarily focus on earth acoustic parameters inversion under specific layered structures. However, they face challenges with experimental data from unknown seabed stratification, hindering accurate parameter inversion.

Methods: To address this, a novel algorithm combines Back Propagation Neural Network (BPNN) for distinguishing seabed stratification and inverting acoustic parameters. Simulated sound pressure data disturb seabed parameters as input, enabling feature recognition for training the neural network inversion model. Acoustic parameters are then estimated under identified stratification using the sound field model.

Results: The inversion model is validated using simulation and pool shrinkage data. Results show the neural network model effectively stratifies simulation and experimental data, providing accurate inversion results for acoustic parameters corresponding to distinct layers.

Discussion: The neural network model's accuracy and practicality are confirmed through hierarchical judgment of scale test data and acoustic parameter inversion. This approach introduces a new perspective for shallow sea acoustic parameter inversion, offering a promising application scenario.

KEYWORDS

seabed stratification, acoustic parameters inversion, BPNN, fast field method (FFM), scale test

1 Introduction

Underwater acoustic parameters are important physical parameters for studying acoustic propagation characteristics in shallow seas. With the development of acoustic technology and the popularity of machine algorithms, it is convenient and efficient to in-vert underwater acoustic parameters using acoustic methods (Li and Zhang, 2017). Aiming at parameters inversion, a large number of underwater acoustic parameters inversion works have been carried out by predecessors, and many methods for underwater acoustic parameters in-version have been developed, such as the acoustic parameters inversion method based on transmission loss (Dragna and Blanc-Benon, 2017), the seafloor sediment parameters were estimated using the genetic algorithm and bottom reflection loss curve with wide grazing angles (Wang et al., 2023), a solution to the problem of stratified seabed parameters estimation using the sound field matching meth-od with vertical angle spectra (Xue et al., 2023),three-dimensional sediment modelling and inversion of geoacoustic parameters using a cross-dimensional Bayesian approach with wideband acoustic sources (Ke et al., 2013), the acoustic signal arrival time (Wang et al., 2023), and the waveguide dispersion characteristics (Kerzhakov and Kulinich, 2016). Among them, the most widely used methods can be summarized as the matching field inversion of underwater acoustic parameters by using the physical characteristics of underwater acoustic signals combined with the global optimization algorithm (Potty et al., 2017). Different physical characteristics are used as the forward model, and then through various optimization algorithms, such as Genetic algorithms (GA) and Simulated Annealing algorithms (SA), the objective function is solved to obtain the parameters results to be in-verted. Existing studies focus on the inversion calculation of geoacoustic parameters un-der specific stratification structures (Wang, 2008; Zhu et al., 2013). The statistical characteristics of the hydroacoustic echo signal reflected from the seabed, which produces a sudden change, reflect the existence of the seabed boundary. That is, on behalf of the seabed, there is a layered structure. There is no coverage of the sedimentary layer in the base. The seabed can be regarded as a semi-infinite seabed (Gerstoft, 1994). Such as the base is covered by a layer of sedimentary layer. As a result of this time, the physical characteristics of the physical characteristics of the seabed and semi-infinite seabed have a large difference and cannot be used to solve a similar problem of the semi-infinite seabed (Zhao et al., 2023). And so the previous introduction of the stratified seabed model, due to the oceanic motion and crustal movement of a stochastic, and so the base covered by the number of sedimentary layers is also not given to a fixed number of layers, it needs to be reflected according to the characteristics of the signal (Zhu et al., 2023b). When processing experimental data under unknown seabed stratification, the inversion results of geoacoustic parameters following the stratification cannot be accurately given. In addition, when the existing classical optimization algorithms are applied, the iterative optimization calculation between the input data and the optimal solution not only consumes a lot of computing time, but also easily falls into the optimal local solution (Zhu et al., 2023a).

Given the performance advantages of neural network algorithms in data processing, many scholars have tried to apply neural network models in classification research and parameters inversion in recent years. Chen, Yang et al. used migration learning and convolutional neural network to study the classification of sediments such as sand, reef and mud (Song and Wang, 2022). Huang has applied the Convolutional neural network (CNN) model to the geo-physical inversion, successfully realizing the inversion of some geological parameters, and proving that the neural network model has strong generalization applicability in the inversion problem (Feng et al., 2022). Wu et al. used a single-hidden layer feedforward neural network an extreme learning machine to perform inversion in shallow water depth remote sensing and obtained relatively accurate inversion results (Chen et al., 2022). Li, Wen et al. applied the back-propagation neural network (BPNN) model to electromagnetic inversion and effective wave height field parameters inversion. They improved the differential evolution algorithm of the BPNN model to achieve a more efficient and accurate inversion target (Yang et al., 2021). Due to BPNN's powerful nonlinear fitting capability, this algorithm can automatically learn and identify hidden patterns and relationships from data. When dealing with the acoustic and geological complexities of shallow waters, it can provide highly accurate and reliable results. Therefore, using BPNN for seabed stratification and acoustic parameters inversion in shallow seas is worth further study (Pang et al., 2021).

Inspired by the numerous successful applications of neural network models in target classification and parameters inversion, this study employs the BPNN for hierarchical structure assessment and geoacoustic parameters inversion. Focused on the shallow-sea sound pressure field, the research employs neural network algorithms to establish a relational model between the predicted sound pressure field and the acoustic parameters to be retrieved (Wang et al., 2022). Subsequently, the model is utilized to achieve accurate hierarchical structure assessment and geoacoustic parameters inversion within a predefined shallow-sea environment. The study is divided into four main sections: the first section provides an overview of geoacoustic parameters inversion methods based on sound pressure fields. The subsequent section introduces the methods of hierarchical assessment and geoacoustic parameters inversion using the BPNN models. The third section assesses the application's effectiveness and the performance of the neural network models through simulation and experimental data. Finally, the research concludes its findings.

2 Method

2.1 Forward modeling of shallow sea sound field

Acoustic parameters in shallow seas are important environmental parameters that determine the distribution characteristics of acoustic fields in shallow sea environments. The change of acoustic parameters in the seabed will have a significant

impact on the distribution characteristics of the acoustic pressure field in water so that the geoacoustic parameters can be retrieved from the measurement data of the acoustic pressure field in the shallow sea (Kerzhakov and Kulinich, 2016).

In the model, the harmonic point source is located on the symmetry axis of the cylindrical coordinate. In shallow sea waveguide environments, the characteristics of seafloor sediments play an indispensable role in influencing marine acoustic fields. Various parameters of sediment layers affect the reflection, transmission, and the paths and directions of sound wave propagation at the seabed boundary. Therefore, investigating the properties of seafloor sediments is crucial for understanding the propagation patterns of marine acoustic fields (Li et al., 2020). In this context, seawater is considered as a homogeneous isotropic fluid medium, and the seafloor sediments as an elastomeric medium. A forward modeling approach for sound field propagation in layered seabed environments is thus established, as depicted in Figure 1. Due to the axial symmetry of the column coordinate system, the three-dimensional problem can be transformed into a two-dimensional (r, z) plane for solving (Wen et al., 2021), z=0 represents the sea surface, and the downward direction of the sea surface is the positive direction of the z-axis of depth, and the positive axis of r represents the propagation direction of the sound field.

In the model, it is assumed that the seafloor is regarded as the superposition of n layers of sediments, and the depth of seawater layer is set as H_1. The sound source with frequency f_0 is located at z_s depth of seawater layer. The density and sound velocity in seawater layer are ρ_1 and c_1, respectively. The depth of the sedimentary layer n is denoted by h_n, and the longitudinal wave sound velocity, shear wave sound velocity, density, longitudinal wave sound velocity attenuation and shear wave sound velocity attenuation of the sedimentary layer are denoted by c_{pn}、c_{sn}、ρ_{bn}、α_{pn}、α_{sn} respectively. The above 6 types of parameters are the submarine acoustic parameters to be retrieved in this study. Under the wave theory, each physical quantity in the above model can be represented by the displacement potential function ϕ. ϕ_1 is the displacement potential function of seawater layer. And the research object of this article sound pressure meets $p=\rho_1\omega^2\phi_1$ (angular frequency $\omega=2\pi f_0$), can be obtained by solving the displacement potential function of fluid, the sound pressure values at various points in the detailed theoretical derivation see literature (Li et al., 2019). Since the displacement potential function in the seawater layer satisfies formula 1 as follows:

$$\begin{cases} \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial^2\phi_1}{\partial z^2}\right)+\frac{\partial^2\phi_1}{\partial z^2}+k_1^2\phi_1=-4\pi\delta(r,z-z_s) \\ 0\le z\le H_1 \end{cases} \tag{1}$$

The sediments of potential function ϕ_n can be represented as follows:

$$\begin{cases} \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi_{pn}}{\partial r}\right)+\frac{\partial^2\phi_{pn}}{\partial z^2}+k_{pn}^2\phi_{pn}=0 \\ \nabla\times\nabla\times\psi_{sn}-k_{sn}^2\psi_{sn}=0, H_n\le z\le H_{n+1} \end{cases} \tag{2}$$

Where $\delta(r,z)$ is the original function, k represents the wave number of each seafloor, where $k=\omega/c_m$, $\omega=2\pi f_0$, pn and sn are the uncertainties contained in the solution, The flow function and potential function of ϕ and ψ correspond to pn and sn respectively. Then the formal solution of formula 2 is as follows:

$$\phi_1(r,z)=\int_0^\infty Z_1(z,\xi)J_0(\xi r)\xi d\xi \tag{3}$$

Where Z is the ordinary differential formula of depth z and ξ of horizontal wavenumber, J_0 is the zero-order Bessel function. According to the derivation results of formula 3, the sound pressure field in the water layer can be expressed as follows::

$$p(r,z)=\rho_1\omega^2\int_0^\infty Z_1(z,\xi)J_0(\xi r)\xi d\xi \tag{4}$$

For the solution of formula (4), the Normal Mode Method (NMM) and Fast Field Method (FFM) can be used to solve formula (4). For shallow sea environments, FFM converts the integral formula in formula (4) into Fourier transform form for the direct solution, which is more suitable for fast calculation of sound field in shallow sea (Frederick et al., 2020). Therefore, FFM is selected in this study to conduct a forward simulation of the sound pressure field in the above parametric model.

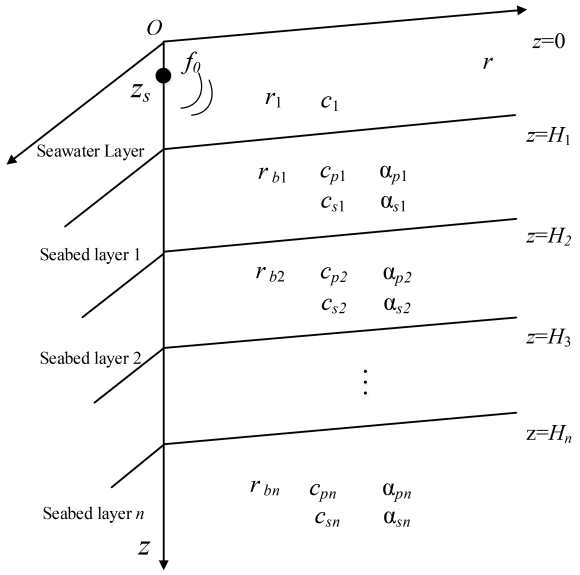


FIGURE 1
Schematic diagram of the sound field model in cylindrical coordinates.

2.2 BPNN inversion model of geoacoustic parameters inversion

Due to the complexity of the marine environment, factors such as sediment layer density, porosity, and average particle size can all influence acoustic parameters. Consequently, establishing a precise functional relationship between acoustic parameters and shallow water acoustic pressure fields presents a significant challenge. To address this challenge, a non-linear mapping approach utilizing a

BPNN is employed (Van Komen et al., 2020). On the other hand, the BPNN is capable of approximating functions through the training of input and output vectors. When feature vectors are input into the network, they can produce results that closely approximate the desired output values. BPNN are a type of multi-layer feedforward neural network that employs both forward signal propagation and backward error propagation to adjust weights and thresholds to minimize the error function value. This iterative process ensures that the modified network output aligns closely with the desired output values (Xu and Pan, 2018).

Therefore, in response to the challenge of inverting seabed bottom properties in the presence of uncertain seabed sediment layering, a method utilizing a BPNN based on shallow water acoustic pressure field data is proposed. Given the inherent coupling between acoustic parameters and the potential for multi-valued solutions, especially in multi-layer seabed environments, training a single neural network directly poses the risk of convergence issues and encounters difficulties due to the vast search space (Zhou et al., 2019).

To address such complexities, this study employs a staged supervised learning approach for multi-layer seabed acoustic parameter inversion. The stepwise seabed layering and parameter inversion process is depicted in Figure 2A. It involves feeding preprocessed and standardized acoustic pressure signals into the NET-1 network, which serves as a classifier for segregating shallow water acoustic pressure data into various seabed layer categories. Based on the classification outcomes, the data is directed to different networks, facilitating targeted acoustic parameter inversion. To accommodate current computational capabilities and practical requirements, this paper primarily discusses scenarios involving seabed layering, specifically semi-infinite seabed (NET-2-1), single sediment layer (NET-2-2), and double sediment layers (NET-2-3). In cases where NET-1 classifies the acoustic pressure data as originating from a single sediment layer, the data is subsequently input into the NET-2-2 network for further acoustic parameter inversion. NET-1 is designed as a single-label classifier solely for layer categorization, without any data alteration. The BPNN models employed for layer categorization in semi-infinite seabed, single sediment layer, and double sediment layer environments are denoted as NET-2-X, where X represents the specific environment (X=1, 2, 3). In total, four BPNN models are trained (Qian et al., 2019).

In the two-step inversion process, the objective of the first step is to determine the seabed type by initially inputting acoustic pressure data into the NET-1 network. The NET model is constructed using a single hidden layer, where neurons within the same layer are not interconnected. In forward propagation, the activation function $f(x)$ links the input signal - acoustic pressure field data $p_j(r_i, z_i)$ - with the hyperparameter matrix $[w, b]$. The activation function $f(x)$ selected is the Sigmoid function. The signal flows forward from the input layer to the output layer, with $w=[w_{kv}, w_{vl}]$ where w_{kv} represents the weight from the input layer to the hidden layer, and w_{vl} represents the weight from the hidden layer to the output layer (Stoll and Kan, 1981). Additionally, b_v signifies the threshold for each neuron in the hidden layer. The backward feedback signal is the error signal E, when the error falls within the predetermined range, network training is terminated. The experimental setup precision is set to 10^{-3}. which reflects the discrepancy between the network model's inversion results and the true values. For network training, the cross-entropy function E_{CE} is utilized. The calculation of the cross-entropy function is performed as shown in formula (5). NET-1 functions as a single-label classifier, generating outputs based on the categorization of seabed layering, distinguishing between scenarios such as semi-infinite seabed (NET-2-1), single sediment layer (NET-2-2), and double sediment layers (NET-2-3).

$$E_{CE} = -\frac{1}{M_t}\sum_{i=1}^{M_t}\sum_{q=1}^{Q} t_{iq} \ln y_{iq} \tag{5}$$

After determining the seabed layering structure, the subsequent step involves the stratified inversion of acoustic parameters. In this second phase, we draw inspiration from the application of matched-

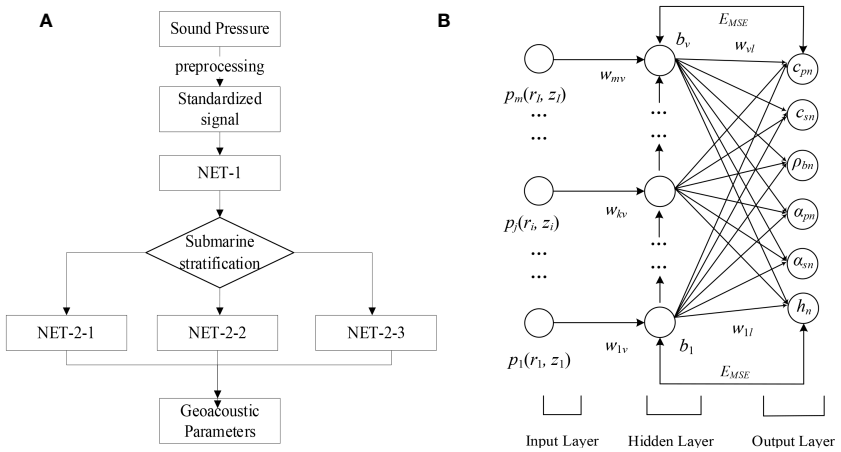


FIGURE 2
Schematic diagram of seabed stratification and parameters inversion calculation based on BPNN model: **(A)** Flow diagram; **(B)** BPNN (NET 2-1) model.

field acoustic parameter inversion methods. For varying seabed layering conditions, the acoustic signals are input into dedicated BPNN models associated with shallow-water multi-layered seabed environments, based on the classification results obtained from NET-1. The architecture of the NET-2-X models closely resembles that of the NET model. As an illustration, consider the NET-2-1 model, as depicted in Figure 2B. i set of acoustic pressure data from different receiver positions (r_i, z_i), where $(1 \le i \le I)$, form the input data, denoted as $\boldsymbol{p} = [p_1(r_1, z_1), \ldots, p_j(r_i, z_i), \ldots, p_m(r_I, z_I)]_{m \times I}$. Corresponding ground sound parameters $\boldsymbol{Y} = [c_{pn}, c_{sn}, \rho_{bn}, \alpha_{pn}, \alpha_{sn}]_{m \times 5}$ are employed as output results for model construction (Zhu et al., 2012).

The number of neurons in each layer can be determined based on formula (6):

$$v = \sqrt{I + l} + \alpha \tag{6}$$

Where n represents the number of input layer nodes, which corresponds to the simulated acoustic pressure data points, and α is a constant coefficient.

The partial derivatives of weight parameters between the input set of acoustic pressures, denoted as p, and the hidden layer are represented as Δw_{jv}, while the partial derivatives of weights between the hidden layer and the ground sound parameters Y are represented as Δw_{vl}. The learning rate, η, is employed during the computation. In this process, the iteration step t is continually updated based on whether the E_{MSE} value satisfies the predetermined accuracy. The adjustment of parameters, w_{jv}、w_{vl} is achieved through the modification of weight values, aiming to minimize errors. The parameter descent is carried out using a gradient descent approach (Zhang et al., 2021). The method for modifying parameters w_{jv} and w_{vl} is as follows, as expressed in the (formulas 7, 8).

$$w_{kv}(t+1) = w_{kv}(t) + \Delta w_{kv} \tag{7}$$

$$w_{vl}(t+1) = w_{vl}(t) + \Delta w_{vl} \tag{8}$$

The assessment of the training network's effectiveness is based on the Root Mean Square Error (RMSE), as described in formula (9).

$$E_{MSE} = \sqrt{\frac{1}{M_t}\sum_{i}^{M_t}[Y_{sim} - Y_{inv}]^2} \tag{9}$$

Where $\boldsymbol{Y}_{rea,pre} = [c_p, c_s, \rho_b, \alpha_p, \alpha_s]$, a matrix comprising the parameters to be inverted, with $\boldsymbol{Y}_{rea}$representing the simulated values and $\boldsymbol{Y}_{pre}$representing the inverted values.

The sound pressure input undergoes a nonlinear transformation within the network, ultimately resulting in the geophysical parameters (Cheng et al., 2021). In this process, the inputs and outputs of the hidden layer are denoted as $I_{kv}$and I_{vl}, respectively. The final inversion result, $\boldsymbol{Y}_{pre}$, is computed as shown in formula (10).

$$Y_{sim} = f(I_{vl}) = f\left[\sum_{k=1}^{v} w_{vl} I_{vl}\right] \tag{10}$$

To determine an appropriate network structure, this paper simplifies the seafloor into two layers, thus establishing the BPNN inversion model (Zhu et al., 2017). In this model, the input layer receives simulated sound pressure data, denoted as "$\boldsymbol{p}$," which includes a set of sound pressure data $p_j(r_i, z_i)$, where the range of i extends from 1 to I. The number of neurons in the hidden layer is set at 9, and this choice is influenced by various factors, one of which is the empirical rule for parameter α, which is set to α = -15 in this context (Zheng et al., 2021). Furthermore, the number of neurons in the output layer, represented as "l," is determined based on the number of geophysical parameters that need to be inverted. For instance, have $l_{\text{NET-1}} = 4$, $l_{\text{NET-2-1}} = 5$, $l_{\text{NET-2-2}} = 11$, $l_{\text{NET-2-3}} = 17$.

2.3 Data generation and fitting verification

The equations should be inserted in editable format from the equation editor. Considering the variation range of ground sound parameters in shallow sea (Li et al., 2019), the parameters training range of the BPNN model for ground sound parameters inversion under a preset environment is set as shown in Table 1.

The simulated sound pressure field data is a set of horizontal equally spaced receiving sound pressure fields under the set sound source depth z_s=20m, receiving depth z_r=10m and seawater depth

TABLE 1 BPNN model training set search range.

Stratification	Geoacoustic parameters	Search range
Sea Layer	c_1(m/s)	1500
	ρ_1(g/cm3)	1.025
	h_1(m)	100
Single sedimentary layer (NET 2-1)	c_{p2} (m/s)	1800-2200
	c_{s2} (m/s)	900-1100
	ρ_b (g/cm^3)	1.4-1.6
	α_{p2} (dB·λ^{-1})	0.1-0.3
	α_{s2} (dB·λ^{-1})	0.1-0.3
	h_2(m)	15-25
Double sedimentary layers(NET 2-2)	c_{p3} (m/s)	2700-3300
	c_{s3} (m/s)	1350-1650
	ρ_{b3} (g/cm^3)	1.8-2.2
	α_{p3} (dB·λ^{-1})	0.1-0.3
	α_{s3} (dB·λ^{-1})	0.1-0.3
	h_3(m)	25-35
Three sedimentary layers (NET 2-3)	c_{p4} (m/s)	3600-4400
	c_{s4} (m/s)	1800-2200
	ρ_{b4} (g/cm^3)	2.25-2.35
	α_{p4} (dB·λ^{-1})	0.1-0.3
	α_{s4} (dB·λ^{-1})	0.1-0.3

H=100m. The receiving points are spaced 2m apart, and a total of I=720 receiving points are set. The model training samples adopted by NET2-X are 2200 groups of sound pressure data randomly generated in each layer within the search range in Table 1, among which 200 groups are randomly divided into training sets and the other 200 groups into test sets. Each group (Layered structure) of the training set and its corresponding environmental sound pressure are mapped into the model one by one for training. When the error function E_{MSE} reaches the set accuracy σ=0.01, the training is completed (Huang et al., 2018).

The verification set is generated using random values. The changes in loss function and prediction accuracy in the training process of NET-1 are shown in Figure 3. As can be seen from Figure 3A, after a certain batch of training, the error of the training curve is reduced to σ, the network stops training, and the confusion matrix further verifies that NET-1 also has a good classification effect on the verification data, and can complete the classification calculation of the submarine stratified structure.

In Figure 3B, the inside of each orange box represents the number of incorrectly predicted samples, the main diagonal represents the number of correctly predicted samples, and the light gray rectangle box at the lower right represents the prediction accuracy of the corresponding sample attributes, that is, the accuracy of 95% in the training process.NET 2-X conducts training for neural networks under three hierarchical structures respectively. To enhance the credibility of the model, the Mean Absolute Error (MAE) was introduced as an evaluation metric to assess the predictive accuracy of the model. The calculation results are shown in Table 2. Taking the NET2-1 scenario as an example, a smaller MAE value indicates better predictive capability of the model. The calculated results demonstrate that the inversion error of the model is small, indicating that the BP neural network performs well in the inversion of shallow-sea acoustic parameters. Consequently, the constructed BP neural network model exhibits good and stable predictive performance in the inversion of shallow seabed acoustic parameters, with high computational efficiency and reliability of the prediction results.

TABLE 2 NET 2-1Parameter setting of the inversion algorithm.

MAE	c_p	c_s	ρ_b	α_p	α_s
value	2.8187	1.5010	0.0175	0.2103	0.0518

The changes in loss function in the training process are shown in Figure 4. Taking NET 2-2, which corresponds to the layered structure of the seabed as a single sedimentary layer, as an example, after the completion of training, the E_{MSE} value reached the setting accuracy after 200 iterations. The error reduction process was relatively stable, indicating that the error reduction speed and training effect of the whole neural network in the training is considerable. Although with the increase of inversion parameters, the number of iteration steps to reach the accuracy setting in the NET 2-X training increased, the target accuracy was reached within 500 iterations, which effectively constructed BPNN model that could meet the inversion accuracy of layered submarine acoustic parameters.

By controlling the training parameters and adjusting the training function, the overfitting phenomenon can be eliminated in the process of network training, and the generalization ability of the model can be improved. The highly generalized multi-output model can solve the multi-value problem caused by the coupling relationship between the seabed parameters in the inversion process to some extent, to realize the purpose of simultaneously inverting multiple seabed acoustic parameters.

2.4 Simulation result analysis

Simulation data and experimental data are used to test the performance of BPNN model respectively, and the network prediction results are used in the seabed stratification calculation and earth acoustic parameters inversion of the measured data of the pool (Yu et al., 2020).

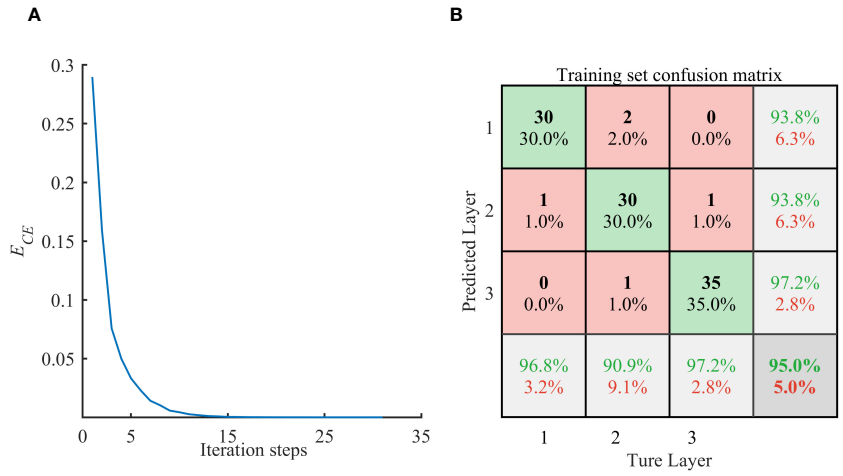


FIGURE 3
NET-1 training analysis: **(A)** Loss function; **(B)** Visual analysis.

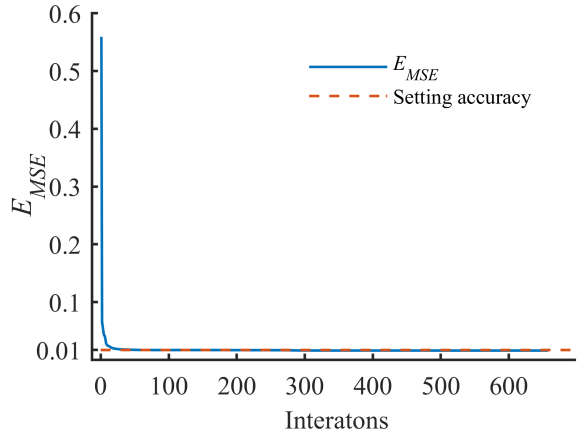


FIGURE 4
NET 2-2 loss function changes.

As a classification model, NET-1 can be directly evaluated based on the classification accuracy of the test set, as shown in formula (11).

$$E_a = \frac{N_t}{N} \times 100\,\% \tag{11}$$

Where, N_t is the number of correct samples for stratification and N= m-M_t, which is the number of total test samples. As a regression analysis problem, to quantify the error between the inversion results of each parameters and the preset truth value, the performance function R^2 was introduced to represent the coincidence degree between the inversion value and the true value numerically. The closer the R^2 value was to 1, the closer the inversion result was to the preset truth value (Zhang, 2023). The calculation method is shown in Formula 12.

$$R^2 = \frac{\left(N\sum_{i=1}^{N} Y_e Y_r - \sum_{i=1}^{N} Y_e \sum_{i=1}^{N} Y_r\right)^2}{\left[N\sum_{i=1}^{N} Y_e^2 - \left(\sum_{i=1}^{N} Y_r\right)^2\right]\left[N\sum_{i=1}^{N} Y_r^2 - \left(\sum_{i=1}^{N} Y_e\right)^2\right]} \tag{12}$$

To verify the robustness of the training completion network, the sound pressure of the test set generated within the search interval and some environmental noise are added as test data. The total number of samples in the test set is 10% of the training set.

Figure 5 shows the layering results of NET-1 on part of the test set. The layering accuracy E_a=99% proves that NET-1 can effectively perform layering calculations on sound pressure field information under different layering structures. In the figure, "x" represents the predicted value, "o" represents the true value, and the Y-axis represents the search range of the number of hierarchical structures.

At the same time, the NET 2-X model is trained. Taking NET 2-2 as an example, Figure 6 shows the results of underwater acoustic parameters inversion of NET 2-2. The solid line and "×" in the figure correspond to the real value and predicted value of the test data set respectively.

Figure 6 shows the degree of fitting between the predicted value and the preset value of each parameter in the test set in the inversion model. During verification, the fitting degree of the inversion results of c_{pn}, c_{sn}, ρ_{bn}, and h_n is maintained above 0.90, showing an excellent inversion effect. The error variation of parameters α_{pn} and α_{sn} is relatively large, but the R^2 value of each parameter is above 0.80, The overall error appears to be within acceptable limits. It can be seen that the BPNN model constructed has good and stable prediction performance for shallow sea floor acoustic parameters inversion, and the prediction results have high reliability.

Figures 6A–F shows the ground sound parameter training results of the first sedimentary layer, and Figures 6G–K shows the ground sound parameter training results of the second sedimentary layer. As can be seen from the figure, the training effect of the first layer is better and the degree of fitting is higher. It can be seen from the literature that different parameters have different influences on acoustic propagation characteristics, especially c_{pn} and c_{sn} have the greatest influence on acoustic field characteristics, so the accuracy of inversion results of these parameters is higher than other parameters. From the fitting degree in the training process, it can be seen that the fitting degree of c_{pn} and c_{sn} parameters is better, which accords with the law of sound field calculation in the forward modeling model, which proves the applicability of the method.

A group of sound pressure data with a single true value is used for inversion calculation. Substitute the true value sound pressure data into the neural network models, and the classification results are shown in Table 3. It can be seen from Table 3 that the classification probabilities of Net 2-X model for the seabed layered structure under the true sound pressure are: X=1, rate=1.21×10^{-3}; X=2, rate=0.99; X=3, rate=1.52×10^{-4},The data is determined to be acoustic pressure data from a two-layer seabed.

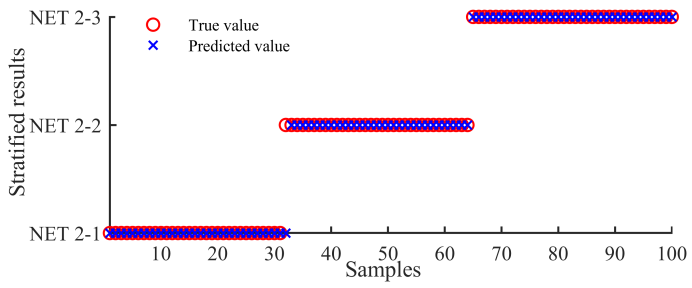


FIGURE 5
Seabed stratification results of NET-1.

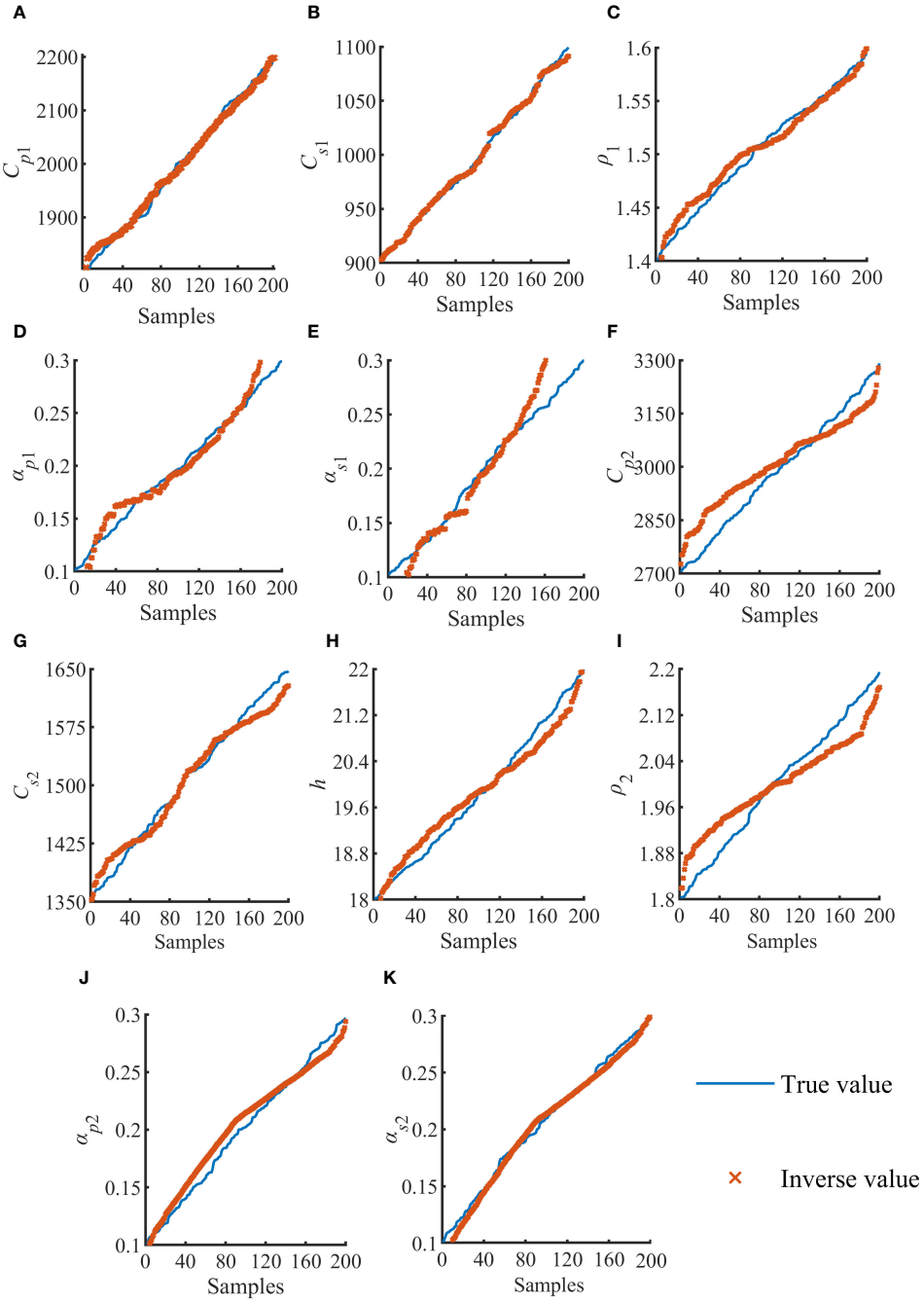


FIGURE 6
NET-2-2 Comparison between the actual and predicted values of 200 samples of single sediment. **(A–K)** are inversion results of 11 parameters in the case of NET 2-2, respectively.

TABLE 3 Prediction results of Net 2-X.

Net 2-X	Number of parameters	Classification Rate
X=1	5	1.21×10^{-3}
X=2	11	0.99
X=3	17	1.52×10^{-4}

After determining the bottom stratification structure, The input data will be brought into the corresponding neural network model Net 2-X for inversion calculation and the acoustic parameters inversion values are shown in Table 3. To verify the correctness of hierarchical judgment, the sound field calculation model at X= 1,2 and 3 is used for inversion discussion and comparison with X=2 respectively. Figure 7 shows the comparison between the Transmission Loss (TL) curve

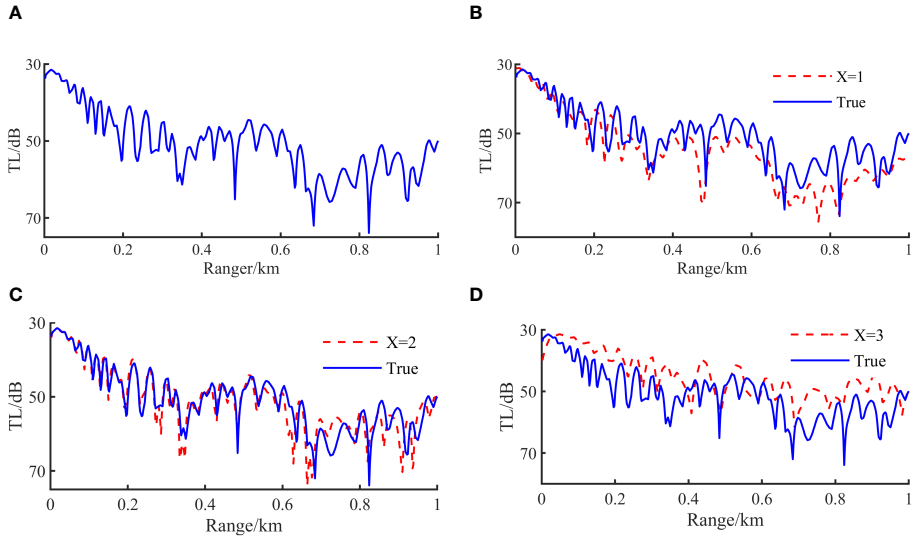


FIGURE 7
Comparison between X=2 simulated TL curve and X=1, 2, 3inversion TL curve: **(A)** TL curve under true value; **(B)** Comparison of TL curves of Net 2-1 models and true value. **(C)** Comparison of TL curves of Net 2-2 and true value. **(D)** Comparison of TL curves of Net 2-3 and true value.

calculated by setting the truth value of acoustic parameters and the TL curve calculated by using inversion results. It can be seen from the comparison that the distribution characteristics of the two curves are consistent, which further proves the accuracy of the inversion results of the subsurface acoustic parameters of the preset model based on BPNN model (Yang, 2023).

Combining the classification results in Table 3, the parameters inversion results in Table 4 and the comparison of TL curves under the three hierarchical structures in Figure 8, the above results show that when the actual model matches the inversion model exactly the prediction is X=2, which proves that the model is the best-parameterized model. At the same time, the TL comparison obtained by the inversion parameters of the X=1, 2, and 3 models also proves the above conclusion. When X=2, the parameters inversion results in Table 4 are consistent with the present. The above simulation results prove that this method can theoretically achieve accurate inversion of seabed layered structures and seabed geoacoustic parameters.

TABLE 4 Simulation parameters setting and search range.

Submarine stratification	Parameters	True value	Inversion results
Single sedimentary layer	c_{p1} (m/s)	2000.00	2040.48
	c_{s1} (m/s)	1000.00	1027.30
	ρ_{b1} (g/cm^3)	1.50	1.52
	α_{p1} (dB·λ^{-1})	0.20	0.19
	α_{s1} (dB·λ^{-1})	0.20	0.20
	h_1(m)	20.00	20.11
Double sedimentary layer	c_{p2} (m/s)	3000.00	3021.44
	c_{s2} (m/s)	1500.00	1468.33
	ρ_{b2} (g/cm^3)	2.00	2.01
	α_{p2} (dB·λ^{-1})	0.20	0.18
	α_{s2} (dB·λ^{-1})	0.20	0.20

3 Measurements and results

3.1 Introduction of experiment

Previous studies have shown that, under the assumption that various acoustic parameters in waveguides are unchanged, the idea of equal ratio can be used to simulate the acoustic propagation test in the actual ocean by using the equal ratio of high-frequency sound sources in the laboratory, and the expression only changes the ratio, while the propagation characteristics of the sound field remain unchanged. Based on simulation verification of the accuracy and applicability of the proposed method, the feasibility of the proposed inversion method in practical application is further verified in this section combined with the experimental data of the muffler pool shrinkage. The experiment was carried out in a hydrating pool, using a uniform and high-hardness polyvinyl chloride (PVC) plate (the density of PVC was 1.20g/cm^{-3}) as a "semi-infinite elastic seabed".

To verify the applicability of the inversion method, two schemes were adopted in the experiment as follows:

1. There is only PVC plate, simulating elastic semi-infinite seabed.
2. The way of laying fine sand on PVC plate simulates the shallow sea waveguide environment with a single layer of elastic sediment and an elastic version of the infinite seabed, and the thickness of the sediment simulated by fine sand is about 250 mm.

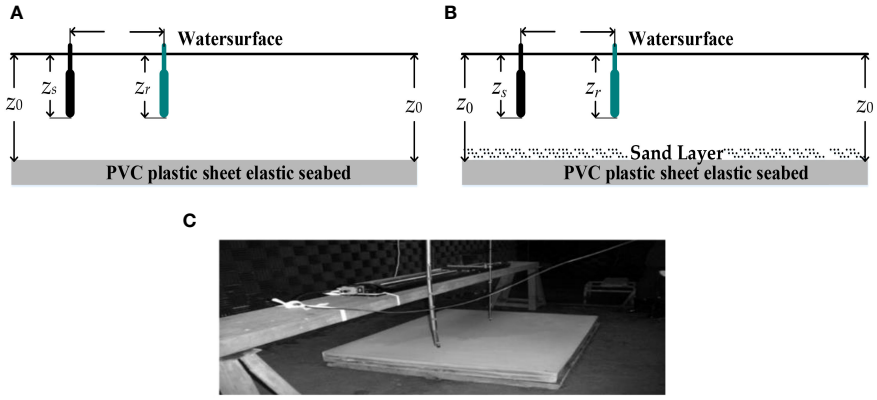


FIGURE 8
The equipment in the experiment: **(A)** Schematic diagram of the experiment under a single sedimentary layer (PVC); **(B)** Schematic diagram of the experiment under a double sedimentary layer (PVC+sand); **(C)** Experimental equipment layout physical map.

Only the stratification of the sea floor was different between the two schemes, and other experimental factors were consistent. The layout of experimental equipment 1 is shown in Figure 8A, The other is shown in Figure 8B. The depth of the sound source and the receiving hydrophone are set to 200mm, and the depth of the fluid layer is 300mm. The speed of sound calculated in the laboratory at room temperature is 1450.212 $m{\cdot}s^{-1}$. High-frequency underwater sound waves are transmitted by a fixed location sound source at a frequency of 155KHz and received by a single TC4038 standard hydrophone at different locations at equal intervals. To improve the measurement accuracy, a high-precision controllable moving platform was selected to limit the error within 2um and the unit accuracy was 2mm. A total of 500 position points were measured during the experiment. Each position was measured 10 times and the average value was taken as the final test data.

3.2 Model selection and inversion results

The computational process of the acoustic field measured under two experimental schemes is illustrated in Figure 2A. Initially, the layered structure is assessed using NET-1. The types of input data include semi-infinite seabed and two-layer seabed, with the layering results depicted in Figure 9, presented in the form of probabilities. It is observed that the NET-1 model's probabilistic assessment for the two input acoustic pressure signals as semi-infinite seabed and single sediment seabed is 99%, consistent with the scale model experiments. Additionally, a comparison between the simulated annealing algorithm and the classical annealing algorithm was conducted, incorporating the respective NET 2-X (X=1,2) models.

The inversion results obtained by BPNN and SA are shown in Table 5. In the pool experiment, PVC boards with a density of 1.20 g/cm^3 were used to simulate the seabed layer, and the thickness of the fine sand layer, representing the sediment layer, was set to 250 mm. Utilizing the data acquired from the pool experiment, the Back Propagation Neural Network (BPNN) inversion results indicated the simulated seabed layer density to be 1.23 g/cm^3, and the inferred thickness of the sediment layer to be 251.13 mm. The relative errors were found to be 2.5% and 0.51%, respectively, demonstrating good accuracy of the inversion process. In addition, the inversion results obtained in this paper are compared with those obtained from PVC plates simulating the same material of semi-infinite seabed. The velocity of P-wave and S-wave in the semi-infinite seabed is 2399.364 m/s and 1242.978 m/s. The relative error is controlled below 5%.

Considering the coupling effect of multiple parameters, to further verify the accuracy of inversion, the comparison between

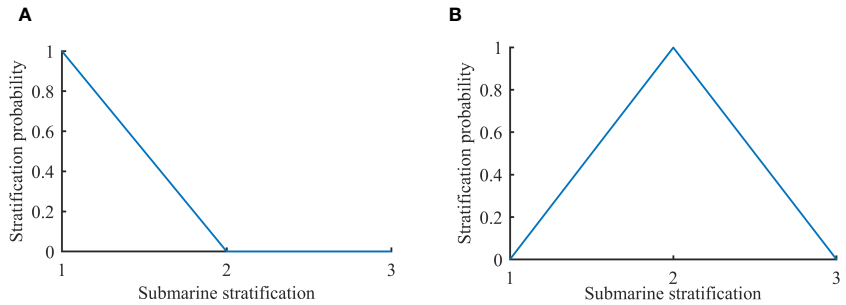


FIGURE 9
Judgment of two types of layered seabed structures: **(A)** NET2-1 judgment for semi-infinite seabed layering, **(B)** NET2-2 judgment for two-layer seabed stratification.

10.3389/fmars.2024.1349478

TABLE 5 Scheme 2 inversion results of measured data.

Submarine stratification	Parameters	BP inversion results	SA inversion results
Single sedimentary layer (Sand)	c_{p1} (m/s)	2074.84	2073.65
	c_{s1} (m/s)	1090.56	1093.76
	ρ_{b1} (g/cm^3)	1.13	1.11
	α_{p1} (dB·λ^{-1})	0.99	0.10
	α_{s1} (dB·λ^{-1})	0.10	0.10
	h_1(m)	251.378	249.69
Semi-infinite layer (PVC)	c_{p2} (m/s)	2436.37	2440.16
	c_{s2} (m/s)	1250.50	1255.81
	ρ_{b2} (g/cm^3)	1.19	1.21
	α_{p2} (dB·λ^{-1})	0.11	0.10
	α_{s2} (dB·λ^{-1})	0.11	0.10

the dispersion curve of BPNN inversion and the measured dispersion curve is shown in Figure 10. And under the two experimental schemes, the comparison curve of propagation loss between BPNN model and SA inversion algorithm on measured data is shown in Figure 11.

Figure 10 shows the frequency-wave number spectrum measured by the water tank. It can be seen from the spectrum that the energy of the received sound pressure signal is mainly distributed in the range of 145kHz-175kHz, and its peak value is around 155kHz, which is consistent with the performance index of the sound source set in the experiment. In order to verify the effectiveness of the BPNN inversion approach and to compare the TL computed from the inversion results for both semi-infinite and two-layer seabed configurations with the TL measured in scale-down experiments, the following findings were observed: For the semi-infinite single-layer seabed scenario, the TL was generally consistent. In the case of the two-layer seabed, taking into account various uncertainties such as the replacement of the sediment layer with fine sand and the homogeneity of the sand, certain discrepancies were observed. However, the overall trend of

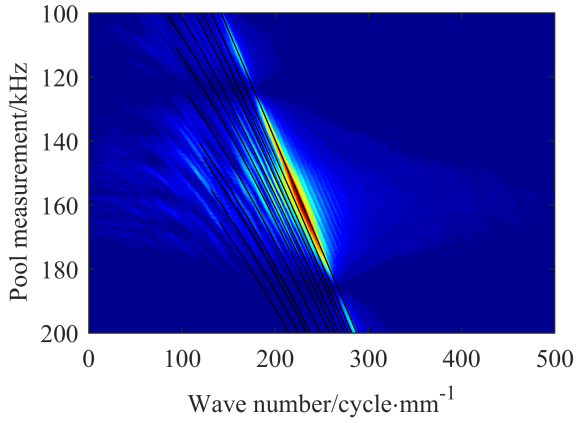


FIGURE 10
Comparison between BP inversion and measured dispersion curve.

the TL was closely aligned. This demonstrates that the inversion method proposed in this article yields results for the inversion of parameters such as longitudinal wave and shear wave velocities in multi-layer seabed structures that are in substantial agreement with the actual data.

3.3 Result analysis¿

Geoacoustic parameters inversion is a nonlinear and multi-parameter optimization problem, so the inversion results obtained by different inversion methods are not the same, but all have the same reliable reference value. The coupling of multiple parameters often leads to the experimental phenomenon of the same result in different parameters combinations, so it is necessary to discuss the sensitivity of different parameters to determine the reference weight of the choice of results.

The applicability of multi-parameter inversion using BPNN can be obtained from spectrum analysis and comparison with SA algorithm. From the comparison and verification of TL curves of the two experiments, it can be seen that when the seafloor structure is inversion for single-layer, the inverse performance results of

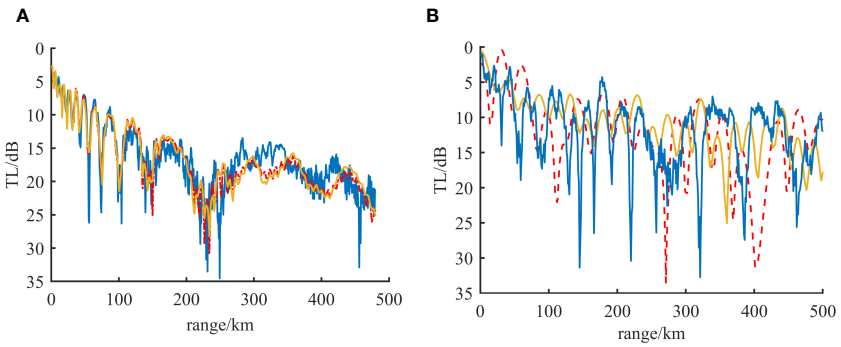


FIGURE 11
BP and SA inversion TL curve compared with measured TL: **(A)** Comparison diagram of TL curve under experimental scheme 1; **(B)** Comparison of propagation loss curves under experimental scheme 2ganxie.

BPNN and SA algorithm have higher accuracy; when the seafloor is inversion for double-layer seafloor, the number of ground acoustic parameters increases, the inversion difficulty increases, and the multi-parameter coupling is enhanced, so there is a relatively obvious deviation. However, in terms of the inversion results of the two algorithms, the root mean square error of the TL curve obtained by the experiment is 121.92 and 160.19, respectively, compared with that of BPNN TL curve and SA TL curve. Relatively speaking, the error of the inversion results of BPNN is smaller and more valuable for reference.

4 Conclusions

By employing the BPNN inversion method to train an extensive dataset of simulated information, a computational model has been established with the capability to effectively evaluate the stratified structure of the seabed along with its acoustic parameters. Within this study, the BPNN inversion algorithm has been applied to the inversion of Earth's acoustic parameters, thus contributing to the advancement of fields in geophysics and ocean acoustics. This approach holds significant promise for practical applications in the realms of seabed resource exploration, ocean environmental investigations, marine engineering, and the exploitation of ocean resources, among others.

Our research findings can be succinctly summarized as follows:

The Fast Field Method (FFM) was utilized to derive theoretical predictions for the shallow-sea sound pressure field. Subsequently, a model was established within the BPNN framework, connecting the predicted sound pressure field with the underlying acoustic parameters. The measured sound pressure field data were then processed through the neural network model to obtain inversion results. Both simulated and experimental data confirm the accuracy of this proposed method in retrieving geoacoustic parameters.

The acoustic pressure field in water is influenced by five crucial acoustic parameters associated with shallow seabed conditions: bottom density, p-wave velocity, S-wave velocity, p-wave velocity attenuation, and S-wave velocity attenuation. Our results reveal that the accuracy of the inversion results for S-wave velocity (c_p), p-wave velocity (c_s), and sedimentary layer density (ρ_b) surpasses that of S-wave attenuation (α_p) and p-wave attenuation (α_s). The first three acoustic parameters exert a more pronounced impact on the propagation characteristics of the shallow-sea sound pressure field, thus demonstrating a more apparent correlation with the acoustic pressure data. This further illustrates the effectiveness of employing the BPNN model for inversion in the context of shallow seabed layering.

Given the practical complexities of shallow seabed conditions, compounded by the influence of various underwater noises on sound field distribution, actual computational results may exhibit some deviation from real-world scenarios. Additionally, the coupling relationships between seabed physical parameters and the sensitivity of each parameter can impact calculation accuracy. To address these limitations, we have optimized the BPNN model by adjusting its neural network structure and introducing random noise in subsequent research, thus enhancing the reliability of the Earth's acoustic parameter inversion model. Moreover, we have conducted an exhaustive exploration and discussion concerning the impact of network configuration parameters on the accuracy of inversion results.

These revisions are aimed at addressing the reviewer's concerns and enhancing the clarity and impact of the conclusion section.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

JW: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. ZC: Conceptualization, Data curation, Methodology, Writing – original draft. HZ: Methodology, Validation, Writing – review & editing. LM: Formal analysis, Supervision, Writing – review & editing. WS: Funding acquisition, Project administration, Resources, Writing – review & editing. XL: Investigation, Methodology, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (Grant No: 12374425), the Stable Supporting Fund of the National Key Laboratory of Underwater Acoustic Technology (JCKYS2023604SSJS016), the Science and Technology program of Zhoushan City (2023C41025), the Fund of Fujian Provincial Key Laboratory of Marine Physical and Geological Processes (KLMPG-23-01) and the Zhejiang University Student Science and Technology Innovation Program (New Miao Talent Program) (2022R411C050, 2022R411C052).

Conflict of interest

Author ZC was employed by the company Hydroacoustics Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Chen, J.-H., Chen, L.-P., Xia, X.-Y., and Zhu, R. (2022). Classification of submarine sonar images based on transfer learning. *Comput. Simul.* 39, 229–233.

Cheng, X. U. E., Zaixiao, G., Yiming, G. U., Yu, W., Peng, L. I. N., and Zhenglin, L. I. (2021). Channel matching of shallow water active detection combined with convolutional neural network. *Acta ACUSTICA* 46 (6), 800–812. doi: 10.15949/j.cnki.0371-0025.2021.06.003

Dragna, D., and Blanc-Benon, P. (2017). Sound propagation over the ground with a random spatially-varying surface admittance. *J. Acoust. Soc. Am.* 142 (4), 2058–2072. doi: 10.1121/1.5006180

Feng, X., Zhou, M., Zhang, X., Ye, K., Wang, J., and Sun, H. (2022). Variational bayesian inference based direction of arrival estimation in presence of shallow water non-gaussian noise. *J. Electron. Inf. Technol* 44 (6), 1887–1896. doi: 10.11999/JEIT211284

Frederick, C., Villar, S., and Michalopoulou, Z.-H. (2020). Seabed classification using physics-based modeling and machine learninga). *J. Acoust. Soc. Am.* 148 (2), 859–872. doi: 10.1121/10.0001728

Gerstoft, P. (1994). Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions. *J. Acoust. Soc. Am.* 95 (2), 770–782. doi: 10.1121/1.408387

Huang, Z., Xu, J., Gong, Z., Wang, H., and Yan, Y. (2018). Source localization using deep neural networks in a shallow water environment. *J. Acoust. Soc. Am.* 143 (5), 2922–2932. doi: 10.1121/1.5036725

Ke, Q. U., Changqing, H. U., and Mei, Z. (2013). Single parameter inversion using transmission loss in shallow water. *Acta ACUSTICA* 38 (4), 472–476. doi: 10.15949/j.cnki.0371-0025.2013.04.017

Kerzhakov, B. V., and Kulinich, V. V. (2016). Retrieval of sea-bed parameters by the method of matching acoustic fields on the basis of vertical angular spectra. *Radiophysics Quantum Electron.* 59 (3), 217–224. doi: 10.1007/s11141-016-9690-x

Li, H., Guo, X.-Y., and Ma, L. (2019). Estimating structure and geoacoustic parameters of sub-bottom by using spatial characteristics of ocean ambient noise in shallow water. *Acta Physica Sin.* 68 (21), 214303–214301-214303-214312. doi: 10.7498/aps.68.20190824

Li, R., Zhang, H., Zhuang, Q., Li, R., and Chen, Y. (2020). BP neural network and improved differential evolution for transient electromagnetic inversion. *Comput. Geosci.* 137, 104434. doi: 10.1016/j.cageo.2020.104434

Li, X., Piao, S., Zhang, M., and Liu, Y. (2019). A passive source location method in a shallow water waveguide with a single sensor based on bayesian theory. *Sensors* 19 (6), 1452. doi: 10.3390/s19061452

Li, Z., and Zhang, R. (2017). Hybrid geoacoustic inversion method and its application to different sediments. *J. Acoust. Soc. Am.* 142 (4_Supplement), 2558–2558. doi: 10.1121/1.5014351

Pang, Y., Xu, F., and Liu, J. (2021). Classification of seafloor sediment based on Gammatone Filter banks time Spectrum and Convolutional Neural networks. *Appl. Acoust.* 40 (04), 510–517. doi: 10.11684/j.issn.1000-310X.2021.04.003

Potty, G. R., Miller, J. H., Dosso, S. E., Bonnel, J., Dettmer, J., and Isakson, M. J. (2017). Sediment parameter inversions in the East China Sea. *J. Acoust. Soc. Am.* 141 (5_Supplement), 3487–3487. doi: 10.1121/1.4987270

Qian, G., Zhang, L., and Wang, Y. (2019). Single-label and multi-label conceptor classifiers in pre-trained neural networks. *Neural Comput. Appl.* 31 (10), 6179–6188. doi: 10.1007/s00521-018-3432-2

Song, W., and Wang, P. (2022). High-resolution modal wavenumber estimation in range-dependent shallow water waveguides using vertical line arrays. *J. Acoust. Soc. Am.* 152 (1), 691–705. doi: 10.1121/10.0012187

Stoll, R. D., and Kan, T. K. (1981). Reflection of acoustic waves at a water–sediment interface. *J. Acoust. Soc. Am.* 70 (1), 149–156. doi: 10.1121/1.386692

Van Komen, D. F., Neilsen, T. B., Howarth, K., Knobles, D. P., and Dahl, P. H. (2020). Seabed and range estimation of impulsive time series using a convolutional neural network. *J. Acoust Soc. Am.* 147 (5), El403. doi: 10.1121/10.0001216

Wang, Z.-J. (2008). Inversion of seabed parameters for vertical array (Harbin Engineering University). Master.

Wang, J., Cao, J.-X., Zhao, S., and Qi, Q.-M. (2022). Inverse prediction method of shear wave velocity based on deep hybrid neural network. *Scientia Sin. Terrae* 52, 1151–1169. doi: 10.1360/SSTe-2021-0128

Wang, Z., Ma, Y., Kan, G., Liu, B., Zhou, X., and Zhang, X. (2023). An inversion method for geoacoustic parameters in shallow water based on bottom reflection signals. *Remote Sens.* 15 (13), 3237. doi: 10.3390/rs15133237

Wen, B.-Y., Tang, W.-C., and Tian, Y.-W. (2021). Significant wave height field inversion of high frequency radar based on BP neural net-work. *J. Huazhong Univ. Sci. Technol.* 49 (4), 114–119. doi: 10.13245/j.hust.210420

Xu, Y.-Y., and Pan, X. (2018). Research on passive geoacoustic inversion. *J. Hangzhou Dianzi Univ. (Nat. Sci. Ed.).* 38, 45–49.

Xue, Y., Zhu, H., Wang, X., Zheng, G., Liu, X., and Wang, J. (2023). Bayesian geoacoustic parameters inversion for multi-layer seabed in shallow sea using underwater acoustic field. *Front. Mar. Sci*, 10. doi: 10.3389.2023/fmars.1058542

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Appl.*, 1–17. doi: 10.1007/s11042-023-16757-0

Yang, L.-F., Zhu, Z.-R., Li, J.-B., and Feng, C.-K. (2021). Seafloor classification based on combined multibeam bathymetry and backscatter using deep convolution neural network. *Acta Geod. Cartogr. Sin.* 50, 71–84. doi: 10.11947/J.AGCS.2021.20200065

Yu, S.-Q., Wang, F., and Zheng, G.-Y. (2020). Progress and discussions in acoustic properties of marine sediments. *J. Harbin Eng. Univ.* 41, 1571–1577. doi: 10.11990/jheu.202007049

Zhang, X. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *Multimedia Tools Appl.*, 1–18. doi: 10.1007/s11042-023-16992-5

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhao, Y., Zhu, K., Zhao, T., Zheng, L., and Deng, X. (2023). Small-sample seabed sediment classification based on deep learning. *Remote Sens.* 15 (8), 2178. doi: 10.3390/rs15082178

Zheng, G.-X., Piao, S.-C., and Zhu, H.-H. (2021). Bayesian inversion method of geo-acoustic parameters in shallow sea using acoustic pressure field. *J. Harbin Eng. Univ.* 42, 497–504. doi: 10.3390/s20072150

Zhou, J., Zhang, M., Piao, S., Iqbal, K., Qu, K., Liu, Y., et al. (2019). Low frequency ambient noise modeling and comparison with field measurements in the South China Sea. *Appl. Acoust.* 148, 34–39. doi: 10.1016/j.apacoust.2018.11.013

Zhu, H., Cui, Z., Liu, J., Jiang, S., Liu, X., and Wang, J. (2023a). A method for inverting shallow sea acoustic parameters based on the backward feedback neural network model. *J. Mar. Sci. Eng.* 11 (7), 1340. doi: 10.3390/jmse11071340

Zhu, H.-H., Hai-Gang, Z., and Wei, L. (2013). Research on the influence of sound speed profile to the dispersion in shallow water waveguide. *Tech. Acoust.* 32, 67–68.

Zhu, H.-H., Piao, S., Zhang, H., Liu, W., and An, X. (2012). The research for seabed parameters inversion with fast field program (FFP). *J. Harbin Eng. Univ.* 33, 648–652 +659. doi: 10.3969/j.issn.1006-7043.201105075

Zhu, H., Xue, Y., Ren, Q., Liu, X., Wang, J., Cui, Z., et al. (2023b). Inversion of shallow seabed structure and geoacoustic parameters with waveguide characteristic impedance based on Bayesian approach. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1104570

Zhu, H.-H., Zheng, G.-X., and Zhang, H.-G. (2017). Study on propagation characteristics of low frequency acoustic signal in shallow water environment. *J. Shanghai Jiao Tong Univ.* 51, 1464–1472. doi: 10.16183/j.cnki.jsjtu.2017.12.009

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 19 March 2024
DOI 10.3389/fmars.2024.1351390

OPEN ACCESS

EDITED BY
Arata Kaneko,
Aqua Environmental Monitoring Limited Liability Partnership (AEMLLP), Japan

REVIEWED BY
Jae-Hun Park,
Inha University, Republic of Korea
Ju Lin,
Ocean University of China, China

*CORRESPONDENCE
Naokazu Taniguchi
ntaniguchi@hiroshima-u.ac.jp;
naokazutaniguchi@gmail.com

RECEIVED 06 December 2023
ACCEPTED 16 February 2024
PUBLISHED 19 March 2024

CITATION
Taniguchi N, Mutsuda H, Arai M, Sakuno Y, Hamada K, Huang C-F, Guo J, Takahashi T, Yoshiki K and Yamamoto H (2024) Application of coastal acoustic tomography: calibration of open boundary conditions on a numerical ocean model for tidal currents.
Front. Mar. Sci. 11:1351390.
doi: 10.3389/fmars.2024.1351390

Application of coastal acoustic tomography: calibration of open boundary conditions on a numerical ocean model for tidal currents

Naokazu Taniguchi[1]*, Hidemi Mutsuda[1], Masazumi Arai[1], Yuji Sakuno[1], Kunihiro Hamada[1], Chen-Fen Huang[2], JenHwa Guo[3], Toshiyuki Takahashi[4], Kengo Yoshiki[4] and Hironori Yamamoto[4]

[1]Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Japan, [2]Institute of Oceanography, National Taiwan University, Taipei, Taiwan, [3]Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei, Taiwan, [4]Fukken Co., LTD, Hiroshima, Japan

Coastal acoustic tomography (CAT), which measures path-averaged currents from reciprocal acoustic transmission experiments and reconstructs velocity fields from the multiple path-averaged current data, is useful for monitoring tidal currents in coastal shallow water, especially if data assimilation is employed. Previous CAT data assimilation studies have focused on state estimation problems, i.e., the reconstruction of tidal currents and following dynamical discussion. In this study, we investigate the use of path-averaged currents in a boundary control problem. Specifically, we aim to use the observed path-averaged currents to determine the parameters of a numerical ocean model, which were tidal amplitudes and phases as the open boundary conditions in this study. We investigate two methods: using the ensemble Kalman filter (EnKF) results and a linearization approach called model Green's function method. Both calibration methods decreased the amplitudes of tidal constituents at the open boundaries. We compare the model performance between the model predictions with and without the calibration of the open boundary conditions. The model predictions with the calibrated open boundary conditions improved the agreement with the observed path-averaged current. We also implemented the sequential updates of EnKF with the two calibrated open boundary conditions. The EnKF results with the independently calibrated two open boundary conditions improved the agreement with the comparison data obtained by acoustic Doppler current profiler measurement compared with the original EnKF result with the initial open boundary conditions.

KEYWORDS

tidal currents, reciprocal acoustic transmission, coastal acoustic tomography, data assimilation, open boundary condition, boundary control problem, ensemble Kalman filter, model Green's function

1 Introduction

Sound waves are a practical tool for remotely sensing the ocean interior where the electromagnetic waves cannot penetrate and thus cannot be used as an observational tool. There are various applications that actively or passively use sound as a tool to know the ocean (or ocean-related issues), and those applications are termed acoustical oceanography (e.g., Howe et al., 2019). One such example of acoustical oceanography is using sound to infer the two- or three-dimensional fields of ocean sound speed and ocean currents (e.g., Worcester, 1977; Elisseeff et al., 1999; Dushaw et al., 2001, 2010). Sound travels faster through the warm (and high-salinity) water and with the direction of the ocean currents than through the cold (and less saline) water and against the ocean currents; thus, one can inversely estimate the sound speed and ocean currents by transmitting a sound pulse and measuring the travel time of the received pulse between multiple sources and receivers located horizontally separately. The effects of sound speed and current on the travel time can be separated by making transmissions in the forward and reverse directions between a pair of transceivers, namely a reciprocal transmission. The sound speed is related to the sum of the travel times of a reciprocal transmission, while the magnitude of ocean current is related to the differential travel times. Sound travels through the ocean at about 1,500 m s^{-1}, which is sufficiently fast compared with the timescales of ocean mesoscale eddies and the speed of observation vessels. Thus, this acoustical method is unique in the sense that one can estimate the nearly instantaneous state of the ocean interior. The method is known as ocean acoustic tomography (OAT; Munk and Wunsch, 1979; Munk et al., 1995). It is often referred to as acoustic thermometry (Dushaw et al., 2009) or coastal acoustic tomography (CAT; Kaneko et al., 2020) when the method is specially used to study large-scale ocean temperature estimation or dynamics of coastal shallow waters with relatively small spatial scales. This paper is closely related to CAT, and we focus on the method as a reconstruction tool of tidal currents in coastal shallow water.

Reconstructing velocity fields of tidal currents from observed travel times corresponds to solving an inverse problem. In CAT inverse problems, it is often the case that there is no unique solution (namely, the problem is ill-posed). Previous studies have solved their inverse problems and found solutions using some prior knowledge or regularization methods (e.g., Yamaoka et al., 2002; Yamaguchi et al., 2005). Researchers have tried to improve the estimations in such ill-posed problems by deploying a relatively large number of transceivers (Zhang et al., 2017a), by distributing transceivers to form a sensor network (Huang et al., 2013; Zhang et al., 2017b), or by using ships or autonomous vehicles to augment the travel time observation on various paths (Huang et al., 2019). Another promising approach for sparse observation (compared with a dimension of states) is data assimilation. In data assimilation, predictions from dynamical models (numerically modeled Navier-Stokes equations, for example) are optimally combined with observation data (e.g., Carrassi et al., 2018). Several CAT studies have implemented the ensemble Kalman filter (EnKF; Evensen, 1994, 2003) as a data assimilation scheme. These CAT data assimilation with EnKF have improved reconstruction compared to the results obtained by solving data-oriented inverse problems (e.g., Park and Kaneko, 2000; Lin et al., 2005; Zhu et al., 2017). The authors also demonstrated the usefulness of CAT data assimilation; strong and rapid spatiotemporal variation in the tidal current, including an island wake with multiple vortex generation at a downstream side of an island, was reconstructed by a CAT data assimilation with EnKF scheme (Taniguchi et al., 2023). As described here, previous CAT studies have mainly focused on reconstructing velocity fields of tidal currents; these may be termed state estimation problems (Munk et al., 1995).

In this study, as opposed to the previous CAT studies for state estimation problems, we investigate the use of reciprocal acoustic transmission data to ask what open boundary conditions are required to drive the ocean model so that it reproduces the sequence of observed states. Such a problem may be referred to as a boundary control problem (Munk et al., 1995) in contrast to the state estimation problem. Specifically, using the observation data and the numerical model used in the previous report (Taniguchi et al., 2023), we control (or calibrate or optimize) the parameters of the open boundary condition, which were tidal amplitudes and phases and determined in a somewhat *ad-hoc* way, by using the observed path-averaged currents. Since open boundary conditions can affect the states throughout the model domain while path-averaged currents obtained by reciprocal acoustic transmission contain the non-local information averaged over the paths, it is expected that path-averaged currents can effectively be used to control open boundary conditions. For this purpose, we investigate two methods: the use of EnKF results and a linearization approach proposed by Menemenlis et al. (2005). In the first method, data assimilation with the EnKF scheme can sequentially update the velocity fields of tidal currents as with the previous report (Taniguchi et al., 2023); then, we apply a harmonic analysis to the time series of normal velocities at open boundary grids and re-compute the amplitudes and phases of tidal constituents there. The second method is termed model green's function approach in Menemenlis et al. (2005). The method assumes that the differences between the observed values and model-predicted values with initial model parameters can be represented by a linear combination of the sensitivity of model predictions to those model parameters. With the assumption, the practical method involves the numerical experiments of model sensitivity to the parameters followed by solving a linear inverse problem to find how much one should deviate the model parameters from the initial values. The method has been used widely as the parameter optimization method in many studies, including tide-related studies with local and regional scales (e.g., Moon et al., 2012; Kobayashi et al., 2016). We compare the model prediction accuracy between the models with and without the calibration of the open boundary conditions. Also, with the calibrated tidal parameters determined by the two methods, we re-perform the sequential data assimilation with EnKF to estimate the velocity fields at each transmission time. We shall show that both EnKF results with the independently calibrated two open boundary

conditions improve the agreement with the acoustic Doppler current profiler observations compared to the original EnKF results with uncertain open boundary conditions.

2 Materials and methods

2.1 Field experiment

An experiment on reciprocal acoustic transmission between four acoustic stations was conducted in an area named Mihara-Seto in the Seto Inland Sea, Japan, from the end of October to December 2020. Figure 1 shows the geographical location of the observation site and the locations of the four acoustic stations named S1, S2, S3, and S4. The coast blocks the transmission between the S1 and S4 stations, i.e., the travel time between them is not observable. During the first two weeks of the experiment, the reciprocal transmissions between the S2 and S3 stations failed almost every low tide, which might be due to the existence of shallow sand banks near the S3 station. Two weeks after the experiment started, we slightly moved the location of S3 to prevent the sound propagation block by the shallow bank. The distances between the five station pairs (S1 and S2, S1 and S3, S2 and S3, S2 and S4, and S3 and S4) were 2,842, 4,930, 2,895, 4,300, and 4,220 m, respectively, after the S3 location was moved.

The acoustic transmitting/receiving system used for this travel time measurement was a further modified version of the system used in a preliminary experiment in the same area in 2019 (Taniguchi et al., 2021a), which was a modification of the system originally used in a moving ship tomography study (Huang et al., 2019). The system consisted of three primary items: a microcontroller with peripheral electrical circuits, a global navigation satellite system (GNSS) receiver module and antenna, and an electro-acoustic transducer. The electro-acoustic transducer used in this experiment was the Model T235 of Neptune Sonar, which can operate over a frequency range of 10–25 kHz, and was used as a transceiver (i.e., both transmitter and hydrophone). The deployments of the transceivers were the same as those in the preliminary experiment (Taniguchi et al., 2021a). The sound transmission circuit consisted of a full-bridge inverter and a step-up voltage transformer. The amplified signal was fed to the transducer with an additional tuning inductor. The source level at full resonance was estimated to be 190 dB re 1 μPa at 1 m. The received signals were amplified and demodulated into in-phase and quadrature components. These two signals were digitized by a 12-bit analog-to-digital converter in the microcontroller. The sampling frequency was twice the carrier frequency of the transmitted signal. The digitized data were then recorded on an SD memory card on the electrical circuit board.

A pulse compression method has been implemented to increase the signal-to-noise ratio (SNR) without sacrificing the time resolution. The transmission signal was a binary phase-shift keyed signal encoded by a pseudo-random binary sequence called a maximal length sequence, which is often referred to as an m-sequence, with a carrier frequency of 18.018 kHz. The length of the m-sequences was 2,047 digits, and each digit contained four carrier cycles. The last 63 and the first 64 digits of the m-sequence were added to the head and tail of the m-sequence, respectively, to achieve the original autocorrelation property of repeated m-sequences during the matched filtering operation (Taniguchi et al., 2021b). The duration of the transmitted signal was then 482 ms. The four stations transmitted the signal in synchronization with the GNSS timing pulse, but with fixed time lags given to each station so that the arrival signals of other stations did not overlap at all stations (Taniguchi et al., 2021b). At the receiver side, the arrival signal appeared as sharp arrival pulses after calculating the cross-correlation between the received (demodulated) signal and the replica of the transmitted m-sequence. A post processing gain of this matched-filtering was 33 dB. Such reciprocal transmissions were performed every two minutes. More information on the system and signal for

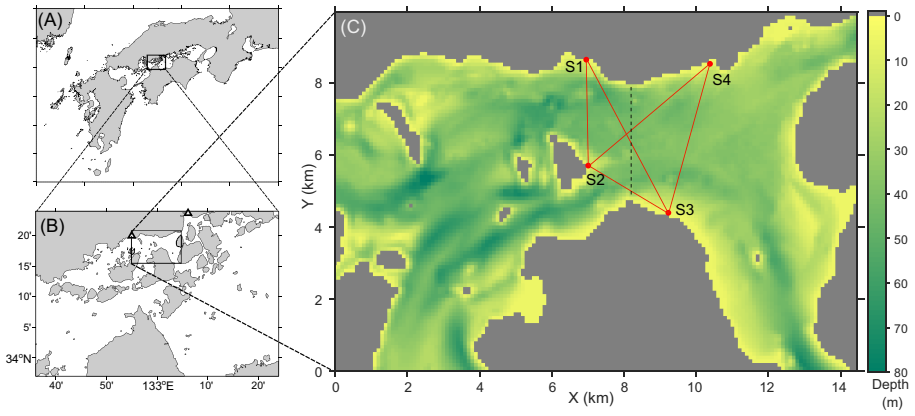


FIGURE 1
(A, B) Geographical location of observation site (Mihara-Seto) in the Seto Inland Sea, Japan; **(C)** domain of a numerical ocean model used in this study with the corresponding locations of acoustic stations (four red circles with labels S1, S2, S3, and S4) while the color indicates the water depth. The two black triangles in panel **(B)** is the location where astronomical tide information is obtained: Takehara for the west and Itozaki for the east. In panel **(C)**, the black dashed line indicates the tracks of shipboard acoustic Doppler current profiling (ADCP) observation performed on Oct. 30. The ADCP observation was also performed nearly along the transmission paths (red solid lines) on Oct. 31 except for the path between the S2 and S3 stations.

the reciprocal acoustic transmission experiment can be found in the authors' previous papers (Taniguchi et al., 2021a, 2021b, 2023).

In the shallow water environment of the observation site, acoustic waves propagating along multiple ray paths arrive at the receiver nearly simultaneously, forming an arrival pulse that is slightly wider than the pulse width of a single ray arrival (Taniguchi et al., 2021a). In such cases, it is difficult to determine the arrival time of individual rays. Thus, the time at which the height of the received arrival pattern first exceeded 14 dB was adopted as the travel time in this study. By adopting such a relatively low threshold, we aimed to capture the arrival time associated with the first arrival, which would be composed of the arrivals of direct and surface-reflected rays (Taniguchi et al., 2021a). We confirmed that 14 dB is generally higher than the noise level and captures the rising edge of the first arrival pulses.

For each paired reciprocal acoustic transmission, differential travel times (τ_d) were computed from the determined travel times and converted to path-averaged current (u) as follows (Worcester, 1977; Howe et al., 1987; Zheng et al., 1997):

$$u = -\frac{c_0^2}{2L}\tau_d\,, \tag{1}$$

where L is the transceiver-to-transceiver distance, and c_0 is the reference sound speed value estimated using the sum of the reciprocal travel times. In this study, we computed the τ_d as, for example, $\tau_d = \tau_{S1\leftarrow S2} - \tau_{S2\leftarrow S1}$, i.e., the travel times of sound pulses from the station with a larger id to the station with a smaller id minus those from a smaller id to a larger id. In this case, with the minus sign in Equation 1, a positive value of u indicates that the direction of the path-averaged current is from the station with a larger id to the station with a smaller id. The erroneous estimates of the path-averaged velocity were removed and linearly interpolated if the data gaps were less than 10 minutes.

The water is well mixed in the sea around the observation site, and there is no density stratification. This condition allows the tidal currents to be nearly vertically uniform at the observation site. The vertically uniform velocity structure was also confirmed using the ADCP observation results (Taniguchi et al., 2021a). Therefore, the path-averaged currents determined from detecting the first arrival can be identified as the depth- and range-averaged currents. The above method of travel time determination made the path-averaged currents consistent with the acoustic Doppler current profiler (ADCP) results, as seen in the Results section.

Hourly shipboard ADCP observations were performed on October 30 and 31 to obtain velocity data for comparison with model and data assimilation results. Readers can refer to Taniguchi et al. (2023) for the ADCP operation parameters. Twenty-one cross-sections of velocity data were obtained along the north-south transect (black dashed in Figure 1) and along the transmission lines (red solid lines in Figure 1) from these ADCP observations. The obtained ADCP velocities were averaged over the depths at each location and also spatially (horizontally) averaged over about 100 m so that the spatial resolution of the ADCP velocity data is nearly the same as that of a numerical ocean model.

2.2 Numerical ocean model

A numerical ocean model used in this study was the same as that used in the authors' previous study (Taniguchi et al., 2023) and was based on shallow water equations for the depth-averaged two-dimensional flows. The prognostic variables are eastward and northward components of the depth-averaged velocity (U and V) and tidal height, which is defined as a sea-surface height with respect to a mean sea level. The model domain with a horizontal grid space of 100 m is shown in Figure 1. The shallow water equations are solved numerically using the finite difference method with the numerical discretization and integration schemes following those used in the Princeton Ocean Model (POM; Blumberg and Mellor, 1987). The time integration with a time step of 1 s was performed using a leap-frog scheme with a Robert-Asselin filter with an additional modification proposed by Williams (2009).

The present model has six open boundaries (two at the west, east, and south), where prescribed tidal forcing (sea surface elevation and normal velocity) drives the model interior. The same boundary conditions were given to the two open boundaries at the west or east; thus, four open boundary conditions drove the model interior. A Flather condition is applied to the tidal height and normal velocity at the four open boundaries, following the form shown by Carter and Merrifield (2007). The tangential velocities at the boundary grids were given by the values of neighboring interior grids (i.e., a zero-gradient condition).

The normal velocity v and tidal height ζ at the open boundary grids were given by the sum of variation due to five tidal constituents:

$$v(t) = \sum_{i=1}^{5} A_i \cos\left(\frac{2\pi t}{T_i} + \theta_i\right), \tag{2}$$

$$\zeta(t) = \sum_{i=1}^{5} B_i \cos\left(\frac{2\pi t}{T_i} + \phi_i\right), \tag{3}$$

where A_i (B_i), T_i, and θ_i (ϕ_i) are the amplitude, period, and phase of the five tidal constituents, respectively. Note that T_i is known value for each tidal constituent. The five tidal constituents are M_2, S_2, N_2, K_1, and O_1 tides in this study. These five constituents are the five largest constituents in the Seto Inland Sea, including the observation site. The following three largest contributions are from the S_a, K_2, and P_1 tides. Since it requires observations with periods of about half or one year to separate the contribution of these tides from other tides by a harmonic analysis (the Rayleigh criterion; e.g., Schureman, 1958), we did not consider these tides because of the shorter duration of our travel time measurement. The forms of Equations 2, 3 are also expressed in the exchangeable forms of Equations 4, 5:

$$v(t) = \sum_{i=1}^{5}\left[a_i \cos\left(\frac{2\pi t}{T_i}\right) + b_i \sin\left(\frac{2\pi t}{T_i}\right)\right], \tag{4}$$

$$\zeta(t) = \sum_{i=1}^{5}\left[c_i \cos\left(\frac{2\pi t}{T_i}\right) + d_i \sin\left(\frac{2\pi t}{T_i}\right)\right]. \tag{5}$$

The prescribed tidal heights $\zeta(t)$ at the west and east boundaries were derived from astronomical tides at the nearest tide stations: Takehara for the west and Itozaki for the east (triangles in Figure 1B). There is no tide station near the two southern boundaries. Thus, the tidal elevations at the southwestern and southeastern boundaries were given by the same tidal heights as the western and eastern boundaries.

Because there is no information on normal velocities at the open boundaries, these were determined via an *ad-hoc* way with the information on the observed path-averaged current as follows. We applied a harmonic analysis to the observed path-averaged current between the S1 and S3 stations and estimated the amplitudes and phases of each tidal constituent. We considered the estimated phases as representative phases for tidal currents at the observation site. Next, we computed the phase differences between the astronomical tides (for the tidal height) and the path-averaged currents for each constituent. Then, we shifted the phases of the path-averaged current by the estimated phase differences and set them to the phases for the normal velocity. As for the amplitudes, we obtained the maximum tidal current at the center of the channel located at the northeastern boundary as about 2.8 m s^{-1} from a nautical chart (not shown here). This maximum current speed was multiplied by 0.8 to roughly convert it to section-averaged current speed, resulting in a maximum section-averaged current of 2.2 m s^{-1}. Here, the relationship between the maximum and the section-averaged flow with coefficient 0.8 was derived by referring to those relationships in turbulent pipe flow with the one-seventh power low. Then, the amplitudes of the estimated five tidal constituents of the path-averaged current were scaled so that the reconstructed tidal currents by the five constituents reached 2.2 m s^{-1} at the maximum during a spring tide. We set these scaled amplitudes as the amplitudes for the normal velocity at the northeastern boundary. Then, amplitudes of the five constituents at other western and southern boundaries were determined so that the volume transport through these boundaries without the tidal heights equaled that at the northeastern boundary. As described here, the open boundary conditions were derived arbitrarily, particularly for the normal velocity. Therefore, controlling (calibrating) those boundary conditions is expected to improve the reconstruction of tidal currents.

2.3 Open boundary condition improvement

2.3.1 Ensemble method

In the previous report (Taniguchi et al., 2023), we demonstrated that data assimilation with EnKF improves the reconstruction of tidal currents compared with the model prediction without data assimilation. The calibration of the open boundary conditions uses the results of the data assimilation with EnKF. Specifically, by applying harmonic analysis to the time series of velocity fields obtained by the data assimilation with EnKF, we re-compute the amplitudes and phases of tidal constituents at the open boundary grids. We focus on a relatively narrow area in the present study (see Figure 1). In such a narrow area, tidal currents are spatially correlated throughout the model domain, and thus, ensemble correlation in the covariance can be used as the physical (or real) correlation between the states. The velocity at the open boundaries would have been reasonably updated by EnKF via those physical correlations, even though the open boundaries are somewhat apart from the transmission paths. Thus, applying harmonic analysis to the results of EnKF data assimilation, one can determine better amplitudes and phases of the tidal constituents.

The EnKF implementation is nearly identical to that in the previous paper (Taniguchi et al., 2023). Readers can refer to Taniguchi et al. (2023) for the EnKF implementation specific to the present study. Below, we describe some key features. Ensembles of 98 members are created by perturbing the amplitudes and phases of the forcing tidal height and normal velocity at the open boundary girds; this method, i.e., perturbing boundary condition by adding noise to each ensemble, is commonly used in CAT data assimilation with EnKF. The model integration started at 00:00 on October 25, 2020, with no motion as the initial condition. The data assimilation step (EnKF updates) started at 16:00 on October 28, which is the approximate time when the path-averaged currents for all the station pairs started to be measured. The model state update was repeated every two minutes. If the number of successful reciprocal transmissions was two or less, the EnKF update was skipped at that time. The state vector contains the east-west and north-south velocity components (U and V) and tidal height at all grids. Both the velocities and tidal height were updated, although the tidal height was not observed in the reciprocal acoustic transmission experiment. The covariance localization was not implemented because the state vector must be updated throughout the model domain, including the open boundary grids. The covariance inflation was implemented with an inflation factor of 1.02.

A difference from the previous implementation was the value of observation error covariance, which is used to make perturbed measurements (Burgers et al., 1998) and to compute the data error covariance matrix in Kalman gain. During the investigation, we found that specifying small observation errors (e.g., 0.025 m s^{-1}; Taniguchi et al., 2021a) led to implausible updates of the state vector and spurious variation apart from the observations sometimes. The value of the observation error 0.025 m s^{-1} was obtained as a time-averaged value on the S1S2 path (Taniguchi et al., 2021a), and this value was likely to underestimate the observation error for some specific periods. Thus, after some trial and error, we set the observation error to 0.05 m s^{-1} (0.05^2 m^2 s^{-2} in a variance), twice the value we obtained as the time-averaged path-averaged current precision. Additional consideration was given to the data error covariance. The present model aimed to reproduce relatively large tidal vortices that are found in the Seto Inland Sea (e.g., Arai, 2004), and it can simulate the generation of an island wake with a size of about 1 km (Taniguchi et al., 2023). By contrast, the observed path-averaged currents may contain contributions from smaller spatial scale features. Thus, there may be representation errors between the numerical model and the observed path-averaged currents. In the presence of representation errors, spurious variations will appear in the model prediction if the model is tightly fitted to the observed data. After some trial and error, we further reduced the impact of the observations by adding a diagonal matrix with a value of 0.08^2 m^2 s^{-2} to the observational error covariance matrix in the Kalman

gain. The added noise was somewhat large, but the EnKF updates keep the model state close to the observed path-averaged currents because of frequent updates of two minutes.

The EnKF updates were continually performed for 30 days. The data for 30 days were used in the harmonic analysis because the separation between M_2 and N_2 constituents needs nearly 30 days (Rayleigh criterion; e.g., Schureman, 1958). The updated tidal heights and normal velocities at the open boundary grids were averaged along the boundary and stacked for 30 days. Then, the harmonic analysis was applied to these time series at each open boundary to determine the amplitudes and phases of five tidal constituents (A_i, B_i, θ_i, ϕ_i); and these obtained amplitudes and phases of the tidal constituents provide the new open boundary conditions. Note that, in the narrow model domain of the present model, the tidal currents have a physical correlation throughout the model domain. Thus, the EnKF can reasonably update the U or V values, although they can easily recover to the values specified as the boundary forcing.

2.3.2 Linearization method

Menemenlis et al. (2005) proposed a method for adjusting control parameters used in a general circulation model by using the results of model parameter sensitivity experiments. The method linearizes an ocean model prediction about a particular model trajectory and was named model Green's function approach following a similar linearization method used in Menemenlis and Wunsch (1997), where model Green's function was defined as the response of a general circulation model to unit perturbations of their state vector. Implementing the method is easy because we only need to repeat a model simulation to determine the model sensitivity to the control parameters. With the results of the sensitivity experiments, the problem is linearized and eventually reduced to solving a linear inverse problem. Following their method, we calibrate the open boundary condition, i.e., tidal amplitudes and phases of the normal velocities at the open boundaries.

In order to apply the method to our present study, we consider the results of harmonic analysis applied to the time series of the observed five path-averaged currents for 30 days as the observation data. Thus, the observation vector $\boldsymbol{y}$ contains the amplitudes and phases (equivalently, the coefficients a and b in Equation 4) of the five tidal constituents for the five paths of the reciprocal transmission as follows:

$$\boldsymbol{y}^T = \left[\, a_{M_2}^{S1S2},\ b_{M_2}^{S1S2},\ a_{S_2}^{S1S2},\ b_{S_2}^{S1S2},\ \cdots,\ a_{O_1}^{S3S4},\ b_{O_1}^{S3S4} \,\right], \tag{6}$$

where we use boldface to represent vector variables as column vectors, and the superscript T indicates the transpose of a matrix (vector); the superscript characters on the right-hand side of Equation 6 indicate the station pairs for the reciprocal transmission experiment.

Observation equivalents must be predicted from the numerical ocean model. To do so, we express the time stepping of the numerical ocean model as

$$\boldsymbol{x}(t+\Delta t) = \boldsymbol{M}(\boldsymbol{x}(t), \boldsymbol{\eta}), \tag{7}$$

where $\boldsymbol{x}$ is the state vector and includes east-west and north-south components of the depth-averaged velocities at all model grids; Function $\boldsymbol{M}$ represents the one-step-ahead integration of our numerical ocean model; $\boldsymbol{\eta}$ is a model parameter vector. We consider the parameters in the boundary forcing as the model parameters; that is, $\boldsymbol{\eta}$ in Equation 7 contains a_i and b_i in Equation 4 for the five tidal constituents and at the four open boundaries:

$$\boldsymbol{\eta}^T = \left[\, a_{M_2}^{W},\ b_{M_2}^{W},\ a_{S_2}^{W},\ b_{S_2}^{W},\ \cdots,\ a_{O_1}^{SE},\ b_{O_1}^{SE} \,\right], \tag{8}$$

where the superscript characters on the right-hand side indicate the locations of the open boundaries; for example, W and SE for west and southeast boundaries, respectively. Note that although there are other model parameters, such as a bottom drag coefficient and kinematic eddy viscosity coefficient, those are fixed in the present study, and we omit them from the model parameters and equations. To relate the outputs from the numerical ocean model with the observation vector $\boldsymbol{y}$, we consider a vector accumulating the state vector at the transmission times (i.e., every two minutes) for 30 days as shown in Equation 9:

$$\boldsymbol{z}^T = \left[\, \boldsymbol{x}_1^T,\ \boldsymbol{x}_2^T,\ \cdots,\ \boldsymbol{x}_{21600}^T \,\right], \tag{9}$$

where the subscript indicates the index corresponding to the transmission time within 30 days. This accumulated state vector $\boldsymbol{z}$ and the above observation $\boldsymbol{y}$ are related by an observation equation:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{\epsilon}, \tag{10}$$

where function $\boldsymbol{f}$ is a composite function corresponding to the operation that converts the model velocity fields to path-averaged currents and then derives the harmonic constants a_i and b_i for the five transmission paths. The residual vector is introduced to represent the error term between $\boldsymbol{y}$ and $\boldsymbol{f}(\boldsymbol{z})$. Since vector $\boldsymbol{z}$ depends solely on $\boldsymbol{\eta}$ provided that the same model is used and other parameters are fixed, Equation 10 can also be written in terms of model parameter vector $\boldsymbol{\eta}$ as follows:

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{\eta}) + \boldsymbol{\epsilon}, \tag{11}$$

where $\boldsymbol{g}$ has a functionality similar to $\boldsymbol{f}$ but with a model run with parameters $\boldsymbol{\eta}$.

Solving Equation 11 for $\boldsymbol{\eta}$ (or, finding $\boldsymbol{\eta}$ that minimizes $\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}$) is a nonlinear inverse problem, but one can linearize $\boldsymbol{g}(\boldsymbol{\eta})$. To linearize $\boldsymbol{g}(\boldsymbol{\eta})$, we assume that optimal parameter values can be expressed as the sum of initial values ($\boldsymbol{\eta}_0$) and control value ($\boldsymbol{\eta}_c$). With the expression $\boldsymbol{\eta} = \boldsymbol{\eta}_0 + \boldsymbol{\eta}_c$, we linearize $\boldsymbol{g}(\boldsymbol{\eta})$ around $\boldsymbol{\eta}_0$, and Equation 11 becomes:

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{\eta}_0) + \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\eta}}\right|_{\eta=\eta_0} \boldsymbol{\eta}_c + \boldsymbol{\epsilon}', \tag{12}$$

or, equivalently,

$$\boldsymbol{y}_d = \boldsymbol{G}\boldsymbol{\eta}_c + \boldsymbol{\epsilon}', \tag{13}$$

where $\boldsymbol{y}_d = \boldsymbol{y} - \boldsymbol{g}(\boldsymbol{\eta}_0)$, and $\left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\eta}}\right|_{\eta=\eta_0}$ in Equation 12 is replaced with matrix $\boldsymbol{G}$. This matrix $\boldsymbol{G}$ is a sensitivity matrix formed by

$$\boldsymbol{G} = \left[\, \cdots,\ \frac{\boldsymbol{g}(\boldsymbol{\eta}_0 + k_j \boldsymbol{e}_j) - \boldsymbol{g}(\boldsymbol{\eta}_0)}{k_j},\ \cdots \,\right], \tag{14}$$

where $\boldsymbol{e}_j$ is a unit vector, and thus $k_j\boldsymbol{e}_j$ is a vector that has a value (k_j) only at j-th element and has 0 at other elements. Thus, constructing matrix $\boldsymbol{G}$ corresponds to evaluating the model sensitivity to each element of $\boldsymbol{\eta}$.

By solving Eqiation 13, we obtain an estimate of the optimal $\boldsymbol{\eta}$ by $\boldsymbol{\eta}_0 + \boldsymbol{\eta}_c$. Assuming that $\boldsymbol{\eta}_c$ and $\boldsymbol{\epsilon}'$ are modeled as Gaussian distributions with zero mean and covariance $\mathbf{S}$ and $\boldsymbol{R}$, we seek $\boldsymbol{\eta}_c$ that minimizes the following cost function:

$$J = (\boldsymbol{y}_d - \boldsymbol{G}\boldsymbol{\eta}_c)^T\boldsymbol{R}^{-1}(\boldsymbol{y}_d - \boldsymbol{G}\boldsymbol{\eta}_c) + \boldsymbol{\eta}_c^T\mathbf{S}^{-1}\boldsymbol{\eta}_c\,. \tag{15}$$

The expected solution that minimizes J of Equation 15 has the following analytical form:

$$\hat{\boldsymbol{\eta}}_c = \left(\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{G} + \boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{y}_d, \tag{16}$$

which corresponds to that called Gauss-Markov estimates (Munk et al., 1995). Optimal tidal amplitudes and phases are then obtained as $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}_0 + \hat{\boldsymbol{\eta}}_c$.

One needs to perform a series of sensitivity experiments to construct matrix $\boldsymbol{G}$ of Equation 14. Kobayashi et al. (2016) mentioned that the results of the sensitivity experiments were independent of the perturbation values in the sensitivity experiments. Yet, the perturbation in the sensitivity experiments should be large enough for the resulting perturbation to appear in the path-averaged currents. In the present study, the sensitivity experiments were performed with a constant value of k_j of +0.1 m s^{-1} for all j (i.e., for all a and b in Equation 8). We confirmed that perturbing the coefficients (a and b) for the normal velocities by +0.1 m s^{-1} caused the change in the harmonic analysis results of the resulting path-averaged currents in the sensitivity experiments.

When we compute the solution of Equation 16, the covariance $\boldsymbol{R}$ and $\boldsymbol{S}$ are required, but the values representing those in the true fields cannot be known precisely, and thus, we specified the values in $\boldsymbol{R}$ and $\boldsymbol{S}$ based on our estimate of the size of the elements in $\boldsymbol{\epsilon}'$ and $\boldsymbol{\eta}_c$. The number of data used in the harmonic analysis was 21,565, 21,576, 17,875, 21,138, and 21,159 for the S1S2, S1S3, S2S3, S2S4, and S3S4 pairs (paths), respectively. With such a sufficient number of data, the expected error variance of the least squares solution ($\hat{a}$ and $\hat{b}$ for all the paths) in the harmonic analysis was about 0.001^2 m^2 s^{-2} assuming measurement noise in the path-averaged currents of 0.1 m s^{-1}. Although the accurate values in $\boldsymbol{R}$ were unknown, they were expected to be somewhat larger than 0.001^2 m^2 s^{-2} because of other sources for $\boldsymbol{\epsilon}'$. In the present study, we then set $\boldsymbol{R}$ as a diagonal matrix with a value of 0.01^2 m^2 s^{-2} for all the elements. Here, each element in $\boldsymbol{\epsilon}'$ was expected to have no correlation with each other. Our expectation for $\boldsymbol{S}$ was that the coefficients of tidal constituents with large amplitude (i.e., M_2 and S_2) would be more uncertain and specified by larger variances than those of the other constituents (N_2, K_1, and O_1). In the present study, we set $\boldsymbol{S}$ as a diagonal matrix with values of 0.05^2, 0.04^2, and 0.03^2 m^2 s^{-2} for the elements associated with M_2, S_2, and other tidal constituents, respectively. The variances for the M_2 and S_2 constituents (0.05^2 and 0.04^2) were about three and two times the variance of the other constituents (3×0.03^2 and 2×0.03^2), respectively. After obtaining a solution, we confirmed that $(\boldsymbol{y}_d - \boldsymbol{G}\hat{\boldsymbol{\eta}})^T(\boldsymbol{y}_d - \boldsymbol{G}\hat{\boldsymbol{\eta}})$ was nearly equal to the trace of $\boldsymbol{R}$ (0.006 and 0.0055, respectively), and at the same time, the elements in $\hat{\boldsymbol{\eta}}$ were not unrealistic values and did not deviate extraordinarily from our expectations.

3 Results

Table 1 summarizes the initial and calibrated normal velocities as the open boundary condition: the amplitudes and phases of the five tidal constituents at the four open boundaries. The most apparent change is that the amplitudes after the calibration are smaller than those before the calibration for both calibration methods, particularly on the M_2 and S_2 constituents. For example, the M_2 amplitude at the eastern boundary had an initial value of 1.30 m s^{-1}, and when the boundary condition was calibrated by using the EnKF results and linearization method, the amplitudes decreased to 1.07 and 1.12 m s^{-1} (i.e., about 18% and 14% reduction), respectively. The amplitude decreased at other boundaries, too. The decrease in the amplitudes is reasonable because the present ocean model would overly predict the velocity of tidal currents compared with ADCP observation (Taniguchi et al., 2023). The amplitude of the S_2 constituent also decreased after the boundary calibration. The ratio of M_2 and S_2

TABLE 1 Summary of the initial and calibrated boundary conditions (amplitude and phase).

		Amplitude (m s^{-1})					Phase (deg.)				
		M_2	S_2	N_2	K_1	O_1	M_2	S_2	N_2	K_1	O_1
West	Initial values	0.58	0.23	0.12	0.10	0.04	71	253	146	197	236
	EnKF	0.49	0.20	0.11	0.09	0.04	79	260	155	201	241
	Linearization	0.42	0.19	0.11	0.10	0.03	73	249	149	195	240
Southwest	Initial values	0.65	0.26	0.13	0.11	0.05	71	253	146	197	236
	EnKF	0.56	0.22	0.11	0.10	0.05	80	261	156	206	245
	Linearization	0.50	0.23	0.12	0.11	0.04	69	251	143	203	236
East	Initial values	1.30	0.52	0.26	0.22	0.10	81	262	155	202	241
	EnKF	1.07	0.43	0.23	0.19	0.09	82	266	157	205	242
	Linearization	1.12	0.48	0.22	0.20	0.09	81	262	155	202	241
Southeast	Initial values	0.80	0.32	0.16	0.14	0.06	81	262	155	202	241
	EnKF	0.68	0.27	0.14	0.12	0.05	83	263	156	205	245
	Linearization	0.69	0.26	0.16	0.13	0.05	85	260	162	201	253

constituents (A_{S_2}/A_{M_2}), which related to the amplitude during the spring and neap tides, was 0.4 at the initial values and remained nearly the same after the calibration; but, slightly large variations are found in the results with the linearization method (0.38–0.46 for the four boundaries). The amplitudes of the N_2, K_1, and O_1 also decreased by the calibration in general. The phases of the tidal constituents are delayed by the calibration with the ensemble method, particularly at the west and southwest boundaries; for example, the phase of the M_2 constituent at the west boundary is 71 and 79 degrees before and after the calibration. As a result, the difference in the phases between the west (and southwest) and east (and southeast) becomes smaller in the calibration results with the ensemble method. On the other hand, there is no clear trend about the phase in the calibration results with the linearization method.

We ran the numerical models with the calibrated open boundary conditions and predicted path-averaged currents. Figure 2 shows the results of harmonic analysis (reconstructed time series) applied to the observed and predicted path-averaged currents. The difference between the predicted and the observed path-averaged current decreased by the calibration of the open boundary condition (Figure 2B). Although the amplitude of the tidal constituents decreased by the boundary condition calibration, the amplitude of the observed path-averaged current is further small. When we compare the two methods, the linearization method resulted in the time series slightly closer to the observed results about the magnitude, which can be seen as the difference from the observed results (Figure 2B). The improvements by the boundary condition calibration are summarized in Table 2, which shows root-mean-squared differences (RMSD) between the observed and model-predicted path-averaged currents and their percent error relative to the observations. The improvements are mainly attained in the paths diagonally crossing the main course of the current (i.e., the S1-S3, S2-S3, and S3-S4 pairs). On the other hand, the improvements are minor for the paths crossing the main course at the right angle (the S1-S2 and S3-S4 pairs). Since the RMS magnitudes of the path-averaged currents for these paths are relatively small, the relative error remains large. Besides, there remains a clear phase error in the S3S4 path (Figure 2A5), while there is less phase error in other paths.

Figure 3A shows the observed and predicted path-averaged currents time series. The figure also shows the ADCP results (black thick lines), where the ADCP velocities were calculated by projecting the ADCP velocities at each observation point and averaging them over each transect. The observed path-averaged currents (gray solid lines) show high-frequency variation superimposed on sinusoidal variation. The path-averaged currents during the ADCP observations are consistent with the corresponding ADCP results. Thus, it is argued that the acoustic differential travel time signal captured the tidal current as a path-averaged quantity, and such high-frequency variations in the path-averaged currents indicate complex spatio-temporal features of the tidal currents.

The amplitude of the modeled path-averaged currents decreased because of the reduced tidal forcing (the normal velocity) at the open boundaries (Figure 3A). However, the high-frequency variation with periods of about 1–2 hours is not well reproduced, as shown in the plots of the difference between the modeled and observed path-averaged currents (Figure 3B). In particular, the magnitude of the high-frequency variation on the S3S4 path (Figure 3A5) is comparable to the tidal sinusoidal variation (with a period of the M_2 constituent); this may be one reason why a phase error appears on this path in the harmonic analysis result (Figure 2A5). Because the periods of the tidal constituents (e.g., about 12.4 hour for the M_2 constituent) are much longer than the periods of fluctuations in the model error (Figure 3B) or the high-frequency (short-period) variations in the path-averaged currents, calibration of the tidal boundary condition is not effective in improving the reproducibility of these high-frequency variations.

Figures 4, 5 show the comparisons of velocity vectors between the model predictions and ADCP observations in the form of spatial velocity fields and scatter plots, respectively. Table 3 summarizes the performance metrics of the comparisons: correlation coefficients and RMSD for the eastward and northward currents (r_U and r_V; RMSD_U and RMSD_V). We also computed fractional error variance (FEV) for the results with the initial and calibrated boundary conditions. The FEV is defined as

$$\mathrm{FEV} = \frac{\langle\, |U_i + jV_i - (U_i^{\mathrm{ADCP}} + jV_i^{\mathrm{ADCP}})|^2 \,\rangle}{\langle\, |U_i^{\mathrm{ADCP}} + jV_i^{\mathrm{ADCP}}|^2 \,\rangle}, \tag{17}$$

where $j \times j = -1$, and $|\cdot|$ and $\langle\cdot\rangle$ indicate computing the absolute value and the averaging over the all ADCP data, respectively. Note that, for the comparisons with the ADCP results with Equation 17, the model results were spatiotemporally interpolated to obtain the velocity at the same locations and times as each ADCP velocity. In general, the model without the boundary condition calibration predicts the flow speed in the east-west component greater than that of the ADCP observations. This is evident in both the spatial velocity fields (Figure 4) and the scatter plots (the black triangles and corresponding regression lines in Figures 5A, C). The slope of the regression between the ADCP results and the model predictions without boundary calibration is 1.11 for the east-west component. The east-west components in the model prediction with the calibrated normal velocity are more compatible with the ADCP observations due to the reduced magnitude of the normal velocity at the open boundaries. With the calibrated boundary condition, the slopes became 0.99 and 0.90, for the ensemble and linearization methods, respectively. The linearization method decreased the slope a little too much. The RMSD_U decreased from 0.29 m s^{-1} to 0.26 and 0.27 m s^{-1} in the results of the boundary-calibrated model predictions (Table 3). In the spatial velocity map, the improvements are evident during the westward flow (transects starting at 13:00 and 14:00 on Oct. 30 and transects at 13:00, 14:00, and 15:00 on Oct. 31; Figure 4). Other examples of improvements in the spatial velocity map are found, for example, along transects starting at 8:00 and 9:00 on Oct. 30, although some degradation is also found (e.g., transects starting at 10:00 on Oct. 30 and 31).

Improvements in the north-south velocity component is not clear. The regression slopes became smaller in the boundary-calibrated model predictions than in the results without the calibrated boundary condition (Figures 5B, D). Also, the

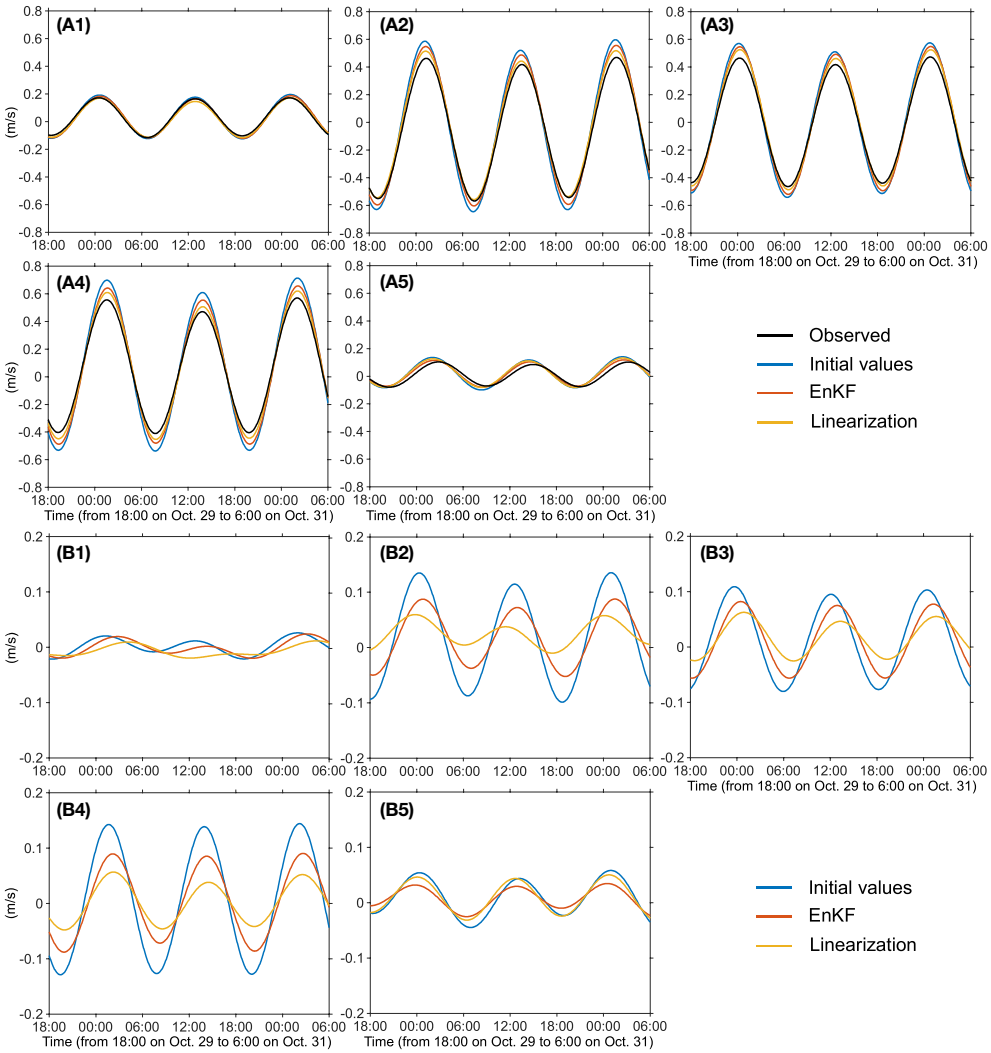


FIGURE 2
(A1–A5) Reconstructed path-averaged currents by using five taidal constituents (M_2, S_2, N_2, K_1, and O_1). The black line is from the observation and colored lines are model-predicted results: the blue, red, and yellow lines are the results with open boundary conditions of initial values, derived from EnKF results, and from linearization method, respectively. The five panels with label 1–5 correspond to the results of the S1-S2, S1-S3, S2-S3, S2-S4, and S3-S4 station pairs. **(B1–B5)** The difference of the model-predicted path-averaged currents (reconstructed by using five tidal constituents) with respect to the observed results. The line colors and panel's label number are the same as those in **(A1–A5)**.

correlation coefficient decreased and RMSD increased (Table 3). Controlling the boundary condition to fit path-averaged currents does not minimize errors in east-west and north-south components independently. When the path-averaged currents are used as the data, the error in the east-west component mainly decreased because the contribution to the path-averaged currents from the east-west component is larger than those from the north-south component. Thus, the calibration does not always improve the north-south component of the velocity. However, it is true that the FEV (i.e., total error variance) is still decreased by the boundary condition calibration.

We evaluated if the boundary condition calibration affects sequential updates with EnKF by re-running EnKF updates with the calibrated boundary conditions. Although there were no distinguishable improvements in vector maps (like Figure 7 in Taniguchi et al., 2023), there were still some changes in EnKF results. Figure 6 exhibits the comparisons of velocity components between EnKF results and ADCP observations. For the eastward component of velocity, the EnKF results without the boundary calibration are still larger than the ADCP observations when the current magnitude is relatively large. This trend is found as that the EnKF results without the boundary condition calibration (black triangles) are higher than the identity line (1:1 line; black dashed line) at the first quadrant and lower than the 1:1 line at the third quadrant. The EnKF results with the new boundary condition (blue dots and red crosses in panels A and C) are more condensed around the 1:1 line than those without the boundary condition. The velocity magnitude of the northward component is small. Thus the difference between the EnKF results with and without the boundary calibration is not clear in the scatter plot. However,

TABLE 2 Root-mean-squared differences (RMSD) between the model-predicted and the observed path-averaged currents for 30 days (m s^{-1}) and its relative error (%).

	Boundary condition	S1-S2	S1-S3	S2-S3	S2-S4	S3-S4
RMSD	initial values	0.103	0.135	0.118	0.151	0.105
	EnKF	0.100	0.117	0.104	0.135	0.098
	Linearization	0.095	0.104	0.898	0.125	0.094
Relative error	initial values	80	39	37	46	106
	EnKF	77	33	33	41	100
	Linearization	74	20	28	38	96

some outliers in the EnKF results without boundary calibration are well suppressed. Table 3 summarizes the comparisons of the EnKF results with ADCP observation. The EnKF results with the calibrated boundary condition improved its agreement with the ADCP observations. The RMSD for both eastward and northward components and for both the ensemble and linearization methods decreased. The FEV is 7.0% and 7.5% for the ensemble and linearization methods. These FEV values are smaller than the

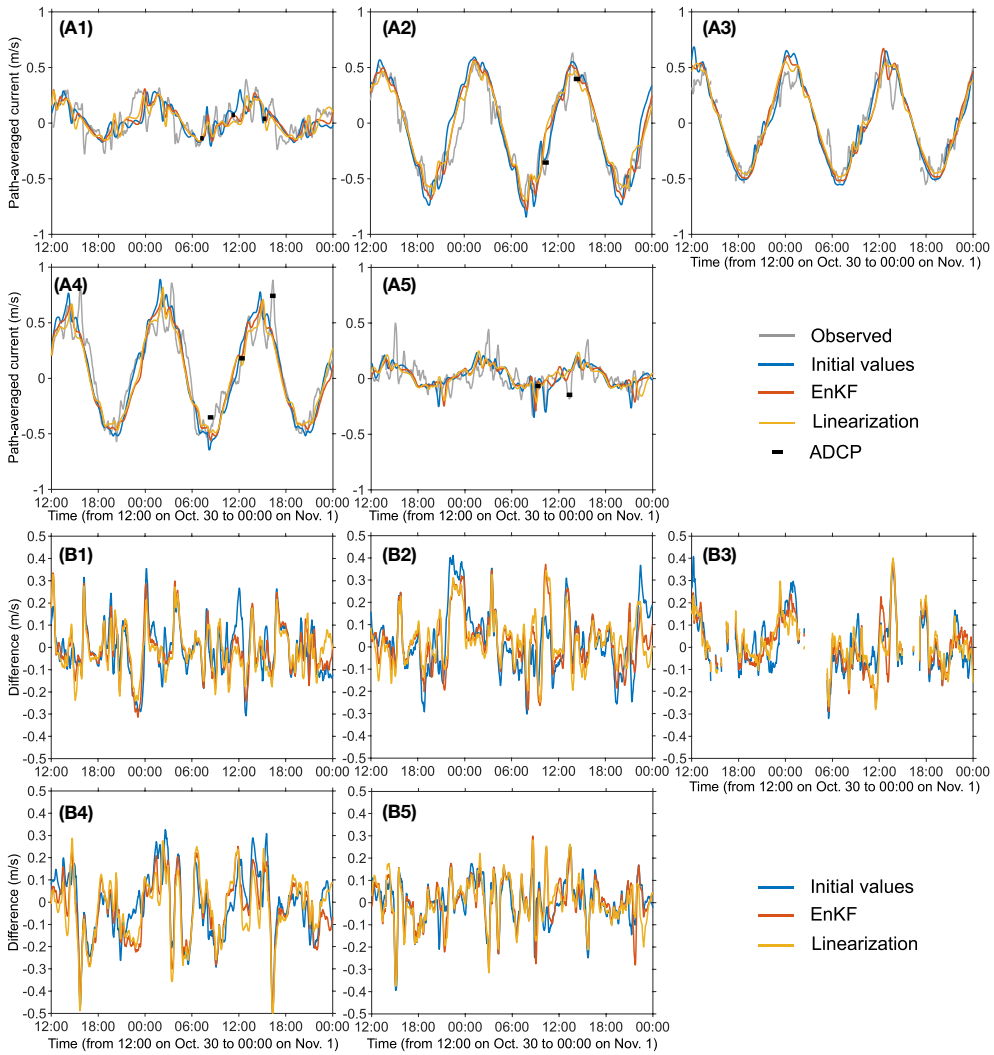


FIGURE 3
(A1–A5) Comparisons of the observed (gray) and model-predicted (colored) path-averaged currents. The blue, red, and yellow lines are the results with the open boundary conditions of initial values, derived from EnKF results, and from linearization method, respectively. The thick black bars are the ADCP results averaged over each transmission path. The five panels with label 1–5 correspond to the results of the S1-S2, S1-S3, S2-S3, S2-S4, and S3-S4 station pairs. **(B1–B5)** The difference of the model-predicted path-averaged currents with respect to the observed results. The line colors and panel's label number are the same as those in **(A1–A5)**.

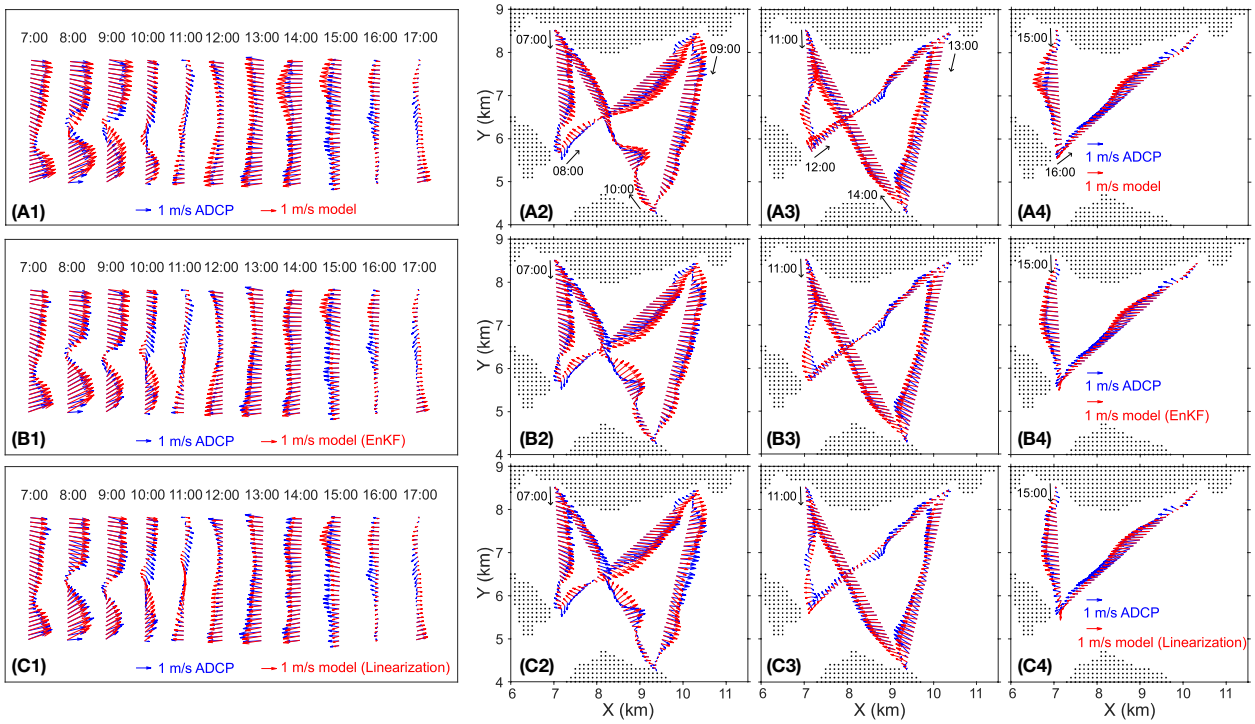


FIGURE 4
Comparisons of velocity vectors between the ADCP results (blue) and model-prediction (red). The model prediction results are obtained from the original boundary condition, the normal velocity calibrated by using the ensemble method, and those by using the linearization method from the top to the bottom, respectively. Panels **(A1, B1, C1)** are the comparisons with the ADCP results on Oct. 30, and panels **(A2–A4, B2–B4, C2–C4)** are the comparisons with the ADCP results on Oct. 31.

result obtained from the EnKF with 980 ensemble members in the previous report (FEV of 8.2%; Taniguchi et al., 2023).

Menemenlis et al. (2005) suggested that the parameter calibrations would be repeated to obtain better parameter values in strongly nonlinear conditions. We iterated the parameter calibration five times by re-computing the sensitivity matrix $\boldsymbol{G}$ in each iteration. Figure 7 shows the calibrated amplitudes of the M_2 constituent at the four boundaries over the iterations. Although the M_2 magnitude gradually changes and converges to some values, the largest change was at the first iteration. A similar trend was found in the magnitudes of the other constituents. In the EnKF for CAT, the boundary condition is perturbed to generate an ensemble, so relatively minor boundary modification after the first iteration might not affect the EnKF results. The FEV for the EnKF result with the boundary condition after the fifth iteration was 7.4, and there is not much additional improvement.

4 Summary and discussion

In this paper, we demonstrated that the CAT path-averaged currents can also be used in a boundary control problem, i.e., to calibrate or tune the open boundary condition, which was the normal velocities at the open boundaries in this study. The two methods were able to derive nearly the same calibration results, particularly for the amplitude; both methods decreased the amplitudes of tidal constituents from the initial values. With the calibrated normal velocities, the agreement of the model-predicted velocity with the ADCP observations improved, although the improvements in the model prediction by the calibration were minor. A reason would be that the initial values of the normal velocities, which were *ad-hoc*ly determined using the observed path-averaged currents, were already reasonable to represent the major patterns of the tidal currents at the observation sites. Also, although we only used the velocity information and calibrated the normal velocities, the tidal height would have to be measured and calibrated as well as the normal velocities because the phase difference between the tidal current and height is one of the key factors controlling the velocity fields at the site. The EnKF data assimilation with the new boundary condition also improved the results in terms of the agreement with ADCP observations; in the EnFK results with the new boundary condition, the overestimated velocity magnitude, which was found in the EnKF results with the original boundary condition (Taniguchi et al., 2023) and due likely to the overly specified amplitude of the initial open boundary condition, decreased so that the agreement with ADCP results become better. Thus, the calibration of the boundary condition would be desired work, particularly when the boundary condition is uncertain. Related to the present study with the CAT, we are purposing developments of a real-time monitoring system of velocity fields of tidal currents. The practical implementation of the present method will be to perform the reciprocal acoustic transmission experiment first to calibrate the boundary

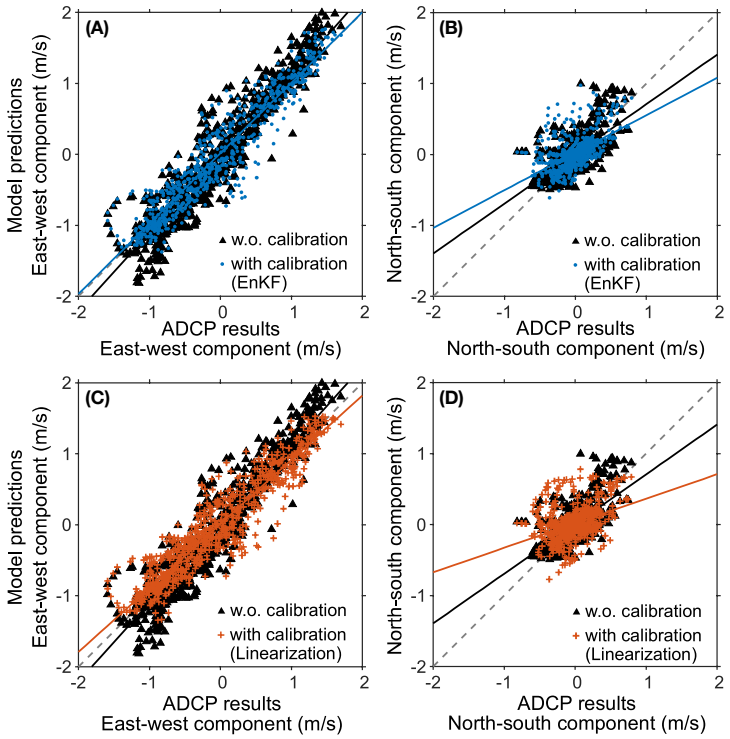


FIGURE 5
Scatter plots of ADCP observation results and model predictions without (black triangles) and with (colored) the calibration of open boundary condition **(A, B)** by applying a harmonic analysis to the EnKF results, and **(C, D)** by the linearization approach. In each panel, gray dashed line is the 1:1 line, and the solid lines indicate the linea regression lines.

condition, followed by the real-time monitoring with sequential assimilation with the calibrated model boundary condition.

Both two methods can control the tidal boundary condition. We do not provide the decision about which is better. The EnKF method is easier in terms of implementation because it performs sequential assimilation twice with the same computation code but with different boundary conditions. However, this method may limit the size of the model domain or target. The EnKF scheme updates the model states using the information on state covariance. Thus, if there is no correlation between the velocity fields at observation locations and at open boundaries, then the EnKF scheme cannot reasonably update the model states at open boundaries. Also, one may need to use relatively large ensemble members to suppress spurious correlation because the covariance localization, which reduces EnKF update impacts at grids far from the observation site, would be prohibited. The linearization method is not limited by the model size and can be used to adjust control parameters in a general circulation model (Menemenlis et al., 2005). However, if we include more constituents (over-tides and compound tides) as the target for the calibration, the necessary computation increases as the number of model parameters increases.

Areas of improvement remain. First, tidal height was not observed in the experiment; thus, the tidal height information was not used to calibrate boundary conditions and in the EnKF updates. The performance of the boundary condition calibration is expected to be improved by adding the tidal height observations. Or, one may observe the tidal height variations at open boundaries independent of reciprocal transmission experiment in order to determine the reliable amplitudes and phases of the tidal constituents at the open boundaries; in this case, one will focus only on updating velocity information by using the reciprocal transmission experiment data like that performed in this study. In the present model, the tidal elevations and normal velocities were constant along each open boundary. The model prediction may improve by specifying variable condition along the boundary and/or adding more

TABLE 3 Summary of the comparison with ADCP observed currents.

	Boundary condition	r_U	r_V	RMSD$_U$ (m/s)	RMSD$_V$ (m/s)	FEV (%)
model	initial values	0.95	0.68	0.29	0.20	19.6
	EnKF	0.94	0.56	0.26	0.23	19.3
	Linearization	0.94	0.53	0.27	0.21	18.5
EnKF	Initial values	0.97	0.88	0.20	0.12	8.5
	EnKF	0.97	0.89	0.17	0.11	7.0
	Linearization	0.97	0.88	0.18	0.12	7.5

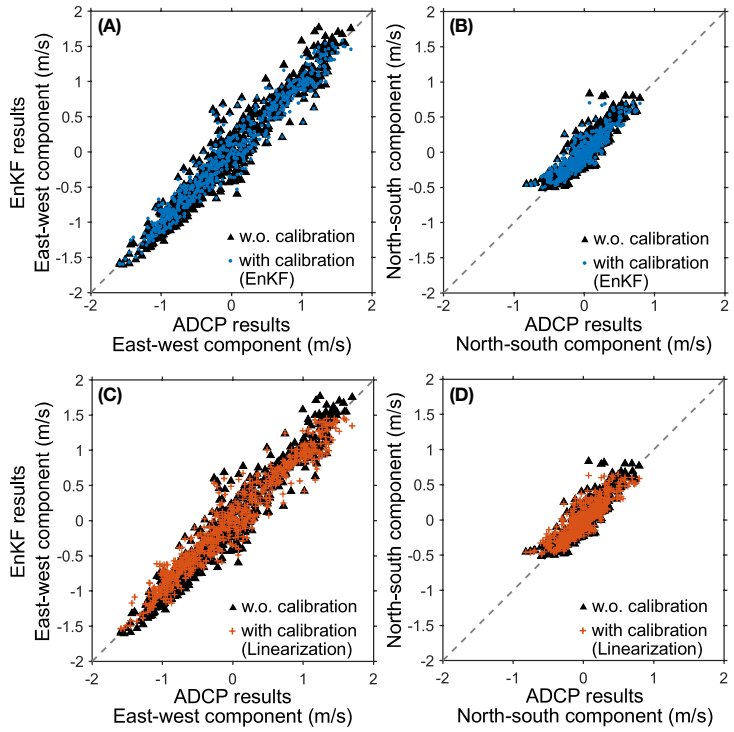


FIGURE 6
Scatter plots of ADCP observation results and EnKF results without (black triangles) and with (colored) the calibration of open boundary condition **(A, B)** by applying a harmonic analysis to EnKF results, and **(C, D)** by linearization approach.

constituents of overtides and compound tides for high frequency variation; however, controlling those additional parameters increases the computation when the linearlization method is employed. Also, controlling other model parameters, such as the coefficients for bottom drag and eddy viscosity, may contribute to further improvements.

The improvement of the EnKF results by the boundary condition calibration is slight (reduction of about 1% in FEV and 0.02–0.03 m s^{-1} in $RMSD_U$ and 0.01 m s^{-1} in $RMSD_V$; Table 3). Figure 8 shows the ensemble spread σ for the east-west and north-south velocity components, computed immediately after the EnKF updates and averaged over two days of Oct. 30 and 31 (the days the ADCP observation was conducted). The ensemble spread σ is about 0.15 m s^{-1} in the east-west component and about 0.1 m s^{-1} in the north-south component in the tomography domain. The standard error of the ensemble mean is estimated as $\sigma/\sqrt{n}$, where n is the number of ensemble members, and the values are $0.15/\sqrt{98} \approx 0.015$ for the east-west component and $0.1/\sqrt{98} \approx 0.01$ for the north-south component. By comparing the RMSD improvements with these estimated standard errors of the mean, the improvement in the east-west component is expected to be statistically meaningful. By repeating the EnKF assimilation with the calibrated boundary conditions multiple times (with randomly generated ensemble members), we also confirmed that the resulting FEV and RMSDs changed slightly in each EnKF run; the EnKF results with the boundary calibration were consistently better (smaller FEV and $RMSD_U$). Although minor improvements do not contribute to the refinement of the instantaneous velocity fields of the tidal currents in a velocity map, it will be important when we focus on residual currents (or time-averaged flow), which are much smaller in magnitude compared to tidal currents and will be the subject when we start monitoring tidal currents operationally with EnKF-CAT.

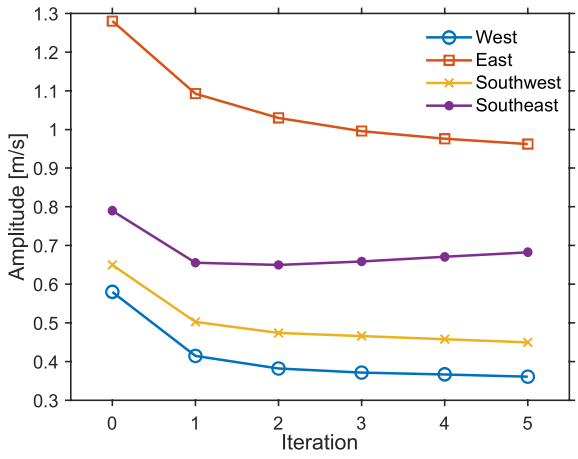


FIGURE 7
The variation of the M_2 amplitude for the normal velocity of the open boundary conditions on each iteration obtained by the linearization method. The values at the 0-th and first iterations correspond to those listed with labels *Initial values* and *Linearization* in Table 1.

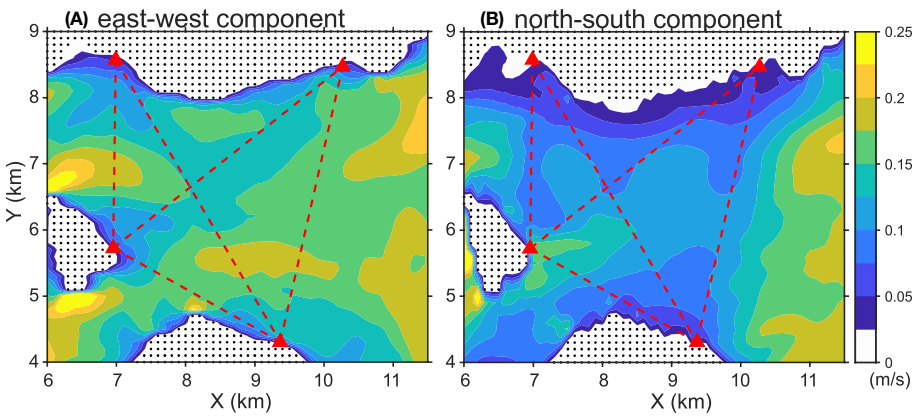


FIGURE 8
Posterior ensemble spread for **(A)** east-west and **(B)** north-south components. The results are obtained by executing the EnKF updates with the calibrated boundary condition (with the ensemble method). The results are averaged over two days on Oct. 30 and 31, the days when the ADCP observations were performed. The red triangles indicate the station locations and the red dashed lines are the acoustic transmission paths.

There is also room for improvement in CAT data assimilation. As shown by Cornuelle and Worcester (1996), path-averaged measurements (path-averaged currents in this study) determine the lower wavenumber information better than the higher wavenumbers, and determining the high wavenumber information requires a sufficiently small measurement error. The present EnKF implementation uses a relatively large data error covariance to suppress implausible EnKF updates. Thus, while the present EnKF reasonably determines the low wavenumber structure, it may lack the ability to determine the high wavenumber variation found in the ADCP observations. The need for a relatively large data error covariance may be due in part to representation errors. If the representation errors in a model are sufficiently small, then the data error covariance can only be given by the measurement noise in the path-averaged currents. Thus, more sophisticated numerical models are suitable for CAT data assimilation to fully exploit the information in the path-averaged currents. It will also be necessary to investigate how precise models are needed to effectively use the information in the path-averaged currents will also be required.

CAT with data assimilation is a promising tool for reconstructing the tidal velocity fields in coastal shallow water. However, there are still space for improving the CAT for the reconstructions of tidal currents. The present study is one such example. Since the velocity fields (and thus tidal currents) in coastal shallow water are essential information for various studies related to the ocean, improving the CAT performance to represent the ocean currents contributes to the progress of those fields.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Data sets used in this study are the same as those used in the previous study (Taniguchi et al., 2023). Requests to access these datasets should be directed to ntaniguchi@hiroshima-u.ac.jp.

Author contributions

NT: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft. HM: Conceptualization, Software, Writing – review & editing. MA: Methodology, Software, Writing – review & editing. YS: Funding acquisition, Resources, Writing – review & editing. KH: Methodology, Project administration, Resources, Writing – review & editing. C-FH: Conceptualization, Methodology, Writing – review & editing. JG: Conceptualization, Methodology, Writing – review & editing. TT: Data curation, Investigation, Resources, Writing – review & editing. KY: Data curation, Investigation, Resources, Writing – review & editing. HY: Data curation, Investigation, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by JSPS KAKENHI Grant Numbers 19H04292, 20H02369, 20K14964, 22K04563.

Acknowledgments

The authors are grateful to Geinan Fisheries Cooperative Association, Mihara-Shi Fisheries Cooperative Association, Ohmishima (Ehime Prefecture) Fisheries Cooperative Association, and TSUNEISHI SHIPBUILDING Co., Ltd. for their support in the reciprocal acoustic transmission experiment.

Conflict of interest

Author TT, KY, and HY are employed by Fukken Co., LTD.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arai, M. (2004). Development of a semi-spectral coastal ocean model and its application to the neko seto sea in the seto inland sea. *J. Oceanography* 60, 597–611. doi: 10.1023/B:JOCE.0000038352.17595.4e

Blumberg, A. F., and Mellor, G. L. (1987). *A Description of a Three-Dimensional Coastal Ocean Circulation Model* (Washington, D.C.: American Geophysical Union (AGU), 1–16. doi: 10.1029/CO004p0001

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly Weather Rev.* 126, 1719–1724. doi: 10.1175/1520-0493(1998)126h1719:ASITEKi2.0.CO;2

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change* 9, e535. doi: 10.1002/wcc.535

Carter, G. S., and Merrifield, M. A. (2007). Open boundary conditions for regional tidal simulations. *Ocean Model.* 18, 194–209. doi: 10.1016/j.ocemod.2007.04.003

Cornuelle, B. D., and Worcester, P. F. (1996). "Ocean acoustic tomography: Integral data and ocean models," in *Modern Approaches to Data Assimilation in Ocean Modeling*, vol. 61 . Ed. P. Malanotte-Rizzoli (Amsterdam: Elsevier Oceanography Series), 97–115. doi: 10.1016/S0422-9894(96)80007-9

Dushaw, B. D., Au, W., Beszczynska-Moller, A., Brainard, R., Cornuelle, B. D., Duda, T., et al. (2010). "A global ocean acoustic observing network," in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, vol. 2. (ESA Publication WPP-306). doi: 10.5270/OceanObs09.cwp.25

Dushaw, B. D., Bold, G., Chiu, C.-S., Colosi, J. A., Cornuelle, B. D., Desaubies, Y., et al. (2001). "Observing the ocean in the 2000's: A strategy for the role of acoustic tomography in ocean climate observation," in *Observing the Oceans in the 21st Century*, Melbourne 391–418.

Dushaw, B. D., Worcester, P. F., Munk, W. H., Spindel, R. C., Mercer, J. A., Howe, B. M., et al. (2009). A decade of acoustic thermometry in the North Pacific Ocean. *J. Geophysical Research: Oceans* 114. doi: 10.1029/2008JC005124

Elisseeff, P., Schmidt, H., Johnson, M., Herold, D., Chapman, N. R., and McDonald, M. M. (1999). Acoustic tomography of a coastal front in Haro Strait, British Columbia. *J. Acoustical Soc. America* 106, 169–184. doi: 10.1121/1.427046

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophysical Research: Oceans* 99, 10143–10162. doi: 10.1029/94JC00572

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* 53, 343–367. doi: 10.1007/s10236-003-0036-9

Howe, B. M., Miksis-Olds, J., Rehm, E., Sagen, H., Worcester, P. F., and Haralabus, G. (2019). Observing the oceans acoustically. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00426

Howe, B. M., Worcester, P. F., and Spindel, R. C. (1987). Ocean acoustic tomography: Mesoscale velocity. *J. Geophysical Research: Oceans* 92, 3785–3805. doi: 10.1029/JC092iC04p03785

Huang, C.-F., Li, Y.-W., and Taniguchi, N. (2019). Mapping of ocean currents in shallow water using moving ship acoustic tomography. *J. Acoustical Soc. America* 145, 858–868. doi: 10.1121/1.5090496

Huang, C.-F., Yang, T. C., Liu, J.-Y., and Schindall, J. (2013). Acoustic mapping of ocean currents using networked distributed sensors. *J. Acoustical Soc. America* 134, 2090–2105. doi: 10.1121/1.4817835

Kaneko, A., Zhu, X.-H., and Lin, J. (2020). *Coastal Acoustic Tomography* (Amsterdam: Elsevier). doi: 10.1016/C2018-0-04180-8

Kobayashi, S., Nakada, S., Takagi, S., and Hirose, N. (2016). Parameter optimization of a 3D coastal model using Green's functions for modelling river plume dynamics. *J. Advanced Simulation Sci. Eng.* 3, 153–164. doi: 10.15748/jasse.3.153

Lin, J., Kaneko, A., Gohda, N., and Yamaguchi, K. (2005). Accurate imaging and prediction of Kanmon Strait tidal current structures by the coastal acoustic tomography data. *Geophysical Res. Lett.* 32. doi: 10.1029/2005GL022914

Menemenlis, D., Fukumori, I., and Lee, T. (2005). Using green's functions to calibrate an ocean general circulation model. *Monthly Weather Rev.* 133, 1224–1240. doi: 10.1175/MWR2912.1

Menemenlis, D., and Wunsch, C. (1997). Linearization of an oceanic general circulation model for data assimilation and climate studies. *J. Atmospheric Oceanic Technol.* 14, 1420–1443. doi: 10.1175/1520-0426(1997)014h1420:LOAOGCi2.0.CO;2

Moon, J.-H., Hirose, N., and Morimoto, A. (2012). Green's function approach for calibrating tides in a circulation model for the east asian marginal seas. *J. Oceanography* 68, 345–354. doi: 10.1007/s10872-011-0097-1

Munk, W., Worcester, P., and Wunsch, C. (1995). Ocean Acoustic Tomography. *Cambridge Monographs on Mechanics* (New York: Cambridge University Press). doi: 10.1017/CBO9780511666926

Munk, W., and Wunsch, C. (1979). Ocean acoustic tomography: a scheme for large scale monitoring. *Deep Sea Res. Part A. Oceanographic Res. Papers* 26, 123–161. doi: 10.1016/0198-0149(79)90073-6

Park, J.-H., and Kaneko, A. (2000). Assimilation of coastal acoustic tomography data into a barotropic ocean model. *Geophysical Res. Lett.* 27, 3373–3376. doi: 10.1029/2000GL011600

Schureman, P. (1958). *Manual of Harmonic Analysis and Prediction of Tides (Revised 1940 Edition Reprinted 1958 with correction). Special Publication No.98* (Washington D.C: Coast and Geodetic Survey, U.S. Department of Commerce).

Taniguchi, N., Mutsuda, H., Arai, M., Sakuno, Y., Hamada, K., Takahashi, T., et al. (2023). Reconstruction of horizontal tidal current fields in a shallow water with model-oriented coastal acoustic tomography. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1112592

Taniguchi, N., Takahashi, T., Yoshiki, K., Yamamoto, H., Hanifa, A. D., Sakuno, Y., et al. (2021a). A reciprocal acoustic transmission experiment for precise observations of tidal currents in a shallow sea. *Ocean Eng.* 219, 108292. doi: 10.1016/j.oceaneng.2020.108292

Taniguchi, N., Takahashi, T., Yoshiki, K., Yamamoto, H., Sugano, T., Mutsuda, H., et al. (2021b). Reciprocal acoustic transmission experiment at Mihara-Seto in the Seto Inland Sea, Japan. *Acoustical Sci. Technol.* 42, 290–293. doi: 10.1250/ast.42.290

Williams, P. D. (2009). A proposed modification to the robert–asselin time filter. *Monthly Weather Rev.* 137, 2538–2546. doi: 10.1175/2009MWR2724.1

Worcester, P. F. (1977). Reciprocal acoustic transmission in a midocean environment. *J. Acoust. Soc Am.* 62, 895–905. doi: 10.1121/1.381619

Yamaguchi, K., Lin, J., Kaneko, A., Yayamoto, T., Gohda, N., Nguyen, H.-Q., et al. (2005). A continuous mapping of tidal current structures in the kanmon strait. *J. Oceanogr.* 61, 283–294. doi: 10.1007/s10872-005-0038-y

Yamaoka, H., Kaneko, A., Park, J.-H., Zheng, H., Gohda, N., Takano, T., et al. (2002). Coastal acoustic tomography system and its field application. *IEEE J. Ocean. Eng.* 27, 283–295. doi: 10.1109/JOE.2002.1002483

Zhang, C., Zhu, X.-H., Zhu, Z.-N., Liu, W., Zhang, Z., Fan, X., et al. (2017a). High-precision measurement of tidal current structures using coastal acoustic tomography. *Estuarine Coast. Shelf Sci.* 193, 12–24. doi: 10.1016/j.ecss.2017.05.014

Zhang, Y., Chen, H., Xu, W., Yang, T. C., and Huang, J. (2017b). Spatiotemporal tracking of ocean current field with distributed acoustic sensor network. *IEEE J. Oceanic Eng.* 42, 681–696. doi: 10.1109/JOE.2016.2582018

Zheng, H., Gohda, N., Noguchi, H., Ito, T., Yamaoka, H., Tamura, T., et al. (1997). Reciprocal sound transmission experiment for current measurement in the seto inland sea, Japan. *J. Oceanogr* 53, 117–127.

Zhu, Z.-N., Zhu, X.-H., Guo, X., Fan, X., and Zhang, C. (2017). Assimilation of coastal acoustic tomography data using an unstructured triangular grid ocean model for water with complex coastlines and islands. *J. Geophysical Research: Oceans* 122, 7013–7030. doi: 10.1002/2017JC012715

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 22 March 2024
DOI 10.3389/fmars.2024.1368317

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Hamada Esmaiel,
Aswan University, Egypt
Dan Song,
Jimei University, China

*CORRESPONDENCE
Zhicheng Tan
tanzc@guat.edu.cn

RECEIVED 10 January 2024
ACCEPTED 15 February 2024
PUBLISHED 22 March 2024

CITATION
Ye K, Tan Z, Wang W, Tian T, Zhou L and Wang Y (2024) A mobile prototype-based localization approach using inertial navigation and acoustic tracking for underwater. *Front. Mar. Sci.* 11:1368317. doi: 10.3389/fmars.2024.1368317

A mobile prototype-based localization approach using inertial navigation and acoustic tracking for underwater

Kun Ye[1], Zhicheng Tan[1,2]*, Wei Wang[1], Tian Tian[1], Lang Zhou[3] and Yongjun Wang[2]

[1]School of Informatics, Xiamen University, Xiamen, China, [2]School of Aeronautics and Astronautics, Guilin University of Aerospace Technology, Guilin, China, [3]School of Electronic Science and Engineering (National Model Microelectronics College), Xiamen University, Xiamen, China

During underwater operations, divers must determine their own trajectories using the Inertial Navigation System (INS) they carry to improve operational efficiency. However, the INS contains a sensor bias that is also incorporated into the quadratic integration process to obtain the displacement, resulting in trajectory drift of the divers during prolonged self-guidance. To overcome the above problem, other aids are needed to correct the accumulated error of the INS. The single-beacon Assisted Inertial Navigation (AIN) method can improve the flexibility of inertial error correction while simplifying the localization equipment, which is suitable for the INS cumulative error correction scenario of divers. However, most of the traditional single-beacon assisted correction methods do not consider the effect of acoustic line bending on hydroacoustic ranging, and at the same time, they do not consider the problem of singular or pathological coefficient matrices introduced by inertial navigation neighbor localization deviations. Based on the above two shortcomings, this paper uses the acoustic velocity profile for acoustic line tracking, combines the localization idea of Mobile Primitives (MP), and proposes an MP-based acoustic line tracking-Assisted Inertial Navigation Localization (AINL) method, which constructs a sliding time window (STW) by taking the historical positioning of divers as a virtual primitive, and combines the nonlinear optimization method for iterative optimization search as a means to improve the accuracy and stability of self-navigation of the divers.

KEYWORDS

underwater active localization, assisted inertial navigation localization, sound source localization, sound velocity profiling, mobile primitives

1 Introduction

Marine localization and navigation technology is widely used in many fields such as diving and salvage, resource exploration, and environmental monitoring [Luo et al. (2021); Zhang et al. (2021); Su et al. (2023); Zhang et al. (2024); Ye et al. (2023)]. Diving is the primary means by which humans explore, exploit and strategize the oceans [Brown and Wang (2013); Kaneko and Kubota (2021)]. The localization and navigation methods for divers mainly include the hydroacoustic localization method, which uses sound waves as information carriers, and the combined navigation method based on INS. Among them, INS is very suitable for diver navigation scenarios due to its good autonomy, continuity, stealth and real-time advantages in the navigation process (Lyu et al. (2022)). However, the drift deviation of the inertial sensor in INS causes the localization error of divers to accumulate over time, and it is difficult for a single INS to meet the application requirements of long-term underwater navigation Liu et al. (2021). Therefore, it is necessary to use other auxiliary methods to correct the accumulated error of the INS in time. Acoustic wave, as the only information carrier that can be transmitted over long distances underwater, has become an effective means to AIN for error correction [Cheng et al. (2022)].

Literature [Zhang et al. (2023)] designed two combined navigation schemes for Autonomous Underwater Vehicles (AUV) for different surface and underwater navigation requirements, respectively. Literature [Xu et al. (2022)] proposed a decentralized co-location method based on Adaptive Cubature Kalman Filter (ACKF) to improve the navigation accuracy of two lead AUVs. Literature [Zhang et al. (2022)] designed a Kalman Filter (KF) method based on a hybrid distribution model derived from acoustic signal round trip delay and pitch angle measurement models to reduce the effect of underwater carrier motion on navigation accuracy, but did not fundamentally eliminate the INS cumulative error. Literature [Ju-Cheng et al. (2017)] proposed an AUV navigation augmentation method based on single beacon ranging, which effectively suppresses the INS cumulative error by introducing distance, speed, and azimuth as measurement information. The single-beacon localization technique, which uses a single acoustic beacon to assist INS in self-navigation, not only simplifies the transponder deployment process of traditional hydroacoustic positioning systems, but also provides a high degree of localization flexibility with navigation accuracy similar to that of INS combined with traditional hydroacoustic localization systems [Hegrens et al. (2009)]. Literature [Jin et al. (2019)] combined acoustic localization and inertial navigation to propose a low-cost single beacon inertial navigation augmentation localization method, but it assumes a constant underwater acoustic velocity and does not consider the problem of acoustic line bending. Literature [Zhang et al. (2019)] proposed an AUV self-localization method based on Virtual Long Baseline (VLBL) and STW, but the method sets the ocean acoustic velocity to a fixed value without sound line correction, and the corresponding coefficient matrix is prone to singular or pathological situations, at which time it is difficult to directly solve the spatial localization of the AUV.

Therefore, this paper focuses on the basic array deployment problem of diver localization scenarios, adopts a single beacon on the water surface for Acoustic-Assisted Inertial Navigation Localization (AAINL), and combines the localization ideas of acoustic velocity field inversion and MP to reduce the deployment cost of the localization system while meeting the scenario requirements of high accuracy, near real-time and high flexibility of diver self-navigation. The localization method in this paper can effectively improve the accuracy, robustness and flexibility of localization estimation of the divers, which is of great practical value and significance for protecting the personal safety for divers and improving the efficiency of underwater operations.

2 Problem and methodologies

2.1 Inertial navigation and localization scenarios

Aiming at the problem of cumulative localization errors generated by divers in long-term inertial navigation scenarios, this paper proposes an AINL method based on acoustic line tracking of MP, and the localization scenario is shown in Figure 1A. Three adjacent locations P_0, P_1, P_2, where the diver receives the acoustic signal from the surface beacon (SB) during the moving process, are selected as virtual primitives to construct the first STW, in which the adjacent displacements of the INS output and the slant distances obtained from the acoustic tracking are combined to solve the spatial localization of the diver at the nearest instant by the system of joint localization equations. The STW is updated and iterated as the diver moves until the corrected localization accuracy meets the requirements.

Figure 1B shows the components of the first two STWs, where the orange dashed box represents the first STW constructed by P_0,

FIGURE 1
Localization scenarios for acoustic tracking inertial navigation methods based on MP. **(A)** Localization Scenarios; **(B)** STW.

P_1, and P_2, and is denoted as STW 1. Assuming the initial localization is $P_0(x_0,y_0, z_0)$, acoustic tracking is used to obtain the slant distance between the diver and the beacon as R_0. After a period of time, the coordinates of divers are $P_1(x_1,y_1, z_1)$, the relative displacement of P_0 and P_1 is recorded in the INS as $\Delta P_{10} = (\Delta x_{10}, \Delta y_{10})$, and acoustic tracking is again used to obtain the slant distance R_1. The third time the acoustic signal is received, the corresponding coordinate is set to $P_2(x_2,y_2, z_2)$, and the displacements of P_1 and P_2, $\Delta P_{21} = (\Delta x_{21}, \Delta y_{21})$, and the slant distance, R_2, are recorded. The two-dimensional coordinates P_0 and P_1 (depth measured by the pressure transducer) are expressed as the difference between the coordinates to be solved, P_2, and the relative displacement of the known INS in the form of Equation 1:

$$\begin{cases} (x_0, y_0) = ((x_2 - \Delta x_{20}), (y_2 - \Delta y_{20})) \\ (x_1, y_1) = ((x_2 - \Delta x_{21}), (y_2 - \Delta y_{21})) \end{cases}, \quad (1)$$

where $\Delta x_{20} = \Delta x_{10} + \Delta x_{21}$ and $\Delta y_{20} = \Delta y_{10} + \Delta y_{21}$.

Combined with the slant distance $R_i(i = 0, 1, 2)$ obtained by acoustic tracking, the system of joint nonlinear localization equations is:

$$\begin{cases} R_0 = \sqrt{[(x_2 - \Delta x_{20}) - x_s]^2 + [(y_2 - \Delta y_{20}) - y_s]^2 + (z_0 - z_s)^2} \\ R_1 = \sqrt{[(x_2 - \Delta x_{21}) - x_s]^2 + [(y_2 - \Delta y_{21}) - y_s]^2 + (z_1 - z_s)^2} \\ R_2 = \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2 + (z_2 - z_s)^2} \end{cases}. \quad (2)$$

In (2), the depth $z_i(i = 0,1,2)$ at each localization is measured by the pressure transducer, which can be considered a known quantity. The coordinate estimation problem for the diver P_2 can be transformed into the problem of solving the corresponding system of localization equations in (2). Considering that the deviation of the adjacent localizations of the INS output may cause the problem of singular or pathological coefficient matrix, this paper adopts a nonlinear optimization method to iteratively search for the optimal solution of the localization of divers Wang et al. (2022).

2.2 Localization p process

Figure 2 shows the overall flow of the localization method in this paper, which includes a delay estimation module, a sound tracking module, an inertial navigation module, a nonlinear optimization module, and a STW iterative update module. The inverse sound velocity profiles are based on Array for Real-time Geostrophic Oceanography (Argo) data at 116.63-119.63E and 17.5-18.5N after Multi-Layer Perceptron (MLP) estimation of the first five orders of Principal Components (PC) estimation for better range.

The INS solution of the Inertial Navigation Module requires the following four steps in sequence: attitude update, coordinate transformation, velocity update, and localization update.

The acoustic line tracking module, as a crucial component of the inertial navigation-based error correction, aims to obtain the acoustic distance measurement information between the diver and the SB, thus providing the slant distance required for the nonlinear optimization process. The acoustic tracking module is first provided with the time of arrival (TOA) of the beacon acoustic signal by the time delay estimation module. Then, an iterative dichotomous search is performed to correct the acoustic line curvature based on the initial grazing angle of the acoustic line. Considering the shallow sea environment where divers operate, the acoustic signal is easily disturbed by multipath channels and environmental noise during propagation, so the TOA method should be selected with both multipath and noise resistance. In this paper, Second Generalized Cross Correlation (SGCC) is selected as the localization method to estimate the TOA of the received signal from the diver in order to avoid pseudo peaks in the cross correlation Yang et al. (2010). The method steps are as follows:

First, the autocorrelation operation is performed on $x_1(t)$ to obtain the autocorrelation function $R_{x1x1}(\tau)$. Second, the mutual correlation function $R_{x1x2}(\tau)$ of $x_1(t)$ and $x_2(t)$ is calculated, followed by the calculation of the mutual power spectral function $G_{RR}(f)$ of R_{x1x1} and R_{x1x2}. Then, the frequency domain filtering of $G_{RR}(f)$ is performed in conjunction with the generalized frequency weighting function $\psi_g(f)$. Finally, the filtered secondary mutual power spectrum is subjected to Fourier Inverse Transform to obtain the Second Generalized Cross Correlation function $R_{y_1y_2}^{(g)}$, as shown in Equation 3:

$$R_{y_1y_2}^{(g)}(\tau) = \int_{-\infty}^{\infty} \psi_g(f) G_{RR}(f) e^{j2\pi f\tau} df \, . \quad (3)$$

Since the method in this paper corresponds to the active localization scenario, when the power of the acoustic signals emitted by SB is large, the function normalizes the amplitude of the reciprocal power spectrum in the frequency domain, and is able

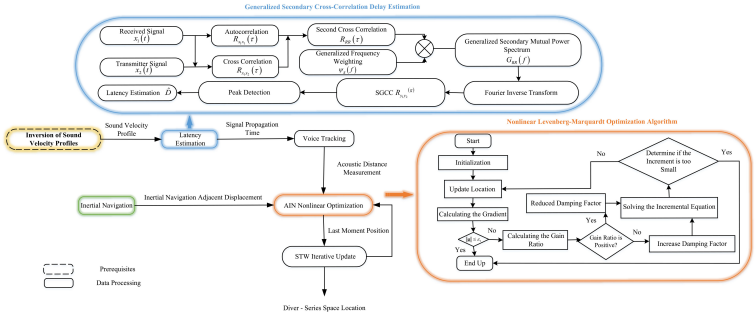


FIGURE 2
The general flowchart of the localization method in this paper.

to effectively sharpen the correlation peaks in active localization scenarios with large signal power, and is suitable for underwater acoustic environments with low or moderate reverberation, and the expression for the phase transformation (PHAT) weighting function is given by $\psi_g(f) = \frac{1}{|G_{RR}(f)|}$.

When searching for peaks for (3), the time corresponding to the peaks is the time delay for the diver to receive the signal as $\hat{D}$, as shown in Equation 4:

$$\hat{D} = arg\left\{\max_{\tau}[R^{(g)}_{y_1y_2}]\right\}. \tag{4}$$

The SGCC-PHAT method in the attempted study combines both good noise immunity and multipath immunity, and is more suitable than other algorithms for delay estimation scenarios of diver-received signals in the noisy environment of the shallow sea.

The nonlinear optimization module is mentioned in the next section.

2.3 Nonlinear optimization

The system of localization Equation (2) is transformed into the following form if the depth of divers, the coordinates of the single beacon on the water surface and the acoustic distance measurements R_1, R_2, R_3 are known:

$$\begin{cases} \sqrt{R_0^2 - (z_0 - z_s)^2} = \sqrt{[x_2 - (\Delta x_{20} + x_s)]^2 + [y_2 - (\Delta y_{20}) + y_s)]^2} \\ \sqrt{R_1^2 - (z_1 - z_s)^2} = \sqrt{[x_2 - (\Delta x_{21} + x_s)]^2 + [y_2 - (\Delta y_{21}) + y_s)]^2} \\ \sqrt{R_2^2 - (z_2 - z_s)^2} = \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2} \end{cases}, \tag{5}$$

where $(\Delta x_{20}, \Delta y_{20})$ is the coordinate change of INS outputs P_0 and P_2 in the inertial reference system.

If the coordinate (x_2, y_2) to be solved is set to (x, y), the Equation (5) is expressed in the form of the following error vector:

$$e(x,y) = \begin{bmatrix} \sqrt{[x - (\Delta x_{20} + x_s)]^2 + [y - (\Delta y_{20}) + y_s)]^2} - \sqrt{R_0^2 - (z_0 - (z_s)^2} \\ \sqrt{[x - (\Delta x_{21} + x_s)]^2 + [y - (\Delta y_{21}) + y_s)]^2} - \sqrt{R_1^2 - (z_1 - (z_s)^2} \\ \sqrt{(x - x_s)^2 + (y - y_s)^2} - \sqrt{R_2^2 - (z_2 - z_s)^2} \end{bmatrix}. \tag{6}$$

Squaring the Equation (6) yields the objective function $E(x, y)$ as follows:

$$E(x,y) = \frac{1}{2}\mathbf{e}(x,y)^T\mathbf{e}(x,y) = \frac{1}{2}\sum_{i=1}^{3}\mathbf{e}_i^2(x,y). \tag{7}$$

The sum of squares of the minimized range error terms according to Equation (5–7), and then use the iterative method to nonlinearly optimize the above least squares problem. The localization method in this paper uses the Levenberg-Marquardt (LM) method for nonlinear optimization of the spatial localization of divers in the following way:

First, the coordinates to be solved are set as a vector $\mathbf{x}(x,y)$. The error term $e(\mathbf{x})$ is linearized based on a first order Taylor series expansion with the expression in Equation 8:

$$e(\mathbf{x} + \Delta\mathbf{x}) \approx e(\mathbf{x}) + \mathbf{J}(\mathbf{x})\Delta\mathbf{x}, \tag{8}$$

where $\mathbf{J}(\mathbf{x})$ is the Jacobi matrix of $e(\mathbf{x})$ with respect to $\mathbf{x}$ and $\Delta\mathbf{x}$ is the iteration step.

The minimization objective function is then approximated as a linear least squares problem to solve for the increment, as shown in Equation 9:

$$\Delta x^* = arg\min_{\Delta\mathbf{x}}\frac{1}{2}\| e(\mathbf{x}) + \mathbf{J}(\mathbf{x})\Delta\mathbf{x}\|^2 \tag{9}$$

The LM method introduces a damping coefficient μ in the construction of the incremental equation, which can be expressed as Equation 10:

$$(\mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}) + \mu\mathbf{I})\Delta\mathbf{x} = -\mathbf{J}(\mathbf{x})^T e(\mathbf{x}), \tag{10}$$

where $\mathbf{I}$ is the unit array; the initial value of the damping coefficient, $\mu_0 = \tau \times \max\{a_{ii}\}$, where $\mathbf{A}_0 = \mathbf{J}(\mathbf{x}_0)^T\mathbf{J}(\mathbf{x}_0)$, a_{ii} is the diagonal elements of $\mathbf{A}_0$; and τ is a constant, usually set to 10^{-3} or 1.

The quality of each iteration step is evaluated according to the change in the goodness of approximation μ, which defines the gain ratio ρ, as shown in Equation 11:

$$\rho = \frac{e(\mathbf{x} + \Delta\mathbf{x}) - e(\mathbf{x})}{\mathbf{J}(\mathbf{x})\Delta\mathbf{x}}. \tag{11}$$

The denominator represents the degree of decrease of the first-order differential component in each iteration, and the numerator represents the decrease of the actual function. When ρ is small, it means that the approximation degree of the iteration step is poor, and μ should be increased accordingly; on the contrary, when ρ is large, μ should be decreased. In this paper, the localization method adopts the μ updating strategy as follows: when $\rho > 0$, $\mu = \mu \times \max\left\{\frac{1}{3}, 1 - (2\rho - 1)^3\right\}$; otherwise, $\mu = \mu \times \nu$ $\nu = 2 \times \nu$, the initial value $\nu_0 = 2$. After adjusting the damping coefficient, it is first substituted into the increment equation to compute the increment amount, and then the increment amount is iterated to update the localization, and then a new round of iteration is started, and the iteration is stopped only when the increment amount is sufficiently small or the gradient of the descent is sufficiently small.

3 Simulation results and analysis

3.1 Simulation analysis of delay estimation module

To investigate the estimation performance of the SGCC-PHAT method in the delay estimation module in shallow sea localization scenarios, this subsection compares the delay estimation effects of different cross-correlation algorithms in terms of noise resistance and multipath resistance, respectively. Assuming that the receiver and transmitter have completed time synchronization, the anti-noise performance of the method is analyzed by adding Gaussian white noise of different power to the transmit signal; and the multipath resistance of the method is explored by simulating the shallow sea multipath channel using the BELLHOP channel model.

First, the noise immunity performance of the four cross-correlation algorithms is analyzed. In this simulation, a single-frequency sine wave with a frequency of 1000 Hz is used as the transmit signal in this 167 simulation. The signal is sampled at 50 kHz, with 1024 samples and a real-time delay value of 100 sample intervals (i.e., 2 ms). The simulation analyzes the noise immunity of the four cross-correlation algorithms from the two scenarios of high signal-to-noise ratio (SNR) (20 dB) and low SNR (-20 dB).

Figure 3 shows the delay estimation results of the Cross Correlation (CC) algorithm, the Generalized Cross Correlation-PHAT (GCC-PHAT) algorithm, the Second Cross Correlation (SCC) algorithm, and the SGCC-PHAT method at high SNR. Overall, all four algorithms can accurately estimate the true delay value of the received signal below 20 dB. From the horizontal comparison, it can be observed that when the signal power is greater than the noise power, the GCC-PHAT method has sharper correlation peaks compared to the CC algorithm, the peak spreading of the CC method leads to a reduction in the resolution of the peaks, and the SGCC-PHAT method and the SCC method have a similar pattern. From the longitudinal comparison, it can be observed that the amplitude of the interference noise around the peaks is significantly reduced in the SGCC-PHAT method compared with the GCC-PHAT algorithm, thus verifying that the two cross-correlation operations can sharpen the peaks and further inhibit the noise from interfering with the signal. The above simulation verifies the good resolution of the SGCC-PHAT method for delay estimation under high SNR conditions, as well as a certain resistance to side flap interference.

Figure 4 compares the estimation performance of the four types of cross-correlation algorithms under low SNR conditions. Among the four algorithms, only CC and SCC estimate accurately, the true value of delay of GCC-PHAT is overwhelmed by a large amount of noise, and the estimate of SGCC-PHAT method contains a small amount of bias (0.00012 s). From the side-by-side comparison, it can be seen that CC has better noise immunity compared to GCC-PHAT, and the true delay value can still be estimated under low SNR conditions. The estimation accuracy of GCC-PHAT decreases sharply with the increase of noise power under the low SNR conditions. This phenomenon occurs because the PHAT weighting function in GCC-PHAT relies on phase delay estimation. When the signal is flooded with noise, the phase of the noise occupies the main component in the estimation result, so the estimation error increases sharply. Compared with GCC, CC has better noise immunity because it preserves the amplitude information in the frequency domain. From the longitudinal comparison, although the estimation accuracy of SGCC-PHAT decreases accordingly under the low SNR condition, its estimation error is generally very small (0.00012 s) compared to GCC-PHAT, which confirms that SGCC-PHAT has better noise suppression ability compared to GCC-PHAT. From the above simulation analysis of the anti-noise performance, it can be seen that under the high SNR condition, the SGCC has a sharper correlation peak compared with the SCC and is more sensitive to the estimation error; under the low SNR condition, the SGCC-PHAT benefits from the two cross-correlation operations to inhibit the interference of the noise on the signal better than the GCC-

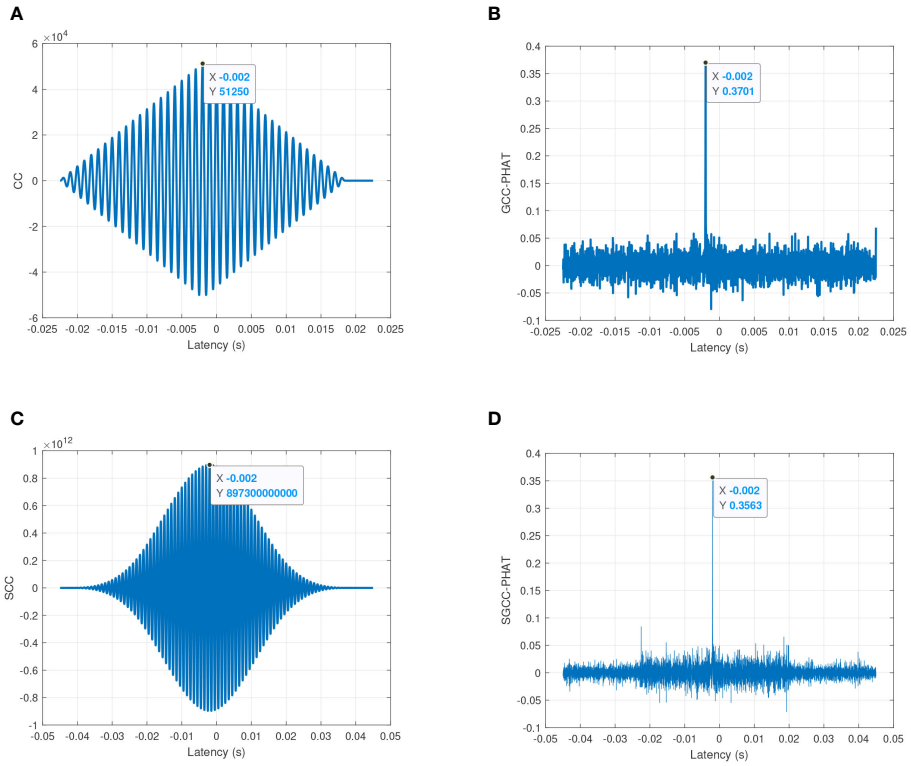


FIGURE 3
Estimation performance comparison of four cross-correlation algorithms under high SNR conditions (20 dB). **(A)** CC; **(B)** GCC-PHAT; **(C)** SCC; **(D)** SGCC-PHAT.

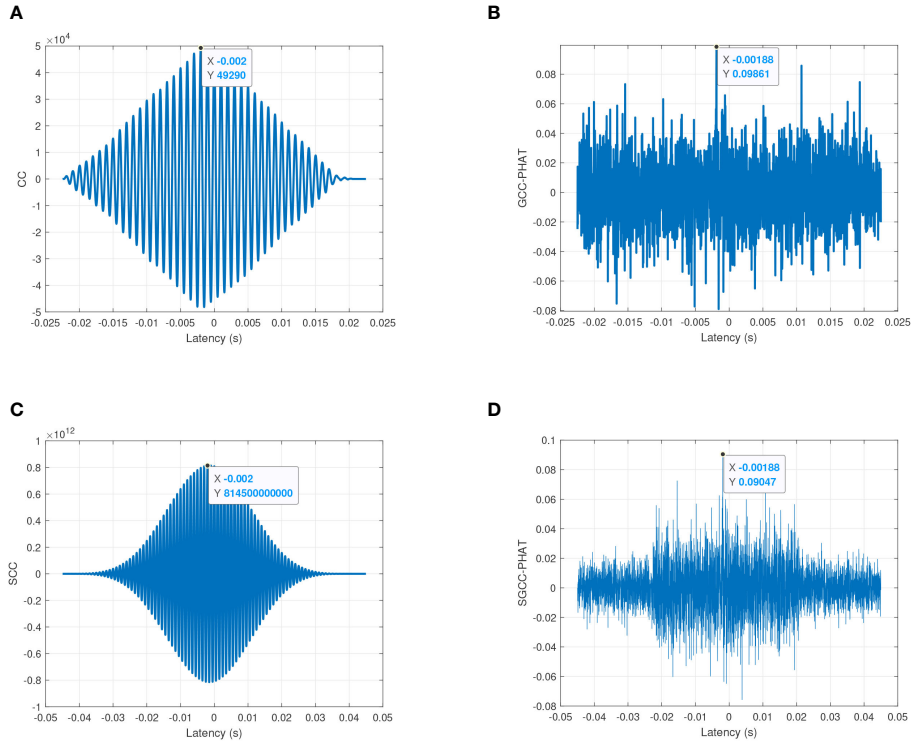


FIGURE 4
Estimation performance comparison of four cross-correlation algorithms under low SNR conditions (-20 dB). **(A)** CC; **(B)** GCC-PHAT; **(C)** SCC; **(D)** SGCC-PHAT.

PHAT, and it has a higher robustness. Considering that the water SB in the localization scenario of this chapter is actively transmitting acoustic signals as a sound source, there is no low SNR environmental condition, in which case it is feasible and superior to choose the SGCC-PHAT as the delay estimation method for localization in this chapter.

Second, simulation analysis is performed to analyze the anti-multipath performance of the four types of cross-correlation algorithms. In this simulation, the BELLHOP model is selected to simulate the underwater multipath channel, and the received acoustic signal is obtained by calculating the convolution of the transmitted signal and the channel impulse response, and the different cross-correlation algorithms are applied to the received signal for time delay estimation. It should be noted that the sound velocity profile used in the BELLHOP model is the inverted sound velocity profile of August 2017 in the China South Sea, and the rest of the parameter settings are shown in Table 1. Considering the complexity of the underwater environment, in this simulation, the linear frequency modulation (LFM) signal is selected as the transmission signal, the center frequency of the signal is set to 30 kHz, the bandwidth is 5 kHz, the duration is 20 ms, and the real propagation time of the acoustic signal can be obtained from the model calculation as 0. 65645 s. The LFM signal before and after passing through the BELLHOP channel is shown in Figure 5, and the change of the waveform shows that the received signal is not only delayed in time, but also the multipath effect leads to different degrees of attenuation of the signal amplitude. Based on the received signals shown in Figure 5, the antimultipath performance of the four types of cross-correlation algorithms is compared, and a total of thirty simulations are performed, which shows that the SGCC-PHAT has a better antimultipath capability, and only one simulation result is shown as shown in Figure 6. The figure reflects the ability of different cross-correlation algorithms to resist the multipath effect in the shallow sea, from which it can be seen that the GCC-PHAT and the SGCC-PHAT have better anti-polymath performance compared to the other two algorithms, and the delay estimation error of both of them is only 0. This verifies that the SGCC-PHAT can resist the multipath effect in the shallow sea, and the estimation error of CC and SCC is extended to 0.00001s. This verifies that SGCC-PHAT can resist the multipath effect in the shallow sea, and SCC is more effective in the shallow sea. SGCC-PHAT can extract the propagation delay of the direct signal in the

TABLE 1 BELLHOP model parameterization.

Parameter Name	Value
Sound Source Depth (m)	40
Receiver Depth (m)	60
Transmitter to Receiver Horizontal Distance (m)	1000
Acoustic Frequency (kHz)	30
Number of voices	30
Sound Line Exit Angle Sector (°)	-15~15
Seabed Sediment Sound Velocity (m/s)	1600
Seafloor Sediment Density (g/cm^3)	1.8

10.3389/fmars.2024.1368317

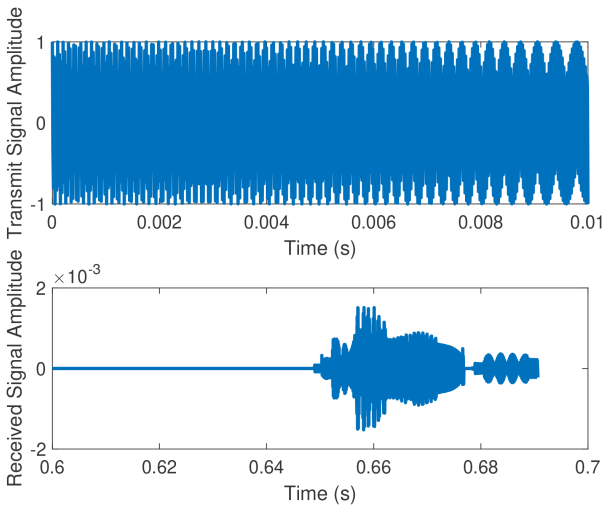


FIGURE 5
Comparison of signal waveforms before and after passing through the BELLHOP channel.

multipath channel better than CC and SGCC, and is more suitable for the localization scenarios the shallow sea operation of divers.

In summary, this subsection verifies, based on simulation, that the SGCC-PHAT has both good anti-noise performance and anti-multipath performance, and is more suitable than other algorithms for the delay estimation scenarios of the received signals of divers in the noise environment of the shallow sea, and thus the methodology in this chapter selects the SGCC-PHAT as a delay estimation method for the voice tracking module with feasibility and superiority.

3.2 Nonlinear module simulation and analysis

This subsection demonstrates the INS cumulative error correction results of the proposed method, and also explores the effect of the transmission period of SB on the INS cumulative error correction. Based on the received signal delay estimation using the SGCC-PHAT algorithm, the INS cumulative error correction performance of the proposed method is simulated and analyzed.

The Inertial Measurement Unit sensor error settings in the simulation are the same as in Table 2, and the measurement period is set to 0.1 seconds. The diver is moving at a speed of approximately 1 knot. The SB is placed at a depth of 0.5 m underwater, and the corresponding latitude and longitude are 17.5016N and 116.6016E. In this simulation, the East-North-Up (ENU) coordinate system is used as the reference coordinate system for navigation, and the initial localizations of divers of 17.5N and 116.6E are used as the origin of the reference system, and the relative localizations of the moving trajectory of divers and the SB are shown in Figure 7. The simulation uses Root Mean Square Error (RMSE) to measure localization accuracy.

Table 3 shows the localization errors for different STWs for a beacon transmission period of 20 seconds, and from the analysis of the graphs, it can be seen that the localization errors of the four STWs

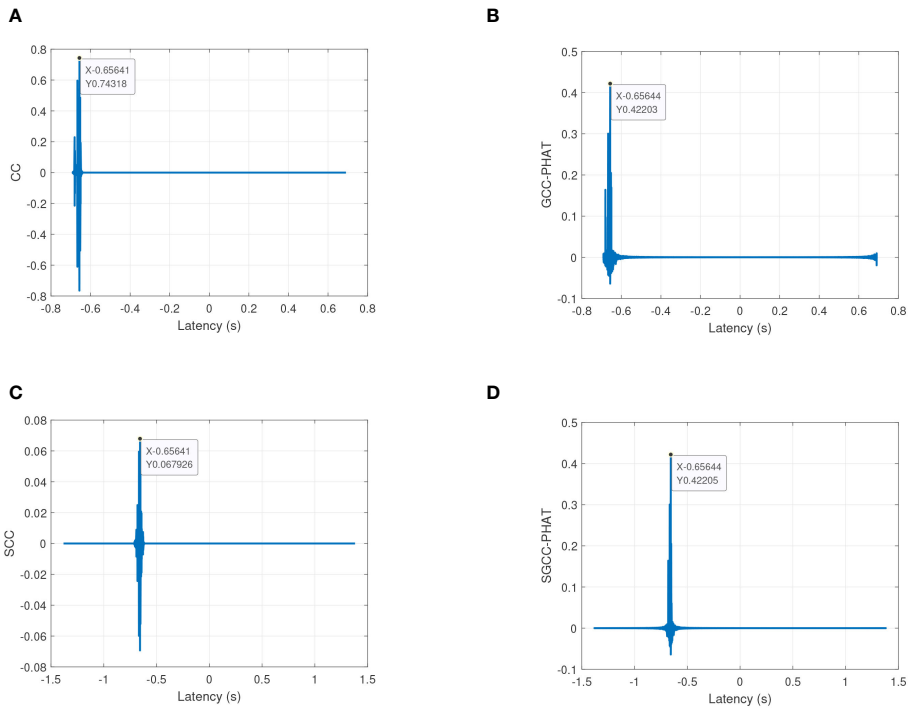


FIGURE 6
Estimation performance comparison of four cross-correlation algorithms under BELLHOP multi-traffic channels. **(A)** CC; **(B)** GCC-PHAT; **(C)** SCC; **(D)** SGCC-PHAT.

TABLE 2 Inertial Measurement Unit (IMU) sensor error parameter settings.

Sensor Type	Error Parameters	Numeric Value
Gyros	Constant Drift (°/h)	0.05
	Angular Velocity Random Walk (°/$\sqrt{h}$)	0.01
Accelerometer	Constant Zero Offset μg √	50
	Speed Random Wandering $\mu g/\sqrt{Hz}$	10

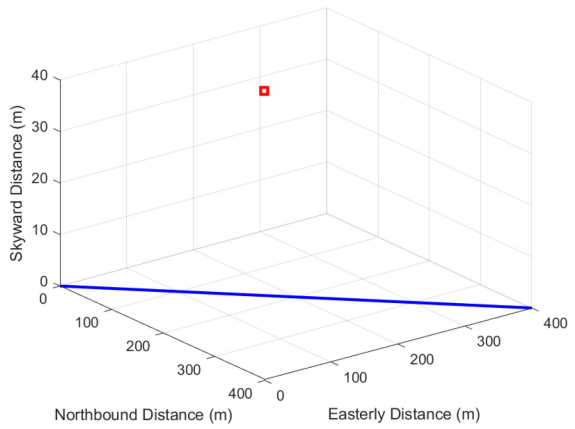


FIGURE 7
Diver motion trajectory and SB localization.

are less than 5 meters, and the corresponding localization errors of the second and third STWs are reduced to about 1 meter, which improves the accuracy of INS localization by about 98% compared with the INS localization accuracy, and the percentage of localization error is less than 0.5%, and the high-precision correction of the cumulative INS error is achieved in a minimum of only three STWs.

Next, the number of STWs is selected as four, and the total simulation time is set to 1000 seconds, of which the first 420 seconds is for Purely Inertial Navigation (PIN), and the AAINL is performed in 420-520 seconds, thus simulating the actual operation scenario in which the diver moves for a certain period of time and then performs error correction. Figure 8 shows the localization trajectory and error of this method when the transmission period of the SB is 20 seconds, 40 seconds, 60 seconds and 80 seconds. The black line in the figure indicates the ideal motion trajectory of divers during 420 seconds to 520 seconds, the blue line indicates the motion trajectory of divers output by the INS, and the red line corresponds to the motion trajectory of divers after AINL by the method of this paper. The depth of divers is measured by the pressure sensor with negligible numerical error.

From Figure 8, it can be observed that the cumulative INS error causes the localization of divers solved by PIN to deviate seriously from the true value, and the acoustic-assisted correction using SB in method of this papers can estimate the localization of divers closer to the true coordinates. The reason why the MP trajectory in Figure 8 is close to the ideal trajectory at the beginning and then gradually moves away from it is that in the iteration of the STW, the accuracy of the localization of divers first gradually improves with the introduction of acoustic ranging, and then, due to the increase in the distance between the diver and the SB, the accuracy of acoustic ranging at this time decreases, leading to the increase in the localization error corresponding to the fourth STW.

From the left vertical comparison of each figure in Figure 8, it can be seen that although the moving trajectories vary under different acoustic distance measurement cycles, they are generally closer to the real value than the trajectories output by the INS, and the above phenomenon is attributed to the fact that method of this papers uses acoustic ranging to introduce the auxiliary information of distance constraints.

Further observation shows that when the diver is closer to the SB, the accuracy of this method is higher, indicating that the distance between the SB and the diver is a key factor affecting the performance of acoustic assisted localization, and the beacon should be deployed as close as possible to the moving trajectory of divers. The localization errors under different acoustic measurement cycles are shown in the right vertical comparison of each figure in Figure 8, from which it can be seen that the localization error of method of this papers is significantly smaller than the INS localization error. As the cumulative error of the INS increases, the localization error of the MP also increases. When the beacon transmission period is 40 seconds, the localization errors of the four STWs are less than 10 meters; when the beacon transmission period is 60 seconds, the maximum localization error corresponding to the STWs is close to 20 meters; when the beacon transmission period further increases to 80 seconds, the maximum error corresponding to the STWs has exceeded 20 meters. The reason for the above increase in the localization error of the MP is that the size of the SB transmission period essentially reflects the distance between adjacent localizations in the STW, i.e., the length of the baseline between the MP. Increasing the beacon transmission period leads to a larger distance between neighboring localizations in the STW, at which time the INS cumulative error also increases, and substituting the relative displacement with a larger deviation into the set of nonlinear localization equations increases the auxiliary localization error. The method in this paper has the problem of choosing the beacon transmission period, because the INS neighbor displacement is introduced as a known quantity, and the localization accuracy improvement is limited by the INS cumulative error.

The localization accuracy improvement values for the four transmission cycles compared with INS in Figure 8 are shown in Table 4, from which it can be seen that as the beacon transmission cycle increases, the localization accuracy improvement of method of this papers as a whole shows a decreasing trend, which is related to the increase of the INS neighborhood localization deviation. It should be noted that the localization accuracy of the method in this paper is also related to the distance of the diver from the SB, so the larger value of the localization accuracy improvement under each transmission period in the table tends to correspond to the STW closest to the beacons. According to the simulation verification, the localization method in this paper has a certain degree of error correction effect in transmission cycles less than 200 seconds.

From the above simulation analysis, it can be seen that the localization method in this paper can effectively correct the

TABLE 3 Comparison of localization errors for a 20-second beacon duration.

STW Number	1	2	3	4
AINL Error(m)	3.2881	1.1322	0.8715	4.8232
INS Localization Error(m)	51.9373	55.9669	60.0731	64.4272
Improved Localization Accuracy(%)	93.67	97.98	98.55	92.51
Percent Localization Error(%)	1.43	0.47	0.35	1.86

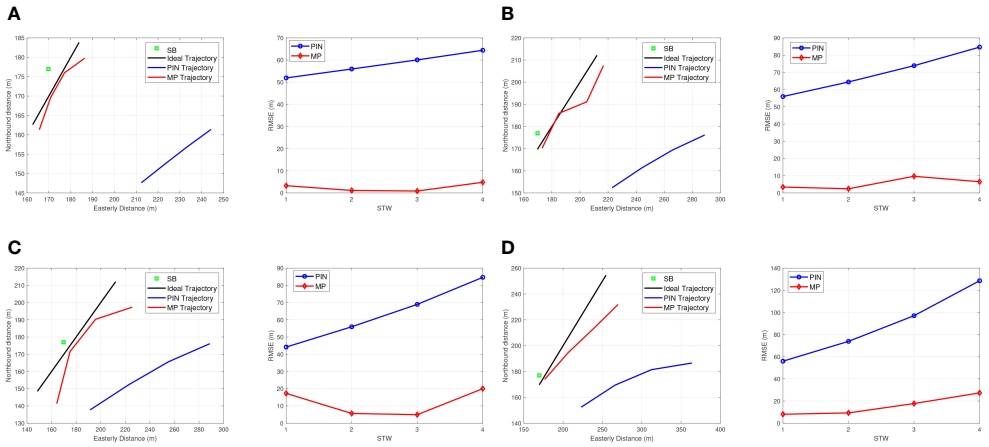


FIGURE 8
Diver trajectories and localization errors under different beacon transmission cycles. **(A)** Beacon transmission time of 20 seconds; **(B)** Beacon transmission time of 40 seconds; **(C)** Beacon transmission time of 60 seconds; **(D)** Beacon transmission time of 80 seconds.

cumulative error of INS under different beacon broadcasting cycles, and the effect of improving the localization accuracy is very significant. The factors affecting the estimation accuracy of the localization method in this paper are mainly the broadcasting period of the SB and the proximity of the diver to the SB. In this paper, the 20-second transmission period is chosen as the acoustic distance measurement period for the AINL, and the SB is placed as close as possible to the trajectory of divers to realize the fast and high-precision INS error correction. Considering that divers need to quickly correct the accumulated INS error during underwater operation, designing too many MP will cause more serious localization deviation and thus reduce the localization accuracy, so this paper chooses three MP to construct a STW In the process of localization, three to four STWs can quickly eliminate the accumulated INS error, and choosing too many STWs not only increases the computational complexity, but also is prone to the interference of acoustic ranging distance and affects the correction effect.

4 Conclusion

In this paper, an acoustic tracking AINL method based on MP is proposed. Considering that it is difficult to lay a large number of acoustic base arrays in a short time for the sudden operation scenarios of divers, acoustic tracking assisted localization using a single beacon on the water surface combined with acoustic line tracking combines the flexibility and simplicity of single beacon localization and improves the aquatic acoustic localization accuracy of divers. The localization idea of MP is introduced, and a STW

TABLE 4 Improving the localization accuracy of the method of this paper over INS with different beacon transmission periods.

Broadcast PeriodSTW Number	1	2	3	4
20 seconds	93.67%	97.98%	98.55%	92.51%
40 seconds	93.88%	96.25%	86.90%	92.31%
60 seconds	61.08%	89.91%	92.86%	76.35%
80 seconds	85.59%	87.45%	81.75%	78.74%

consisting of three virtual primitives is constructed using the current localization of divers and two historical positions, and the spatial localization of divers at the nearest moment is solved based on the acoustic distance measurement and the inertial navigation displacement in the neighboring measurement period. To solve the problem of singularity or pathology in the coefficient matrix of the localization equation system due to the deviation of the neighboring localizations of the inertial guides, the LM method is used for nonlinear iterative optimization search. Then, how to realize the time synchronization between the SB and the diver is investigated and the corresponding underwater experiments are carried out.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

KY: Conceptualization, Project administration, Visualization, Writing – original draft. ZT: Supervision, Writing – review & editing. WW: Supervision, Writing – review & editing. TT: Conceptualization, Supervision, Writing – review & editing, Writing – original draft. LZ: Supervision, Writing – review & editing. YW: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Key Research and Development Program of Guangxi, grant number AB21196066, in part by the Key Research and Development Program of Guilin, grant number 20220113-6, in part by Key Research and Development Program of Guilin, grant number 20210206-3, and in part by the Guangxi Science and Technology Plan Project (Grant No.AD22035141).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Brown, H. C., and Wang, H. (2013). Underwater augmented reality: Navigation and identification. *2013 OCEANS - San Diego*. 1–5.

Cheng, S., Cheng, J., Zang, N., and Cai, J. (2022). "Polar sins/usbl integrated navigation algorithm considering acoustic communication delays," in *2022 5th International Symposium on Autonomous Systems (ISAS)*. 1–6.

Hegrens,, Gade, K., Hagen, O. K., and Hagen, P. E. (2009). "nderwater transponder positioning and navigation of autonomous underwater vehicles," in *OCEANS 2009*. 1–7.

Jin, B., Xu, X., Zhu, Y., Zhang, T., and Fei, Q. (2019). Single-source aided semi-autonomous passive location for correcting the position of an underwater vehicle. *IEEE Sensors J.* 19, 3267–3275. doi: 10.1109/JSEN.7361

Ju-Cheng, Z., Chun-Hao, S., Da-Jun, S., and Yun-Feng, H. (2017). "Auv integrated navigation algorithm based on single beacon ranging," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. 1–5.

Kaneko, K., and Kubota, N. (2021). "Underwater acoustic positioning system using ultrasonic waves for diving support," in *2021 World Automation Congress (WAC)*. 308–312.

Liu, S., Zhang, T., Zhang, J., and Zhu, Y. (2021). A new coupled method of sins/dvl integrated navigation based on improved dual adaptive factors. *IEEE Trans. Instrumentation Measurement* 70, 1–11. doi: 10.1109/TIM.2021.3118090

Luo, J., Yang, Y., Wang, Z., and Chen, Y. (2021). Localization algorithm for underwater sensor network: A review. *IEEE Internet Things J.* 8, 13126–13144. doi: 10.1109/JIOT.2021.3081918

Lyu, X., Hu, B., Wang, Z., Gao, D., Li, K., and Chang, L. (2022). A sins/gnss/vdm integrated navigation fault-tolerant mechanism based on adaptive information sharing factor. *IEEE Trans. Instrumentation Measurement* 71, 1–13. doi: 10.1109/TIM.2022.3214628

Su, R., Gong, Z., Li, C., and Shen, X. (2023). Algorithm design and performance analysis of target localization using mobile underwater acoustic array networks. *IEEE Trans. Vehicular Technol.* 72, 2395–2406. doi: 10.1109/TVT.2022.3211830

Wang, S., Xu, M., Zhang, X., and Wang, Y. (2022). Fitting nonlinear equations with the levenbergmarquardt method on google earth engine. *Remote Sens.* 14, 2055. doi: 10.3390/rs14092055

Xu, B., Hu, J., and Guo, Y. (2022). An acoustic ranging measurement aided sins/dvl integrated navigation algorithm based on multivehicle cooperative correction. *IEEE Trans. Instrumentation Measurement* 71, 1–15.

Yang, Z., Xu, W., Xiao, Z., and Pan, X. (2010). "Passive localization of an autonomous underwater vehicle with periodic sonar signaling," in *OCEANS'10 IEEE SYDNEY*. 1–4.

Ye, K., Cai, Y., Hong, S., and Sun, H. (2023). Direction-of-arrival estimation based on difference-sum co-array of a special coprime array. *Electron. Lett.* 59, e12701. doi: 10.1049/ell2.12701

Zhang, T., Wang, Z., Li, Y., and Tong, J. (2019). A passive acoustic positioning algorithm based on virtual long baseline matrix window. *J. Navigation* 72, 193206. doi: 10.1017/S0373463318000590

Zhang, X., He, B., and Gao, S. (2023). An integrated navigation method for small-sized auv in shallow-sea applications. *IEEE Trans. Vehicular Technol.* 72, 2878–2890. doi: 10.1109/TVT.2022.3216003

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver sas imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., Yang, P., Wang, Y., Shen, W., Yang, J., Wang, J., et al. (2024). A novel multireceiver sas rd processor. *IEEE Trans. Geosci. Remote Sens.* 1–1. doi: 10.1109/TGRS.2024.3362886

Zhang, L., Zhang, T., and Wei, H. (2022). A novel robust inertial and ultra-short baseline integrated navigation strategy under the influence of motion effect. *IEEE Trans. Intelligent Transportation Syst.* 23, 19323–19334.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 April 2024
DOI 10.3389/fmars.2024.1378396

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Xing Chuanxi,
Yunnan Minzu University, China
Zhiping Xu,
Jimei University, China

*CORRESPONDENCE
Fengqin Zhu
fqzhu_07@163.com

RECEIVED 29 January 2024
ACCEPTED 26 March 2024
PUBLISHED 11 April 2024

CITATION
Qu K, Yin W, Zhu F and Meng L (2024) Modeling sound speed profile based on ocean normal mode.
Front. Mar. Sci. 11:1378396.
doi: 10.3389/fmars.2024.1378396

Modeling sound speed profile based on ocean normal mode

Ke Qu[1], Weifeng Yin[1], Fengqin Zhu[1]* and Lei Meng[2]

[1]College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, China, [2]Unit 92578 of the People's Liberation Army, Beijing, China

Introduction: Statistical methods such as empirical orthogonal functions (EOFs) are often used to model the sound speed profile (SSP). However, their statistical nature often leads to the sample dependence and physical fuzziness.

Method: This study proposes a technique for modeling the SSP from the perspective of ocean dynamics. It employs the ocean normal mode, which is the mode of fluid particles motion, to deduce perturbations in the SSP, which is called the ocean mode basis (OMB).

Result: The results of SSP reconstruction of in-situ samples showed that a few leading orders of the OMB can provide a compact representation of the SSP. Oscillations of the contours and gradient of the sound speed in thermocline were analyzed by using the first two orders of the projection coefficients of the relationship between the OMB and the baroclinic mode. As a physical model, this technique can also be used to characterize the dynamics of internal solitary waves. Furthermore, the OMB derived from archival date was used for SSP inversion. The results showed that the OMB can reconstruct SSP of a reasonable resolution without requiring in-situ samples.

Discussion: Compared with statistical models, the OMB can better explain the ocean dynamics underlying variations in the SSP while requiring fewer samples.

KEYWORDS

sound speed profile, empirical orthogonal function, internal solitary wave, inversion, basis function

1 Introduction

Sound speed profile (SSP) is the basic acoustic characteristic of the water column in the ocean. The properties of sound propagation are strongly influenced by temporal and spatial changes in the SSP due to the ocean dynamics. Conversely, the water column can be observed and analyzed by examining the SSP and perturbations in it. Information on the ocean, ranging from the large-scale marine environmental monitoring of the global climate to the fine-scale analysis of internal waves and turbulence in local seas, can be obtained from SSP inversion (Behringer et al., 1982; Yang and Liu, 2017). SSP plays an important

role in underwater applications (Xing et al., 2021, Zhang et al., 2021; Xing et al., 2023; Yang, 2024; Zhang, 2024, Zhang et al., 2024).

To provide constraints on the search space for inversion, it is necessary to apply a dimensionality reduction technique to model a refined SSP. As a technique of principal component analysis, empirical orthogonal functions (EOFs) have been the most widely used method for modeling SSPs in recent decades. LeBlance first proved that the EOF can describe the SSP without losing much information by using a few groups of basis vectors and projection coefficients (LeBlanc and Middleton, 1980). The EOF was subsequently used in a considerable amount of research on SSPs, including ocean tomography (Li et al., 2015), uncertainty analysis of inversion (Jiang and Chapman, 2009), perturbation analysis of the sound field (Hjelmervik et al., 2012), and rapid environmental assessment (Chen et al., 2018). Moreover, Bianco and Cheng used machine learning to reduce error in the reconstruction of SSPs while using fewer coefficients (Bianco and Gerstoft, 2017; Cheng et al., 2022). This effort indicated that the resolution of the SSP model can be improved by lifting the orthogonality restriction. However, it remains difficult to satisfy certain demands of ocean observation in light of the statistical nature of the SSP model currently in use. To form the basis vectors, the abovementioned SSP models obtain the rules of perturbation of the sample data mainly by statistical methods: The basis vectors are entirely the products of numerical analysis. Early studies on ocean observation and analysis suggested that statistical models do not necessarily correspond to the true dynamical characteristics or modes of the ocean physical behavior (Dommenget and Latif, 2002; Behera et al., 2003). In addition, there are certain requirements on the sample size and temporal–spatial coverage of SSPs for basis generation that pose a barrier to SSP modeling in scenarios where in-situ samples are lacking. This paper proposes a technique to model the SSP from the perspective of its physical mechanism. Based on the ocean normal mode (ONM), which represents the dynamic characteristics of fluid particles, the SSP is represented by a basis derived from the stratification characteristics of the water column. This study shows that the physical model can better explain the dynamic activity of the ocean, and mitigate to a greater extent, the reliance on samples than the statistical model.

The ONM refers to the dynamic mode that describes the vertical velocity of fluid particles. According to Gill's definition (Gill, 1982), numerous normal modes can be obtained through the motion equation of a stratified fluid. Mode zero is the barotropic mode, which is unrelated to the depth and is a response to fluctuations on the sea surface. Baroclinic modes, which start with mode one, originate from fluctuations in the density interface. These modes are vital for describing the dynamic process in the ocean interior, and are defined as the ONM. The ONM has been used in oceanography to analyze the process of transformation of marine energy, and to explain the dynamic activity of the ocean interior at various scales, even linking it to variations in the climate (Liu, 1999; Zhang and Liu, 1999; Moon et al., 2004; Qiu et al., 2007). Therefore, a derived basis for SSPs based on the ONM has the potential to explain the dynamic mechanism of the ocean environment. The ONM is obtained from the stratification characteristics of the water column, and reduces the dependence on samples for the basis acquisition.

This paper proposes a sound speed profile model based on the ONM. The SSPs are reconstructed by using the bases derived from in-situ data and archival date, and the results are evaluated against EOF-based methods. In Section 2, theories related to the SSP model are presented. In Section 3, SSP reconstruction is carried out based on the in-situ data, and the dynamic processes of the ocean are analyzed according to the bases and projection coefficients. Section 4 contains a description of the calculation of the basis using climatological mean data, following by SSP inversion. Finally, Section 5 offers the conclusions of the paper.

2 Theory

To provide a compact presentation, the SSP model is usually expressed as

$$c(z) = c_0(z) + \sum_{n=1}^{M} a_n \psi_n(z), \tag{1}$$

where $c(z)$ is the reconstruction of the SSP model, z represents the discrete point in depth, and $c_0(z)$ is the invariant component of the SSP. The corresponding perturbation component is approximated by the superposition of M orders of oscillation pattern $\psi_n(z)$, the amplitude of which is weighted by the corresponding coefficient a_n. There are different statements for Equation (1) in different studies, and this paper claims that the term "basis" refers to the vertical oscillation pattern and "projection coefficient" refers to the weight coefficient.

For EOF technology, SSP modeling is implemented by extracting the principal component of a sufficient number of samples. As SSP samples subtract the average value of the profiles, the anomaly vectors X can be obtained. The result is a $p \times q$ matrix, where p and q represent discrete points of depth and sample size, respectively. The singular value decomposition (SVD) is

$$(R - \lambda I)K = 0, \tag{2}$$

where R is the covariance matrix of X, λ is the eigenvalue of R, I is an identity matrix, and $K = [k_1, k_2, \cdots, k_q]$ is a $p \times q$ matrix with EOF columns. The leading perturbation feature k_n in Equation (2), which corresponds larger eigenvalue and describes more total variance of samples, is selected as the basis in Equation (1), and the corresponding projection coefficient is usually obtained as a result of inversion.

To elaborate the dynamic mechanism and the method used to determine the physical basis, the derivation begins from the equation of a continuously stratified fluid. The motion equation of fluid particles in the vertical direction is:

$$\frac{1}{\rho_0}\frac{d}{dz}\left(\rho_0 \frac{dW}{dz}\right) + \frac{N^2 W}{\kappa^2} = 0, \tag{3}$$

where W is the velocity, ρ_0 is the density, N is the buoyancy frequency, and κ is the phase speed along the horizontal direction.

After introducing the Boussinesq approximation (i.e.,ρ_0 varies more slowly than W), Equation (3) can be simplified to the simpler Sturm–Liouville problem:

$$\frac{dW^2}{dz^2} + \frac{N^2 W}{\kappa^2} = 0. \tag{4}$$

Based on the boundary condition $W = 0$ on the sea surface and the seabed for Equation (4), the movement of the fluid particles can be expressed by the superposition of many ONMs:

$$W(z) = \sum_{n=1}^{M} B_n \phi_n(z) e^{i(k_n x + l_n y - \sigma t)}, \tag{5}$$

where $\phi_n(z)$ is the $n-th$ order ONM with amplitude B_n, x and y are the eastern and northern directions along the horizontal, respectively, k_n and l_n are their corresponding wavenumbers, and σ is the frequency in Equation (5). The displacement ξ of fluid particles in time t causes a variation in sound speed (Munk and Zachariasen, 1976):

$$\Delta c = \xi \frac{dc}{dz}. \tag{6}$$

The sound speed is a function of temperature T, salinity S, and depth (Kim et al., 2015):

$$\frac{dc(T,S,z)}{dz} = \beta T_z + \gamma S_z + 0.016, \tag{7}$$

where

$$\beta = 4.95 - 0.11T + 0.00087T^2 - 0.01S$$

$$\gamma = 1.34 - 0.01T$$

$$\frac{\partial T}{\partial z} = T_z, \frac{\partial S}{\partial z} = S_z .$$

The change in sound speed with temperature, salinity, and depth is approximately linear in Equation (7). For a fixed depth, once the temperature rises by 1°C, the sound speed increases by about 4 m/s, while for a 1 psu increase in salinity, the corresponding increase in sound speed is only 1.1 m/s. Considering that the interval of changes in salinity in most seas across the world is much smaller than that in temperature, the latter is often a crucial factor influencing sound speed. Thus, only the temperature, i.e., $dc/dz \approx \beta T_z$ is considered. The variations in profiles of the sound speed and temperature are often consistent, and the approximation has been proved to be reasonable in previous applications (Song et al., 2014). In a follow-up study, the approximation was shown to ensure enough precision in most seas. Equation (6) can then be expressed as $\Delta c = \beta T_z \xi$. The vertical profile of the ocean has remarkable time-variant characteristics that can be decomposed into a steady background profile and perturbation. The background profile is often stable at a large time scale, and disturbance is caused mainly by the short-term meso- and micro-scale dynamic activities, and do not change properties of the background profile. Therefore, SSP $c(t)$ can be expressed as $c_0 + \Delta c$, where c_0 is the background SSP corresponding to fluid particles in equilibrium $\xi = 0, \Delta c$ is the perturbation of the SSP caused by the integration of fluid particle motion. Then, Equation (6) can be expressed as:

$$c(t) = c_0 - \beta \frac{\partial T}{\partial z} \sum_{n=1}^{M} \frac{B_n \phi_n}{i\sigma} e^{i(k_n x + l_n y - \sigma t)} . \tag{8}$$

The reconstitution Equation (8) of the sound speed model corresponding to Equation (1) is:

$$\psi_n(z) = \frac{\partial T}{\partial z} \phi_n, \tag{9}$$

$$a_n = -\beta \frac{B_n}{i\sigma} e^{i(k_n x + l_n y - \sigma t)} . \tag{10}$$

Equation (9) is the basis derived from the ONM reflecting the perturbation in the SSP caused by the dynamic activity of the water column, that is, the ocean mode basis (OMB). In Equation (10), the real component is the projection coefficient of the corresponding OMB. In ocean observation, it usually serves as the result of inversion extracted from the acoustic signal.

It is clear from Equation (9) that the OMB is the basis of the physical modes. After SSP inversion, the results can be translated into the ONMs and their corresponding amplitudes by Equations (9) and (10), and such dynamic parameters as pressure, flow velocity, Ursell number, and Ostrovsky coefficient can then be discussed (Farmer et al., 2009; Yang et al., 2009). In addition, the vertical mode is an effective tool to explain the interior fluctuations in the ocean. From the vertical motion of the fluid particle, the OMB can explain dynamic ocean phenomena.

It is also clear that the OMB does not require a large number of samples. The OMB is deduced from the stratification characteristics of the water column—that is, the buoyancy frequency of the background profile. On the one hand, the stratification information can be extracted from the in-situ measurement. On the other hand, they can be obtained from historical data, and even from ocean numerical models. When using background profile from archival data or ocean numerical models for OMB calculation, in-site samples are not required. In comparison with statistical model, the requirements related to sample size are laxer. In the following sections, the acquisition and application of the OMB using in-situ data and archival data are detailed. Considering the current understanding of varied statistical basis functions and their effectiveness in inverse problem applications, EOF is selected as the counterpart in our paper for subsequent research.

3 OMB application based on *in situ* data

3.1 Datasets

The SSP data were calculated from two sets of temperature data, collected in August 2011 in the Yellow Sea and May 2001 in the South China Sea. The mooring locations are shown in Figure 1.

The SSP data were calculated from two sets of temperature data, collected in August 2011 in the Yellow Sea and May 2001 in the

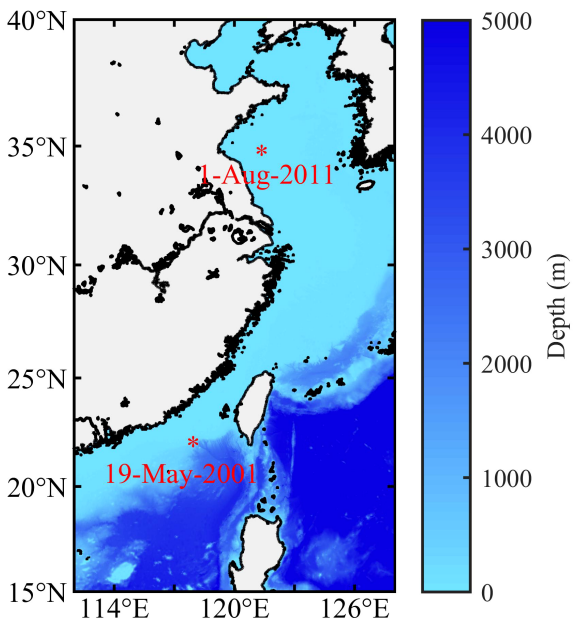


FIGURE 1
Locator map of the thermistor string. The asterisks show the anchor positions in the Yellow Sea and the South China Sea.

South China Sea. The SSP data for the two experiments are shown in Figure 2. In August 2011, a thermistor string was anchored offshore Tsingtao (35.66°N 121.00°E) in the Yellow Sea, where this was a semi-closed continental shelf and shallow sea. The depth was 40 m. The thermistor string was composed of 19 units, located at depths ranging from 8.9 to 31.6 m. The sampling interval was 0.5 min. In this area, cold water mass was the characteristic. Owing to intense radiation on the sea surface in the summer, the sea surface temperature rises to the annual maximum. Under the impact of these two factors, the span of mean temperature measured by the thermistor string was about 16°C, forming a strong thermocline. At the same time, The strong linear internal wave activity was noted. Under the influence of local circulation, with the measuring points located in the low-salt zone along the coast, the salinity of the water column stabilized at about 31 psu. The thermistor string, located near the continental slope (21.55°N, 117.35°E) in May 2001, was a part the measurement of the Asia Sea International Acoustic Experiment (ASIAEX). The depth was 139 m, the thermistor string was composed of 13 units ranging in depth from 21.7 to 135.1 m, and the sampling interval was 1 min. As the measuring location was in the tropics, the water temperature was higher than that in the Yellow Sea. Due to a combination of the bottom slope and transbasin waves, strong internal tidal and internal solitary wave were formed. Moreover, the variety in the amplitudes of the sound speed contour reached up to 70 m owing to internal solitary waves. The salinity was stylized at approximately 34 psu.

Because diurnal tidals were dominant in both areas, the cycle of a diurnal tide (24 h) was set as the time window to test using SSP models. The averaging method was applied to continuously measured in-situ data to extract the background profile. The salinity values were taken from CTD measurements near anchor positions of the thermistor string. As the variation in the salinity of multiple groups of CTD measurements was small, the salinity profile closest to the thermistor string was used. The buoyancy frequency and the ONM were calculated using the mean temperature profiles from the thermistor string and the salinity profile from the CTD. The OMB was then obtained by Equation (9). The mean temperature profiles of the thermistor did not cover the depth of the mixing layer, which might have led to extrapolation errors in the background profile under the sea surface. However, the

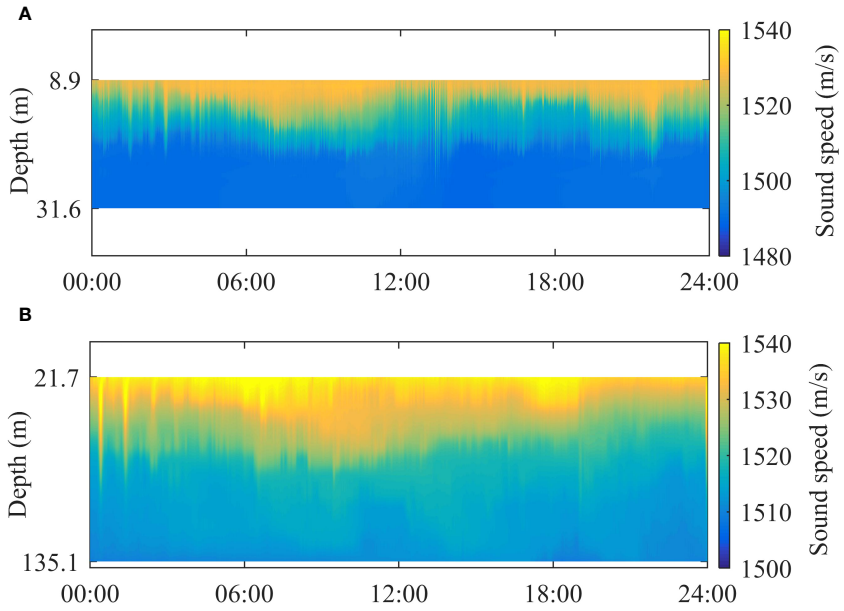


FIGURE 2
Variation in the SSP over 24 hours calculated from the thermistor string. **(A)** Yellow Sea in 2011. **(B)** South China Sea in 2001.

CTD measurements in two experiments indicated that the mixing layer under the sea surface was thin, and linear extrapolation errors were thus acceptable. The OMN for mixing layers of different depths were also simulated, and the results showed that even in the presence of a thick mixing layer, the ONM values were nearly identical to those with a thin layer. Therefore, linear extrapolation was used on the depth data obtained from the thermistor string. In the subsequent analysis, the range of depth refers to the measured range of the thermistor string at a resolution of 0.1 m.

3.2 SSP reconstruction compared with EOF

The OMB and the EOF were calculated using the background profile and the SSP samples, respectively, as shown in Figure 3. The first three order on the Yellow Sea had good consistency while the first two orders for the South China Sea were similar. As the order increased, the difference between the bases increased. In general, the first three orders described a large part of the total variance. This suggests that the main statistical characteristics of SSP perturbation as determined using the EOF were consistent with the motion laws of the fluid particles determined using the OMB. Due to the impact of circulation and the geometric, dynamic activities in the South China Sea are more complex than in the Yellow Sea. This might have led to more factors contributing to the statistical perturbation component compared with that using the OMB. There were certain differences in the higher-order models, but did this not have a significant impact on the reconstruction owing to their smaller weights than in the first two orders. Some vertical perturbation features of the ONM were retained to determine the distribution of the OMB. For example, the OMB values showed an increasing number of changes in sign as the number of modes increased, and the maximum first-order amplitude appeared in the depth interval of the maximum change in sound speed.

Table 1 lists the cumulative proportions of variance for the different orders. It shows small differences in the proportions of the leading orders between the OMB and the EOF for each set of experimental data. The proportion of the EOF was slightly higher than that of the OMB. In case the number of samples is sufficient, the random and fine disturbances caused by some factors unrelated to the ONM can be better embodied by the EOFs. For both the EOF and the OMB, the first five orders of reconstruction exceeded 95% of the commonly used threshold, which means the resolution of the SSP reconstruction of the OMB was close to that of the EOF with a sufficient number of samples.

To analyze the reconstructions of the OMB and the EOF, the mean reconstruction error *ME* of a set of *p* discrete points and *q*

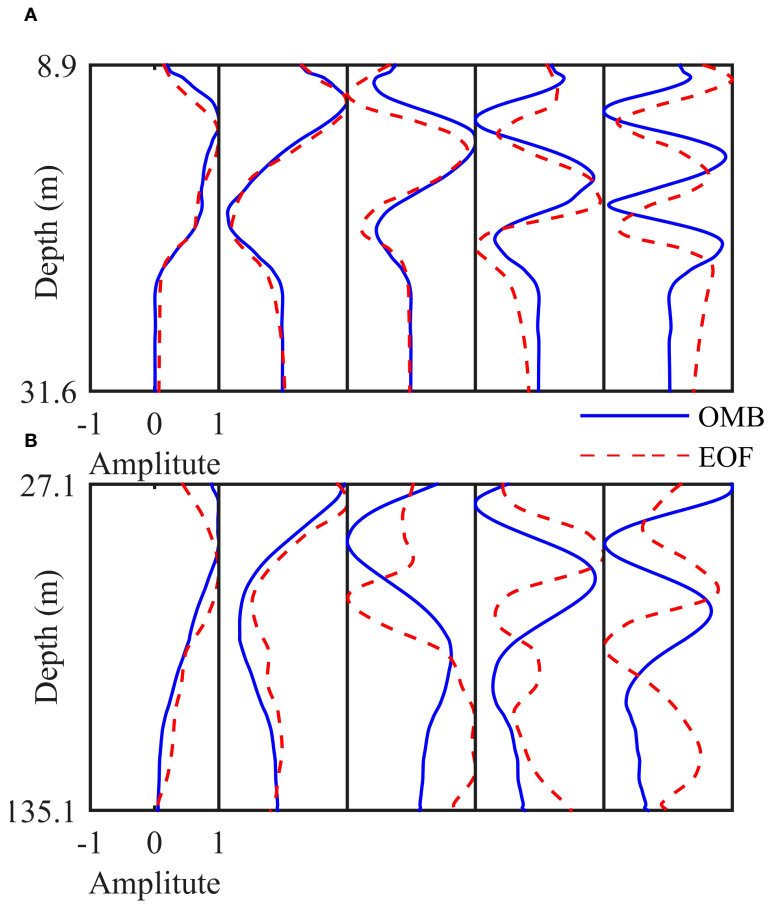


FIGURE 3
Comparison of the first five orders between the OMB and the EOF. **(A)** Yellow Sea in 2011. **(B)** South China Sea in 2001: from left to right are, in order, the first to the fifth orders.

TABLE 1 Cumulative proportions of variance for different orders.

	1 order	2 orders	3 orders	4 orders	5 orders
OMB (2011)	78.6%	88.3%	93.5%	95.1%	96.5%
EOF (2011)	81.6%	90.6%	94.5%	97.6%	98.4%
OMB (2001)	80.5%	90.7%	92.4%	94.7%	96.6%
EOF (2001)	84.6%	92.3%	96.2%	97.5%	98.6%

samples is defined as follows:

$$ME = \frac{1}{q}\sum_{j=1}^{q}\sqrt{\frac{1}{p}\sum_{i=1}^{p}(c_j(z_i) - c_j'(z_i))}, \tag{11}$$

where c_j (z_i) represents the SSP of the $j-th$ sample at the $i-th$ depth, and the apostrophe refers to the reconstruction value. The values of *ME* in Equation (11) of different orders are shown in Table 2. The results correspond to the results in Table 1. The EOF, with a sufficient number of samples, yielded a smaller error than the OMB. However, there was little difference in the error between them. For both seas, the OMB provided reasonable results. To illustrate the representational resolution of the OMB, the SSPs were reconstructed using the five orders of the OMB and EOF, and compared with the measured values in Figure 4. The SSP could be reconstructed accurately for the two sets of data using five orders bases. In particular for data from the South China Sea, the precision of reconstruction could still be guaranteed by the OMB even in case of significant anomalies caused by internal solitary waves. An analysis of results with large errors revealed that samples with large reconstruction errors in the two reconstruction methods were nearly identical. SSP perturbation caused by turbulence or water mass was difficult to express using the OMB, and could not constitute the principal component of perturbation (EOF). When the cumulative proportions of variance is low, their reconstruction accuracy will be affected.

3.3 Ocean dynamic analysis based on OMB

As a physical basis, the most appealing feature of the OMB is that the coefficient obtained from inversion can directly explain the ocean environment. In this section, dynamic changes in the sound speed contours (isotherm) and internal solitary waves are analyzed.

According to oceanographic analysis, different baroclinic modes correspond to different dynamic processes. The relative proportion of the ONM in an area of the sea reflects the leading dynamic activity. As it is the leading mode, it is important for the adjustment of ocean dynamics for the first two baroclinic modes. According to work by Liu (Liu, 1999; Zhang and Liu, 1999), the fluctuation resulting from Ekman pumping is mainly reflected in the first baroclinic mode, that is, changes in the depth of the thermocline. The second baroclinic mode occurs mainly due to the anomaly originating in fluctuation in buoyancy, and is manifested as a variation in the thickness of the thermocline. For the first-order ONM, the symbols were consistent in the water column, showing that the fluid particles moved in the same direction at different depths. The second-order ONM manifested as a variation in the inverse symbols of the upper and lower boundaries of the thermocline, that is, a variation in the thickness of the thermocline. Owing to the high variance of the first two orders in SSP construction, the projection coefficient of the first-order OMB can be used to represent changes in the depth of the sound speed contours in the thermocline. Furthermore, the variation in the sound speed gradient with the depth of the thermocline can be described by the projection coefficient of the second-order OMB.

Figure 5 shows a comparison of the projection coefficients and the structural parameters of the thermocline of the SSP. Figures 5A–C show that the first-order projection coefficient was in accordance with the trend of variation in the sound speed contour with depth. Figures 5B, D show similar results. The Pearson coefficients of the structural parameter of the thermocline and the corresponding projection coefficient are shown in Table 3. They suggest that the first two coefficients were highly relevant to changes in the depth of the sound speed contour and the sound speed gradient, which could be used to monitor them. The difference between reconstruction and the measured samples was small, which confirms the precision of reconstruction.

The relationships among the OMBs can also help explain the dynamic activity of the ocean. $CR1 = da_1/dt$ is used to represent the change rate of the first-order projection coefficient, and $CR2 = da_2/dt$ is that in the second-order projection coefficient. According to the 2.5-dimensional internal wave Lamb model and experimental observations of the Strait of Messina, Casagrande formulated the

TABLE 2 Mean reconstruction error *M E* for different orders.

M E (m/s)	1 order	2 orders	3 orders	4 orders	5 orders
OMB (2011)	1.55	1.15	0.86	0.74	0.63
EOF (2011)	1.43	1.03	0.79	0.64	0.52
OMB (2001)	1.15	0.79	0.72	0.60	0.48
EOF (2001)	1.02	0.70	0.50	0.41	0.31

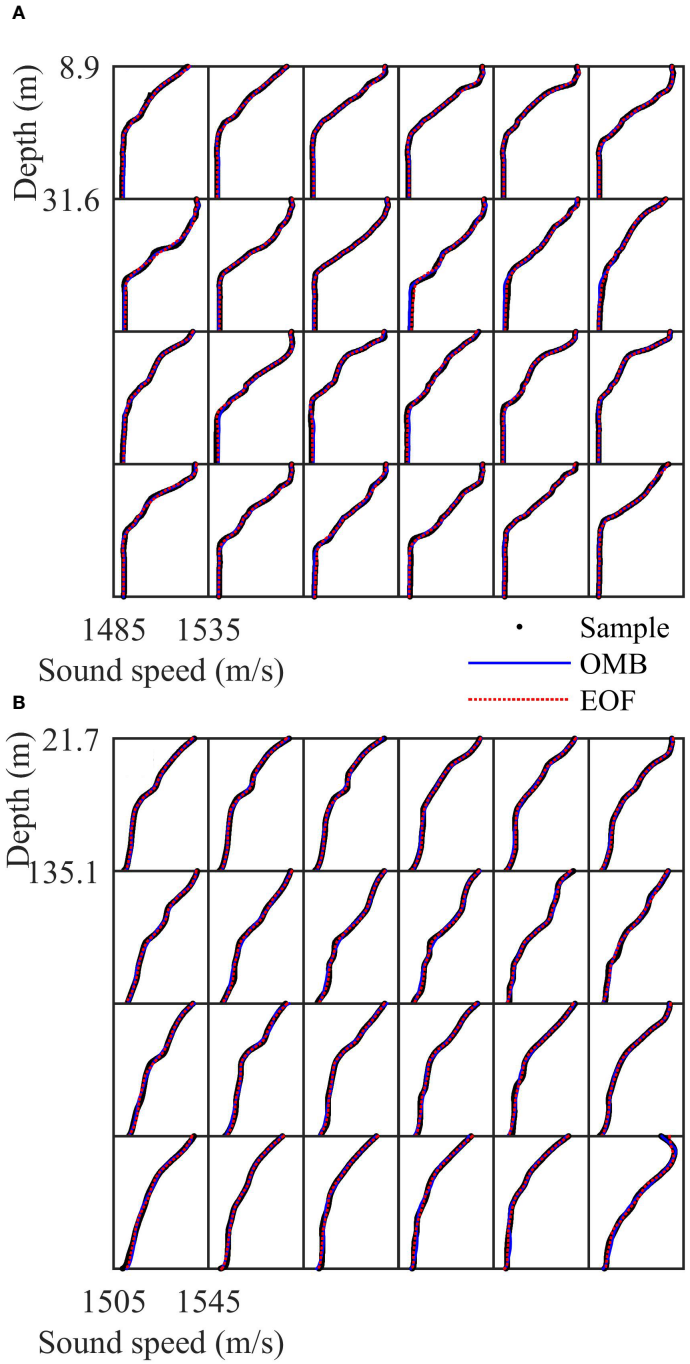


FIGURE 4
Comparison of SSP reconstruction of 24 example profiles on the hour from 01:00 hrs to 24:00 hrs. **(A)** Yellow Sea in 2011. **(B)** South China Sea in 2001.

following dynamic laws of internal solitary waves (Casagrande et al., 2009): In the first half of an internal solitary wave, the disturbance of the pycnocline at a depth is controlled by the variation of the first-order baroclinic mode, corresponding to the change in *CR*1. The deviation in the pycnocline can lead to opposite circulation in the upward and downward sides of its depth, where this mainly manifests as a variation in the second-order baroclinic mode, which corresponds to a change in *CR*2. As the first two ONMs are orthogonal, the signs of *CR*1 and *CR*2 are opposite to each other. At the end of first half of the wave, the variation in the pycnocline reaches its peaks, and the corresponding values of *CR*1 and *CR*2 are zero. In the first half of an internal solitary wave, *CR*1 and *CR*2 undergo an irregular-arch change and have opposite signs. In the second half of the wave, *CR*1 and *CR*2 undergo inverse processes to those in the first half of the wave. Thus, the variations in *CR*1 and *CR*2 are similar during an internal solitary wave. Their zero points are consistent but their signs are opposite, which indicates a double oscillation pattern.

10.3389/fmars.2024.1378396

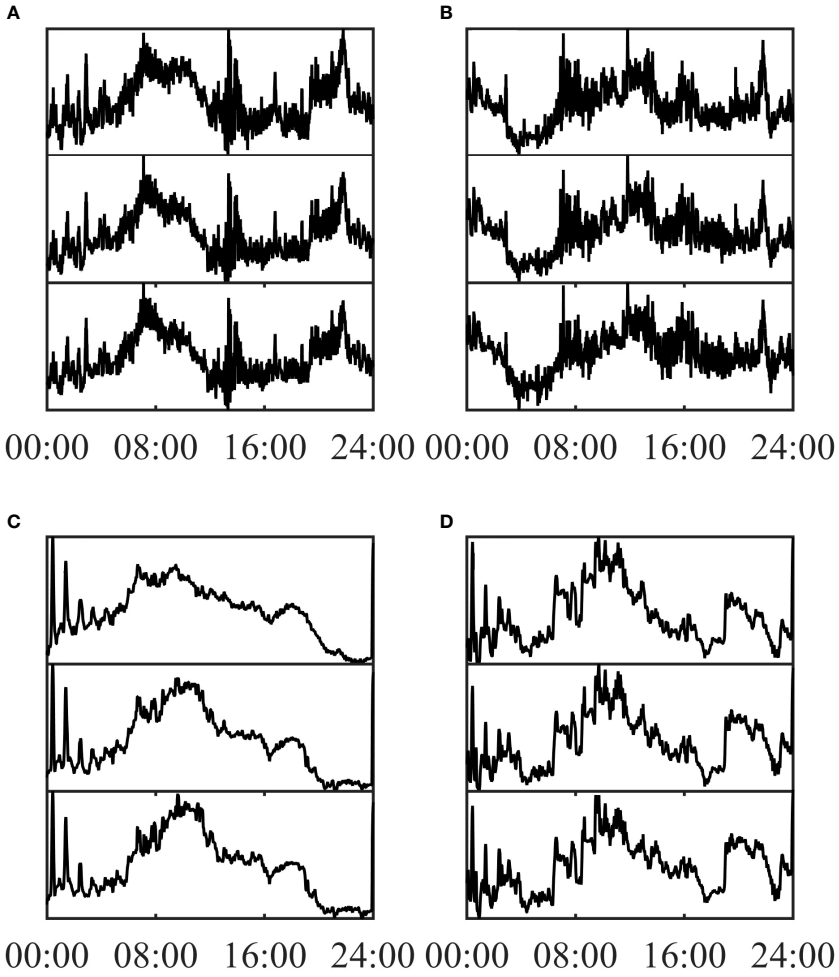


FIGURE 5
Comparison of the structural parameters of the thermocline and projection coefficients. **(A)** Depth of the sound speed contour (1518 m/s) and the first-order projection coefficient in the Yellow Sea in 2011. **(B)** Sound speed gradient at depths ranging from 11.4 m to 19.2 m, and the second-order projection coefficient in the Yellow Sea in 2011. **(C)** Depth of the sound speed contour (1528 m/s), and the first-order projection coefficient in the South China Sea in 2001. **(D)** Sound speed gradient at depths ranging from 24 m to 78.1 m, and the second-order projection coefficient in the South China Sea in 2001. From top to bottom, the plots show the corresponding projection coefficient, the structural parameters of the thermocline of the reconstructed SSP, and those of the sample SSP, respectively. Min–max normalization was carried out.

Figure 6 shows an example of the analysis of the double oscillation pattern. The SSPs had prominent internal solitary wave trains. In the first two internal solitary waves, the *CR*1 and *CR*2 exhibited clear double oscillation patterns with a large amplitude. Although the third wave was smaller, its start and end times were identified, and it too was determined to have a double oscillation pattern. The fourth wave was visually identical but its dynamic characteristics were significantly different, and it did not have a double oscillation pattern. This analysis of the OMB confirmed waves of the internal solitary train. Furthermore, the amplitudes and wavelength characteristics could also be estimated. Double oscillation was also observed when analyzing the density EOF but not in the velocity EOF (Vázqueze et al., 2006; Casagrande et al., 2010). By contrast, the connection of the OMB to physical

TABLE 3 Pearson correlation coefficient of the projection coefficient (PC) and the structural parameters of the thermocline.

	2011		2001	
	1st PC	2st PC	1st PC	2st PC
Depth of sound isospeed line				
(Sample/reconstruction)	0.95/0.96	\	0.93/0.95	\
Sound speed gradient				
(Sample/reconstruction)	\	0.95/0.96	\	0.93/0.97

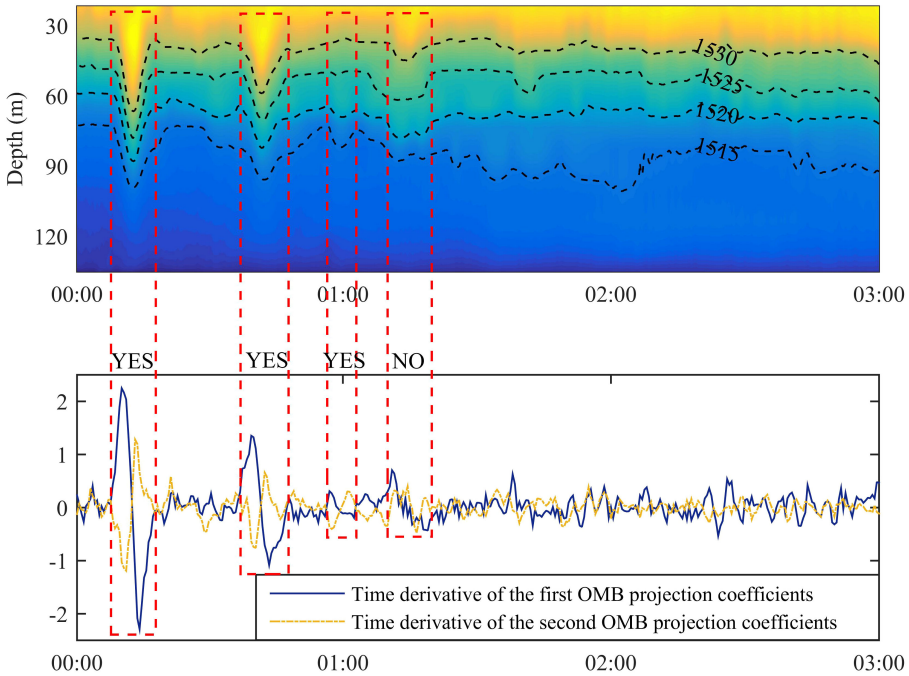


FIGURE 6
Internal solitary wave from 00:00 hrs to 03:00 hrs on May 19, 2001 in the South China Sea. The figure on top shows the SSPs and that at the bottom shows the corresponding *CR*1 and *CR*2. During an internal solitary wave, *CR*1 increased first and then returned to zero at the maximum wave amplitude. As the wave amplitude decreased, *CR*1 increased with the opposite sign, and then decreased to zero. The change in the amplitude of *CR*2 was similar to that of *CR*1 but opposite in sign. The patterns of changes in *CR*1 and *CR*2 are defined as double oscillation.

properties was closer. Based on the OMB and its coefficient, the dynamic activity of the ocean could be analyzed using a combination of baroclinic modes, which is useful for interpreting the results of inversion.

4 Application of OMB based on archival data

Another attractive feature of the OMB is its lax requirements of sample size. Theoretically, a representative background profile is sufficient to calculated the OMB. With the development of Argo, underwater gliders, and other measurement methods, a large number of samples of the global profile have been accumulated. In combination with ocean numerical models, many database products have been developed to provide the statistical mean and objectively analyzed mean temperature–salinity profile. In this section, the effectiveness of the OMB extracted from archival data is tested in case of the absence of in-situ data. The background profile was extracted from climatological filed data of the World Ocean Atlas 2013 (WOA13), published by the National Oceanographic Data Center (NODC, https://www.nodc.noaa.gov/OC5/woa13), and data on acoustic propagation were from ASIAEX2001 in the East China Sea.

The acoustic propagation experiment of ASIAEX2001 in the East China Sea was carried out at a depth of 105 m. A total of 32 vertical arrays were hung over a receiving ship. On the course away from the receiving ship, a launching ship released the broadband explosive sources of 38 g TNT with a rated depth of 50 m. CTD measurements were carried out many times in this experiment, and the SSPs are shown in Figure 7. Some SSPs around a depth of 60 m were low, possibly because of the cold water mass. The invariant component of the SSPs was calculated from the mean profile of all measured CTD values. As the CTD survey was not conducted during the acoustic propagation experiment, the effectiveness of the OMB was evaluated through matched field tomography, obtained by the EOF extracted from CTD samples and the OMB extracted from the archival data.

Considering that there were clear seasonal characteristics in the oceanic background profile, the objectively analyzed mean summer profile (1955–2012), recorded by WOA13 at a spatial resolution of 0.25°, was used. The climatological profiles of the experimental areas are shown in Figure 8. The thermocline in summer covered almost the entire water column, and salinity changed by little. At a depth of 20–70 m, the buoyant frequency was high. This corresponded to the thermocline in CTD measurements, reflecting seasonal background characteristics.

Based on the results of the reconstruction test, three orders with the highest reconstruction accuracy was adopted for the SSP inversion. A comparison of the first three orders of the basis between the OMB calculated by archival data and the EOF calculated by CTD samples is shown in Figure 9. A certain similarity was noted in the distribution between them, but large differences were also noted in terms of the depth of the extrema and fine structure. This indicates that the seasonal background can embody the macroscopic dynamic characteristics. However, there

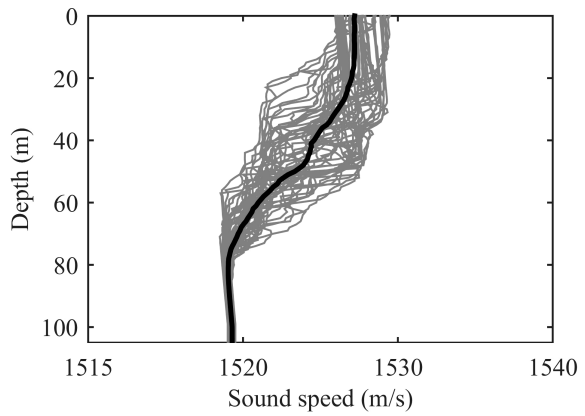


FIGURE 7
SSPs measured by the CTD. The dark line represents the mean SSP, and the other lines show the samples.

was a large difference between the mean climate and in-situ measurements. The key focus of the OMB application was to determine whether the resolution was sufficient to describe the SSP based on the OMB extracted from historical data products.

SSP inversion was carried out using conventional matched field processing. All environmental parameters except the projection coefficient were set as known quantities. Based on the broadband Bartlett processor, 35 frequency points in the frequency band 99–201 Hz were processed in the inversion. The optimal values of the first three orders of vector quantity a were searched using the genetic algorithm in the optimization space to implement the minimum cost function $E(a)$:

$$E(a) = \frac{1}{L}\sum_{l=1}^{L}\frac{\left|\sum_{n=1}^{N} p_{nl}^{e}(a)p_{nl}^{c}(a)^{*}\right|}{\sum_{n=1}^{N}\left|p_{nl}^{e}(a)\right|^{2}]\left[\sum_{n=1}^{N}\left|p_{nl}^{c}(a)\right|^{2}\right]}, \tag{12}$$

where L is the number of frequency points, N is the number of hydrophones, $p_{nl}^{e}(a)$ and $p_{nl}^{c}(a)$ are the measured sound pressure

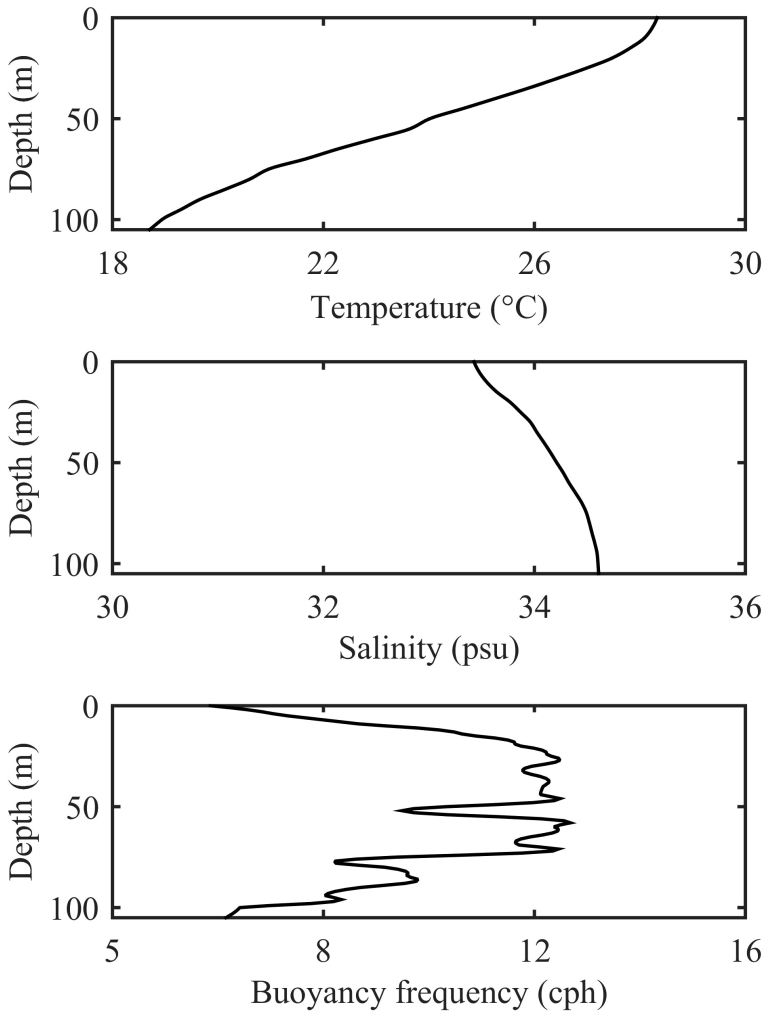


FIGURE 8
The background profiles extracted from WOA13. Buoyancy frequency was calculated based on temperature and salinity.

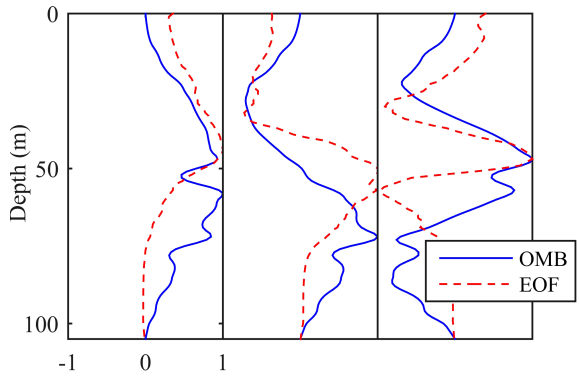


FIGURE 9
Comparison of the first three orders between OMB and EOF.

and the sound pressure of the replica field at the $a-th$ frequency point, respectively, when the projection coefficient was a. The asterisk in Equation (12) indicates the conjugation. According to the experimental results, a density of 1.86 g/cm^3 was set on the half space seabed. The sound speed and attenuation coefficient were 1610 m/s and 0.15 dB/λ, respectively.

Figure 10 shows the inversion results for two explosive sources 10.2 km from the receiver. It is clear that the inversion results of the OMB and the EOF were similar. The absolute errors of the two inversion results are 0.15 m/s and 0.22 m/s, respectively. The root mean square error of the two inversion results are 0.21 m/s and 0.27 m/s, respectively. This indicates that the OMB can guarantee a similar resolution to that of the EOF. In general, features of the seasonal stratification were stable, and the principal characteristics

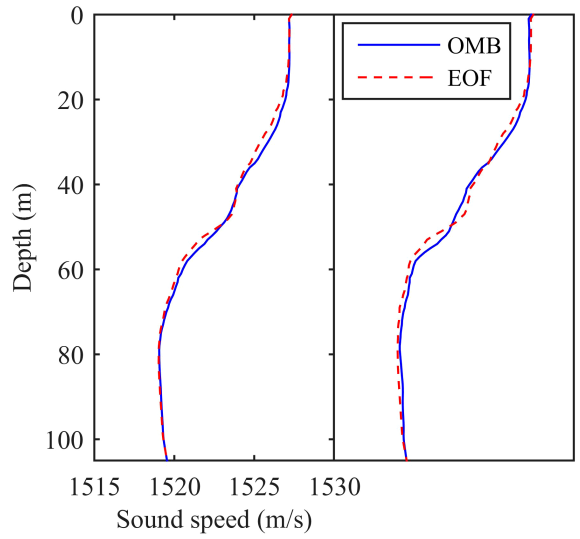


FIGURE 10
Two inversion results at 10.2 km. Due to the uncertainty of source depth, acoustic focalization was carried out before the SSP inversion. The actual source depth was 48 m and 49 m, respectively. The agreement between OMB and EOF is better than that in the former inversion using the rated depth of 50 m (Qu et al., 2019). Both methods have resolution to reflect the relief of source depth mismatch.

of the ocean dynamics were controlled by it. This shows that the effects of the other factors of the sea area were small. The application of the OMB without in-situ data had different effects at different times and in different spaces. In the sea area where the dynamic activities were controlled mainly by the baroclinic mode, for which reliable historical data were available, the OMB should have reasonable precision. However, in areas with complex dynamic activities, such as those influenced by air–sea interaction and water mass, and where perturbations in the sound speed are short term, in-situ measurements are indispensable to obtain an accurate background profile. Drastic perturbations at a small range of depth occurring randomly are difficult to express, whether by the physical or the statistical SSP model.

As currently available ocean data products can provide profile-related information worldwide, OMB can be theoretically calculated for applications without any in-situ measurement, where the statistical model cannot do this. However, when features of the water column stratification obtained from archival data are not consistent with the empirical situation, in-situ samples are needed for accurate SSP modeling. A compromise is to measure the profile during the slack tide to obtain stratification-related information for the OMB. This not only reduces the sample size needed, but also ensures the real-time determination of the stratification features.

5 Conclusions

The ONM is an important means of explaining the kinetic energy and heat transfer of the water column, and can be used to describe the structure of the profile. A typical example is the analysis of internal waves,

where Ursell number and Ostrovsky number calculated by ONM can effectively explain the motion and transport status of water column (Farmer et al., 2009; Yang et al., 2009). Based on this physical mode, an SSP model was proposed and tested in this paper. In contrast to the statistical model applied widely in research, the SSP perturbation described by the OMB is based on the motion law of fluid particles.

The OMB can explain the ocean environment directly through the ONM. In tests involving SSP reconstruction in the Yellow Sea and the South China Sea, the OMB yielded reasonable precision. In combination with the barotropic mode, it made possible analyses of the thermocline structures and internal solitary waves using the first two orders of the projection coefficient.

Another feature of the OMB is that it has no rigid requirement on sample size compared with the statistical model. The regional OMB was extracted through the objectively analyzed mean profile of WOA13. SSP inversion was carried out for data from ASIAEX2001 on the East China Sea, and the results of SSP inversion were similar to those of the EOF. Although the global OMB could be obtained using archival data without in-situ measurements, precision was difficult to guarantee due to the inconsistency between the mean stratification features obtained through historical data and those determined in real time. A more feasible method is to obtain the real-time background

profile by few measurements. The archival data can then serve for scenarios where measurements are unavailable.

The improvement to the SSP model effected by the OMB is in providing a bridge to directly link the dynamic mechanism of the ocean with the perturbation of SSP. By using few leading orders of basis vectors and projection coefficients, it can provide a compact representation of SSPs for inversion and the determination of ocean dynamics. It presents an alternative to the statistical model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KQ: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. WY: Project administration, Validation, Visualization, Writing – original draft. FZ: Conceptualization, Methodology, Writing – review & editing. LM: Project administration, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Natural Science Foundation of Guangdong Province grant number [2022A1515011519].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Behera, S. K., Rao, S. K., and Saji, H. N. (2003). Comments on "A cautionary note on the interpretation of EOFs". *J. Clim.* 16 (17), 1087–1093. doi: 10.1175/1520-0442(2003)016<1087:COACNO>2.0.CO;2

Behringer, D., Birdsall, T., Brown, M., Cornuelle, B., Heinmiller, R., Knox, R., et al. (1982). A demonstration of ocean acoustic tomography. *Nature* 299, 121–125. doi: 10.1038/299121a0

Bianco, M., and Gerstoft, P. (2017). Dictionary learning of sound speed profiles. *J. Acoust. Soc Am.* 141, 1749–1758. doi: 10.1121/1.4977926

Casagrande, G., Varnas, A. W., Folegot, T., and Stephan, Y. (2010). An original method for characterizing internal waves. *Ocean Model.* 31, 1–8. doi: 10.1016/j.ocemod.2009.07.007

Casagrande, G., Varnas, A. W., Stéphan, Y., and Folégot, T. (2009). Genesis of the coupling of internal wave modes in the strait of messina. *J. Mar. Syst.* 78, S191–S204. doi: 10.1016/j.jmarsys.2009.01.017

Chen, C., Ma, Y., and liu, Y. (2018). Reconstructing sound speed profiles worldwide with sea surface data. *Appl. Ocean Res.* 77, 26–33. doi: 10.1016/j.apor.2018.05.002

Cheng, L., Cheng, X., Zhao, H., Li, J., and Xu, W. (2022). Tensor-based basis function learning for three-dimensional sound speed fields. *J. Acoust. Soc Am.* 151, 269–285. doi: 10.1121/10.0009280

Dommenget, D., and Latif, M. (2002). A cautionary note on the interpretation of eofs. *J. Clim.* 15, 261–225. doi: 10.1175/1520-0442(2002)015⟨0216:ACNOTI⟩2.0.CO;2

Farmer, D., Li, Q., and Park, J. H. (2009). Internal wave observations in the south China sea: The role of rotation and non-linearity. *Atmos.-Ocean* 47, 267–280. doi: 10.3137/OC313.2009

Gill, A. E. (1982). *Atmosphere-ocean dynamic* (New York: Academic Press).

Hjelmervik, K. T., Jensen, J. K., Østenstad, P., and Ommundsen, A. (2012). Classification of acoustically stable areas using empirical orthogonal functions. *Ocean Dyn.* 62, 253–264. doi: 10.1007/s10236-011-0499-z

Jiang, Y., and Chapman, N. R. (2009). The impact of ocean sound speed variability on the uncertainty of geoacoustic parameter estimates. *J. Acoust. Soc Am.* 125, 2881–2895. doi: 10.1121/1.3097770

Kim, J., Kim, H., Paeng, D. G., Bok, T. H., and Lee, J. (2015). Low-salinity-induced surface sound channel in the western sea of jeju island during summe. *J. Acoust. Soc Am.* 137, 1576–1585. doi: 10.1121/1

LeBlanc, L. R., and Middleton, F. H. (1980). An underwater acoustic sound velocity data model. *J. Acoust. Soc Am.* 67, 2055–2062. doi: 10.1121/1.384448

Li, Z., He, L., Zhang, R., Li, F., Yu, Y., and Lin, P. (2015). Sound speed profile inversion using a horizontal line array in shallow water. *Sci. China Phys. Mech. Astron.* 58, 014301. doi: 10.1007/s11433-014-5526-x

Liu, Z. (1999). Forced planetary wave response in a thermocline gyre. *J. Phys. Oceanogr.* 29, 1036–1055. doi: 10.1175/1520-0485(1999)029⟨1036:FPWRIA⟩2.0.CO;2

Moon, B. K., Yeh, S. W., Dewitte, B., Jhun, J. G., Kirtman, B. P., and Kang, I. S. (2004). Vertical structure variability in the equatorial pacific before and after the pacific climate shift of the 1970s. *Geophys. Res. Lett.* 31, L03203. doi: 10.1029/2003GL018829

Munk, W. H., and Zachariasen, F. (1976). Sound propagation through a fluctuating stratified ocean: Theory and observation. *J. Acoust. Soc Am.* 59, 818–838. doi: 10.1121/1.380933

Qiu, B., Chen, S., and Hacker, P. (2007). Effect of mesoscale eddies on subtropical mode water variability from the kuroshio extension system study (kess). *J. Phys. Oceanogr.* 37, 982–1000. doi: 10.1175/JPO3097.1

Qu, K., Piao, S., and Zhu, F. (2019). A noval method of constructing shallow water sound speed profile based on dynamic characteristic of internal tides. *Acta Phys. Sin.* 68, 124302. doi: 10.7498/aps.68

Song, W., Hu, T., Gou, S., Ma, L., and Lu, L. (2014). A methodology to achieve the basis for the expansion of the sound speed profile. *Acta Acustica* 39, 11–18. doi: 10.15949/j.cnki.0371-0025.2014.01.017

Vázqueze, A., Stashchuk, N., Vlasenko, V., Bruno, M., Izquierdo, A., and Gallacher, P. C. (2006). Evidence of multimodal structure of the baroclinic tide in the strait of Gibraltar. *Geophys. Res. Lett.* 33, L17605. doi: 10.1029/2006GL026806

Xing, C., Dong, S., and Wan, Z. (2023). Direction-of-arrival estimation based on sparse representation of fourth-order cumulants. *IEEE Access* 11, 128736–128744. doi: 10.1109/ACCESS.2023.3332991

Xing, C., Wu, Y., Xie, L., and Zhang, D. (2021). A sparse dictionary learning-based denoising method for underwater acoustic sensors. *Appl. Acoust.* 180, 108140. doi: 10.1016/j.apacoust.2021.108140

Yang, P. (2024). An imaging algorithm for high-resolution imaging sonar system. *Multimed. Tools Appl.* 83, 31957–31973. doi: 10.1007/s11042-023-16757-0

Yang, T. C., and Liu, C. H. S. H. H. J. (2017). Frequency striations induced by moving nonlinear internal waves and applications. *IEEE J. Ocean. Eng.* 42, 663–671. doi: 10.1109/JOE.2016.2593865

Yang, Y. J., Fang, Y. C., Chang, M. H., Steven, S. R., Kao, C. C., and Tang, T. Y. (2009). Observations of second baroclinic mode internal solitary waves on the continental slope of the northern south China sea. *J. Geophys. Res.-Oceans.* 114, C10003. doi: 10.1029/2009JC005318

Zhang, X. (2024). An efficient method for the simulation of multireceiver SAS raw signal. *Multimed. Tools Appl.* 83, 37351–37368. doi: 10.1007/s11042-023-16992-5

Zhang, R., and Liu, Z. (1999). Decadal thermocline variability in the north pacific ocean: two pathwaysaround the subtropical gyre. *J. Clim.* 12, 3273–3296. doi: 10.1175/1520-0442(1999)012⟨3273:DTVITN⟩2.0.CO;2

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver sas imagery based on monostatic conversion. *IEEE J-STARS.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X., Yang, P., Wang, Y., Shen, W., Yang, J., Wang, J., et al. (2024). A novel multireceiver SAS RD processor. *IEEE T. Geosci. Remote* 62, 4203611. doi: 10.1109/TGRS.2024.3362886

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 16 May 2024
DOI 10.3389/fmars.2024.1375766

OPEN ACCESS

EDITED BY
Likun Zhang,
University of Mississippi, United States

REVIEWED BY
Ze-Nan Zhu,
Ministry of Natural Resources, China
Xiao-Hua Zhu,
Ministry of Natural Resources, China
Ryan McCarthy,
University of California, San Diego, United States

*CORRESPONDENCE
Zhenyi Ou
ouzhenyi@stu.gdou.edu.cn

RECEIVED 24 January 2024
ACCEPTED 25 April 2024
PUBLISHED 16 May 2024

CITATION
Zhao Y, Xu P, Li G, Ou Z and Qu K (2024) Reconstructing the sound speed profile of South China Sea using remote sensing data and long short-term memory neural networks.
Front. Mar. Sci. 11:1375766.
doi: 10.3389/fmars.2024.1375766

Reconstructing the sound speed profile of South China Sea using remote sensing data and long short-term memory neural networks

Yu Zhao[1], Pan Xu[1], Guangming Li[2], Zhenyi Ou[3]* and Ke Qu[3]

[1]College of Meteorology and Ocean, National University of Defense Technology, Changsha, China, [2]Innovation Institute of defense technology, Beijing, China, [3]College of Electronic and Information Engineering, Guangdong Ocean University, Zhanjiang, China

Introduction: Sound waves are refracted along the direction of their propagation owing to spatial and temporal fluctuations in the speed of sound in seawater. Errors are compounded when sound speed profiles (SSPs) with low precision are used to detect and locate distant underwater targets because an accurate SSP is critical for the identification of underwater objects based on acoustic data. Only sparse historical spatiotemporal data on the SSP of the South China Sea are available owing to political issues, its complex atmospheric system, and the unique topography of its seabed, because of which frequent oceanic movements at the mesoscale affect the accuracy of inversion of its SSP.

Method: In this study, we propose a method for the inversion of the SSP of the South China Sea based on a long short-term memory model. We use continuous-time data on the SSP of the South China Sea as well as satellite observations of the height and temperature of the sea surface to make use of the long-term and short-term memory-related capacities of the proposed model.

Result: It can achieve highly accurate results while using a small number of samples by virtue of the unique structure of its memory. Compared with the single empirical orthogonal function regression method, the inversion accuracy of this model is improved by 24.5%, and it performed exceptionally well in regions with frequent mesoscale movements.

Discussion: This enables it to effectively address the challenges posed by the sparse sample distribution and the frequent mesoscale movements of the South China Sea.

KEYWORDS

sound speed profile, remote sensing observation data, long short-term memory, sound speed disturbance, empirical orthogonal function

1 Introduction

The sound speed profile (SSP) is an important oceanic parameter that is used in a variety of marine acoustic applications, such as underwater target identification, underwater communication, and marine environmental monitoring (Teymorian et al., 2009; Xu et al., 2013; Liu L. et al., 2021; Luo et al., 2022; Su et al., 2022; Zhan et al., 2023). The speed of sound varies significantly even in adjacent areas of the sea due to the complexity and variability of the marine environment. Even though it is the largest territorial marine area in China, research on the characteristics of the South China Sea began relatively late owing to political and territorial issues. According to the most recent map of the global seabed published in 2023, only about one-third of the seabed of the South China Sea has been surveyed thus far. The speed of sound is among the most significant factors that currently limit the accuracy of detection of underwater targets. Researchers have spent a considerable amount of time and effort in reducing errors in the speed of sound and ray tracing to improve the accuracy of detection of underwater engineering (Xu et al., 2005). By denoising the signal and optimizing the algorithm, the researchers reduce the impact of low precision sound speed on underwater engineering applications (Li et al., 2022b; Li et al., 2022a; Li et al., 2022c).

Researchers have identified links between the parameters of profiles of the sea surface and subsurface, and have proposed a number of methods to satisfy the increasingly stringent demands on the precision and speed of marine data in ocean engineering (Carnes et al., 1990; Stammer, 1997; Wunsch, 1997; Liu Y. et al., 2021; Yan et al., 2022). Remote sensing technology can be used to capture near-real-time and large-scale data on the ocean surface, where this enables the rapid acquisition of SSPs in the ocean. The corresponding techniques have provided us with a better understanding of the underlying processes of deep ocean motion (Klemas and Yan, 2014). Initial research in the area used linear approaches to infer the SSPs from the parameters of remote sensing data obtained from satellites. The empirical orthogonal function (EOF) was used as the basic function in this process. It plays a critical role in limiting the dimensionality of the parameters, reducing the computational load during inversion, and filtering out minor errors during computations (LeBlanc and Middleton, 1980). Carnes discovered that the parameters of satellite remote sensing, such as the height and temperature of the sea surface, are essential for inferring the temperature profiles of water bodies (Carnes et al., 1994). This insight led to the development of the EOF-based method of inversion called the single empirical orthogonal regression function (sEOFr). Chen et al. used this approach to invert the global SSPs, and showed that the sEOFr method can be used to directly infer the SSP without converting the temperature (Chen et al., 2018). The United States Navy successfully used this method in a modular ocean data assimilation system (Rahaman et al., 2016). While these methods are effective, the relationship between the parameters of the sea are not linear, and errors are thus inevitably generated when using the linear sEOFr method to describe the physical relationship between the relevant parameters. Jain found that errors in data on the inverted SSP primarily converged at depths ranging from 40 to 125 m owing to intense oceanic movements in the South China Sea at the mesoscale. Linear methods struggle to resolve such parametric relationships (Jain and Ali, 2006).

Su et al. used machine learning-based techniques instead of linear methods to investigate the relationship between parameters of the ocean. They used classical machine learning methods and support vector regression to predict global ocean temperatures beyond 1000 m by using satellite remote sensing data (Su et al., 2015; Su et al., 2019). Machine learning methods not only have advantages over conventional techniques in inferring the temperature profiles, but also in inferring the SSP. Ou used a tree-based algorithm along with parameters of remote sensing to invert the SSP, and reported a 25% improvement in the accuracy of the outcomes (Ou et al., 2022). Furthermore, Li et al. successfully inverted the SSP of the South China Sea by using a non-linear approach based on self-organizing maps (Li et al., 2021).

Inverting the SSP by using machine learning methods in conjunction with the parameters of remote sensing remarkably improves the accuracy of the results. However, the sEOFr method as well as other currently used techniques require a large number of training samples to deliver accurate results, and deliver subpar performance in the presence of intense activity at the mesoscale. The underwater terrain of the South China Sea is characterized by a deep ocean basin surrounded by sloped land, where the southwest slopes are higher than those in the northeast. The water bodies in the central and northern basins of the sea exchange water with the Pacific Ocean via certain straits, while the southern shelf near the Equator exchanges water with the Java Sea via the Malay Peninsula and the Borneo passage. Hence, the South China Sea contains water masses with varied origins and, thus, different hydrological characteristics. The tropical oceanic climate of the region is notable for its alternating rotation of southwestern winds in the summer and northeastern winds in the winter, and this leads to the formation of a complex atmospheric system. Scant historical data on the South China Sea have been accumulated for political reasons, which makes it challenging to invert its SSP. This task is rendered more onerous owing to the complex mechanism of disturbance in the SSP caused by the atmospheric system and the unique terrain of the area.

The authors of this study propose a long short-term memory (LSTM) based algorithm to invert the SSP of the South China Sea by using the parameters of remote sensing. The linear constraints in the relation between the parameters of the surface and the ocean can be eliminated by introducing an artificial neural network. The unique memory structure of the LSTM network can be used to overcome the problem of the small number of samples as well as the complex mechanism of disturbance in the SSP in the area. We used the root mean-squared error (RMSE) and mean absolute error (MAE) to compare the proposed method with the sEORr method, and the results showed that it is more accurate, and requires a smaller number of data samples. Moreover, it delivers better performance in regions featuring greater disturbances.

2 Data

We chose the South China Sea as the location for the inversion of the SSP because it is particularly challenging in this regard owing to frequent oceanic movements in the region. We used the LSTM model in conjunction with remote sensing data to precisely invert the SSP. We used a variety of datasets, including remote sensing data to create a regression database, data from Argo to construct SSP fields under water, and WOA18 data to compute the background profiles. The data used here had been collected from 2009 to 2018.

The remote sensing dataset included information on the height and temperature of the sea surface from the L4 satellite observation product of the Copernicus project (https://resources.marine.copernicus.eu). The data had a one-day temporal resolution and a spatial resolution of 0.25°. The experiments involved computing the mean values of all the data on the height and temperature of the sea surface, and then deriving the sea surface height anomaly and the sea surface temperature anomaly from them to establish a regression database.

The Argo data were obtained from the Argo dataset on the global ocean (2009–2018), and were preprocessed to remove anomalous data while retaining data within the undistorted range of depth of 5–1000 m. The Argo data had been obtained by using Argo floats, which are capable of simultaneously measuring the temperature and salinity profiles of seawater. The SSP is a function of the temperature, salinity, and hydrostatic pressure, and can be calculated by using the empirical formula proposed by Del Grosso to determine the SSP (Del Grosso, 1974). Figure 1 shows the entire set of 3,883 samples used for this study. A segment of continuously measured data was selected to train and test the LSTM model, and was called the TEST dataset. It is represented by the black dots in Figure 1. The TEST dataset contained 269 samples that were arranged chronologically from July 9, 2014 to April 2, 2015.

The background profiles represented the stable and unchanging portion of the SSP, and are typically represented by the average values of all profiles. WOA18 data were used to calculate the background profiles in this study. These data were obtained from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (https://www.nodc.noaa.gov/OC5/woa18/), and combined multiple datasets with measurements of the temperature, salinity, density, and other climate-averaged data from various global oceanic regions. The experiments made use of annually averaged data that were obtained at a spatial resolution of 0.25° ×0.25°from 2009 to 2018. In this paper, WOA data at the center point of the inversion region (15.5° N, 145.5°E) is selected as the background profile of this experiment, and the specific profile values are shown in Figure 2.

3 Inversion of sound speed profiles

3.1 Construction of the basis function of the sound speed profile

Basis functions serve as a method of dimension reduction in the context of the problem of inverting the field of sound speed, and their accuracy has a significant influence on the precision of calculation of the field of sound speed. Figure 3 shows how to extract the basis function EOF from the historical data and obtain the corresponding disturbance coefficient. Finally, the reliability of SSP reconstructed by perturbation coefficient is verified. The shift of the ssp sample relative to the mean is called a disturbance. $SSP^{disturbance}_{Argo\ SSP}$ represents the difference between the SSP field and the background profile, and is denoted by the perturbation in the field of the sound speed.

$$COV \times EOF = EOF \times \lambda \quad (1)$$

We calculated the covariance matrix, COV, of the disturbance in the speed field and performed orthogonal decomposition by Equation 1. In this equation, EOF represents the basis functions of the SSP while λ stands for the eigenvalue matrix.

The EOF can be used to identify the primary modes of changes in water. The role of EOF is to reduce the dimensionality of the data, reducing the amount of computation while avoiding the introduction of additional noise. Figure 4 shows the amplitudes of

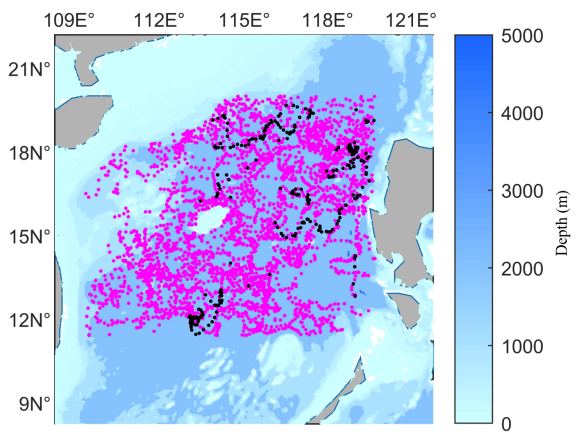


FIGURE 1
Distribution of the experimental samples (the black dots are test samples and the rest are training samples).

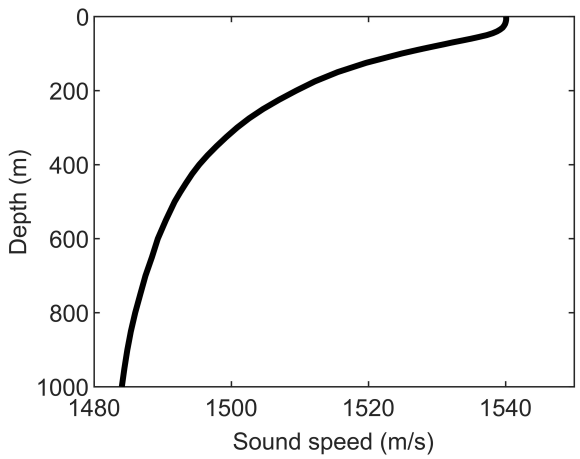


FIGURE 2
Background profile.

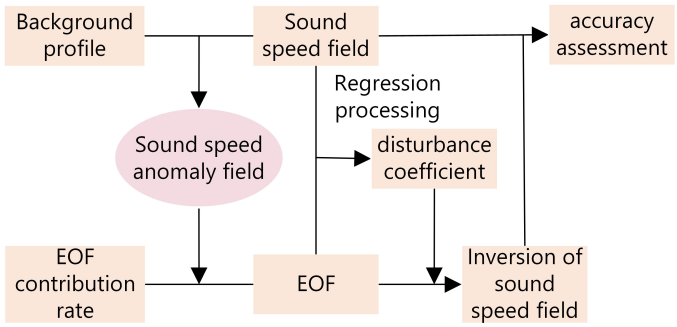


FIGURE 3
Flowchart of preprocessing of the sound speed profile.

the first 5 orders of EOF for this experiment. It shows that most of the disturbance in the water is concentrated in the depth of 100–200m, and there is basically no disturbance below 500m.It is widely assumed that a contribution of 95% can represent a majority of disturbances in water. Based on λ, the contribution rate of each mode of EOF can be calculated. Figure 5 illustrates the distribution of the contributions of the first five modes, accounting for 70.69%, 16%, 4.86%, 3.30%, and 1.58% of the total, for an overall contribution of 96.43%. We thus used the first five orders of the EOF as the basis functions for the experiments in this study.

The least square method is used to fit the EOF and the sound speed field, and the disturbance coefficient is obtained. Then the perturbation coefficient and EOF are used to calculate the sound speed field to ensure the accuracy of the perturbation coefficient and EOF. A comparison between the reconstructed values obtained from this inversion and the actual values yielded an RMSE of 0.62 m/s. Such a small error indicates that the shape functions of the EOF adequately represented a significant part of the variance in disturbances within the region, thus ensuring a relatively accurate reconstruction.

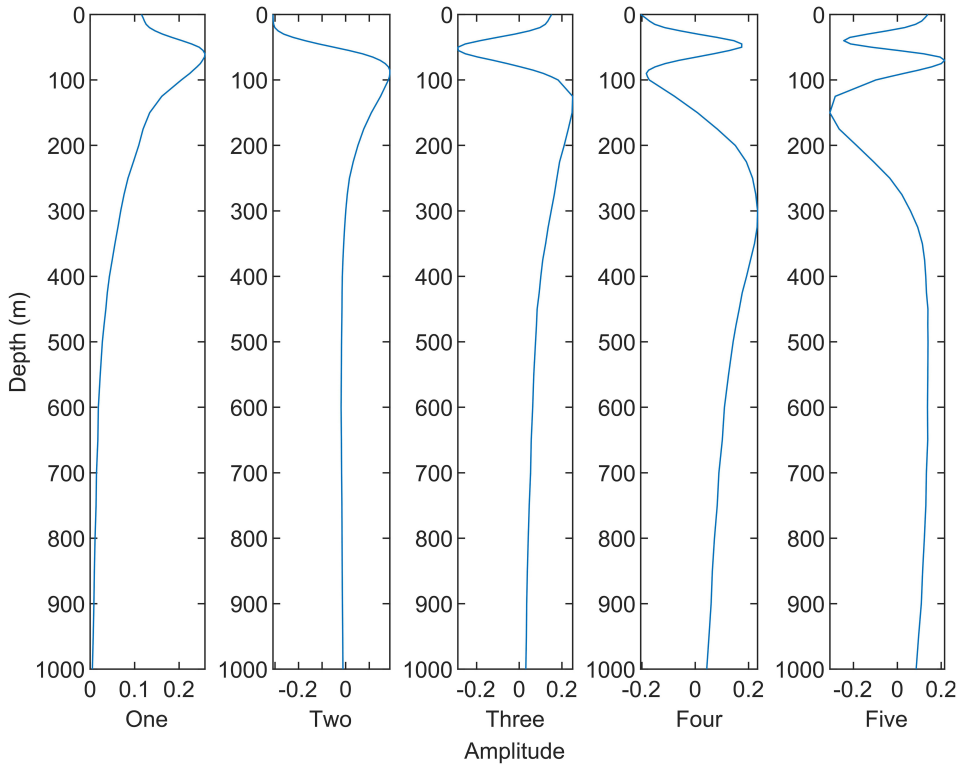


FIGURE 4
Amplitude of the first 5 EOFs.

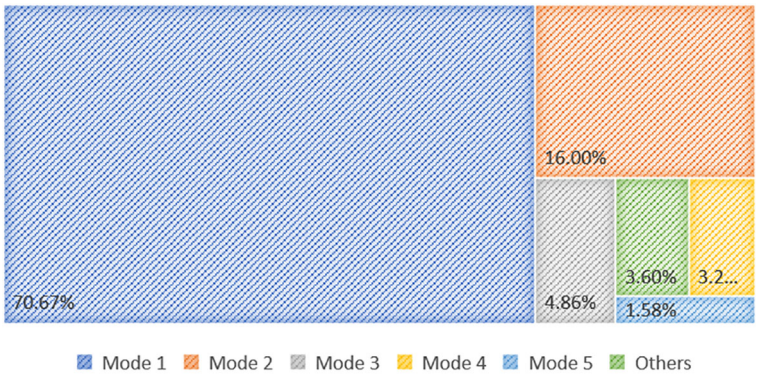


FIGURE 5
Distributions of contributions of the basis function.

3.2 Single empirical orthogonal regression function

The parameters of remote sensing at the same time and at the same location can be linearly related to those of the seabed. We created a regression database by using a large amount of historical data to establish a regression relationship among the temperature of the sea surface, its height, and the coefficients of perturbation.

$$\omega_{i,j} = \alpha + \beta \times A + \gamma \times B + \delta \times A \times B \quad (2)$$

This procedure entailed fitting a linear equation by using the database, expressed as Equation 2. where $\omega_{i,j}$ denotes the *j*-th order coefficient of perturbation of the *i*-th sample, and A and B denote anomalies in the height and the temperature of the sea surface, respectively. Linear fitting was used to obtain the coefficients α, β, γ, and δ. The corresponding coefficients of perturbation were obtained by entering the parameters of remote sensing, and the field of sound field of the South China Sea could then be inverted. The sEOFr method is based on linear regression between the parameters of remote sensing and the coefficients of projection. This linear relationship is derived from statistical results obtained from a large number of samples collected from the sea. In general, the errors tended to be concentrated in cases involving prominent differences between individual characteristics and statistical features.

3.3 Inversion of sound speed profile by using LSTM model

Given that the relationships between the parameters of the ocean were not purely linear, error was concentrated in regions featuring conspicuous perturbations. We propose a method of SSP inversion based on the LSTM neural network to improve the accuracy of inversion. Hochreiter created the LSTM model, which is an iterative version of the RNN model (Hochreiter and Schmidhuber, 1997). The LSTM model contains a memory cell that enables it to incorporate historical data, assess the relevance of information, improve its retention of valid information, filter out irrelevant information, and generate an output (Jain et al., 2019; Khataei Maragheh et al., 2022).

Figure 6 shows the structure of the LSTM model. It is composed of a forget gate, an input gate, and an output gate. Based on the previous output and the current input, the forget gate decides whether to forget the previous information or add it to the current memory cell.

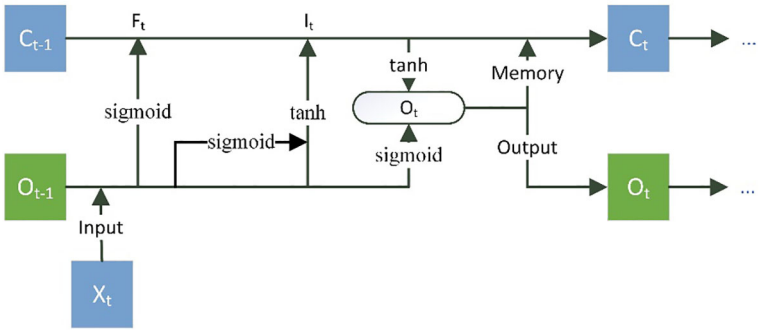


FIGURE 6
Diagram of the LSTM model.

$$F_t = \text{sigmoid}(\epsilon_f[O_{t-1}, X_t] + \theta_1) * C_{t-1} \quad (3)$$

Equation 3 is the calculation principle of the forgetting gate. where O_{t-1} denotes the output data from the previous time step, X_t denotes the input data in the current time step, C_{t-1} denotes the memory cell of the previous time step, and sigmoid denotes the activation function used to screen information within the range (0,1). In this experiment, Xt refers to sea surface height data and sea surface temperature data. θ_1 is a bias term that serves as an additional input for the corresponding neuron, and ϵ_f is the weight that represents the strength of the connection between units of the corresponding gate. The forget gate allows for the reinforcement of useful information while discarding irrelevant information, thus avoiding such problems as gradient explosion and the vanishing gradient that are caused by multiple iterations (Wang et al., 2020).

$$I_t = \text{sigmoid}(\epsilon_i[O_{t-1}, X_t] + \theta_2) * \tanh(\epsilon_c[O_{t-1}, X_t] + \theta_3) + F_t \quad (4)$$

The input gate is used to validate information and update the memory cell I_t, which is calculated by Equation 4. where ϵ_i and ϵ_c are weights, and θ_2 and θ_3 are bias terms that are used with the hyperbolic tangent (tanh) activation function in the interval (-1, 1). The tanh function is used by the input gate to generate the memory cell for the current time step. There are two steps involved, the first is to control the value between (0, 1) by the σ function, and the second is to generate the cell state of the current input by a tanh function. Following this, the information is filtered and added to the memory cell from the previous time step to enable it to be updated.

$$O_t = \text{sigmoid}(\epsilon_o[O_{t-1}, X_t] + \theta_4) * \tanh(I_t) \quad (5)$$

The output gate determines the output data and passes them to the next time step. The relevant calculation formula is shown in Equation 5. where O_t represents the output from the hidden layer at time t, ϵ_o is its weight, and θ_4 is a bias term. In this experiment, Ot is the EOF coefficient.

Figure 7 shows the training and testing of the LSTM model. S1-S269 in Figure 5 is the input data corresponding to sample No. 1–269 in the test example, including sea surface height data and sea surface temperature data. "C1-C269" refers to the output data corresponding to samples 1–269 in the test sample, and the output data is the EOF coefficient. We used the parameters of remote sensing as the input to the model and obtained the coefficients of perturbation as the output in the experiments. To train the LSTM neural network model, the parameters of remote sensing were fed to the input gate. The specific operational procedure entails utilizing the actual values of samples 1–8 as inputs for training the model, while the predicted value of sample 9 is generated as the output. Subsequently, the model undergoes training with the true values of samples 2–9, leading to the prediction of the value for sample 10. This iterative process continues until the model output yields the predicted value for sample 269, thereby culminating in the prediction of values for samples 9 through 269.The model was continually adjusted by being trained on temporally sequential data, and the RMSE was used as the loss function. Following this, the parameters of remote sensing for the next time step were entered to yield the corresponding coefficients of perturbation for SSP inversion. The EOF coefficient of the output is tested and then brought into Equations 6, 7. The SSP based on LSTM model inversion is calculated. The SSP can be expressed as the background profile plus the disturbance value. The background profile was obtained from WOA data. The sound speed disturbance value is obtained by multiplying the EOF coefficient calculated by the model with the EOF extracted previously. M is the order of EOF selected in the experiment.

$$SSP = \widehat{SSP} + SSP^{\text{disturbance}} \quad (6)$$

$$SSP^{\text{disturbance}} = \sum_{j=1}^{M} \omega_j EOF(j), M \in (1, 2, 3, 4, 5) \quad (7)$$

LSTM model is a nonlinear model, which has the advantage of preventing gradient vanishing and gradient explosion when dealing with long series data. Compared with linear sEOFr model, it is more

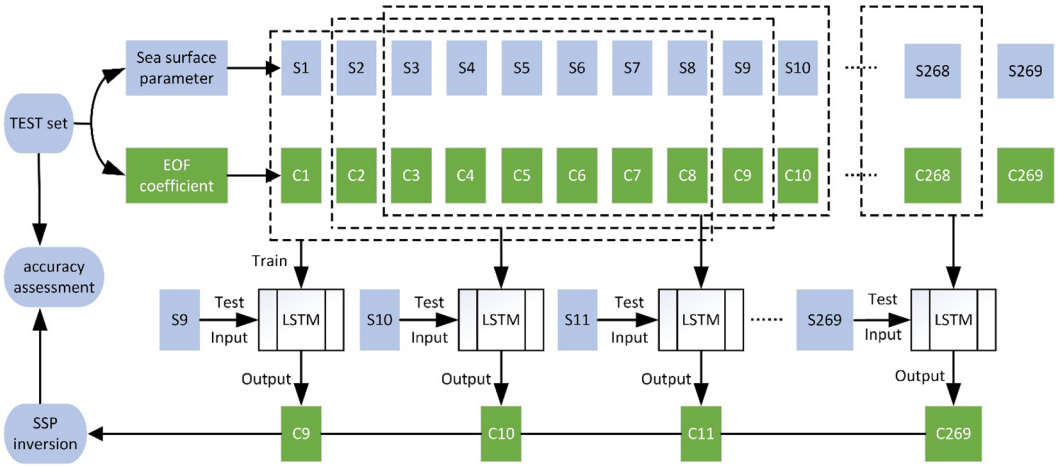


FIGURE 7
Flowchart of inversion of the field of sound speed by using the LSTM model.

suitable for complex ocean dynamic model. The use of the LSTM model allowed us to apply an incremental approach to invert the sound field of the TEST dataset. SSP has a strong time correlation, LSTM algorithm can learn the correct time pattern by memorizing the structure, and predict the subsequent data. When we used a continuous temporal duration of eight for training, the model was able to maintain a relatively high accuracy of training with a small number of training samples in the experiment. This method reduced the reliance of the model on a large number of samples while maintaining a high accuracy.

4 Results and analysis of inversion

4.1 Evaluation of results of inversion of sEOFr and LSTM

Temperature, salinity, and pressure are the primary determinants of the speed of sound. Its speed increases by approximately 4.2 m/s for every 1°C increase in the temperature of water, an increase of 0.1% in the salinity of water corresponds to that of 0.13 m/s in the speed of sound, while a 1 atm increase in water pressure corresponds to a 0.17 m/s increase in the sound speed.

From the sea surface to the 100 m underwater, the seawater is referred to as the mixed layer because it receives sunlight exposure, allowing it to absorb solar heat, resulting in relatively higher temperatures and minor temperature variations. The thermocline is a layer located approximately 100 m beneath the mixed layer. The temperature drops rapidly with depth at this thermocline. The thermocline in the South China Sea, which is located in a medium-to-low-latitude region, was assumed to be 100 m deep in our experiments. The rapid change in the temperature of this layer led to prominent fluctuations in the sound speed. Furthermore, eddies at the mesoscale, internal waves, and other oceanic activities occur frequently in this area (Hu et al., 2000; Sun et al., 2020). The complex combination of these factors contributes to the difficulty of inverting the SSP. Figure 8 shows the variation of the error in the direction of depth. The factors mentioned earlier cause the error in the model inversion results to be concentrated at a depth of about 100 m. The variance in the temperature gradually stabilized below the thermocline, thus reducing the errors in modeling. In this study, Figure 9 shows the average inversion errors of the two models in different seasons. The results indicate that the errors of the LSTM model increase in July, December, and April, corresponding to seasons with substantial variations. This is attributed to the poor performance of the LSTM model during seasonal changes, as it only utilizes the preceding 8 time steps of predicted samples for training. As the seasonal transitions stabilize, the errors of the LSTM model decrease. In contrast, the sEOFr model, being based on a linear model statistically derived from annual sound speed profile data, exhibits larger errors during winter due to significant sound speed disturbances.

Figure 10 shows the spatial distributions of errors in the fields of sound obtained by the two models. The average error of the LSTM model was 1.76 m/s while that of the sEOFr model was 2.33 m/s.

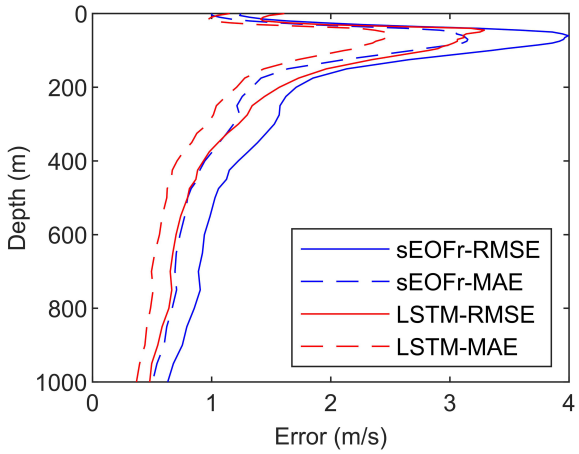


FIGURE 8
Distribution of error incurred by the model with depth below the sea surface.

Errors incurred by the latter were mostly concentrated in the blue-framed area in the figure (119° E–119.5° E, 18° N–18.5° N). Mesoscale eddies were frequently active at this location, especially with intense Ekman aspiration activity (Xiao et al., 2013) that led to the mixing of deep and surface waters to thicken the mixed layer of the ocean. However, the sound speed field in the sea area where the thickness of the mixed layer is large will produce a large disturbance. Figure 11 shows the spatial distribution of sound speed disturbance values. The area in which error was concentrated and that in which the disturbance was large significantly overlapped, indicating that the disturbance-related values were a key factor influencing the accuracy of inversion of the model. The RMSE of the linear sEOFr model in the error concentration area is 3.83 m/s, 1.50 m/s higher than the overall RMSE, and the accuracy is reduced by 64% The RMSE of the LSTM model is 2.16m/s, which is only 0.40m/s higher than the overall RMSE, and the accuracy is only reduced by 23%.In this area, the

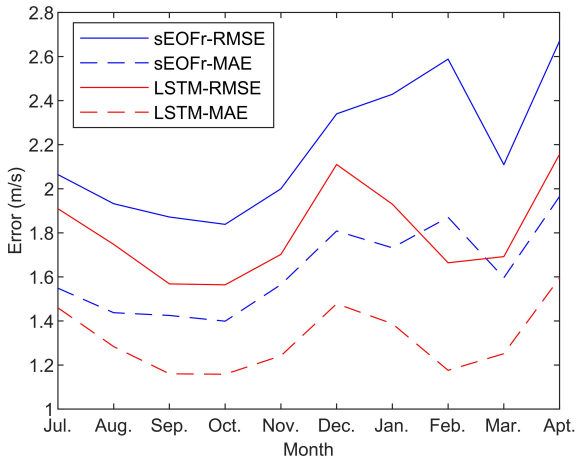


FIGURE 9
The distribution of errors in different seasons. (from July 9, 2014 to April 3, 2015).

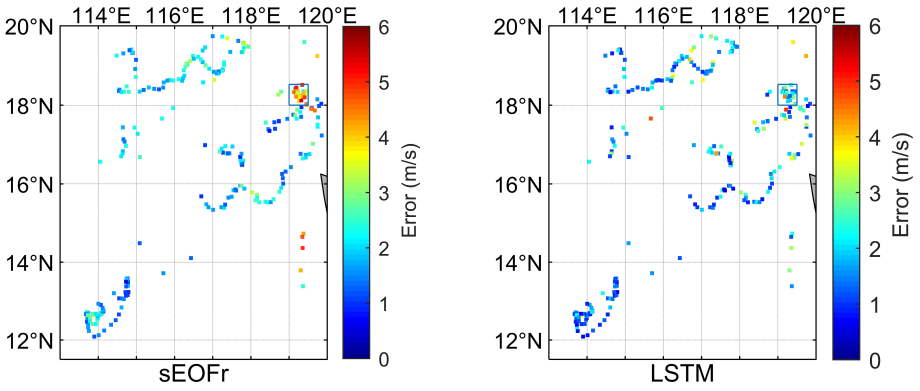


FIGURE 10
Spatial distribution of the error incurred in the inversion of the SSP.

inversion accuracy of LSTM model is improved by 43.6% compared with sEOFr model. It also shows that the linear model was unable to handle disturbances in this area, where this led to the concentration of error, while the LSTM model continued to deliver better performance and higher robustness in such scenarios.

4.2 Analysis of area of error concentration

We further investigated the relationship between the RMSE of the reconstructed profile and the values of disturbance in the profile obtained by the Argo data, as shown in Figure 12.

$$SSP^{rmse}_{sEOFr} = 0.30 \times SSP^{disturbance}_{Argo\ SSP} + 1.161 \quad (8)$$

$$SSP^{rmse}_{LSTM} = 0.12 \times SSP^{disturbance}_{Argo\ SSP} + 1.323 \quad (9)$$

The blue line in the figure represents the results of fitting of the sEOFr model according to Equation 8. Equation 9 shows the results of fitting of the LSTM model, which are represented by the red line in Figure 12. For a deviation of 1 m/s between the profile of the Argo data and the background profile, the error of the sEOFr model increased by 0.30 m/s while that of the LSTM model increased by only 0.12 m/s. The average speed of disturbance in the concentrated area was 8.49 m/s. When Equation 8 is applied to this average disturbance, the calculated RMSE was 3.71 m/s, which is smaller than the actual value of 3.83 m/s. When this average disturbance is substituted into Equation 9, the RMSE was 2.34 m/s, which is greater than the actual error of 2.16 m/s.

The sEOFr model exhibited an advantage over the LSTM model when the disturbance was minor. This is because it is based on a statistical relationship derived from a large amount of historical data. Conversely, the LSTM model delivered superior performance when handling profiles featuring substantial disturbances, with an accuracy that was 43.6% higher on average. Furthermore, the greater the disturbance was (in areas where oceanic activity was

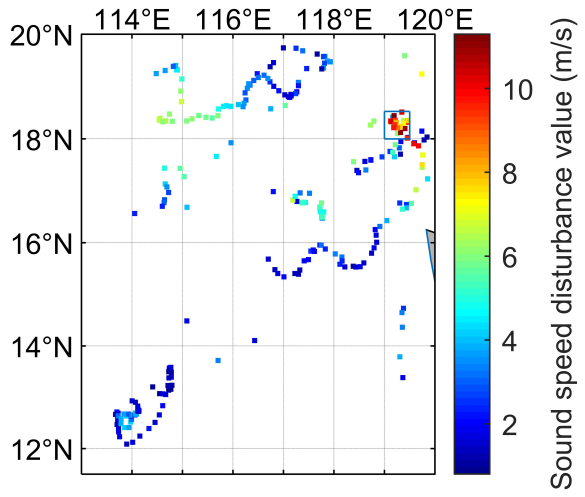


FIGURE 11
Spatial distribution of the values of disturbance in the sound speed.

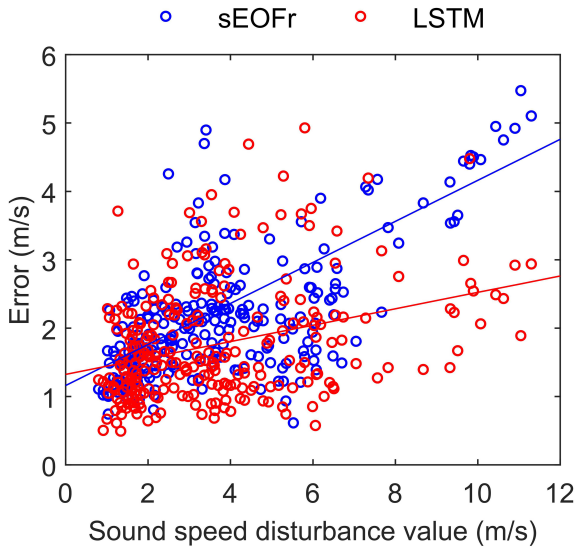


FIGURE 12
Relationship between Argo profile data reconstruction error and disturbance value.

frequent and disturbances were substantial) the better was the performance of the LSTM model.

4.3 Validity of results of inversion of acoustic field

The proposed LSTM method guarantees a high accuracy of reconstruction of the SSP, but may not explicitly reveal certain errors in the fine structure of its results. The primary goal of reconstructing the SSP is to calculate the acoustic field, which enables the observation of the fine structure of the SSP. It is important to predict acoustic fields in sonar systems so that targets can be accurately detected. In this section, we report the use of Kraken software to calculate the loss of acoustic transmission in the profile reconstructed by the LSTM model (Model Kraken software is referenced from http://oalib.hlsresearch.com/AcousticsToolbox/), with a significant improvement in the accuracy of reconstruction. The structure of these profiles is illustrated in Figure 13A. Given that the inversion results of the models are only valid within 1000 meters, and the sound speed disturbances below 1000 meters are relatively small, the difference in inversion errors between the two models is not significant. We utilized WOA18 data to fit the sound speed distribution in waters deeper than 1000 meters using empirical formulas. The sound source was 80 m deep, the receiver depth is 80 m with a frequency of 100 Hz, a density of seafloor of 1.5 g/cm3, the seabed sound speed is 1550 m/s, an attenuation coefficient of 0.15 dB/λ, and a depth of water of 3500 m. Figure 13B displays the transmission loss calculated by using the SSPs obtained under these conditions. The sEOFr model recorded an RMSE of 3.83 dB in its calculation of the non-coincident loss of transmission, with 90% of the points yielding errors of 7.75 dB or smaller. The RMSE of the LSTM model was 1.68 dB, with 90% of the points yielding errors of 3.65 dB or smaller. The loss of transmission of both models peaked at 5.5 km, but the loss incurred by the sEOFr model was different by 14.58 dB from the Argo profile, while the LSTM model yielded a difference of only 6.24 dB from it by comparison. After 15 km, the transmission loss calculated by the sEOFr method exhibited a prominent shift in the structure of interference, while the interference structure of LSTM method is basically consistent with Argo. This suggests that the results of inversion of the LSTM model accurately described actual changes in the transmission loss. In most cases, its error was consistently below 3.65 dB.

Table 1 summarizes the reconstruction results of the two models. The sEOFr model used 3,620 samples to train the model. LSTM trains the model using only 268 samples, of which 8 are the number of samples trained at one time. LSTM model can reconstruct SSP with fewer samples, and its reconstructed RMSE is 1.76m/s, which is more accurate than sEOFr model. In the disturbed area, the accuracy of sEOFr decreases significantly, while the accuracy of LSTM model decreases only a little. The error of sEOFr model is more than twice that of LSTM in predicting propagation loss, and the absolute error range is also twice that of LSTM. Compared with sEOFr model, LSTM model can better solve the problem of sparse sample, large disturbance in sea area, and forecast transmission loss.

5 Conclusions

In this paper, we proposed a method of SSP inversion based on the LSTM network. By using the parameters of remote sensing as inputs to the model, this method can be used to derive the coefficients of disturbance for SSP inversion. We tested the

TABLE 1 Comparison of inversion results of the two models.

Model	sEOFr	LSTM
Number of training samples	3620	268(8)
The RMSE of reconstruct SSP (m/s)	2.33	1.76
The RMSE of the disturbed concentration area (m/s)	3.83	2.16
The RMSE of forecast transmission loss (dB)	3.83	1.68
The bounds absolute error of forecast transmission losses (dB)	14.58	6.24

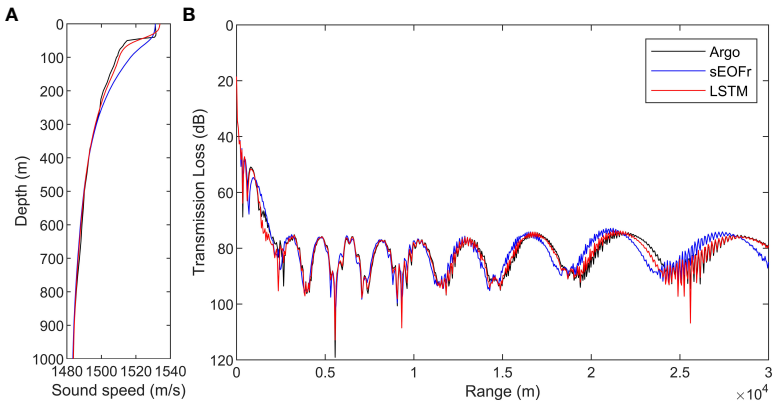


FIGURE 13
(A, B) Analysis of transmission loss calculated by using the sound speed profile.

proposed method on data from the South China Sea and compared its performance with that of the sEOFr model. The results revealed that it had a higher accuracy of inversion of the SSP. It recorded an RMSE that was smaller than that of the sEOFr model by 0.57 m/s, with a 24.46% improvement in accuracy. The concentration of disturbances complicates inversion and reduces the accuracy of the model. However, the memory structure of the proposed LSTM model enabled it to perform well in areas with concentrated disturbances in the sound speed. Furthermore, it delivered excellent performance when the disturbances were large. It reduced the RMSE by 1.67 m/s for such areas in comparison with the sEOFr model, resulting in a 43.60% higher accuracy. This demonstrated its superior performance and robustness in regions with a high concentration of disturbances.

The acoustic field for the profile with the highest improvement in accuracy in inversion based on the LSTM model was calculated by using Kraken software. Its RMSE for the non-coincident loss of transmission was 1.68 dB, with 90% of the error points falling below 3.65 dB. This constituted an improvement of greater than 50% over the sEOFr model, and shows that the proposed LSTM method of SSP inversion can accurately predict changes in the TL.

The proposed non-linear method of SSP inversion is better suited to non-linear relationships between the parameters of the ocean, and yields more accurate outcomes for areas in which traditional models struggle to address concentrated disturbances. The transmission loss in the SSP derived from its approach to inversion more closely approximates the actual profile. This method is important for quickly obtaining the underwater sound field, where this is important for predicting the acoustic field for target detection in sonar systems and underwater acoustic communication.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YZ: Writing – review & editing, Funding acquisition, Project administration, Resources, Validation. PX: Formal analysis, Resources, Supervision, Writing – review & editing. GL: Investigation, Validation, Visualization, Writing – original draft. ZO: Data curation, Project administration, Software, Writing – original draft. KQ: Conceptualization, Funding acquisition, Methodology, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Natural Science Foundation of Guangdong Province under contract No. 2022A151501151.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Carnes, M. R., Mitchell, J. L., and De Witt, P. W. (1990). Synthetic temperature profiles derived from geosat altimetry: comparison with air-dropped expendable bathythermograph profiles. *J. Geophys. Res.* 95, 17979–17992. doi: 10.1029/JC095iC10p17979

Carnes, M. R., Teague, W. J., and Mitchell, J. L. (1994). Inference of subsurface thermohaline structure from fields measurable by satellite. *J. Atmos. Oceanic Technol.* 11, 551–566. doi: 10.1175/1520–0426(1994)011<0551:IOSTSF>2.0.CO;2

Chen, C., Ma, Y., and Liu, Y. (2018). Reconstructing sound speed profiles worldwide with sea surface data. *Appl. Ocean Res.* 77, 26–33. doi: 10.1016/j.apor.2018.05.002

Del Grosso, V. A. (1974). New equation for the speed of sound in natural waters (with comparisons to other equations). *J. Acoust. Soc Am.* 56, 1084–1091. doi: 10.1121/1.1903388

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural. Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hu, J., Kawamura, H., Hong, H., and Qi, Y. (2000). A review on the currents in the south China sea: seasonal circulation, south China sea warm current and kuroshio intrusion. *J. Oceanogr.* 56, 607–624. doi: 10.1023/A:1011117531252

Jain, S., and Ali, M. M. (2006). Estimation of sound speed profiles using artificial neural networks. *IEEE Geosci. Remote Sens. Lett.* 3, 467–470. doi: 10.1109/LGRS.2006.876221

Jain, G., Sharma, M., and Agarwal, B. (2019). Optimizing semantic LSTM for spam detection. *Int. J. Inf. Tecnol.* 11, 239–250. doi: 10.1007/s41870-018-0157-5

Khataei Maragheh, H., Gharehchopogh F.S. Majidzadeh, K., and Sangar, A. B. (2022). A new hybrid based on long short-term memory network with spotted hyena optimization algorithm for multi-label text classification. . *Math* 10, 488. doi: 10.3390/math10030488

Klemas, V., and Yan, X.-H. (2014). Subsurface and deeper ocean remote sensing from satellites: an overview and new results. *Prog. Oceanogr.* 122, 1–9. doi: 10.1016/j.pocean.2013.11.010

LeBlanc, L. R., and Middleton, F. H. (1980). An underwater acoustic sound velocity data model. *J. Acoust. Soc. Am.* 67, 2055–2062. doi: 10.1121/1.384448

Li, Y., Gao, P., Tang, B., Yi, Y., and Zhang, J. (2022a). Double feature extraction method of ship-radiated noise signal based on slope entropy and permutation entropy. *Entropy* 24, 22. doi: 10.3390/e24010022

Li, Y., Geng, B., and Jiao, S. (2022b). Dispersion entropy-based lempel-ziv complexity: A new metric for signal analysis. *Chaos Solitons Fractals* 161, 112400. doi: 10.1016/j.chaos.2022.112400

Li, H., Qu, K., and Zhou, J. (2021). Reconstructing sound speed profile from remote sensing data: nonlinear inversion based on self-organizing map. *IEEE Access* 9, 109754–109762. doi: 10.1109/ACCESS.2021.3102608

Li, Y., Tang, B., and Yi, Y. (2022c). A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy. *Appl. Acoustics* 196, 108899. doi: 10.1016/j.apacoust.2022.108899

Liu, Y., Chen, W., Chen, Y., Chen, W., Ma, L., and Meng, Z. (2021). Ocean front reconstruction method based on K-means algorithm iterative hierarchical clustering sound speed profile. *JMSE* 9, 1233. doi: 10.3390/jmse9111233

Liu, L., Silver, D., and Bemis, K. (2021). Visualizing acoustic imaging of hydrothermal plumes on the seafloor. *IEEE Comput. Grap. Appl.* 41, 63–75. doi: 10.1109/MCG.2020.2995077

Luo, P., Song, Y., Xu, X., Wang, C., Zhang, S., Shu, Y., et al. (2022). Efficient underwater sensor data recovery method for real-time communication subsurface mooring system. *JMSE* 10, 1491. doi: 10.3390/jmse10101491

Ou, Z., Qu, K., Shi, M., Wang, Y., and Zhou, J. (2022). Estimation of sound speed profiles based on remote sensing parameters using a scalable end-to-end tree boosting model. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1051820

Rahaman, H., Behringer, D. W., Penny, S. G., and Ravichandran, M. (2016). Impact of an upgraded model in the NCEP global ocean data assimilation system: the tropical Indian ocean: NCEP-GODAS ANALYSIS WITH MOM4P1. *J. Geophys. Res. Oceans* 121, 8039–8062. doi: 10.1002/jgrc.v121.11

Stammer, D. (1997). Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. *J. Phys. Oceanogr.* 27, 1743–1769. doi: 10.1175/1520–0485(1997)027<1743:GCOOVE>2.0.CO;2

Su, J., Li, Y., and Ali, W. (2022). Underwater passive manoeuvring target tracking with isogradient sound speed profile. *IET Radar Sonar Navi* 16, 1415–1433. doi: 10.1049/rsn2.12269

Su, H., Wu, X., Yan, X.-H., and Kidwell, A. (2015). Estimation of subsurface temperature anomaly in the Indian ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach. *Remote. Sens. Environ.* 160, 63–71. doi: 10.1016/j.rse.2015.01.001

Su, H., Yang, X., Lu, W., and Yan, X.-H. (2019). Estimating subsurface thermohaline structure of the global ocean using surface remote sensing observations. *Remote Sens.* 11, 1598. doi: 10.3390/rs11131598

Sun, S., Fang, Y., Zu, Y., Liu, B., and Tana Samah, A. (2020). A seasonal characteristics of mesoscale coupling between the sea surface temperature and wind speed in the south China sea. *J. Clim.* 33, 625–638. doi: 10.1175/JCLI-D-19-0392.1

Teymorian, A. Y., Cheng, W., Ma, L., Cheng, X., Lu, X., and Lu, Z. (2009). 3D underwater sensor network localization. *IEEE Trans. Mobile Comput.* 8, 1610–1621. doi: 10.1109/TMC.2009.80

Wang, R., Peng C. Gao, J., Gao, Z., and Jiang, H. (2020). A dilated convolution network-based LSTM model for multi-step prediction of chaotic time series. *Comp. Appl. Math.* 39, 30. doi: 10.1007/s40314-019-1006-2

Wunsch, C. (1997). The vertical partition of oceanic horizontal kinetic energy. *J. Phys. Oceanogr.* 27, 1770–1794. doi: 10.1175/1520–0485(1997)027<1770:TVPOOH>2.0.CO;2

Xiao, J., Xie Q. Liu, C., Chen, J., Wang, D., and Chen, M. (2013). A diagnostic model of the South China Sea bottom circulation inconsideration of tidal mixing, eddy-induced mixing and topography. *Acta Oceanol Sin. (in Chinese)* 35, 01–13. doi: 10.3969/j.issn.0253–4193.2013.05.001

Xu, P., Ando, M., and Tadokoro, K. (2005). Precise, three-dimensional seafloor geodetic deformation measurements using difference techniques. *Earth Planet Sp* 57, 795–808. doi: 10.1186/BF03351859.\

Xu, C., Li, H., Chen, B., and Zhou, T. (2013). Multibeam interferometric seafloor imaging technology. *J. Harbin Eng.* 34, 1159–1164. doi: 10.3969/j.issn.1006–7043.201302007

Yan, K., Wang, Y., and Xiao, W. (2022). A new compression and storage method for high-resolution SSP data based-on dictionary learning. *JMSE* 10, 1095. doi: 10.3390/jmse10081095

Zhan, D., Wang, S., Cai, S., Zheng, H., and Xu, W. (2023). Acoustic localization with multi-layer isogradient sound speed profile using TDOA and FDOA. *Front. Inform Technol. Electron Eng.* 24, 164–175. doi: 10.1631/FITEE.2100398

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 29 July 2024
DOI 10.3389/fmars.2024.1351327

OPEN ACCESS

EDITED BY
Kai Zhao,
Nanjing Tech University, China

REVIEWED BY
Chuanlin He,
Harbin Institute of Technology, China
Hiroyuki Matsumoto,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Shaowei Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Yanliang Pei
peiyanliang@fio.org.cn

RECEIVED 06 December 2023
ACCEPTED 05 July 2024
PUBLISHED 29 July 2024

CITATION
Wei Z, Pei Y, Zhu X, Liu K, Zhang X, Zong L and Li X (2024) A high-precision positioning method for deep-towed multichannel seismic arrays.
Front. Mar. Sci. 11:1351327.
doi: 10.3389/fmars.2024.1351327

A high-precision positioning method for deep-towed multichannel seismic arrays

Zhengrong Wei[1], Yanliang Pei[2,3]*, Xiangqian Zhu[4], Kai Liu[4,5], Xiaobo Zhang[1], Le Zong[5] and Xinyu Li[4]

[1]College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China, [2]First Institute of Oceanography of Ministry of Natural Resources, Qingdao, China, [3]Laboratory for Marine Geology, Qingdao Marine Science and Technology Center, Qingdao, China, [4]Key Laboratory of Deep Sea Mineral Resources Development, Shandong (Preparatory), Qingdao, China, [5]Key Laboratory of Submarine Acoustic Investigation and Application of Qingdao (Preparatory), Qingdao, China

Kuiyang-ST2000 is a deep-towed multichannel seismic system that provides high-resolution exploration of sub-seabed geological formations. Due to the uncertainty of the sound speed at full ocean depth, the travel-time positioning of sea surface reflected waves still has flaws in positioning arrays. This research reveals that the average sound speed of seawater selected for computing the array position only affects the vertical displacement of the arrays. thus, a polynomial fitting method is proposed to position the arrays. Because the nonuniform mass distribution complicates the array shape, first, the weight of the digital transmission unit is balanced by one designed floater so that the array shape becomes a simple convex curve during towing conditions. Afterward, one general sound speed is used to calculate the initial array position; then, the polynomial fitting method is used to tune the sound speed so that the seismic source and hydrophones are on the same convex curve. Finally, an accurate array position is calculated by the proposed positioning method, and the submarine shallow strata are imaged at a high resolution.

KEYWORDS

deep-towed multichannel seismic system, array shape, array positioning, sound speed at full ocean depth, polynomial fitting

1 Introduction

The accuracy of array positioning is closely related to the data processing quality of deep-towed seismic exploration systems (Howard and Syck, 1992; Lu et al., 2003). Generally, vertical accuracy is required to reach a sampling interval (Δt), and horizontal accuracy is required to reach the 1/4 wavelength of the seismic wave (Krödel et al., 2015). However, the towing depth of deep-towed seismic systems is greater than 1000 m; therefore, navigation via a global positioning cannot be used to determine the geometric

relationship between the source and the arrays. Additionally, due to the influences of retracting and releasing the towing cable, the instability of the towing speed and the variation in the deep-sea environment, the seismic source and arrays inevitably float up and down during the data acquisition process. Many researchers have focused on methods for aligning arrays (Varypaev and Kushnir, 2020; Nanni et al., 2022), positioning arrays (Vickery, 1998; Lan and Ding, 2012), and calibrating arrays (Fikes et al., 2019; Che et al., 2021).

The Deep Tow Array Geophysical System (DTAGS) is the first deep-towed multichannel seismic survey system developed by the Naval Ocean Research and Development Activity to measure the geoacoustic parameters of the ocean floor (Fagot et al., 1980; Chapman et al., 2002; Ker et al., 2010). The array positioning system of DTAGS relies mainly on depth sensors installed on towed vehicles and hydrophone 28 (at 138 m), hydrophone 38 (at 288 m) and hydrophone 48 (at 438 m) of the arrays. The water pressure and temperature are measured in a timely manner, and the values are converted to depth values according to an empirical formula for sea area measurements. Initially, the arrays between adjacent depth sensors are simplified as one straight line, and the position of each hydrophone in the arrays is calculated via linear interpolation (Rowe and Gettrust, 1993). However, this assumption does not conform to the real shape of the arrays, resulting in large errors in seismic data processing. These errors reduce the signal-to-noise ratio of the seismic data after stacking imaging. The travel time of a sea surface reflection (SSR) wave has been used to locate the arrays, improving the accuracy of the relative positions between the source and hydrophone in single-shot gathering (Walia and Hannay, 1999). The array shape calculated with depth sensors is simultaneously constrained by the travel time of direct waves and SSR waves, and the genetic algorithm can be used to optimize the sound speed by He (He et al., 2009). Although the positioning accuracy is improved, it is still unsatisfactory because the water depths measured by depth sensors have errors of 3-5 m. Following that, it was determined that direct waves and SSR waves can be picked up more accurately from waveform envelope lines (Kong et al., 2012). However, due to the time delay in data acquisition, direct wave signals can be measured only at the last 3~4 hydrophones, and the positioning of the front hydrophones is not accurate.

A microelectromechanical system (MEMS) is installed on each array channel (hydrophone) in the deep-towed seismic system SYSIF, developed by the French oceanographic institute IFREMER. The array shape can be identified by attitude angles, such as the pitch, roll and yaw angles (Marsset et al., 2014). One convex curve was expected; however, the ethernet switch installed in a digital transmission unit (DTU) had a mass of 1.3 kg. The weights of these DTUs were greater than their buoyancy, which deformed the straight line of the arrays and generated a "W" shape (Colin et al., 2020). Additionally, the inversion accuracy is limited by the accuracy of the MEMS. Bathymetry mapping was established through multibeam data, and the array position calculated by MEMS sensors was used for forward modeling. The relationships between the calculated travel times of seafloor reflections and actual travel-time records were analyzed, and the array positions calculated from MEMS data were obviously optimized (Marsset et al., 2018). The accuracy of this positioning method depends on the bathymetry results. However, accurate bathymetry in areas with complex seabed topographies is difficult to establish. With the assumption that the seafloor was locally flat, the filtered MEMS data were optimized by one local optimization method (Colin et al., 2020). The travel times of direct and seafloor reflection waves were the constraint conditions, and the positioning results were consistent with the array shape in seawater. However, this method does not work when the seafloor slope is high, which is the general case in reality.

Array positioning relies mainly on the travel time of the direct wave and the SSR wave in the Kuiyang-ST2000 deep-towed seismic system. The travel-time positioning results of sea trials in 2019 revealed that the arrays exhibited a "W" shape (Wei et al., 2020). Moreover, numerical simulations demonstrated that the "W" shape was generated by the unbalanced weight of the DTUs and the insufficient drag force provided by the drogue (Zhu et al., 2020). To balance the additional weight of the DTU, a floater was designed and installed on each DTU. In a sea trial in 2021, both the numerical simulation and the travel time of the SSR wave demonstrated that the array shape was a simple convex curve under general towing conditions. Therefore, a weighted least squares polynomial fitting method is proposed to optimize the sound speed at a full ocean depth so that the seismic source and the arrays form a simple convex curve. High-precision array positioning for the Kuiyang-ST2000 deep-towed seismic exploration system was achieved with the corrected sound speed, and the energy groups of the velocity spectrum were well concentrated in both deep and shallow layers.

This paper is organized as follows: the Kuiyang-ST2000 system, the travel-time positioning, and the array shape correction method, is introduced in Section 2. A deep-towed array location modification method is presented in Section 3. A typical application is presented in section 4. The conclusions and prospects are given in Section 5.

2 Method

2.1 Kuiyang-ST2000 system

The Kuiyang-ST2000 deep-towed high-resolution seismic system is shown in Figure 1 (Pei et al., 2022). The plasma spark source sound level is 216 dB, and the dominant frequency is 750 Hz (frequency bandwidth: 150–1200 Hz). The source wavelet is a pulse. The arrays are composed of a zero-buoyancy front section (Figure 1H), a working section (Figure 1G), a balance section (Figure 1I) and a drogue (Figure 1J). The length of the front section is 12.5 m. The working section consists of three segments connected by DTUs (Figure 1F), which are used to convert analog signals to digital signals. Each segment is 50 m in length and contains 16 channels with spacings of 3.125 m. The system is equipped with an ultra-short-baseline (USBL) (Figure 1B), a conductivity-temperature-depth (CTD) sensor module (Figure 1K), a depth sensor (Figure 1C), and an altimeter (Figure 1D). The Kuiyang-ST2000 system has a maximum towing speed of 3 knots and can operate at depths of up to 2056 meters.

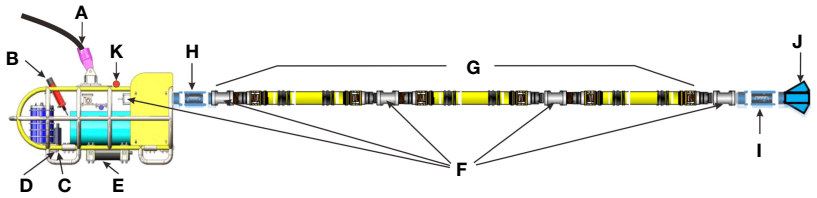


FIGURE 1
Components of the Kuiyang-ST2000 deep-towed high-resolution seismic system. (**A**: electro-optical cable; **B**: USBL; **C**: depth sensor; **D**: altimeter; **E**: plasma spark source; **F**: DTU; **G**: working section; **H**: front section; **I**: balance section; **J**: drogue; **K**: CTD).

2.2 Travel-time positioning

The travel-time positioning method is based on the geometric seismic principle, as shown in Figure 2.

According to the geometric relationship (Eli, 2010), the position of hydrophones can be calculated by the following equations:

$$d_r^i = \frac{(t_s^i v_w)^2 - (t_d^i v_f)^2}{4d_s}, \tag{1}$$

$$x_h^i = \sqrt{(t_d^i v_f)^2 - (d_r^i - d_s)^2}\,. \tag{2}$$

where d_r^i and x_h^i are the depth and horizontal offset of the i^{th} hydrophone, respectively; t_d^i and t_s^i are the travel times of the direct wave and SSR wave picked by the i^{th} hydrophone, respectively; v_w is the average sound speed of seawater at full ocean depth; v_f is the sound speed of seawater near the seafloor as measured by the CTD sensor; and d_s is the seismic source depth measured by the depth sensor.

Figure 3 shows the one-shot seismic record (S2750) collected in the South China Sea in 2019. The seismic records include the travel times of the direct waves, reflected waves from the strata and reflected waves from the sea surface, as shown in Figures 3A–C, respectively. Finally, the array shape calculated by using the travel-time positioning method is shown in Figure 3D.

2.2.1 Positioning analysis

Because the sound speed of seawater varies with water depth, temperature and salinity, the assumption that the average sound speed of seawater is 1484.5 m/s is not accurate for every shot. Figure 4 shows the positioning results at velocities of 1483 m/s, 1485 m/s, and 1487 m/s. The positioning differences generated at velocities of 1483 m/s and 1487 m/s compared to 1485 m/s are illustrated in Figure 5. Even though the sound speed of seawater exhibits a small deviation, the array shape or inclination angle of the arrays with a still water level remains unchanged. The sound speed only affects the vertical position of the arrays. Fortunately, the vertical position of the seismic source is determined in a timely manner by the depth sensor, and the geometrical relationship between the towed vehicle and the arrays indicates that the vertical position of the seismic source can be used to adjust the sound speed of seawater. However, because the DTU between two adjacent working segments is made of stainless steel, the weight of the DTU is greater than its buoyancy. These redundant gravity forces the arrays to form a “W” shape, which makes it difficult to determine the geometrical relationship.

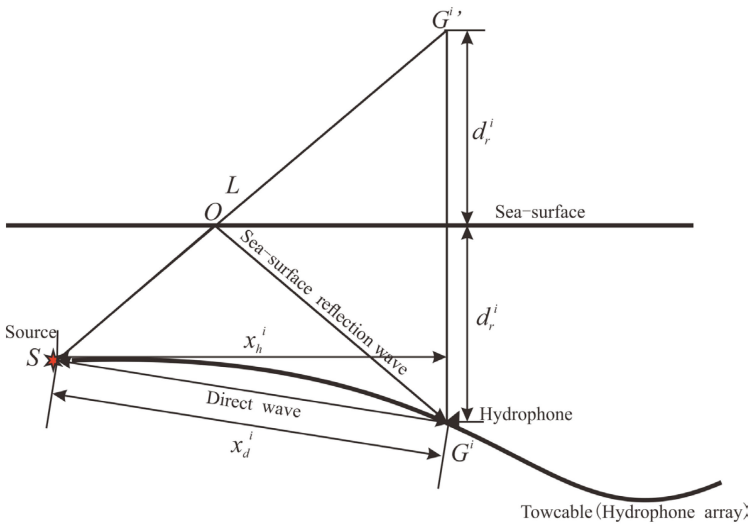


FIGURE 2
Travel-time positioning (direct and reflected wave paths) schematic diagram. (*S*, seismic source; $G^{i'}$, hydrophone mirror; G^i, hydrophone; d_r^i, depth of hydrophone; d_s, depth of seismic source; x_h^i, horizontal offset; x_d^i, travel distance of direct wave; *L*, travel distance of sea-surface reflection wave).

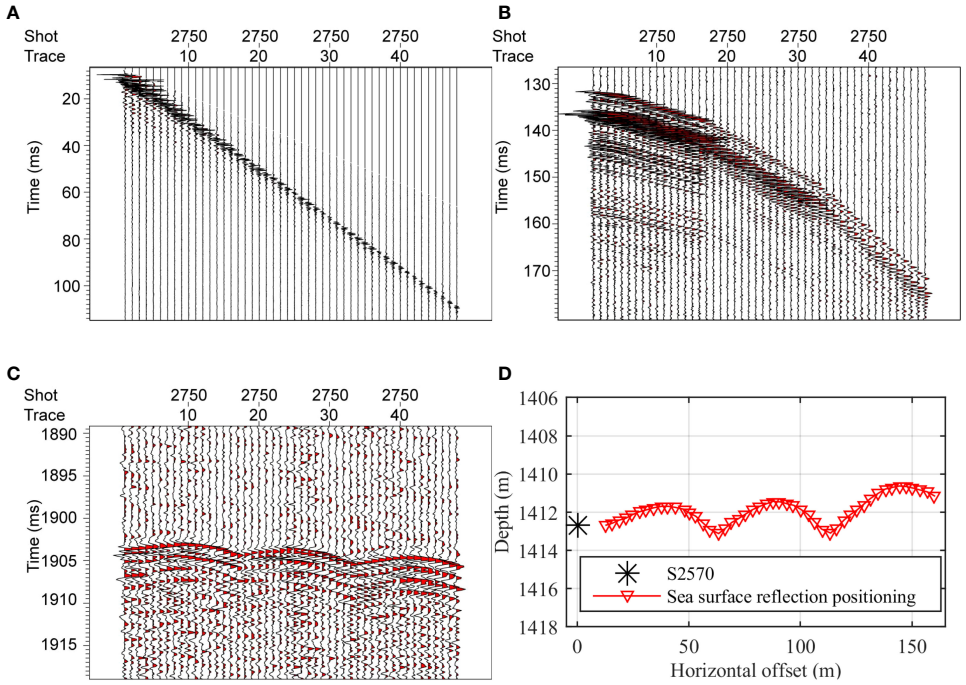


FIGURE 3
Examples of seismic records of sea surface reflection waves and travel-time positioning results. **(A)** Direct wave, **(B)** Seafloor reflection wave, **(C)** Sea surface reflection wave, **(D)** Array shape by travel-time positioning.

2.3 Array shape correction

2.3.1 Floaters

The segment joint (DTU) is made of stainless steel and has a mass of 1.85 kg in air. The mass of occupied seawater is approximately 1.1 kg, so an excess of 0.75 kg should be balanced to reach zero buoyancy. Therefore, a floater is designed, as shown in Figure 6A. These devices can withstand pressures reaching a water depth of 3000 m. Thus, each floater provides equivalent buoyancy compared to the weight of the DTU. A photograph of the floater is shown in Figure 6B, and the arrays equipped with the floaters are shown in Figure 6C.

2.3.2 Numerical model

To verify the performance of the floaters, an array model is established. The numerical modeling of the arrays and floaters are shown below. Via the lumped mass method, the array model is established with respect to the relative velocity element frame, as shown in Figure 7. The array is divided into 21 elements. The node

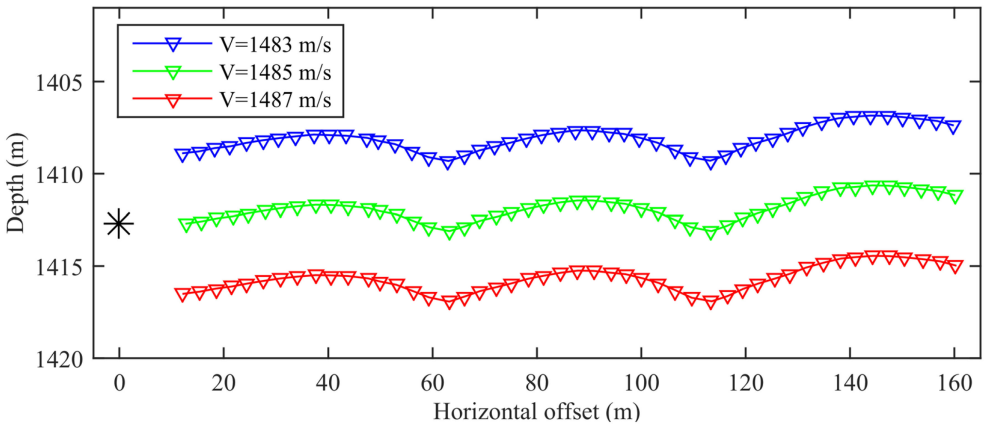


FIGURE 4
Travel-time positioning results with different acoustic velocities at shot point S2750.

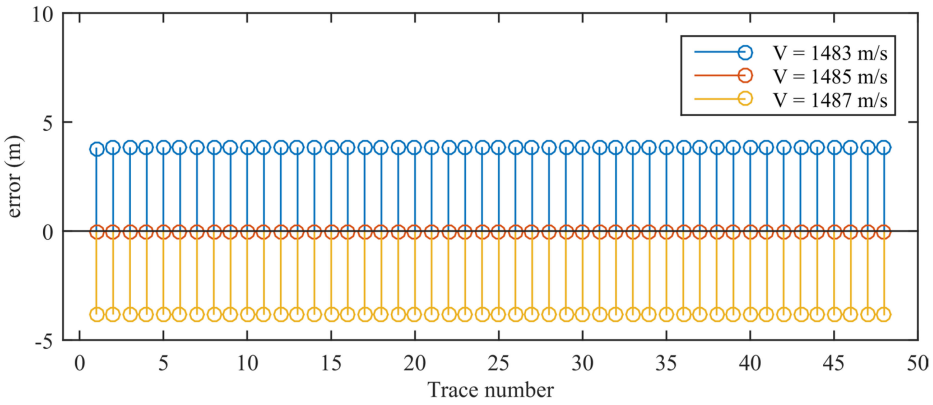


FIGURE 5
Depth differences with respect to the sound speed of seawater 1485 m/s.

connected to the towed vehicle is defined as the 1^{st} node, while the end node is defined as the 22^{nd} node.

The arrays are subjected mainly to internal tension and damping forces, hydrodynamic forces, and self-inertial forces in water. The tension and damping forces are calculated from the axial extension and the extension velocity, respectively, and are oriented in the same direction as the element orientation vector. According to the Morison formula, the drag resistance is proportional to the square of the relative velocity of the cable in water. These forces are simply expressed in Equation (3), as per.

$$\begin{cases} \underline{T}_b^i = \frac{\pi d_c^2}{4} E \varepsilon_b^i, \\ \underline{D}_b^i = C_d \underline{\underline{A}}^{iT} (\dot{\underline{N}}_g^{i+1} - \dot{\underline{N}}_g^i) \bar{z}, \\ \underline{W}_g^i = (\frac{\pi d_c^2}{4} l_0 \rho_c - \frac{\pi d_c^2}{4} l_0 \rho_f) g, \\ \underline{F}_b^i = -\frac{1}{2} C_n \rho_f d_c l^i \| \underline{V}_b^i \| \underline{V}_b^i, \end{cases} \tag{3}$$

where the cable tension $\underline{T}_b^i$ is defined by the axial strain ε_b^i, elastic modulus E, and cable diameter d_c. ε_b^i is defined by the current length of the i^{th} element l^i and the initial length l_0, as shown in Equation (4). The damping force $\underline{D}_b^i$ is expressed by the velocity difference between the terminal nodes in the element orientation direction. C_d is the damping coefficient of the cable, and $\dot{\underline{N}}_g^i$ is the node velocity with respect to the global frame. $\bar{z}$ is the unit vector of the i^{th} element orientation vector $\underline{E}_g^i$, which is obtained from the node position vectors $\underline{N}_g^{i+1}$ and $\underline{N}_g^i$, as shown in Equation (4). The transformation matrix $\underline{\underline{A}}^{iT}$ is used to convert the velocity from the global frame to the element-fixed frame. $\underline{W}_g^i$ is the weight of the cable in water, and $\underline{g}$ indicates that the force is oriented in the same direction as gravity. ρ_c and ρ_f are the densities of the cable and seawater, respectively. $\underline{F}_b^i$ is the hydrodynamic force acting on the i^{th} element, and C_n is the normal drag coefficient of the cable. $\underline{V}_b^i$ and $\underline{V}_g^i$ are the relative velocities of the i^{th} element with respect to the element-fixed frame and global frame, respectively. The average

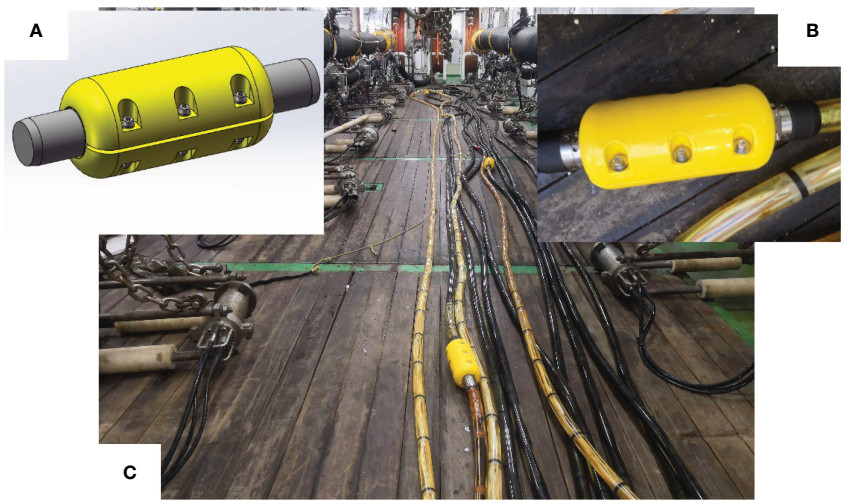


FIGURE 6
Floaters and arrays. **(A)** Buoyancy block model, **(B)** Installation of buoyancy blocks, **(C)** Array with buoyancy blocks installed.

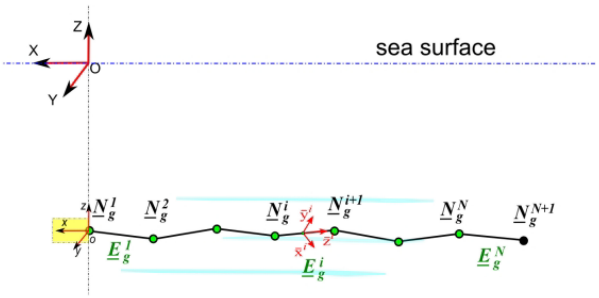


FIGURE 7
Modeling of arrays using the lumped mass method (Zhu et al., 2020).

velocity of the terminal nodes is used to express the velocity of the given element, as shown in Equation (4). $\underline{V}_f^i$ is the velocity of the water particle at the i^{th} node.

$$\begin{cases} \underline{E}_g^i = \underline{N}_g^{i+1} - \underline{N}_g^i, \\ \varepsilon_b^i = \frac{l^i - l_0}{l_0}, \\ l^i = \sqrt{\underline{E}_g^{it}\underline{E}_g^i}, \\ \underline{V}_b^i = \underline{A}^{iT}\underline{V}_g^i, \\ \underline{V}_g^i = \frac{\dot{\underline{N}}_g^{i+1} + \dot{\underline{N}}_g^i - \underline{V}_f^{i+1} - \underline{V}_f^i}{2}. \end{cases} \tag{4}$$

To consider the inertia force of the added mass, the mass matrix of the i^{th} element in the element-fixed frame is given by Equation (5), where C_A is the added mass coefficient. As the cable is continuous in the z-direction, the added mass is ignored along the cable axis.

$$\underline{\underline{M}}_b^i = \begin{pmatrix} \frac{\pi d_c^2}{4} l_0 \rho_c + C_A \frac{\pi d_c^2}{4} l_0 \rho_f & 0 & 0 \\ 0 & \frac{\pi d_c^2}{4} l_0 \rho_c + C_A \frac{\pi d_c^2}{4} l_0 \rho_f & 0 \\ 0 & 0 & \frac{\pi d_c^2}{4} l_0 \rho_c \end{pmatrix}. \tag{5}$$

The i^{th} node mass matrix $\underline{\underline{M}}_g^i$ is expressed using the mass matrix of the elements $\underline{\underline{M}}_b^{i-1}$ and $\underline{\underline{M}}_b^i$, as shown in Equation (6). The transformation matrix $\underline{\underline{A}}^i$ is used to convert the mass matrices from the element-fixed frame to the global frame.

$$\underline{\underline{M}}_g^i = \frac{1}{2}\underline{\underline{A}}^{i-1}\underline{\underline{M}}_b^{i-1}\underline{\underline{A}}^{(i-1)T} + \frac{1}{2}\underline{\underline{A}}^i\underline{\underline{M}}_b^i\underline{\underline{A}}^{iT}. \tag{6}$$

Finally, the element loads are shared equally by two terminal nodes. Moreover, the governing equation for the i^{th} node is determined by the loads acting on the $(i\text{-}1)^{th}$ and i^{th} elements, as given in Equation (7). $\ddot{\underline{N}}_g^i$ represents the acceleration of the i^{th} node.

$$\begin{aligned} \underline{\underline{M}}_g^i \ddot{\underline{N}}_g^i = \underline{\underline{A}}^i(\underline{T}_c^i + \underline{D}_c^i + \tfrac{1}{2}\underline{F}_D^i) + \tfrac{1}{2}(\underline{F}_W^i + \underline{F}_W^{i-1}) - \\ \underline{\underline{A}}^{i-1}(\underline{T}_c^{i-1} + \underline{D}_c^{i-1} - \tfrac{1}{2}\underline{F}_D^{i-1}). \end{aligned} \tag{7}$$

Because the weight is balanced with the buoyancy, the floater introduces only additional mass effects and hydrodynamic forces at the node position where the floater is mounted. Considering the mass of the DTU and floater, an additional mass of 2.87 kg is added to the mass matrix. The ocean current in the deep sea is very small, so the velocity and acceleration of the water particles are ignored here. Because the towing speed changes smoothly during formal towing, the additional mass effect of seawater is also ignored here. Finally, the additional drag force introduced by the floater is expressed by Equation (8), which is a function of the cross area and towing speed. C_{tb} and C_{nb} are the tangential and normal drag coefficients, respectively. A_{tb} and A_{nb} are the cross areas of the floater in the tangential and normal directions, respectively. The node velocity in the element frame is used to carry out the drag force, and $V_b^{i,x}$, $V_b^{i,y}$, and $V_b^{i,z}$ are the x-, y- and z-components of the node velocity at the i^{th} floater, respectively. $F_{bd}^{i,x}$, $F_{bd}^{i,y}$, and $F_{bd}^{i,z}$ are the x-, y-, and z-components of the additional drag generated by the floater in the element frame, respectively. Finally, these forces are converted to the global frame.

$$\begin{cases} F_{bd}^{i,x} = -\frac{1}{2}C_{tb}\rho_f A_{tb}\left|V_b^{i,x}\right|V_b^{i,x}, \\ F_{bd}^{i,y} = -\frac{1}{2}C_{nb}\rho_f A_{nb}\left|V_b^{i,y}\right|V_b^{i,y}, \\ F_{bd}^{i,z} = -\frac{1}{2}C_{nb}\rho_f A_{nb}\left|V_b^{i,z}\right|V_b^{i,z}. \end{cases} \tag{8}$$

The modeling parameters of the arrays and floaters are listed in Tables 1 and 2, respectively. The drag coefficients of the floaters are measured via computational fluid dynamics simulation.

TABLE 1 Modeling parameters of the arrays.

Parameter	d_c (m)	L (m)	ρ_c (kg/m^3)	E	C_d	C_n	C_t	C_A
Value	0.046	210	1025	2.1e8	100	1	0.01	1

TABLE 2 Cross-sectional area and drag coefficient of the floaters.

Parameter	A_{tb} (m^2)	A_{nb} (m^2)	C_{tb}	C_{nb}
Value	0.012	0.015	0.74	1.24

2.3.3 Correction effect

An array model with floaters was established, and the array shape under various towing conditions was studied via numerical simulation. The towing conditions include rising, diving and flat towing, as shown in Figure 8A. Because the floaters balance the weight of the DTUs, the arrays no longer form a "W" shape but become a simple convex curve, as illustrated in Figure 8. Because of the seismic source on the convex curve, a reasonable sound speed can be easily adjusted with the simple convex curve.

The arrays equipped with floaters were used in the 2021 sea trial, and the location is indicated by the red point shown in Figure 9. The sea conditions were calm with ripples, and the water depth was approximately 1400-1600 m. The seismic records from the 2019 sea trial are shown in Figures 10A and C. Because floaters are not installed, the seismic records indicate that the sea surface reflection waves have broken lines and that the sea seafloor reflection waves exhibit a "W" shape. However, when the arrays are equipped with floaters, the seismic records exhibit simple convex curves, as shown in Figures 10B and D.

Two examples of sea surface reflection (SSR) records (Figures 11B, C) and travel-time positioning results (Figures 11D, F) are presented in Figure 11, revealing smooth curves in both the

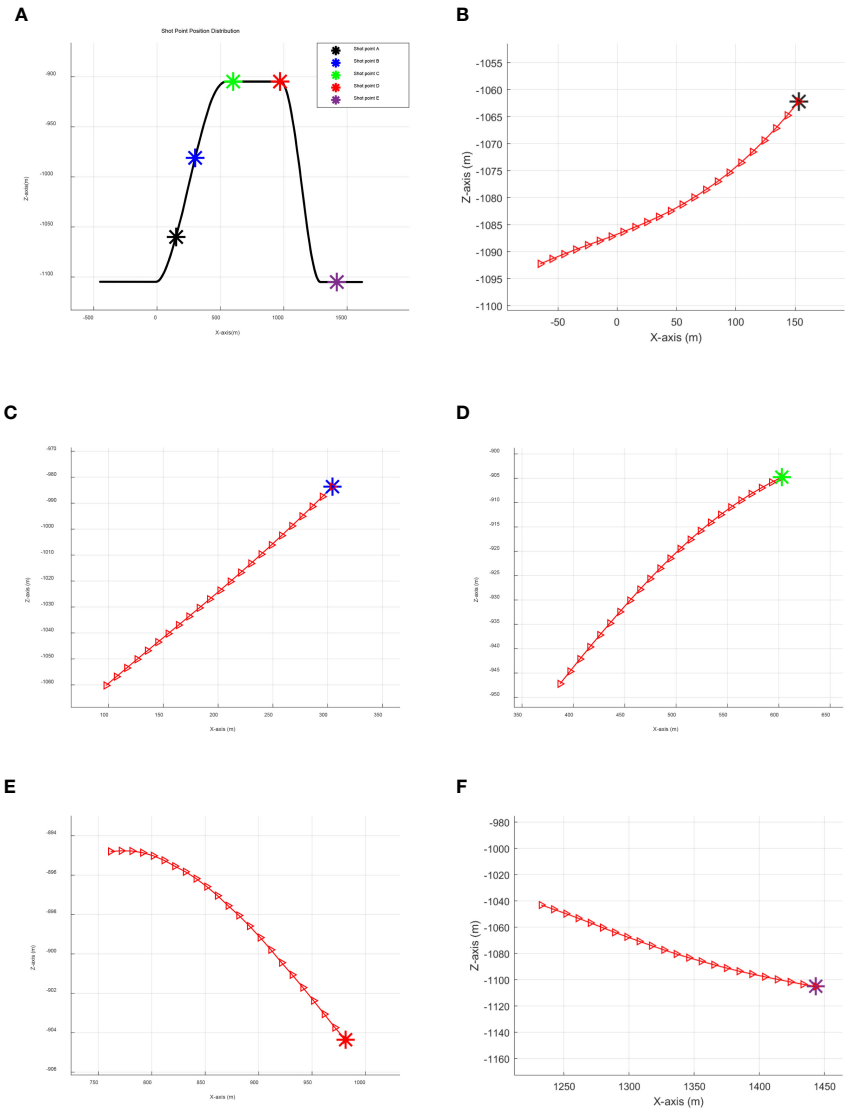


FIGURE 8
Array shape with floaters under different towing conditions. **(A)** Shot point distribution, **(B)** Array shape at shot **(A)**, **(C)** Array shape at shot **(B)**, **(D)** Array shape at shot **(C)**, **(E)** Array shape at shot **(D)**, **(F)** Array shape at shot **(E)**.

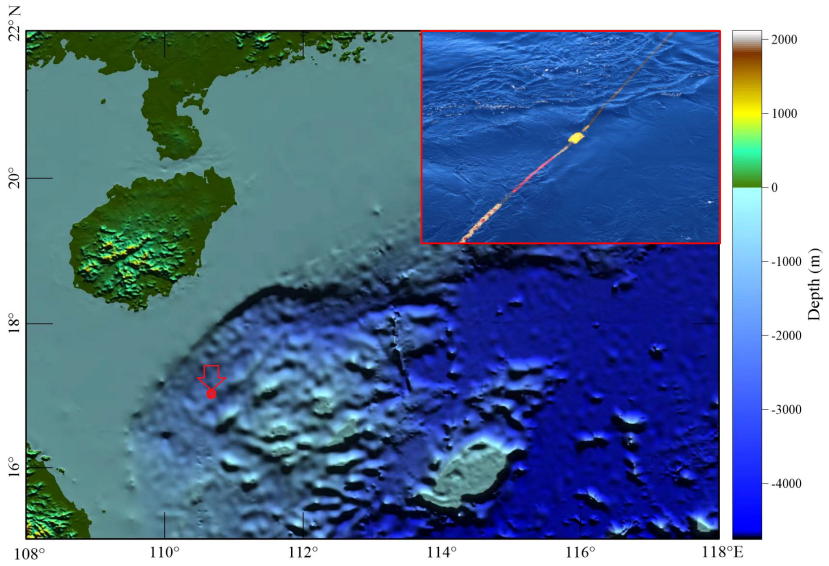


FIGURE 9
Location and water depth in the 2021 sea trial.

morphology of the sea surface reflection waves and the positioning structure of the towed array. Although few hydrophones jump out (S2129), the arrays equipped with floaters indeed exhibit a convex curve. Abnormal hydrophones are generated by phase distortion (waveform change), which is the general case in seismic exploration. Importantly, the seismic source is not on this convex curve (green dashed line), which indicates that the sound speed of seawater should be optimized and that abnormal hydrophones should be eliminated when forming the shape curve of seismic arrays.

3 Location modification

Both the numerical simulation and travel-time positioning results reveal that the array shape equipped with floaters is a simple convex curve under stable towing conditions. Because the depth of the seismic source is determined by the depth sensor and because the sound speed affects the vertical displacement of the arrays, if the assumed sound speed is greater than or less than the real velocity, then the seismic source is not on the convex curve. According to the weighted least squares polynomial fitting method, the sound speed of seawater can be optimized so that the seismic source and the arrays are on the same convex curve.

For a series of given points, which are the position coordinates of the source and hydrophones, the n^{th}-order polynomial function $f(x)$ used for fitting is shown by Equation (8b) (Parrilo and Sturmfels, 2003).

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n . \tag{8b}$$

where a_0, a_1, … a_n are the polynomial coefficients. Except for the abnormal points, herein, all the hydrophones are used to fit this convex curve. The differences between the polynomial function $f(x)$ and the y_i values of the given points should be minimized, as shown in Equation (9) (Burden and Faires, 2010). w_i represents the weight coefficient, x_i represents the horizontal displacement of the hydrophone relative to the towed vehicle, and y_i represents the vertical depth of the towed vehicle and hydrophones.

$$s(a_0, a_1, \cdots, a_n) = \min\left(\sum_{i=1}^{N} w_i [f(x_i) - y_i]^2\right). \tag{9}$$

Because the weighted least squares method is used to determine the polynomial function $f(x)$, this method is called weighted least squares polynomial fitting (Chen and Guo, 2023; Ke et al., 2023). The fitting quality can be evaluated by the goodness of fit R^2 and the root mean square error S, as shown in Equations (10) and (11), respectively.

$$R^2 = \frac{\sum_{i=1}^{N} [f(x_i) - \bar{y}]^2)}{\sum_{i=1}^{N} (y_i - \bar{y})^2)}, \tag{10}$$

$$S = \sqrt{\sum_{i=1}^{N} [y_i - f(x_i)]^2)}, \tag{11}$$

where $\bar{y}$ is the average value of the hydrophone depth calculated by the travel-time positioning method $f(x_i)$ is the vertical depth of the hydrophone after fitting. The polynomial fitting becomes optimal when the goodness of fit R^2 is close to 1 or the root mean square error S is the minimum.

To demonstrate the feasibility of the proposed method, two shot points, S2329 and S2350, were selected for positioning optimization. The sea floor is approximately flat at these two points. The seismic records of the SSR waves from the sea trial are shown in Figure 12. Based on the sound speed at the sea floor measured by the CTD sensor, the positioning results of the sea floor reflection (SFR) waves are shown in Figure 13. Additionally, the SSR positioning results of

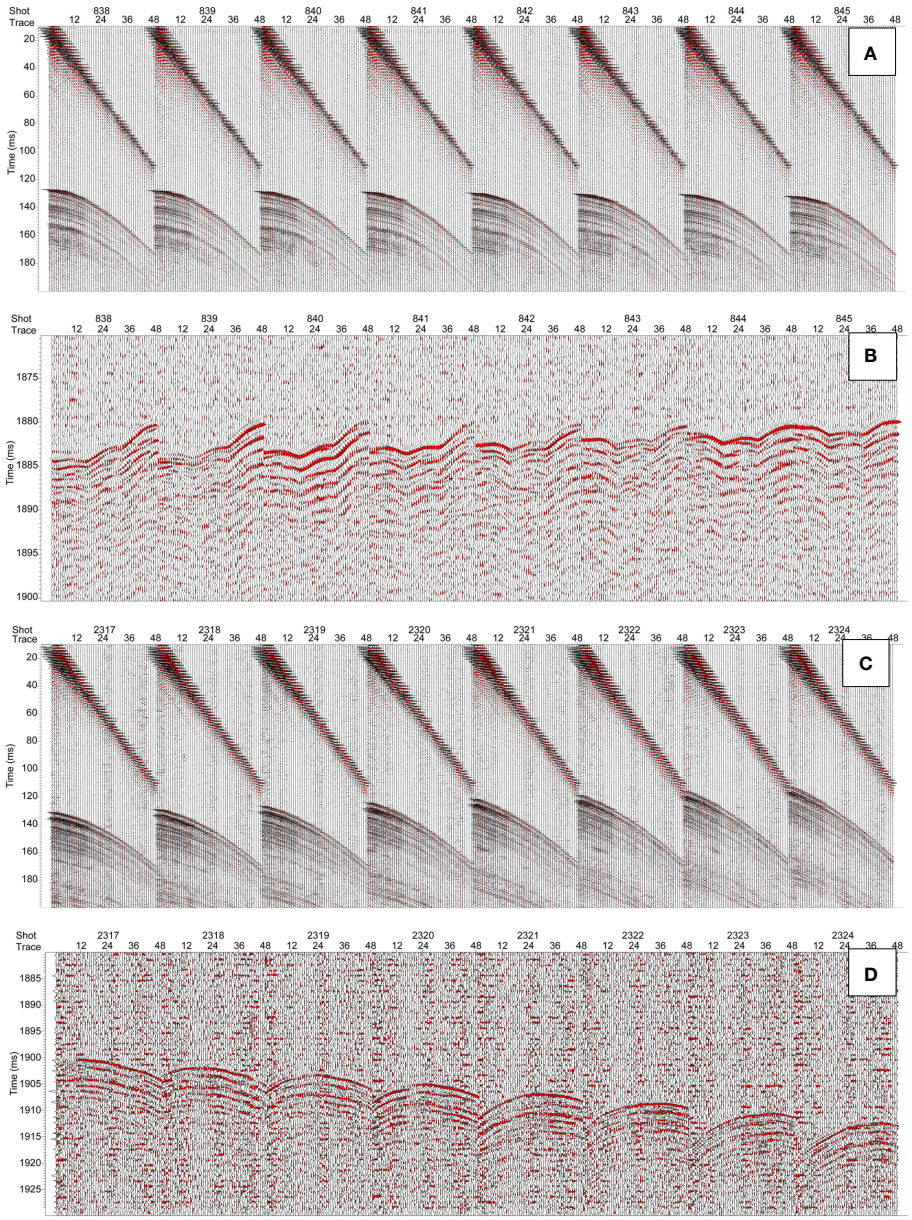


FIGURE 10
Recorded patterns in the 2019 (**A**: direct and seafloor reflection waves; **B**: sea surface reflection waves) and 2021 (**C**: direct and seafloor reflection waves; **D**: sea surface reflection waves) sea trials.

the SSR waves are also illustrated with the assumption that the average sound speed of seawater is 1484.5 m/s at the full depth of the survey area. The shape of the array cable generated by the SSR is similar to that generated by the SFR, but the seismic source is not on the extension line of the convex curve generated by the SSR. Because the sea floor is approximately flat at these two shots and because of the high speed of seawater at the sea floor measured by the CTD, the positioning results of the SFR are more accurate than those of the SSR. Therefore, the assumed average sound speed of seawater (v_w) at the full depth of the survey area, i.e., 1484.5 m/s, needs to be adjusted.

The steps to optimize the travel-time positioning are as follows:

Step 1: The range of the sound speed of seawater at full depth (v_{min}, v_{max}) is defined according to experience (v_{min}= v_w-5 m/s, and v_{max}= vw+5 m/s). The initial position of the arrays (x_h^i, d_r^i) is determined by using Equations (1) and (2) with v_w and v_f.

Step 2: Source coordinates (0, d_s) and hydrophone coordinates (x_h^i, d_r^i) are combined in a data set {(0, d_s), (x_h^1, d_r^1), (x_h^2, d_r^2), …, (x_h^{48}, d_r^{48})}, and the weighted least squares polynomial is subsequently used for fitting. The weight coefficient w is set to 0.01 for the abnormal hydrophones, which are indicated by phase distortion in Figure 11. However, the weight coefficient for the other hydrophones is 1.0. One large root mean square error S is derived if the seismic source is not on the curve of the arrays calculated with the current v_w, for example, ‘curve a’ or ‘curve c’, as shown in

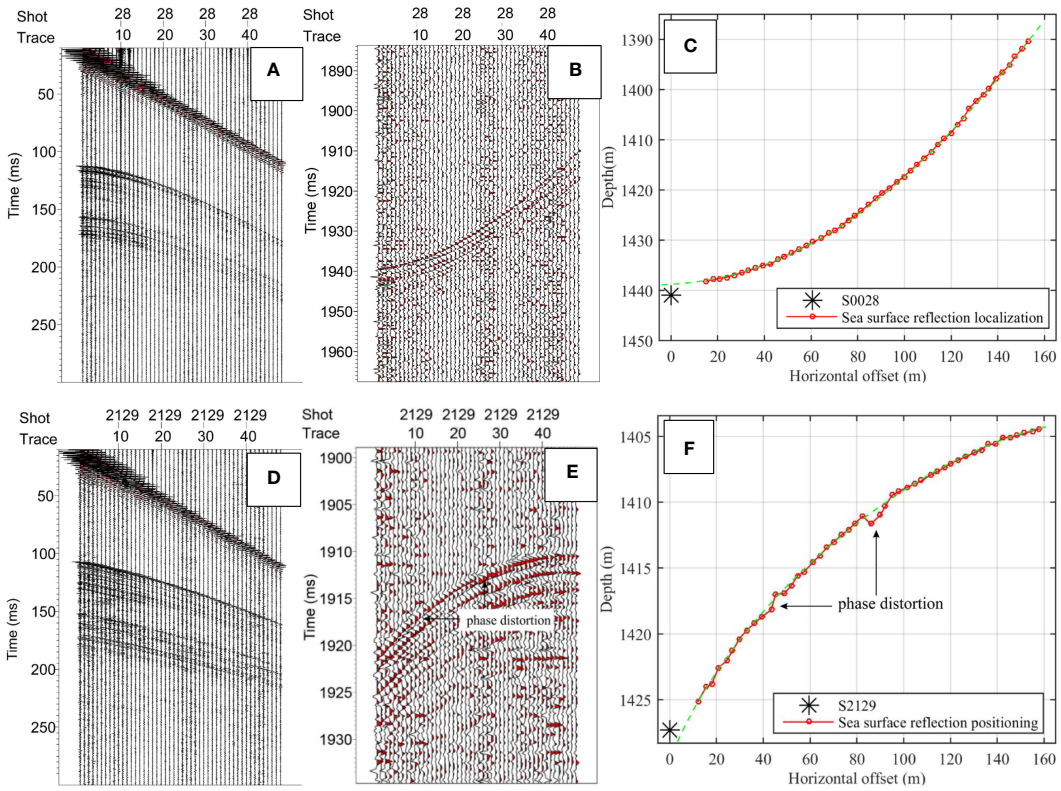


FIGURE 11
Wave field records and sea-surface reflection positioning results after array correction (**A**: direct and seafloor reflection wave at S0028; **B**: sea-surface reflection wave at S0028; **C**: sea-surface reflection positioning result at S0028; **D**: direct and seafloor reflection wave at S2129; **E**: sea surface reflection wave at S2129; and **F**: sea surface reflection positioning result at S2129).

Figure 14. The optimized accuracy depends on the order of the polynomial fit, which is defined in this step. Taking S2329 and S2350 as examples, Table 3 shows the relationships among the polynomial fitting order, goodness of fit R^2 and root mean square error S. As the fitting order increases, the goodness of fit approaches 1.0, and the root mean square error decreases. However, a long computation time is needed, and the stability of the fitting calculation decreases when the fitting order is larger than 4; thus, a fourth-order polynomial is sufficient herein. The polynomial function is $f(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4$.

Step 3: The sound speed v_i is increased by a step of 0.001 m/s from $v_{\min}$ to $v_{\max}$, and the goodness of fit R^2 and root mean square error S are calculated.

Step 4: The velocity v_i with the maximum goodness of fit (R^2) or the minimum root mean square error (S) is taken as the optimal sound speed for the final calculation of the array position (horizontal offset $x_{ro}{}^i$ and vertical depth $d_{ro}{}^i$). Moreover, the fitted value $f(x_i)$ is obtained. Because all the points $\{(0, d_s), (x_{ro}{}^i, f(x_i))\}$ are on a convex curve, such as curve b, whose extension (dotted green line) passes through the shot point, $f(x_i)$ is taken as the final hydrophone depth

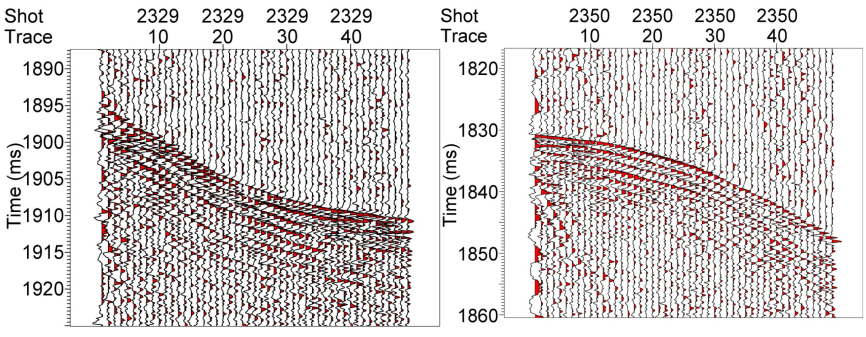


FIGURE 12
Seismic records from sea surface reflections at shots S2329 and S2350.

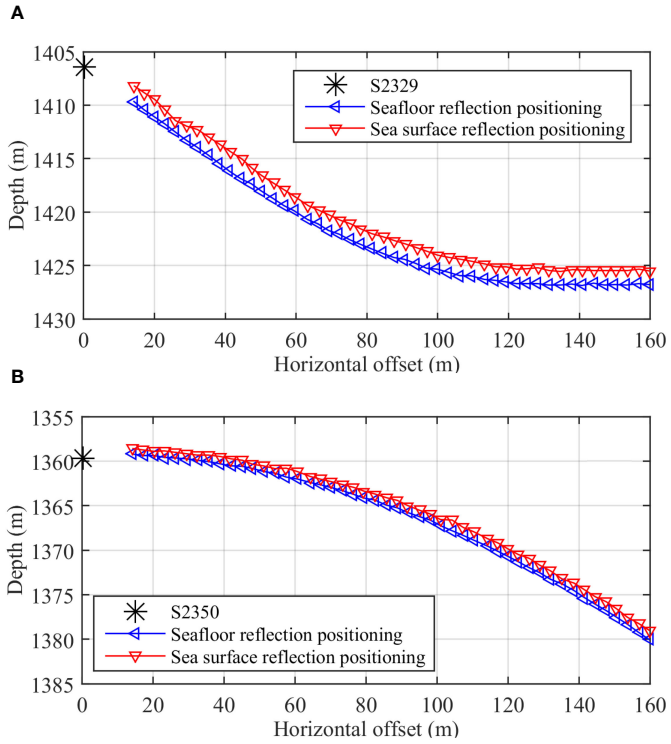


FIGURE 13
Positioning results calculated by the seafloor reflection wave and sea surface reflection wave. **(A)** Comparison of sea-surface reflection wave and seafloor reflection wave positioning results for shot point S2329, **(B)** Comparison of sea-surface reflection wave and seafloor reflection wave positioning results for shot point S2350.

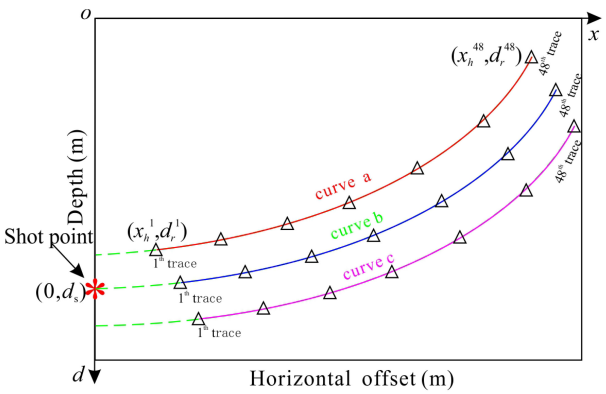


FIGURE 14
Polynomial fitting diagram.

TABLE 3 Relationships between the fitting order and fitting quality for shots S2329 and S2350.

Order of Polynomial		*n*=2	*n*=3	*n*=4	*n*=5
S2329	R^2	0.9991	0.9991	0.9995	0.9995
	S	0.2062	0.2051	0.1455	0.1453
S2350	R^2	0.9997	0.9998	0.9998	0.9998
	S	0.1046	0.1001	0.0988	0.0987

TABLE 4 Polynomial fitting parameters at shot 2329.

R^2	S	v_o (m/s)	d_s (m)	a_0	a_1	a_2	a_3	a_4
0.9995	0.1355	1485.305	1406.5	1406.499	0.216	-0.012	1.993×10^{-5}	5.782×10^{-8}

value, which also solves the phase distortion problem of abnormal hydrophones and reduces the influence of ocean waves.

According to the above processes, the array shape at shot S2329 was fitted by using a fourth-order polynomial, i.e., $f(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4$. The coefficients and fitting parameters of the polynomial are shown in Table 4. The goodness of fit is 0.9995, and the root mean square error is only 0.1355. Theoretically, the seismic source depth ds is the intercept of the fitting polynomial on the depth axis; thus, it is equal to the constant term a_0. The fitting result a_0 shown in Table 4 is approximately equal to the seismic source depth d_s, indicating that the fitting quality is high.

The comparison between the optimization results and the initial positioning results is shown in Figure 15, and the root mean square error and goodness of fit are shown in Figure 16. Compared to that of seafloor reflection positioning, the depth errors in SSR positioning before and after polynomial fitting are shown in Figures 17A and B, respectively. These results indicate that the polynomial fitting method is accurate and that the absolute values of the errors are less than 0.15 m (~0.1 ms).

Similarly, the array shape at shot point S2350 was fitted by the proposed method. The parameters of the fitted polynomials are shown in Table 5. The goodness of fit is 0.9940, and the root mean square error is only 0.0660. In addition, the fitted polynomial a_0 is equal to the seismic source depth d_s, indicating that the fitting is optimal.

The array shapes obtained by the polynomial fitting method and by the assumed average speed of 1484.5 m/s are shown in Figure 18. The root mean square error and goodness of fit are shown in Figure 19. Compared to those of the seafloor reflection positioning, the depth errors in SSR positioning before and after polynomial fitting are shown in Figures 20A and B, respectively. The absolute values of the depth errors are less than 0.14 m (~0.09 ms) after polynomial fitting.

Additionally, the positions of the arrays at shot points S0028 and S2129 were also corrected according to the above processes. The

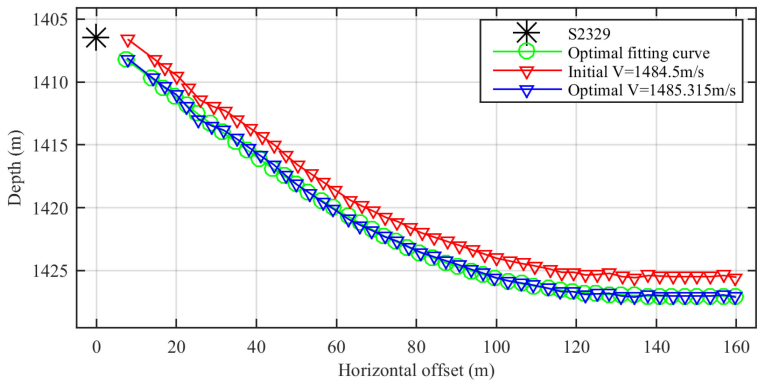


FIGURE 15
Travel-time positioning before and after correction at shot S2329.

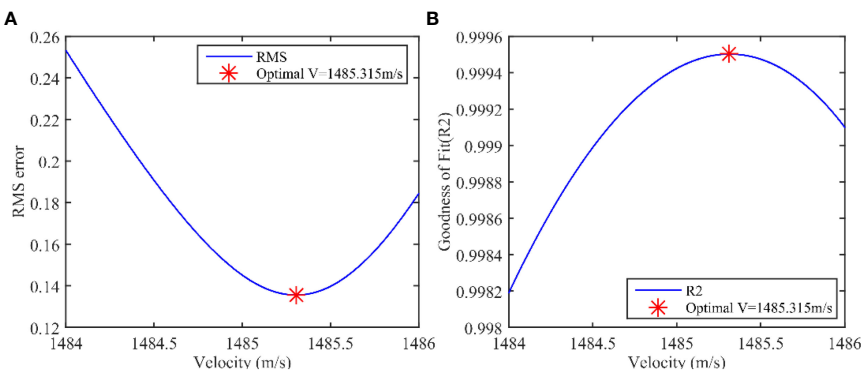


FIGURE 16
Fitting quality at shot S2329. **(A)** Root mean square error, **(B)** Goodness of fit.

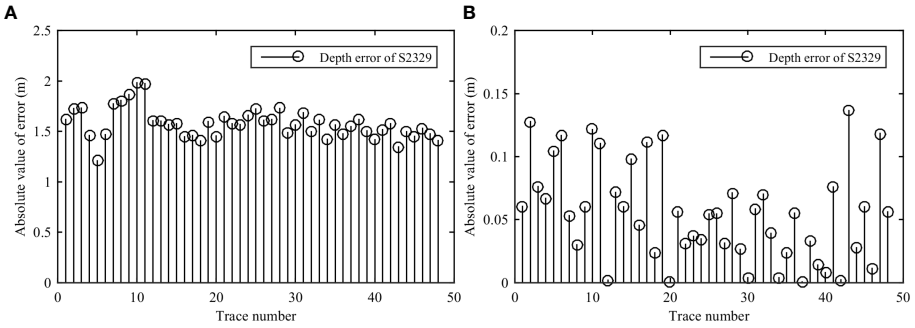


FIGURE 17
Depth error of sea surface reflection positioning compared to seafloor reflection positioning at shot S2329. **(A)** Before optimization, **(B)** After optimization.

TABLE 5 Polynomial fitting parameters at shot point S2350.

R^2	S	v_o (m/s)	d_s (m)	a_0	a_1	a_2	a_3	a_4
0.9997	0.092	1485.241	1359.7	1359.699	0.236	0.010	4.177×10^{-4}	-5.615×10^{-9}

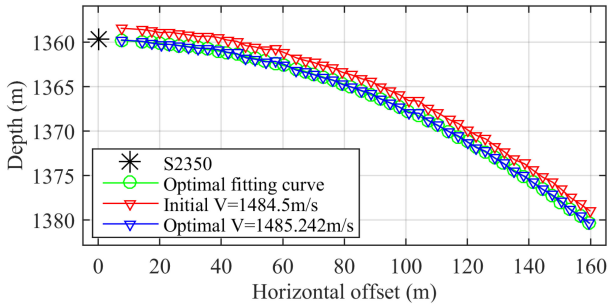


FIGURE 18
Travel-time positioning before and after correction at shot S2350.

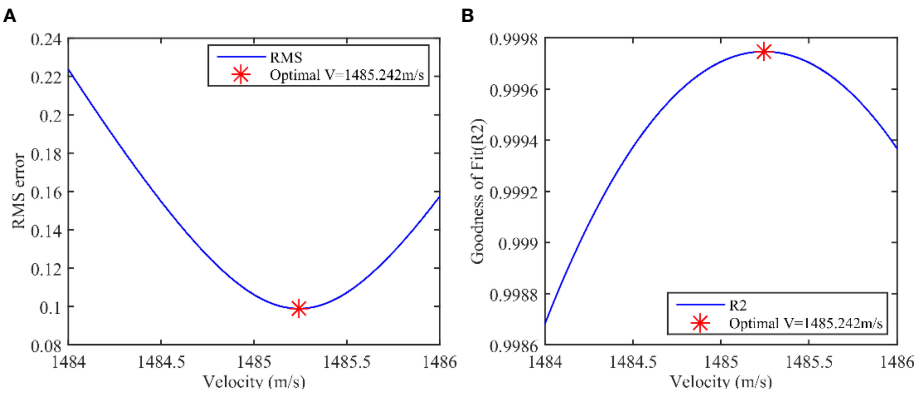


FIGURE 19
Fitting quality at shot point S2350. **(A)** Root mean square error, **(B)** Goodness of fit.

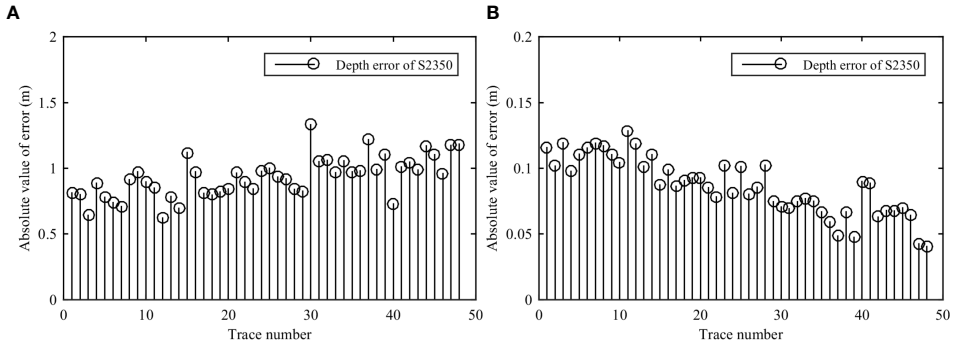


FIGURE 20
Depth error of sea surface reflection positioning compared to seafloor reflection positioning at shot point S2350. **(A)** Before optimization, **(B)** After optimization.

TABLE 6 Polynomial fitting parameters at shot points S0025 and S2129.

Shot point	R^2	S	d_s (m)	a_0	a_1	a_2	a_3	a_4
S0028	0.9999	0.1668	1440.9	1440.907	-0.008	-0.002	1.645×10^{-6}	-2.610×10^{-10}
S2129	0.999	0.2322	1427.3	1427.319	-0.271	9.305×10^{-4}	-8.836×10^{-7}	-7.133×10^{-10}
S2390	0.9994	0.0920	1434.2	1434.200	-0.052	–	–	–

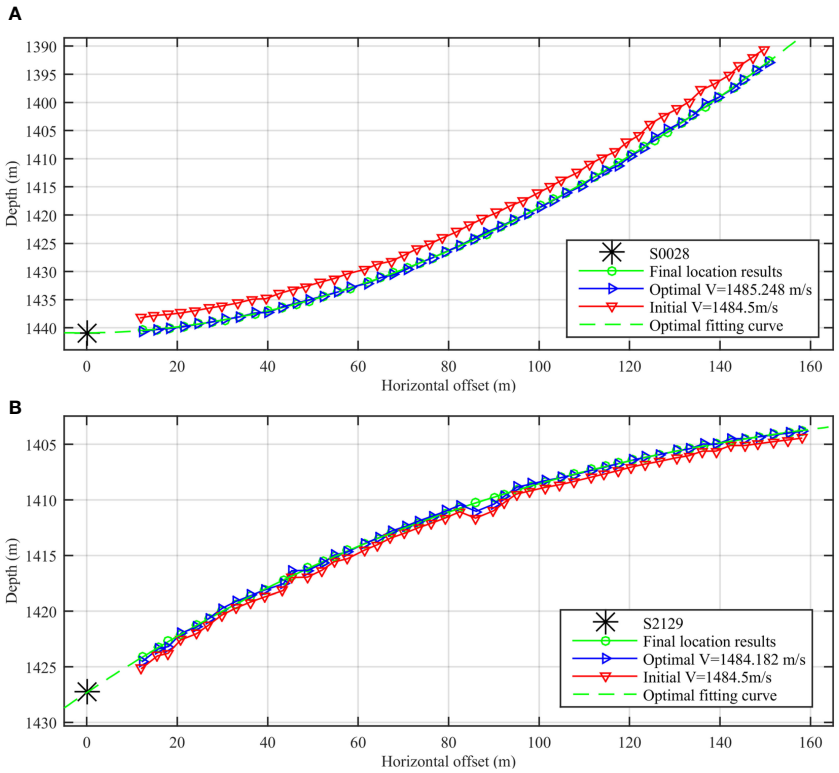


FIGURE 21
Travel-time positioning before and after correction. **(A)** Comparison of the corrected sea-surface reflection wave positioning results and seafloor reflection wave positioning results for shot point S0028, **(B)** Comparison of the corrected sea-surface reflection wave positioning results and seafloor reflection wave positioning results for shot point S2129.

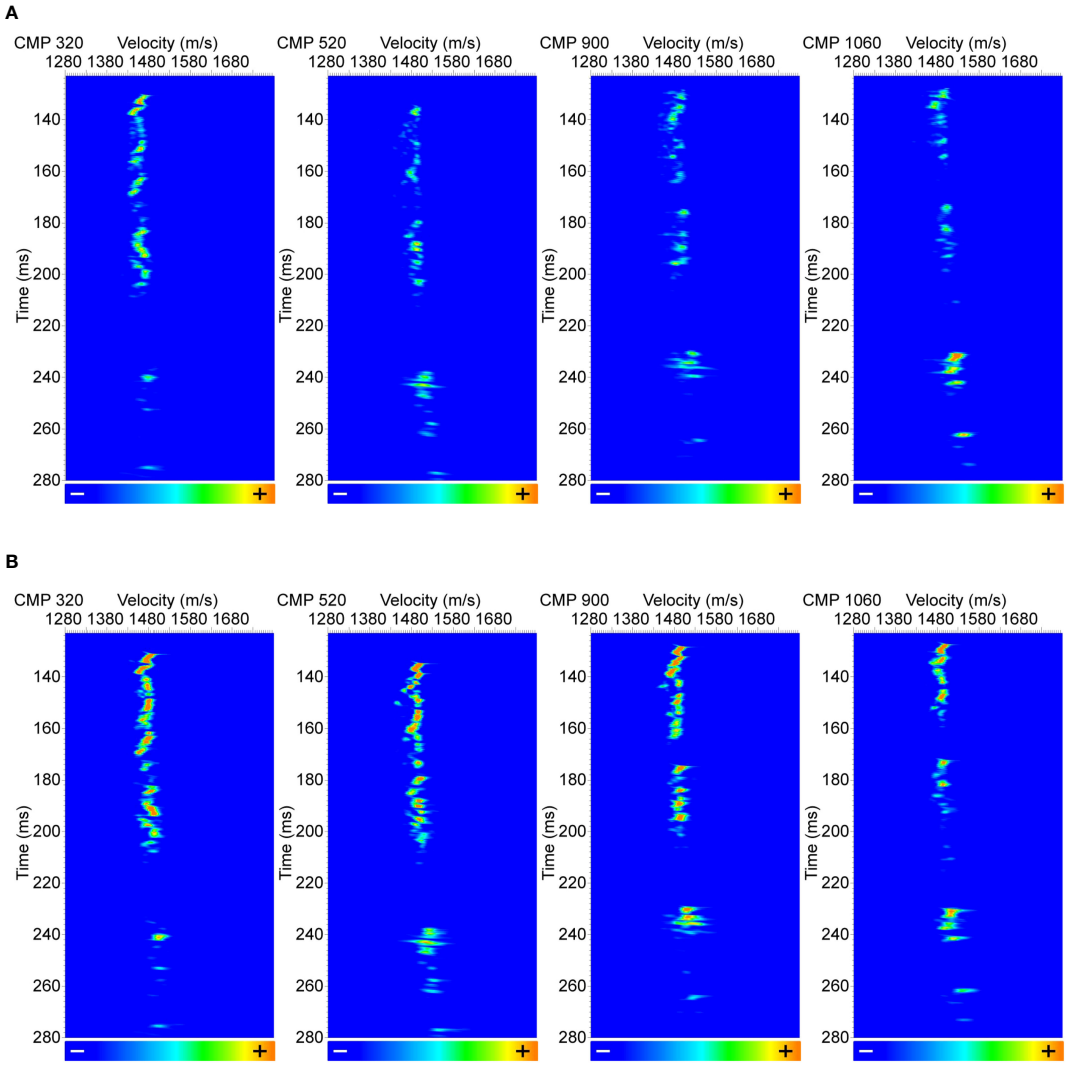


FIGURE 22
Velocity semblance panels before and after travel-time positioning correction. **(A)** Before travel-time positioning correction. **(B)** After travel-time positioning correction.

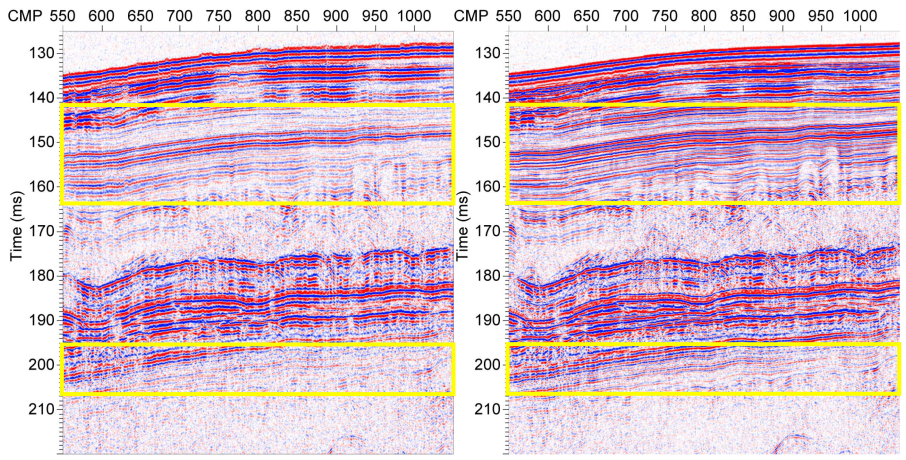


FIGURE 23
Seabed stack profile before (left) and after (right) travel-time positioning correction.

polynomial fitting parameters, goodness of fit R^2 and root mean square error S for the array positioning are listed in Table 6. The seismic source is not on the convex curve, generating an assumed velocity of 1484.5 m/s, as shown in Figure 21. However, the sound speed of seawater is optimized by the polynomial fitting method, and the seismic source and the positioning results of the arrays are on the same convex curve (dashed green line). In addition, although there are phase distortions in some channels at shot S2129, the weighted least squares polynomial method overcomes inaccurate positioning.

4 Application

The prestack shot gathers in the sea trial with Kuiyang-ST2000 conducted in 2021 in the Qiongdongnan area in the South China Sea are selected, as shown in Figure 8. The main acquisition parameters of the deep-towed multichannel seismic datasets are as follows: the source energy is 3000 J, shot spacing is 6.25 m, trace spacing is 3.125 m, minimum offset is 12.5 m, trace number is 48, recording duration is 3000 ms, and sampling rate is 8 kHz.

We utilize the SSR travel-time positioning method described in Section 2.2 to obtain the initial values of the towed array. Based on the characteristics of the towed array structure described in Section 2.3, we applied the correction method outlined in Chapter 3 to determine the accurate positions of the hydrophone array in the towed cable. Figure 22A shows the velocity spectra of common middle point (CMP) gathered on survey lines corrected by floating datum (with reference to the datum setting as land complex surface data processing; the shot point and hydrophone point are placed on a relatively smooth float datum) in seismic data processing (Yilmaz, 2001). The focusing of the energy group of the velocity spectrum calculated by the semblance coefficient method (Kirlin, 1992) is not highly concentrated in either the deep layer or the shallow layer before positioning correction. The focusing of the energy group in the velocity spectrum is highly concentrated in both the deep layer and shallow layer after correction, as shown in Figure 22B, and the resolution of the velocity spectrum is notably improved. The preliminary stack profiles before and after correction are shown in Figure 23. The phase continuity and signal-to-noise ratio are obviously improved after correction, as shown in Figure 23B, compared with the uncorrected (Figure 23A), especially within the yellow boxes, and this approach can accurately describe the geological structure morphology under the seabed.

5 Conclusions

To correct for the "W" shape of seismic arrays, floaters were installed at the DTUs of the Kuiyang-ST2000 deep-towed seismic system. The corrected arrays presented a simple smooth convex curve during towing conditions, which was demonstrated by both the numerical simulation and travel-time positioning of the sea trial. Given the geometrical relationship between towed vehicles and seismic arrays, the weighted least squares polynomial fitting method is proposed herein to optimize the average sound speed of seawater at full ocean depth. The final positioning accuracy (0.15 m/~0.1 ms) was less than the system sampling interval of 0.125 ms, which met the accuracy requirements of the Kuiyang-ST2000 deep-towed seismic system. Additionally, the energy groups of the velocity spectrum were also highly concentrated in both the deep layer and shallow layer after positioning correction, and the signal-to-noise ratio and resolution of superposition imaging were effectively improved in the seabed stack profile. The conclusions of general validity are listed here for reference.

(1) The array shape follows a simple smooth convex curve when the additional weights of the DTUs are balanced by floaters.

(2) Weighted least squares polynomial fitting is an efficient method for obtaining the accurate average sound speed of seawater at full ocean depth under the constraint that the seismic source is located on a smooth convex curve.

(3) Because seafloor reflection positioning is limited by the constraint that the seafloor is flat, sea surface reflection positioning can be widely used when the average sound speed at full ocean depth is known.

(4) The phase distortions generated by abnormal hydrophones can be repaired via the proposed polynomial fitting method.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

ZW: Data curation, Writing – original draft, Writing – review & editing, Formal analysis, Investigation, Methodology, Resources, Visualization. YP: Data curation, Funding acquisition, Project administration, Resources, Writing – original draft, Writing – review & editing. XQZ: Formal analysis, Methodology, Writing – review & editing. KL: Data curation, Resources, Writing – review & editing. XBZ: Formal analysis, Investigation, Visualization, Writing – review & editing. LZ: Writing – review & editing. XL: Investigation, Software, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key Research and Development Program of China (No. 2023YFC2811200, No. 2016YFC0303901); National Natural Science Foundation of China (Grant Nos. 51909145 and 42106072); Laoshan Laboratory (No. LSKJ202203604).

Acknowledgments

The authors would like to acknowledge FunctionBay Inc. and Prometech Software Inc. for providing license files of RecurDyn and Particleworks, respectively.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Burden, R. L., and Faires, J. D. (2010). *Numerical Analysis. 9th ed* (Brooks Cole). Boston: Cengage Learning Press.

Chapman, N. R., Gettrust, J. F., Walia, R., Hannay, D., Spence, G. D., Wood, W. T., et al. (2002). High-resolution, deep-towed, multichannel seismic survey of deep-sea gas hydrates off western Canada. *Geophysis* 67, 1038–1047. doi: 10.1190/1.1500364

Che, Z., Wang, J., Zhu, J., Zhang, B., Zhang, Y., and Wu, Y. (2021). Real-time array shape estimation method of horizontal suspended linear array based on non-acoustic auxiliary sensors. *IEEE Access.* 9, 90500–90509. doi: 10.1109/ACCESS.2021.3061446

Chen, H., and Guo, L. (2023). Convergence of adaptive MPC for linear stochastic systems. *Sci. CHINA-INFORMATION Sci.* 66, (5). doi: 10.1109/ACCESS.2021.3061446

Colin, F., Ker, S., and Marsset, B. (2020). Fine-scale velocity distribution revealed by datuming of very-high-resolution deep-towed seismic data: Example of a shallow-gas system from the western Black Sea. *Geophysics* 85, B181–B192. doi: 10.1190/geo2019-0686.1

Eli, M. (2010). *The Pythagorean Theorem: A 4,000-Year History* (New Jersey: Princeton University Press).

Fagot, M. G., Gholson, N. H., Moss, G. J., and Milburn, D. A. (1980). Deep-towed geophysical array system development program review. *Navel Ocean Res. Dev. Activity NSTL Station MS.*

Fikes, A. C., Safaripour, A., Bohn, F., Abiri, B., and Hajimiri, A. (2019). Flexible, conformal phased arrays with dynamic array shape self-calibration. 2019 IEEE MTT-S International Microwave Symposium (IMS). *IEEE*, 1458–1461. doi: 10.1109/MWSYM.2019.8701107

He, T., Spence, G. D., Wood, W. T., Riedel, M., and Hyndman, R. D. (2009). Imaging a hydrate-related cold vent offshore Vancouver Island from deep-towed multichannel seismic data. *Geophysics* 74, B23–B36. doi: 10.1190/1.3072620

Howard, B. E., and Syck, J. M. (1992). Calculation of the shape of a towed underwater acoustic array. *IEEE J. ocea. eng.* 17, 193–203. doi: 10.1109/48.126976

Ke, K., Zhang, C. M., Wang, Y. Q., Zhang, Y. J., and Yao, B. L. (2023). Single underwater image restoration based on color correction and optimized transmission map estimation. *Masure. Sci. Techol.* 34, (5). doi: 10.1088/1361-6501/acb72d

Ker, S., Marsset, B., Garziglia, S., Le Gonidec, Y., Gibert, D., Voisset, M., et al. (2010). High-resolution seismic imaging in deep sea from a joint deep-towed/OBH reflection experiment: application to a Mass Transport Complex offshore Nigeria. *GEOPHYS. J. Int.* 182, 1524–1542. doi: 10.1111/j.1365-246X.2010.04700.x

Kirlin, R. L. (1992). The relationship between semblance and eigenstructure velocity estimators. *Geophysics*. 57 (8), 1027–1033. doi: 10.1190/1.1443314

Kong, F., He, T., and Spence, G. D. (2012). Application of deep-towed multichannel seismic system for gas hydrate on mid-slope of northern Cascadia margin. *Sci. China-Earth Sci.* 55, 758–769. doi: 10.1007/s11430-012-4377-4

Krödel, S., Thomé, N., and Daraio, C. (2015). Wide band-gap seismic metastructures. *Extreme Mech. Lett.* 4, 111–117. doi: 10.1016/j.eml.2015.05.004

Lan, H., and Ding, Y. (2012). Ordering, positioning and uniformity of quantum dot arrays. *Nano Today* 7, 94–123. doi: 10.1016/j.nantod.2012.02.006

Lu, F., Milios, E., Stergiopoulos, S., and Dhanantwari, A. (2003). New towed-array shape-estimation scheme for real-time sonar systems. *IEEE J. ocea. eng.* 28, 552–563. doi: 10.1109/JOE.2003.816694

Marsset, B., Ker, S., Thomas, Y., and Colin, F. (2018). Deep-towed high resolution seismic imaging II: Determination of P-wave velocity distribution. *Deep-Sea Res. Part I-Oceanogr. Res. Pap.* 132, 29–36. doi: 10.1016/j.dsr.2017.12.005

Marsset, B., Menut, E., Ker, S., Thomas, Y., Regnault, J. P., Leon, P., et al. (2014). Deep-towed High Resolution multichannel seismic imaging. *Deep-Sea Res. Part I-Oceanogr. Res. Pap.* 93, 83–90. doi: 10.1016/j.dsr.2014.07.013

Nanni, U., Roux, P., Gimbert, F., and Lecointre, A. (2022). Dynamic imaging of glacier structures at high-resolution using source localization with a dense seismic array. *Geophys. Res. Lett.* 49, e2021GL095996. doi: 10.1029/2021GL095996

Parrilo, P. A., and Sturmfels, B. (2003). Minimizing polynomial functions. Algorithmic and quantitative real algebraic geometry, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 60, 83–99.

Pei, Y., Wen, M., Zhang, L., Yu, K., Kan, G., Zong, L., et al. (2022). Development of a high-resolution deep-towed multi-channel seismic exploration system: Kuiyang ST2000. *J. Appl. Geophys.* 198, 104575. doi: 10.1016/j.jappgeo.2022.104575

Rowe, M. M., and Gettrust, J. F. (1993). Faulted structure of the bottom simulating reflector on the Blake Ridge, Western North-Atlantic. *Geology* 21, 833–836. doi: 10.1130/0091-7613(1993)021<0833:FSOTBS>2.3.CO;2

Varypaev, A., and Kushnir, A. (2020). Statistical synthesis of phase alignment algorithms for localization of wave field sources. *Multidimens. Syst. Signal Process.* 31, 1553–1578. doi: 10.1007/s11045-020-00722-3

Vickery, K. (1998). "Acoustic positioning systems. A practical overview of current systems," in *Proceedings of the 1998 Workshop on Autonomous Underwater Vehicles (Cat. No. 98CH36290)* (IEEE), 5–17. doi: 10.1109/AUV.1998.744434

Walia, R., and Hannay, D. (1999). Source and receiver geometry corrections for deep towed multichannel seismic data. *Geophys. Res. Lett.* 26, 1993–1996. doi: 10.1029/1999GL900402

Wei, Z., Pei, Y., and Liu, B. (2020). A new deep-towed,multi-channel high-resolution seismic system and its preliminary application in the South China Sea. *Oil Geophys. Prospect.* 55, 965–972. doi: 10.13810/j.cnki.issn.1000-7210.2020.05.004

Yilmaz, Ö. (2001). Seismic data analysis: Processing, inversion, and interpretation of seismic data. *Soc. Explor. geophys.* doi: 10.1190/1.9781560801580

Zhu, X., Wei, Z., Pei, Y., Yu, K., and Zong, L. (2020). Dynamic modeling and position prediction of deep-towed seismic array. *J. Shandong Univers. Eng. Sci.* 50, 9–16.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 31 July 2024
DOI 10.3389/fmars.2024.1365095

OPEN ACCESS

EDITED BY
Lei Kou,
Qilu University of Technology (Shandong Academy of Sciences), China

REVIEWED BY
Xiao Feng,
Nanjing University of Posts and Telecommunications, China
Jian Wang,
Kunming University of Science and Technology, China

*CORRESPONDENCE
ZhiWen Qian
zhiwenqian@tju.edu.cn

RECEIVED 03 January 2024
ACCEPTED 26 June 2024
PUBLISHED 31 July 2024

CITATION
Li J, Qian Z, Hong D and Zhai J (2024) Precise and low-complexity method for underwater Doppler estimation based on acoustic frequency comb waveforms. *Front. Mar. Sci.* 11:1365095. doi: 10.3389/fmars.2024.1365095

Precise and low-complexity method for underwater Doppler estimation based on acoustic frequency comb waveforms

Jie Li[1], ZhiWen Qian[1,2]*, DeYue Hong[1] and JingSheng Zhai[1]

[1]School of Marine Science and Technology Tianjin University, Tianjin, China, [2]Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Nature Resources, Guangzhou, China

Ocean observation has advanced rapidly in recent decades due to its crucial role in resource exploration and scientific research, with the Doppler factor being widely utilized. However, the precision of Doppler estimation is frequently constrained by frequency resolution. Traditional frequency estimation methods using single-tone signals face considerable challenges with low accuracy and poor robustness. In response, this paper introduces a novel Doppler-sensitive Acoustic Frequency Comb (AFC) for estimating the Doppler factor, enabling multiple measurements with a single transmission and reception of the signal. The proposed Combined Uneven Uncertainty (CUU) method based on AFC achieves a bias of less than 1.1×10^{-5}, significantly surpassing the optimal result of 3.2×10^{-5} attained by other frequency estimation methods in the absence of noise. Compared to traditional single-tone methods, the AFC approach improves spectral leakage performance and enhances estimation accuracy without increasing computational complexity. Experimental results demonstrate that the CUU method realizes a difference performance of less than 3.4×10^{-6}, notably lower than that of 3.2×10^{-5} induced by coherent spectral leakage in fast Fourier Transform (FFT).

KEYWORDS

underwater acoustic communication, acoustic frequency comb, Doppler estimation, spectral leakage, Fourier transform

1 Introduction

Merely 5% of the ocean is currently understood by humanity, necessitating the advancement of acoustic applications to address escalating human ocean activities, encompassing mineral mining and scientific research (Mikhail et al., 2014). Sonar, as the sole critical instrument facilitating long-distance transmission in the ocean, assumes a crucial role in ocean observation, but its capability is impeded by the complex oceanic environment, characterized by undulating sea surfaces, turbulent flows, ambient noise,

uneven seabeds, and the pervasive underwater Doppler effect. Wherein, underwater Doppler can induce drastic time-frequency shifts, which can seriously limit the performance of acoustic applications such as dynamic underwater location (Chan and Jardine, 1990), synthetic aperture sonar (SAS) imaging (Zhang et al., 2021; Zhang et al., 2024), underwater communication (Chen et al., 2015; Ahmad et al., 2018), and sound monitoring (Greene and Hendricks, 2015; Yang and Fang, 2021). Thus, accurate Doppler estimation and compensation are of great importance for the above acoustic applications. However, the Doppler estimation is difficult to measure precisely mainly due to the low sound speed underwater, the influence of ocean currents and waves, marine environmental noise, and multipath effects (Gong et al., 2020; Wan et al., 2020). Thus, accurate and strongly robust Doppler estimation methods remain the primary challenges for underwater sonar applications.

In this hot topic, various methods have been used to overcome challenges. In the time domain, some methods such as the maximum likelihood estimation algorithm (Rife and Boorstyn, 1974), the phase information of autocorrelation functions (Kay, 1989), the block Doppler method (Sharif et al., 2000), and the constructed ambiguity function (Sen and Nehorai, 2010) are successively proposed. These approaches are simple and easy to implement, but balancing estimation accuracy with computational complexity remains a further improvement. Another focus is the use of the fast Fourier transform (FFT) with the transformation into the frequency domain, thereby significantly enhancing the efficiency and performance of ocean observation systems (Yang, 2023; Zhang, 2023). To overcome the coherent pitfall of spectral leakage in FFT (Li and Chen, 2008), a rough estimation followed by a fine estimation is proved to be an effective Doppler estimation method, but with limited estimation accuracy, particularly in large deviation scenarios (Quinn, 1994; Macleod, 1998; Jacobsen and Kootsookos, 2007; Candan, 2011). Zero padding (Fang et al., 2012) and the iterative method (Aboutanios et al., 2005) were used to improve the estimation accuracy, but associated with a higher computational cost. Therefore, reconciling the contradiction between high-accuracy Doppler estimation and low-complexity computational processing proves to be a daunting challenge. There is a great need to explore an algorithm that provides precise Doppler estimation with low computational cost.

The optical frequency comb (OFC) invented by the Nobel Prize winners Hänsch and Hall (Jones et al., 2000; Hänsch, 2006; Hall, 2006), consists of frequency components evenly distributed in the frequency domain. It was successfully used in the wavelength calibration of astronomical spectrometers, and the measurement accuracy was significantly improved, making it possible to observe the Doppler phenomenon in astronomy, including the movement of planets and even the expansion of the universe (Braje et al., 2008; Steinmetz et al., 2008). Similar to the OFC generation, the acoustic frequency comb (AFC) has been extended to the distance measurement in the underwater acoustic field, achieving a precision of less than 50 μm (Wu et al., 2019). Despite that, the employment of the AFC signal for Doppler estimation remains to be discovered. In our work, the AFC signal is performed to estimate the Doppler factor. The quantitative relationship between the AFC and the Doppler factor in the time-frequency domains is derived theoretically. At the cost of bandwidth, conducting one measurement can acquire the phonon frequency shifts of multiple frequency components. Thus, the Doppler factor at the same time can be calculated multiple times, resulting in a precise Doppler estimation after carrying out mathematical statistics, and without any increment in computational expense.

2 Preliminaries

2.1 Properties of AFC signal

Unlike traditional narrow-band wave signals (e.g., continuous waves, CW) and wide-band wave signals (e.g., linear frequency modulation, LFM), AFC consists of a series of modes with the same amplitude, phase, and evenly distributed frequencies, as shown in Figure 1. In the time domain, the AFC signals have narrow pulse width, high stability of frequency, large instantaneous power and good coherence, which can be mathematically expressed by Equations 1–3:

$$s(t) = rect(\frac{t}{T})\sum_{n=1}^{m} AE_n cos(2\pi f_n t + \theta_n), \tag{1}$$

$$rect(\frac{t}{T}) = \begin{cases} 1, & 0 \le t \le T, \\ 0, t < 0 \ \ or \ \ t > T, \end{cases} \tag{2}$$

$$f_n = f_{ceo} + (n-1)f_{rep}, \tag{3}$$

where *rect* is the rectangular pulse function, t is the time, T is the pulse period, n is an integer, m is the number of modes, A is the power amplification factor, E_n is the amplitude of individual mode, f_n is the frequency of individual mode, θ_n is the initial phase, f_{ceo} is the initial frequency, f_{rep} is the interval frequency.

The strictly evenly distributed frequency intervals and the extremely precise amplitude and phase lead to the stability of the measurement results with f_{ceo} and f_{rep} well referenced to an Rb clock. Thus, through the coherent superposition of each highly stable frequency component, the pulse signal with a narrow pulse width and high peak power in the time domain can be formed. Each pulse contains all harmonic components and the corresponding frequency and phase information can be calculated by Fourier transform. The echo waveforms of AFC can be expressed as Equation 4:

$$r(t) = a \cdot s[(1+\Delta)(t - 2R/c)] + w(t), \tag{4}$$

where a is the gain coefficient of the echo intensity, Δ is the Doppler factor, R is the distance between the signal source and the moving target, c is the velocity of sound, and $w(t)$ is white noise (Equation 5).

10.3389/fmars.2024.1365095

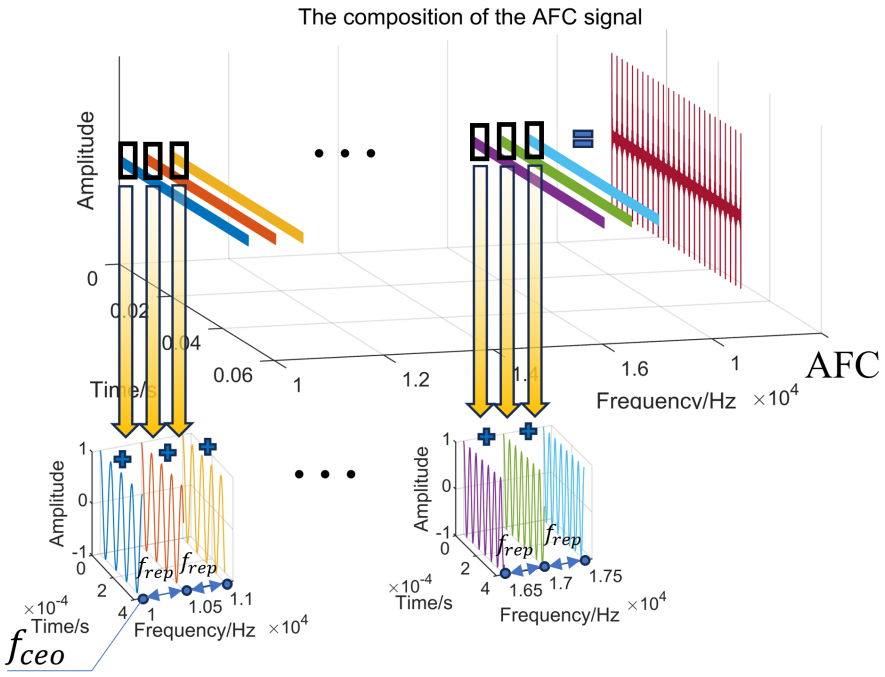


FIGURE 1
The composition of the AFC signal in the time and frequency domains.

$$\Delta = \frac{2v}{c - v} \approx \frac{2v}{c} \tag{5}$$

where v is the radial velocity of a moving target.

2.2 Time domain characterization of the Doppler factor

Assume that at time t_0, the pulse front of the signal hits a fast-moving target with speed v and reflects it back. After the time interval of t_d, i.e. at moment t_1, the pulse back edge hits the moving target. During this time, the target moves the distance vt_d. According to the quantitative relationship in Figures 2A, B, the following equations can be obtained:

$$c\tau = ct_d + vt_d, \tag{6}$$

$$c\tau' = ct_d - vt_d, \tag{7}$$

where c is the velocity of sound, τ and $\tau'$are the duration of the emitted and received pulse. From Equation 6 and Equation 7, it can be easily deduced as

$$\frac{\tau}{\tau'} = \frac{c + v}{c - v}. \tag{8}$$

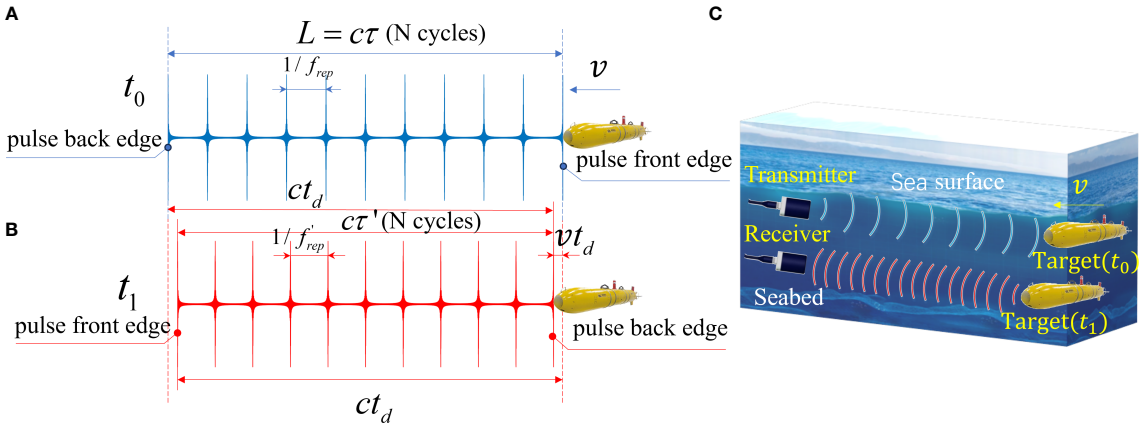


FIGURE 2
(A) The AFC waveform of the transmitted signal when it first touches the moving target; **(B)** The AFC waveform of the echo signal when it just leaves the moving target; **(C)** The time-domain compression phenomenon caused by the Doppler effect.

Suppose the number of periods of the transmitted signal is N_T, obviously with no change in two moments. The f_{rep} and $f_{rep}^{'}$ can be determined using the following equations:

$$f_{rep} = \frac{N_T}{\tau}, \tag{9}$$

$$f_{rep}^{'} = \frac{N_T}{\tau'}. \tag{10}$$

According to Equations 8–10, the equation can be deduced as Equation 11

$$f_{rep}^{'} = \frac{c+v}{c-v} f_{rep} = (1+\Delta) f_{rep}. \tag{11}$$

The direction of movement of the target is set as positive. As can be seen from Figure 2C, the Doppler effect leads to compression of the AFC signal in the time domain when the target moves toward the transmitter and the receiver. The severity of the Doppler effect can be reflected via the peak intervals of the signal envelope.

3 CUU estimator for the Doppler factor based on AFC waveforms

3.1 Spectrum analysis

Fourier transform is performed on the received signal,

$$X(\omega) = \mathcal{F}\{r(t)\}(\omega) = \int_{-\infty}^{\infty} r(t) e^{-j\omega t} dt. \tag{12}$$

According to the linear property and time-shift property of Fourier transform, the Equation 12 can be written as

$$\mathcal{F}\{r(t)\}(\omega) = a \cdot e^{-j\omega\frac{2R}{c}} \cdot \mathcal{F}\{s[(1+\Delta)t]\}(\omega). \tag{13}$$

Due to the time-domain scaling property of the Fourier transform, Equation 13 can be further rewritten as

$$\mathcal{F}\{r(t)\}(\omega) = e^{-j\omega\frac{2R}{c}} \cdot \frac{a}{1+\Delta} \cdot \mathcal{F}\{s(t)\}\left(\frac{\omega}{1+\Delta}\right). \tag{14}$$

Set $e^{-j\omega\frac{2R}{c}} \cdot \frac{a}{1+\Delta}$ as β, and make $\frac{\omega}{1+\Delta}$ as ω', and the Equation 14 can be written as

$$\mathcal{F}\{r(t)\}(\omega) = \beta \cdot \sum_{n=1}^{m} \mathcal{F}\left\{\cos\left[\frac{2\pi(1+\Delta) f_n t}{1+\Delta}\right]\right\}(\omega'). \tag{15}$$

ω' is changed to ω, and Equation 15 is simplified to

$$\mathcal{F}\{r(t)\}(\omega) = \beta \cdot \sum_{n=1}^{m} \mathcal{F}\{\cos(2\pi(1+\Delta) f_n t)\}(\omega). \tag{16}$$

Applying the discrete Fourier transform to Equation 16, we can get

$$X(k) = \beta \cdot \sum_{n=1}^{m} \sum_{i=0}^{N-1} \cos(2\pi(1+\Delta) f_n i T_s) e^{-j\frac{2\pi ki}{N}}, \tag{17}$$

where T_s is the sampling interval.

It can be seen from Equation 17, the frequency domain characteristics of the echo signal are the linear superposition of individual phonon signals. Since the phonon modes are Doppler-sensitive single-tone signals, the Doppler factor can be determined by each independent component. This is equivalent to obtaining multiple Doppler factor measurements from just one set of transmitting and receiving pulse signals. The frequency of the received signal can also be expressed as Equation 18:

$$f_{rn} = (1+\Delta_n) f_n. \tag{18}$$

According to the relationship between the repetition frequency and period of the AFC signal, it is also consistent with Equation 11.

The AFC signal consists of multiple signals with separate single-tone components, each affected by the Doppler effect, resulting in the Doppler frequency shift phenomenon. The frequency shift of each phonon mode can be directly calculated by $f_{dn} = f_{rn} - f_n$. f_{rn} can be quickly determined by the FFT of the received signals followed by peak detection, and f_n is the given information of the broadcast signal.

As can be seen from Figure 3A, the frequency shift between the first frequency component of the transmitted waveform and the

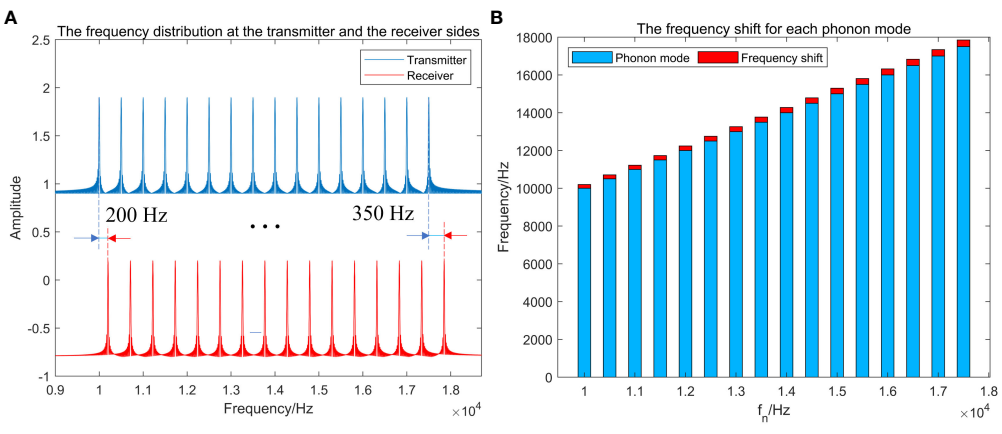


FIGURE 3
(A) The frequency distribution at both the transmitter and the receiver sides; **(B)** The frequency shift for each phonon mode.

received waveform is 200 Hz, and the last is 350 Hz. It varies linearly with the frequency of the emitted phonon mode (Figure 3B). The phonon modes of the transmitted signals differ from each other. Still, the variable size of the frequency shift f_{dn} for each frequency component can be determined, corresponding to the same Doppler factor. This means the Doppler factor can be calculated multiple times with just one measurement, leading to a more accurate estimate.

3.2 System modeling

Different frequency components with the same frequency resolution lead to different deviations for different phonons (Figure 4A). The variance of Δ_n can be deduced as: $D(\Delta_n) = 1/(12f_n^2)$, assuming that the frequency resolution is 1 Hz and the f_{dn} is evenly distributed in the range from -0.5 Hz to 0.5 Hz. To obtain a more precise estimate, we calculate the weight of each measurement by taking the inverse of the variance of Δ_n. The sum of these inverses is then utilized to standardize the weight of each measurement. Further derivation of the Doppler factor is as Equation 19

$$\Delta = \sum_{n=1}^{m} \frac{f_n^2}{\sum_{n=1}^{m} f_n^2} \Delta_n . \tag{19}$$

To further clarify the principle of the AFC-based Doppler estimation method, an AFC signal with an initial frequency of 100 kHz, an interval frequency of 500 Hz, a cutoff frequency of 200 kHz, and an initial phase of 0 rad for all phonon modes is constructed. The Doppler factor is set to 0.02 with a frequency resolution of 1 Hz, and the duration is set at 50 ms. The received signal in the time domain is transformed by FFT. The uncertainty of each individual phonon mode can be induced by the frequency resolution, and all Doppler shifts correspond to the same Doppler factor. This distortion can be eliminated by multiple unequal precision measurements (Figure 4B), a mode similar to the principle of the CUU estimation method.

As shown in Figure 5, as δ ranges from 0.05 to 0.45, which is the ratio of offset frequency to frequency resolution, the bias of single-tone signal after FFT does not exceed 3.2×10^{-5}. While the AFC-based method keeps a bias of less than 1.1×10^{-5}. Traditional Doppler factor estimation using single-tone waveforms is limited by the number of samples of the Fourier transform N, which inevitably leads to the spectral leakage. However, the use of AFC waveforms can enable the acquisition of multiple different Doppler factor values across different phonon modes, which is equivalent to carrying out numerous measurements of the identical Doppler factor simultaneously. In this way, the utilization of the spectrum has been significantly improved. It can be further reduced by increasing the number of phonon modes, and there is almost no additional computational effort.

4 Numerical section

4.1 Experimental parameters

To demonstrate the effectiveness of the proposed method, two sets of Monte Carlo experiments were conducted using the methods listed in Table 1, which were repeated 1000 times. In the experiments, the duration of the single-tone signals and the AFC signal were both set at 50 ms. The AFC was set with a starting frequency of 10 kHz, a cutoff frequency of 17.5 kHz, and an interval frequency of 0.5 kHz, the frequency of the single-tone signal was set to 14 kHz, the sampling rate and the sampling number were set to 96 kHz. The conditions for two series of tests were set as follows:

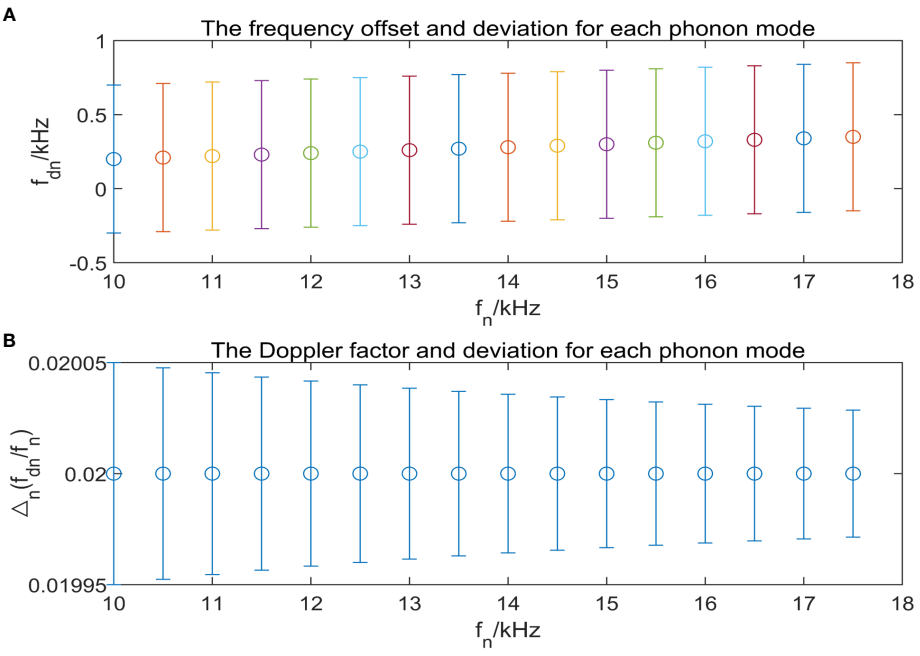


FIGURE 4
(A) The frequency offset and deviation for each phonon mode; **(B)** The Doppler factor and deviation with respect to the frequency of phonon mode.

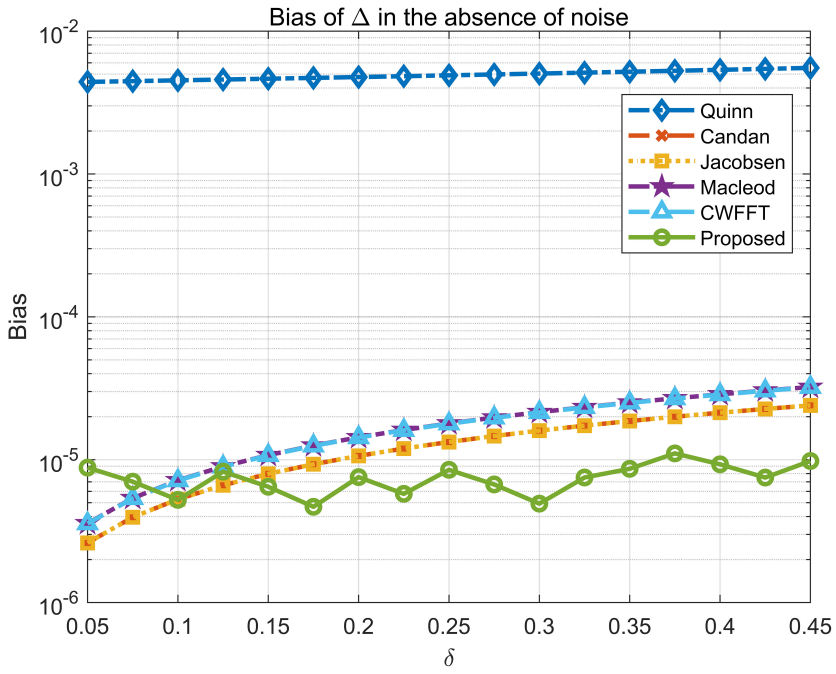


FIGURE 5
Bias of different estimators in the absence of noise.

1) Keep δ of the single-tone signals unchanged at 0.45, convert it to an equivalent Doppler factor, and gradually increase the input signal-to-noise ratio (SNR) from -2 dB to 10 dB. Observe the bias and the root mean square error (RMSE);

2) The number of sampling points was reduced to half of the previous, and other parameters remained the same as in experiment 1. Observe the RMSE with SNR.

The Cramer-Rao bound (CRB) can be used to indicate the best-estimated performance, although it is not typically applicable for biased estimators. It can be expressed by Equation 20

$$var(\hat{f}) \geq CRB = \frac{3f_s^2}{2\pi^2 N(N^2-1)SNR}, \tag{20}$$

where f_s is the sampling frequency. The variance of the Doppler factor can be obtained through the variance transfer formula.

TABLE 1 Estimation expressions of different methods.

Methods	Expressions
Quinn (Quinn, 1994)	$\alpha_1 = Real\{X(k-1)/X(k)\}$, $\alpha_2 = Real\{X(k+1)/X(k)\}$, $\delta_1 = \alpha_1/(1-\alpha_1)$, $\delta_2 = \alpha_2/(1-\alpha_2)$if $\delta_1 > 0$ and $\delta_2 > 0$, $\hat{\delta} = \delta_2$ else $\hat{\delta} = \delta_1$ $\Delta = \hat{\delta}/f_{cw}$
Macleod (Macleod, 1998)	$d = \frac{Real(X(k-1)X(k)^* - X(k+1)X(k)^*)}{Real(2\lvert X(k)\rvert^2 - X(k-1)X(k)^* - X(k+1)X(k)^*)}$ $\hat{\delta} = (\sqrt{1+8d^2}-1)/(4d)$ $\Delta = \hat{\delta}/f_{cw}$
Jacobsen (Jacobsen and Kootsookos, 2007)	$\hat{\delta} = Real\left\{\frac{X(k-1)-X(k+1)}{2X(k)-X(k-1)-X(k+1)}\right\}$ $\Delta = \hat{\delta}/f_{cw}$
Candan (Candan, 2011)	$\hat{\delta} = \frac{\tan(\pi/N)}{\pi/N} Real\left\{\frac{X(k-1)-X(k+1)}{2X(k)-X(k-1)-X(k+1)}\right\}$ $\Delta = \hat{\delta}/f_{cw}$
Proposed	$\Delta_n = \frac{f_{rn}-f_n}{f_n}, n = 1, 2\ldots, m$ $\Delta = \sum_{n=1}^{m} \frac{f_n^2}{\sum_{n=1}^{m} f_n^2} \Delta_n$

Real denotes the real part.

4.2 Numerical results

As shown in Figure 6, the deviation of the true Doppler factor by the CUU method based on AFC does not exceed 1.1×10^{-5} when δ is 0.45. In contrast, alternative frequency estimation methods have a minimal bias of 1.5×10^{-5} and a maximal bias of 5.3×10^{-5}. It is noteworthy that the CUU method has a bias with a maximum value that is smaller than the minimum value observed in other frequency estimation methods.

Figure 7 demonstrates that at a frequency resolution of 1 Hz, the RMSE of the proposed method is no greater than 1.5×10^{-5}, significantly lower than 2.9×10^{-5} observed with other frequency estimation methods. At a frequency resolution of 2 Hz, the RMSE for the proposed method is lower than 1.8×10^{-5}, better than 6.0×10-5 for other methods, as shown in Figure 8. In both scenarios, the RMSE of the AFC-based method is closer to the CR bound, showing the high accuracy and robustness of the CUU method based on AFC.

4.3 Experimental results

In order to provide further validation for the effectiveness of the proposed method, Watermark, a widely accessible benchmark for physical-layer techniques in underwater acoustic communications,

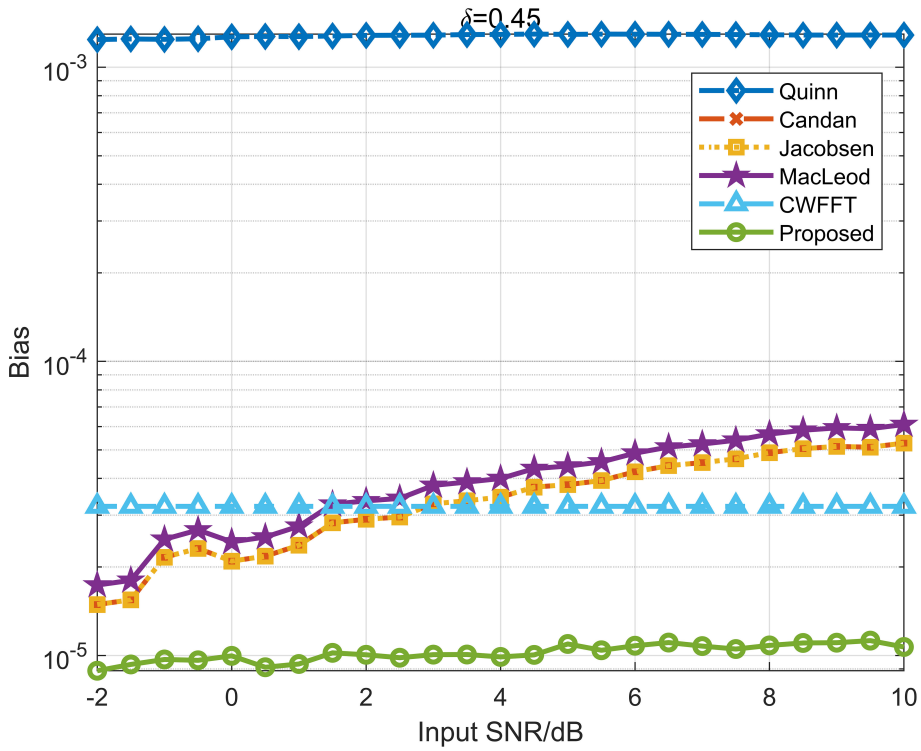


FIGURE 6
Variation of bias with respect to SNR for δ = 0.45 with a frequency resolution of 2 Hz.

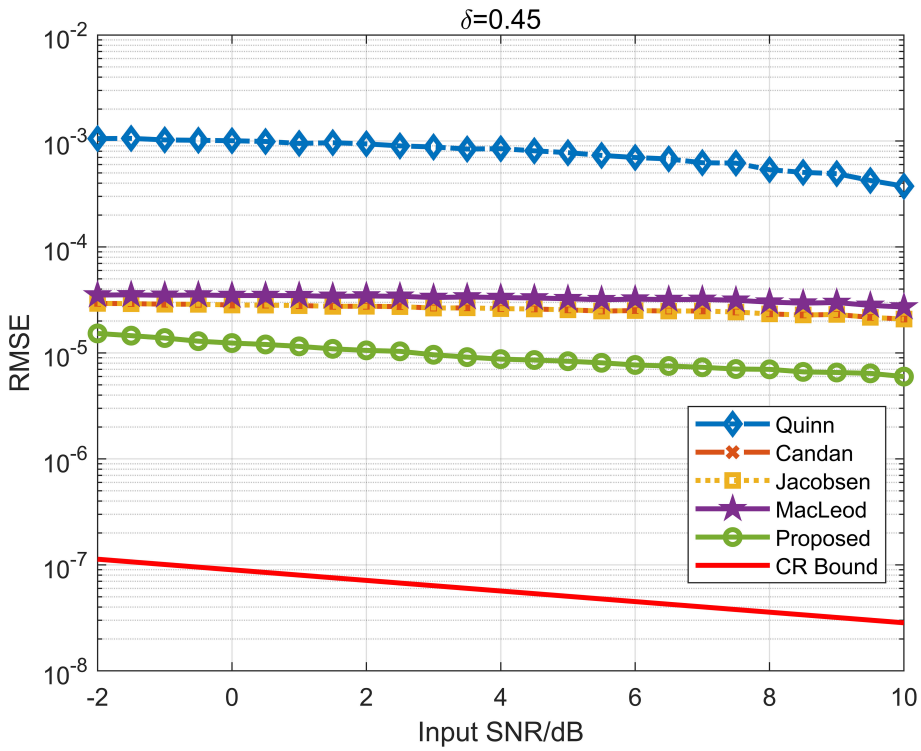


FIGURE 7
Variation of RMSE with respect to SNR for δ = 0.45 with a frequency resolution of 1 Hz.

10.3389/fmars.2024.1365095

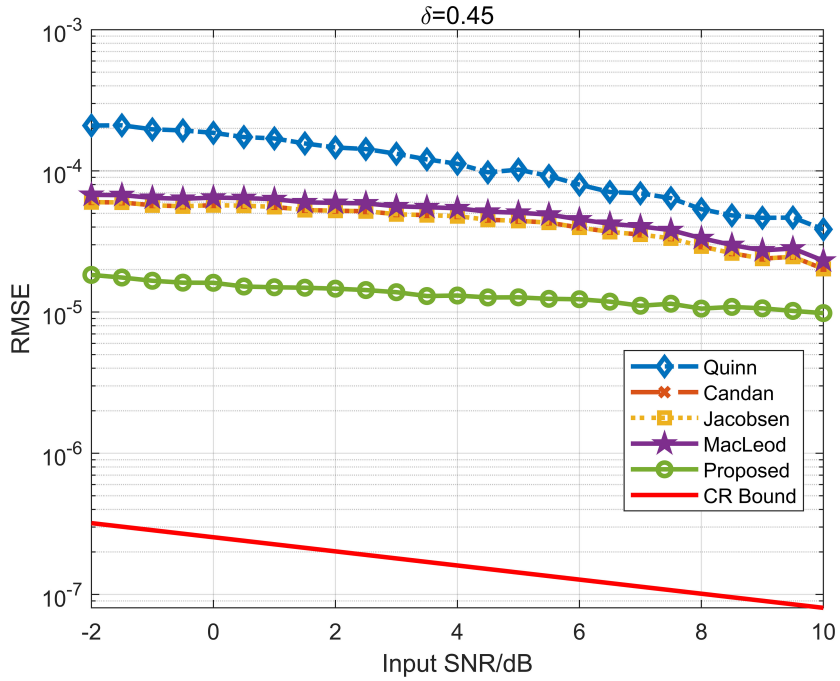


FIGURE 8
Variation of RMSE with respect to SNR for δ = 0.45 with a frequency resolution of 2 Hz.

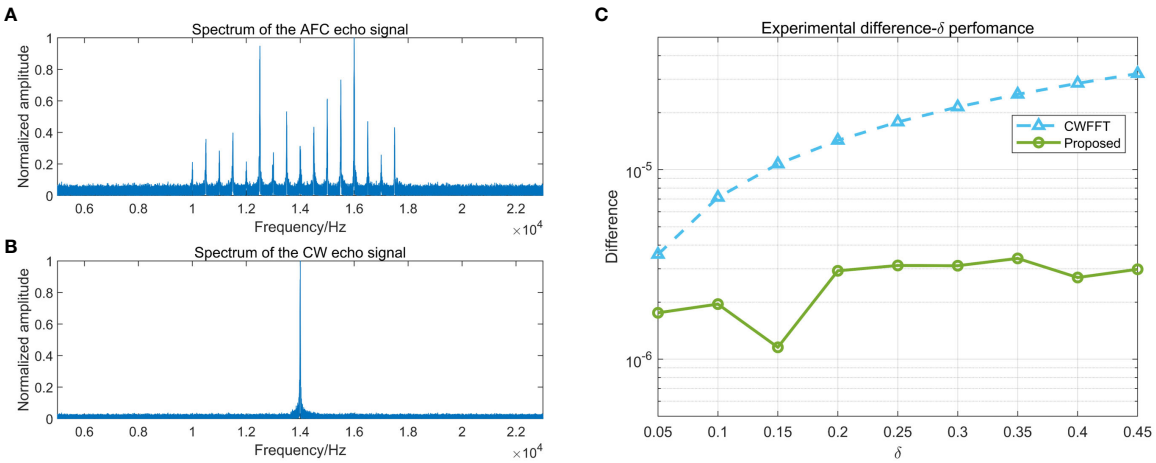


FIGURE 9
(A) The spectrum of the AFC echo signal; **(B)** The spectrum of the CW echo signal; **(C)** The experimental difference-δ performance.

TABLE 2 The computational requirements of different methods.

Methods	Complex multiplications (×)	Complex Additions (+)	Complex multiplications (×) (N = 128×1024)	Complex Additions (+) (N = 128×1024)
Quinn (Quinn, 1994)	(N/2)*$\log_2$N+2	N*$\log_2$N	1114114	2228224
Macleod (Macleod, 1998)	(N/2)*$\log_2$N+5	N*$\log_2$N+3	1114117	2228227
Jacobsen (Jacobsen and Kootsookos, 2007)	(N/2)*$\log_2$N+1	N*$\log_2$N+3	1114113	2228227
Candan (Candan, 2011)	(N/2)*$\log_2$N+1	N*$\log_2$N+3	1114113	2228227
Proposed	(N/2)*$\log_2$N	N*$\log_2$N	1114112	2228224

was employed. This benchmark is based on empirical measurements of the time-varying impulse response collected at sea (van Walree et al., 2017).

The Norway-Oslofjord (NOF1) channel with the large available bandwidth was selected to validate the practical efficacy of underwater communication. In the single-input single-output (SISO) scenario, the time-varying impulse response (TVIR) of the transmitted signal affected by the Doppler effect is obtained, and the signal packet is retrieved through serial acquisition processes. The signal has been established with a SNR of 20 dB, a pulse length of 50 ms, a frequency resolution of 1Hz, and other parameters consistent with the first Monte Carlo experiment in Section 4.1. Figure 9A demonstrates that due to the frequency-selective fading of underwater acoustic channels, the magnitude of each frequency component varies, while the AFC signal spectrum remains uniformly spaced. By calculating the Doppler shift based on frequency rather than phase, the impact of multipath effects is limited. Contrary to conventional methods that primarily focus on spectral peaks and their adjacent lines while disregarding the remaining spectrum, our proposed approach considers signals from different frequencies under the influence of the same Doppler effect, leading to distinct spectral peaks at various positions on the spectrum. Unlike the singular spectral peak observed in CW signal circumstance (Figure 9B), each spectral peak of the AFC signal effectively reflects the potency of the Doppler effect. The cumulative effect of multiple measurements leads to the estimation of the Doppler factor aligning closely with the true value, thereby enhancing the accuracy of estimation and resulting in improved estimation precision (Figure 9C). The difference of the CUU method utilizing the AFC signal is below 3.4×10^{-6}, notably lower than that of the 3.2×10^{-5} induced by spectral leakage.

4.4 Computational burden

Table 2 shows the calculation scales of the methods. It can be seen that the computational cost of the proposed algorithm is lower than that of other frequency estimation methods for single-tone signals. The proposed method does not require additional complex multiplication and complex addition, and it achieves a better Doppler estimation performance. Essentially, this is because the CUU method makes full use of spectrum information.

5 Conclusion

In this paper, we proposed a novel acoustic broadband signal, AFC, to estimate the Doppler factor of underwater acoustic applications. The quantitative relationship between the AFC and the Doppler factor in the time and frequency domains was derived theoretically. At the expense of bandwidth, the phonon frequency shifts of multiple frequency components can be determined by one measurement, allowing the Doppler factor to be calculated multiple times simultaneously, resulting in a precise Doppler estimate. The proposed method maintains a low computational cost and improved spectral leakage performance by means of CUU. Besides, the uncertainty can be calculated from a single measurement, not available with other methods.

The AFC-based method keeps a bias of less than 1.1×10^{-5}, while the bias of single-tone signal after performing FFT does not exceed 3.2×10^{-5} in the absence of noise. In the presence of additive white Gaussian noise, the bias of the AFC-based CUU method is less than 1.1×10^{-5}, significantly lower than the observed bias of 5.3×10^{-5} with other estimation methods at a frequency resolution of 1 Hz. The RMSE of the proposed method is no greater than 1.5×10^{-5}, significantly lower than 2.9×10^{-5} observed with other frequency estimation methods. At a frequency resolution of 2 Hz, the RMSE for the proposed method is 1.8×10^{-5}, compared to 6.0×10^{-5} for other methods. In the SISO scenario of Watermark, The difference of the CUU method utilizing the AFC signal is below 3.4×10^{-6}, notably lower than that of the 3.2×10^{-5} induced by spectral leakage in FFT. Both numerical simulations and experimental results demonstrate that the CUU method based on AFC outperforms traditional frequency estimation methods for single-tone signals in terms of accuracy and computational efficiency. This introduces a new platform for acoustic applications and enhances the accuracy of Doppler estimation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JL: Visualization, Writing – original draft, Writing – review & editing. ZQ: Formal analysis, Supervision, Writing – original draft. DH: Investigation, Visualization, Writing – original draft. JZ: Conceptualization, Methodology, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Open Fund Projects of Key Laboratory of Marine Environmental Survey Technology and Application (MESTA-2023-B006), Basic Science Centre Project of the National Natural Science Funding Council (62388101) and the Natural Science Foundation of Tianjin (22JCQNJC00270).

Acknowledgments

Special thanks to the funding that provided support for this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aboutanios, E., and Mulgrew, B. (2005). Iterative frequency estimation by interpolation on Fourier coefficients. *IEEE Trans. Signal Processing* 53, 1237–1242. doi: 10.1109/TSP.2005.843719

Ahmad, A. M., Kassem, J., Barbeau, M., Kranakis, E., Porretta, E., and Garcia-Alfaro, J. (2018). Doppler effect in the acoustic ultra low frequency band for wireless underwater networks. *Mobile Networks Applications* 23, 1282–1292. doi: 10.1007/s11036-018-1036-9

Braje, D. A., Kirchner, M. S., Osterman, S., Fortier, T., and Diddams, S. A. (2008). Astronomical spectrograph calibration with broad-spectrum frequency combs. *Eur. Phys. J. D.* 48, 57–66. doi: 10.1140/epjd/e2008-00099-9

Candan, Ç. (2011). A method for fine resolution frequency estimation from three DFT samples. *IEEE Signal Process. Letters* 18, 351–354. doi: 10.1109/LSP.2011.2136378

Chan, Y. T., and Jardine, F. L. (1990). Target localization and tracking from Doppler-shift measurements. *IEEE J. Oceanic Engineering* 15, 251–257. doi: 10.1109/48.107154

Chen, Y., Yin, J., Zou, L., Yang, D., and Cao, Y. (2015). Null subcarriers based Doppler scale estimation with polynomial interpolation for multicarrier communication over ultrawideband underwater acoustic channels. *J. Syst. Eng. Electronics* 26, 1177–1183. doi: 10.1109/JSEE.2015.00128

Fang, L., Duan, D., and Yang, L. (2012). A new DFT-based frequency estimator for single-tone complex sinusoidal signals. *In MILCOM 2012-2012 IEEE Military Commun. Conference* 8, 1–6. doi: 10.1109/MILCOM.2012.6415812

Gong, Z. J., Li, C., and Jiang, F. (2020). Analysis of the underwater multi-path reflections on Doppler shift estimation. *IEEE Wireless Commun. Letters* 9, 1758–1762. doi: 10.1109/LWC.5962382

Greene, A. D., and Hendricks, P. J. (2015). Using an ADCP to estimate turbulent kinetic energy dissipation rate in sheltered coastal waters. *J. Atmospheric Oceanic Technol* 32, 318–333. doi: 10.1175/JTECH-D-13-00207.1

Hall, J. L. (2006). Nobel lecture: defining and measuring optical frequencies. *Rev. Modern Physics* 78, 1279–1295. doi: 10.1103/RevModPhys.78.1279

Hansch, T. W. (2006). Nobel lecture: passion for precision. *Rev. Modern Physics* 78, 1297–1309. doi: 10.1103/RevModPhys.78.1297

Jacobsen, E., and Kootsookos, P. (2007). Fast, accurate frequency estimators [DSP Tips & Tricks]. *IEEE Signal Process. Magazine* 24, 123–125. doi: 10.1109/MSP.2007.361611

Jones, D. J., Diddams, S. A., Ranka, J. K., Stentz, A., Windeler, R. S., Hall, J. L., et al. (2000). Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* 288, 635–639. doi: 10.1126/science.288.5466.635

Kay, S. (1989). A fast and accurate single frequency estimator. *IEEE Trans. Acoustics Speech Signal Processing* 37, 1987–1990. doi: 10.1109/29.45547

Li, Y., and Chen, K. (2008). Eliminating the picket fence effect of the fast Fourier transform. *Comput. Phys. Commun.* 178, 486–491. doi: 10.1016/j.cpc.2007.11.005

Macleod, M. D. (1998). Fast nearly ML estimation of parameters of real or complex single tones or resolved multiple tones. *IEEE Trans. on Signal Process* 46, 141–148. doi: 10.1109/78.651200

Mikhail, Z., David, S., Paola, G.-P., and Carolina, G. (2014). Efficient CO_2 fixation by surface Prochlorococcus in the Atlantic Ocean. *ISME J.* 83, 2280–2289. doi: 10.1038/ismej.2014.56

Quinn, B. G. (1994). Estimating frequency by interpolation using Fourier coefficients. *IEEE Trans. Signal Processing* 42, 1264–1268. doi: 10.1109/78.295186

Rife, D., and Boorstyn, R. (1974). Single-tone parameter estimation from discrete-time observations. *IEEE Trans. Inf. Theory* 20, 591–598. doi: 10.1109/TIT.1974.1055282

Sen, S., and Nehorai, A. (2010). Adaptive design of OFDM radar signal with improved wideband ambiguity function. *IEEE Trans. Signal Processing* 58, 928–933. doi: 10.1109/TSP.2009.2032456

Sharif, B. S., Neasham, J., Hinton, O. R., and Adams, A. E. (2000). A computationally efficient Doppler compensation system for underwater acoustic communications. *IEEE J. Oceanic Engineering* 25, 52–61. doi: 10.1109/48.820736

Steinmetz, T., Wilken, T., Araujo-Houck, C., Holzwarth, R., Haensch, T. W., Pasquini, L., et al. (2008). Laser frequency combs for astronomical observations. *Science* 321, 1335–1337. doi: 10.1126/science.1161030

van Walree, P. A., Socheleau, F. X., Otnes, R., and Jenserud, T. (2017). The watermark benchmark for underwater acoustic modulation schemes. *IEEE J. oceanic engineering* 42, 1007–1018. doi: 10.1109/JOE.2017.2699078

Wan, L., Jia, H., Zhou, F., Muzzammil, M., Li, T., and Huang, Y. (2020). Fine Doppler scale estimations for an underwater acoustic CP-OFDM system. *Signal Processing* 170, 107439. doi: 10.1016/j.sigpro.2019.107439

Wu, H., Qian, Z., Zhang, H., Xu, X., Xue, B., and Zhai, J. (2019). Precise underwater distance measurement by dual acoustic frequency combs. *Annalen der Physik* 531, 1900283. doi: 10.1002/andp.201900283

Yang, P. X. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Applications* 83, 1–16. doi: 10.1007/s11042-023-16757-0

Yang, Y., and Fang, S. (2021). Dynamic optimization method for broadband ADCP waveform with environment constraints. *Sensors* 21, 3768. doi: 10.3390/s21113768

Zhang, X. B. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *Multimedia Tools Applications* 83, 1–17. doi: 10.1007/s11042-023-16992-5

Zhang, X. B., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sensing* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X. B., Yang, P., Wang, Y., Shen, W., Yang, J., Wang, J., et al. (2024). A novel multireceiver SAS RD processor. *IEEE Trans. Geosci. Remote Sensing*. 62, 1-11. doi: 10.1109/TGRS.2024.3362886

frontiers | Frontiers in Marine Science

TYPE Methods
PUBLISHED 17 September 2024
DOI 10.3389/fmars.2024.1362416

OPEN ACCESS

EDITED BY
Haiyong Zheng,
Ocean University of China, China

REVIEWED BY
JianBo Zhou,
Northwestern Polytechnical University, China
Shuqing Ma,
National University of Defense Technology, China
Haitao Zhao,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
Yanling Ran
ranyanling0810@gmail.com

RECEIVED 28 December 2023
ACCEPTED 29 August 2024
PUBLISHED 17 September 2024

CITATION
Xing C, Ran Y, Lu M, Tan G and Meng Q (2024) A TMSBL underwater acoustic channel estimation method based on dictionary learning denoising.
Front. Mar. Sci. 11:1362416.
doi: 10.3389/fmars.2024.1362416

A TMSBL underwater acoustic channel estimation method based on dictionary learning denoising

Chuanxi Xing[1,2], Yanling Ran[1,2]*, Mao Lu[1,2], Guangzhi Tan[1,2] and Qiang Meng[1,2]

[1]School of Electrical and Information Technology, Yunnan Minzu University, Kunming, China, [2]Yunnan Key Laboratory of Unmanned Autonomous System, Yunnan Minzu University, Kunming, China

The shallow sea underwater acoustic channel exhibits a significant sparse multipath structure. The temporally multiple sparse Bayesian learning (TMSBL) algorithm can effectively estimate this sparse multipath channel. However, the complexity of the algorithm is high, the signal-to-noise ratio (SNR) of shallow-sea underwater acoustic communication is low, and the estimation performance of the TMSBL algorithm is greatly affected by noise. To address this problem, an improved TMSBL underwater acoustic channel estimation method based on a dictionary learning noise reduction algorithm is proposed. Firstly, the K-Singular Value Decomposition (K-SVD) dictionary learning method is used to reduce the noise of the received pilot matrix, reducing the influence of noise on the signal. Then, the Generalized Orthogonal Matching Pursuit (GOMP) channel estimation method is combined to obtain *a priori* information such as the perceptual matrix and hyperparameter matrix for TMSBL channel estimation; and the noise variance is obtained by using the null subcarrier calculation instead of iteratively updating the noise variance in the TMSBL, to improve the estimation accuracy and reduce the algorithmic complexity. Finally, the TMSBL channel estimation method is used to estimate the underwater acoustic channels of different symbols jointly. The simulation results show that the normalized mean square error of the channel estimation of the improved TMSBL method is reduced by about 92.2% compared with the TMSBL algorithm, obtaining higher estimation accuracy; running time is reduced by about 45.6%, and there is also better performance in terms of the running speed, which provides a reference for underwater acoustic channel estimation.

KEYWORDS

underwater acoustic channel estimation, temporally multiple sparse Bayesian learning, K-SVD dictionary learning, underwater sparse channel estimation, orthogonal frequency division multiplexing

1 Introduction

Underwater acoustic communication is a crucial means of transmitting information in the ocean due to its high reliability and data transmission rates (Xu et al., 2016; Xing et al., 2021b; Zhang et al., 2021; Zhang et al., 2024). The use of Orthogonal Frequency Division Multiplexing (OFDM) technology is widespread in high-speed underwater acoustic communication due to its effectiveness against frequency-selective fading, efficient band utilization, robust resistance to multipath propagation, and straightforward implementation of channel equalization (Jia et al., 2022). The underwater acoustic channel is considered one of the most complex channels due to its multipath, time-varying, frequency-varying, and null-varying characteristics (Xing et al., 2022; Xing et al., 2023). In the underwater acoustic channel of shallow seas, reflections and scattering from the seafloor and sea surface cause significant delay extension and multipath effects. These effects result in a sparse multipath structure at the receiving end (Tong et al., 2022; Yang, 2023; Zhang, 2023). The multipath structure is formed at the receiving end. The shallow sea underwater acoustic channel's complex and variable nature significantly impacts the OFDM communication system (Yin et al., 2021). To ensure communication quality, it is necessary to estimate the channel state at the receiving end. The key characteristic parameters obtained through channel estimation are used to adjust the signal processing method, which serves as a crucial basis for achieving channel matching and improving the quality of signal recovery. This is of great importance in improving the performance of underwater acoustic OFDM communications in shallow sea environments.

The underwater acoustic channel exhibits significant sparse characteristics. However, traditional channel estimation algorithms, such as Least Square (LS) and Minimum Mean Square Error (MMSE), fail to leverage the sparsity of the underwater acoustic channel, necessitating a large number of pilot signals for accurate channel estimation, resulting in serious occupation of spectral resources (Meng and Liu, 2023). To achieve accurate channel estimation that exploits the sparsity of the channel, compressed sensing techniques are employed for sparse channel estimation with a large number of zero taps in the time domain response (Wu and Tong, 2017; Jiang et al., 2021; Meng and Liu, 2023). Matching Pursuit (MP) is a greedy iterative algorithm widely employed in compressed sensing for sparse channel estimation. It exhibits higher estimation accuracy compared to methods such as LS and MMSE (Cotter and Rao, 2002). Another category of compressed sensing reconstruction algorithms, including Least Absolute Shrinkage And Selection Operator (LASSO) (Tibshirani, 1996) and Basis Pursuit (BP) (Chen et al., 2001), constitutes convex optimization tracking algorithms grounded in paradigm constraints. They seek the approximation of sparse signals by converting a non-convex problem into a solvable convex problem. These algorithms, in comparison with traditional methods, harness the sparse characteristics of the underwater acoustic channel, leading to higher estimation accuracy. However, their performance is significantly influenced by sparsity selection, and their computational complexity is high, rendering practical applications challenging. To enhance the channel estimation technique, we leverage the priori knowledge of the sparse signal. Introducing the sparse Bayesian learning class algorithm, based on the Bayesian criterion, into the sparse channel estimation problem (Chen et al., 2020; Lyu et al., 2021) yields improved estimation performance and has been extensively researched.

Algorithms for sparse signal reconstruction using sparse Bayesian learning have been extensively researched in recent years (Wipf and Rao, 2004; Wipf and Rao, 2007; Zhang and Rao, 2011). Wipf and Rao, 2004 (Wipf and Rao, 2004) introduced the Sparse Bayesian Learning (SBL) algorithm for sparse signal reconstruction in single-measurement models; Subsequently, in 2007 (Wipf and Rao, 2007), they extended it to multi-measurement models and derived the Multiple Sparse Bayesian Learning (MSBL) algorithm for sparse signal reconstruction. In (Zhang and Rao, 2011) the Temporal Sparse Bayesian Learning (TSBL) algorithm and its extension, the TMSBL algorithm based on the MSBL algorithm, are derived. Among these algorithms, the TMSBL algorithm not only leverages the channel sparsity property but also explores the correlation between channels. It considers the priori distribution of the channel and incorporates space-time information, resulting in high channel estimation accuracy. Consequently, the TMSBL algorithm has found widespread use in underwater channel estimation (Qiao et al., 2018; Hong et al., 2022). Consequently, the TMSBL algorithm finds extensive application in underwater channel estimation. In (Qiao et al., 2018), the TMSBL algorithm is incorporated into the channel estimation of slow time-varying underwater acoustic OFDM communication systems. Correlation is utilized to jointly estimate the channels of several consecutive blocks. This approach achieves optimal performance in strongly time-correlated channels and maintains robustness in weakly time-correlated channels. However, it is more sensitive to noise, which leads to degraded estimation accuracy and increased computational complexity in low signal-to-noise ratio scenarios. In (Hong et al., 2022), singular value decomposition noise reduction is performed to address the above challenges. Using LS channel estimation to obtain *a priori* information such as perception matrix and hyperparameter matrix of TMSBL for high-precision and low-complexity underwater acoustic OFDM communication. However, due to the increased noise sensitivity of the LS channel estimation method and the limited effectiveness of the singular value decomposition for noise reduction, the accuracy of the channel estimation is reduced under low signal-to-noise ratio conditions.

Dictionary learning algorithms provide effective noise reduction and are widely used in image denoising, active sonar target classification, and weak signal detection in underwater acoustic (Wang et al., [[NoYear]]; Zhu et al., 2020; Xing et al., 2021a). It is common for existing channel estimation methods to incorrectly identify noise as channel tap coefficients in environments with a low signal-to-noise ratio (SNR). This reduces the accuracy of channel estimation and increases the computational complexity. To address the limitations of the previously mentioned channel estimation methods and to account for the sparse multipath structure of the signal in the shallow sea underwater acoustic channel, we utilize the dictionary learning algorithm for TMSBL underwater acoustic

channel estimation. As a result, we propose an enhanced TMSBL underwater acoustic channel estimation method based on the dictionary learning noise reduction algorithm. Initially, the enhanced K-SVD dictionary learning algorithm is employed to reduce the noise of the received pilot matrix, thereby enhancing the accuracy of channel estimation under low signal-to-noise ratio conditions. Subsequently, the initialization parameter matrix and perception matrix of TMSBL are acquired by integrating the GOMP channel estimation method. This integration alleviates the limitation of the TMSBL method, where noise is erroneously estimated as a channel tapping coefficient. Lastly, the null subcarrier of the OFDM system is utilized to obtain a more precise noise variance, replacing the step of updating the noise variance in TMSBL. This modification reduces the complexity of the TMSBL algorithm and enhances estimation accuracy. The paper's contributions can be summarized as follows.

1. The K-SVD dictionary learning algorithm is employed in the domain of underwater acoustic communication to denoise the received signal pilot matrix, thereby mitigating the impact of noise on channel estimation accuracy and enhancing the performance of underwater acoustic OFDM communication systems.
2. The GOMP channel estimation method is employed to derive the time-domain underwater acoustic channel impulse response and the initialization and perception matrices for the TMSBL algorithm, thereby reducing the number of iterations and the computational complexity of the TMSBL algorithm. Furthermore, the incorporation of the priori knowledge addresses the limitation of the TMSBL method, which is prone to misestimating noise as a channel tap coefficient. This enhances the overall performance of channel estimation.

The rest of the article is organized as follows. In Section 2, the received signal model and the TMSBL channel estimation method are introduced. Section 3 presents enhanced TMSBL channel estimation methodologies, including a K-SVD dictionary learning-based noise reduction technique and a method for obtaining TMSBL priori knowledge using the GOMP algorithm. The efficacy of the proposed algorithm is substantiated through simulations in Section 4. Further validation of the algorithm using sea trial experimental data is provided in Section 5, demonstrating its effectiveness in real marine environments. The paper concludes in Section 6.

2 TMSBL-based underwater acoustic channel estimation method

2.1 Received signal model

In underwater acoustic communications in shallow seas, signal propagation is significantly affected by reflections, diffraction, and scattering from both the sea surface and seafloor. This results in a complex multipath structure of the underwater acoustic channel. The underwater acoustic channel exhibits a significant sparse characteristic due to signal energy absorption by seawater during most of the multipath propagation. The mathematical expression for the channel impact response of the underwater acoustic time-varying channel is given by (Cheng and Wang, 2022).

$$h(t,\tau)=\sum_{i=1}^{L}h_i(t)\delta\left(\tau-\tau_i(t)\right) \tag{1}$$

where $h(t)$ is the channel impulse response at time t, and L is the multipath number, the $h_i(t)$ and $\tau_i(t)$ are denoted as the gain and delay of the ith path at time t, respectively.

Consider an OFDM system with N subcarriers, L pilots, and channel coherence time significantly exceeding the OFDM symbol period. If the impulse response of the channel remains time-invariant within one OFDM symbol period, (Equation 1) can be expressed as follows:

$$h(\tau)=\sum_{i=1}^{L}h_i\delta(\tau-\tau_i) \tag{2}$$

Assuming the cyclic prefix of OFDM symbols exceeds the maximum multipath delay of the channel, the frequency domain expression for the OFDM communication system is:

$$\boldsymbol{y}=\boldsymbol{XFh}+\boldsymbol{v} \tag{3}$$

where $\boldsymbol{y}\in\mathbb{C}^{N\times1}$ is the received signal, $\boldsymbol{X}\in\mathbb{C}^{N\times N}$ is the diagonalization matrix with diagonal elements representing the transmitted signals. $\boldsymbol{F}\in\mathbb{C}^{N\times M}$ is the DFT matrix, and $\boldsymbol{h}\in\mathbb{C}^{M\times1}$ is the time domain channel impulse response. The $\boldsymbol{v}$ follows $\mathcal{CN}(0,\lambda\boldsymbol{I}_N)$ of Gaussian white noise. From the received signal $\boldsymbol{y}$ out of the pilot signal, the received model of the pilot signal is:

$$\boldsymbol{y}_p=\boldsymbol{X}_p\boldsymbol{F}_p\boldsymbol{h}+\boldsymbol{v}_p \tag{4}$$

where $\boldsymbol{y}_p\in\mathbb{C}^{p\times1}$ is the received pilot signal, $\boldsymbol{X}_p\in\mathbb{C}^{p\times p}$ is the diagonalization matrix with diagonal elements representing the known pilot signals. $\boldsymbol{F}_p\in\mathbb{C}^{p\times M}$ is the corresponding DFT matrix at the pilot position. The system model described in (Equation 4) is a single-measurement model. A multi-measurement model is considered: several different OFDM symbols are modeled with the following expressions:

$$\boldsymbol{Y}_p=\boldsymbol{X}_p\boldsymbol{F}_p\boldsymbol{H}+\boldsymbol{V}_p=\boldsymbol{\Phi}_p\boldsymbol{H}+\boldsymbol{V}_p \tag{5}$$

Among them. $\boldsymbol{Y}_p=[\boldsymbol{y}_{p,1},\boldsymbol{y}_{p,2},\ldots,\boldsymbol{y}_{p,L}]\in\mathbb{C}^{p\times L}$ represents the received pilot matrix of L OFDM symbols, and $\boldsymbol{\Phi}_p\in\mathbb{C}^{p\times M}$ is the perception matrix, $\boldsymbol{H}=[\boldsymbol{h}_1,\boldsymbol{h}_2,\ldots,\boldsymbol{h}_L]\in\mathbb{C}^{M\times L}$.

2.2 TMSBL underwater acoustic channel estimation

The article uses the TMSBL algorithm (Qiao et al., 2018), using temporal correlation to jointly estimate (Equation 5) of $\boldsymbol{H}$ the estimation reconstruction problem. Firstly, the priori probability of each $\boldsymbol{H}_i$ is modeled as:

$$p(\boldsymbol{H}_i;\gamma_i,\boldsymbol{B}_i)\sim N(0,\gamma_i\boldsymbol{B}_i),\quad i=1,\ldots,M \tag{6}$$

where $\boldsymbol{H}_i$ is the ith row of $\boldsymbol{H}$, i.e., the number of channels tapping coefficients for different OFDM symbols at the same moment; γ_i is the nonnegative hyperparameter matrix that controls the sparsity of each row in $\boldsymbol{H}$. When $\gamma_i = 0$, $\boldsymbol{H}_i = 0$; $\boldsymbol{B}_i$ is a positive definite matrix, which represents the temporal correlation structure among the elements within $\boldsymbol{H}_i$ and can be estimated using the TMSBL algorithm for the positive definite matrix $\boldsymbol{B}$. (Equation 6) can be written as:

$$p(\boldsymbol{H}_i;\boldsymbol{\Gamma},\boldsymbol{B}_i)=\prod_{i=1}^{M}p(\boldsymbol{H}_i;\gamma_i,\boldsymbol{B}_i) \tag{7}$$

where $\boldsymbol{\Gamma}$ is the hyperparameter matrix $\boldsymbol{\Gamma}=\mathrm{diag}(\boldsymbol{\gamma})=\mathrm{diag}([\gamma_1,\gamma_2,\ldots,\gamma_M]^T)$.

$\boldsymbol{H}_i$, Obeying the mean Gaussian probability distribution, its posterior probability can be written as:

$$p\left(\boldsymbol{h}_l|\boldsymbol{y}_{p,l};\boldsymbol{\Gamma}\right)\sim N(\boldsymbol{\mu}_l,\Sigma),\quad l=1,2,\ldots,L \tag{8}$$

where the mean and covariance can be expressed as:

$$\boldsymbol{\Sigma}=\left(\sigma^{-2}\boldsymbol{\Phi}_p^{\mathrm{H}}\boldsymbol{\Phi}_p+\Gamma^{(r)-1}\right)^{-1} \tag{9}$$

$$\boldsymbol{M}=[\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_L]=\sigma^{-2}\Sigma\boldsymbol{\Phi}_p^{\mathrm{H}}\boldsymbol{Y}_p \tag{10}$$

Of these, $\boldsymbol{\mu}_l$ and $\mathcal{M}$ are respectively the estimated values of $\boldsymbol{H}_i$ and $\boldsymbol{H}$. $\Gamma^{(r)}$ is the estimated value of the $\boldsymbol{\Gamma}$ update matrix for the first r iteration of the Expectation Maximization (EM) algorithm. The hyperparameters are estimated using the EM algorithm. The E-step update rule of the EM algorithm is given in (Equations 9, 10). The M-step update rule is given in:

$$\boldsymbol{\gamma}_i=\frac{1}{L}\boldsymbol{\mathcal{M}}_iB^{-1}\boldsymbol{\mathcal{M}}_i^{\mathrm{H}}+\Sigma(i,i) \tag{11}$$

$$\boldsymbol{B}=\left(\frac{1}{M}\sum_{i=1}^{M}\frac{\Sigma(i,i)}{\boldsymbol{\gamma}_i}\right)\boldsymbol{B}+\frac{1}{M}\sum_{i=1}^{M}\frac{\boldsymbol{H}_i^H\boldsymbol{H}_i}{\boldsymbol{\gamma}_i} \tag{12}$$

$$\sigma^2=\frac{1}{pL}\|\boldsymbol{Y}_p-\boldsymbol{\Phi}_p\boldsymbol{H}\|_F^2+\frac{\sigma^2}{p}Tr\left(\boldsymbol{\Phi}_p\boldsymbol{\Gamma}\boldsymbol{\Phi}_p^{\mathrm{H}}(\sigma^2\boldsymbol{I}+\boldsymbol{\Phi}_p\boldsymbol{\Gamma}\boldsymbol{\Phi}_p^{\mathrm{H}})^{-1}\right) \tag{13}$$

where $\|\cdot\|_F^2$ denotes the quadratic of the F-parameter of the vector $Tr(\cdot)$ is the trace of the matrix. The joint estimation of the channel impulse response after the iteration of the EM algorithm is completed $\hat{\boldsymbol{H}}=\mathcal{M}$.

(Equation 13) is the updated formula for the noise variance, which is calculated using the null subcarriers of the OFDM system:

$$\sigma^2=E[|\boldsymbol{Y}_n|^2] \tag{14}$$

where $\boldsymbol{Y}_n$ is the frequency domain null subcarrier. Using (Equation 14) to obtain a more accurate noise variance, a more accurate channel estimation can be obtained. This reduces the influence of the noise variance by the TMSBL input parameters such as the number of iterations, the threshold, and the received frequency-conducting matrix. At the same time, it can reduce the TMSBL algorithm for the σ^2 update step in the TMSBL algorithm, reducing the complexity of the algorithm.

3 Improved TMSBL channel estimation method

The conventional TMSBL channel estimation method is susceptible to misidentifying noise as channel tapping coefficients in low SNR conditions of underwater acoustic channels, resulting in reduced channel estimation accuracy. Simultaneously, the increase in channel length leads to heightened computational complexity. Furthermore, the TMSBL algorithm inadequately utilizes the characteristics of the underwater acoustic channel for selecting the initial parameters of the EM algorithm, leading to excessive iterations and slower convergence in computation.

Aiming at the limitations of the traditional TMSBL algorithm, the K-SVD dictionary learning algorithm is used on the receiver side to perform noise reduction and reconstruction of the received pilot matrix $\boldsymbol{Y}_p$, and obtains the noise-reduced receiver pilot matrix $\boldsymbol{Y}_p^{'}$, enhancing channel estimation accuracy under low SNR conditions. Following this, the GOMP algorithm is utilized to estimate the underwater acoustic channel, acquiring *a priori* knowledge for the TMSBL algorithm. This knowledge involves removing invalid atoms and smaller hyperparameters from the dictionary. Finally, the TMSBL algorithm, combined with the *a priori* knowledge from the GOMP algorithm, conducts joint channel estimation for different OFDM symbols. The block diagram of the receiver system based on the improved TMSBL channel estimation method is depicted in Figure 1.

3.1 K-SVD-based noise reduction of the received pilot matrix

According to the theory of sparse decomposition, $\boldsymbol{Y}_p$ can be decomposed into $\boldsymbol{Y}_p=\boldsymbol{X}_p\boldsymbol{F}_p\boldsymbol{H}+\boldsymbol{V}_p=\boldsymbol{A}(\boldsymbol{X}_s+\boldsymbol{X}_n)$. Where $\boldsymbol{A}\in\mathbb{C}^{p\times J}$ is the redundant dictionary matrix, X_s and X_n are the sparse coefficient matrices corresponding to $\boldsymbol{X}_p\boldsymbol{F}_p\boldsymbol{H}$ and $\boldsymbol{V}_p$, respectively. As the signal is sparse, whereas the noise is not sparse, and the coefficient values are generally very small, existing only in a finite number of non-zero coefficients, the approximated signal obtained by the linear combination of these sparse counterparts of the atoms contains the vast majority of the information about the signal, while the vast majority of the noise is discarded, thus realizing the purpose of noise cancellation. Let $\boldsymbol{X}=\boldsymbol{X}_s+\boldsymbol{X}_n$, then $\boldsymbol{Y}_p=\boldsymbol{A}\boldsymbol{X}$. Where $\boldsymbol{X}\in\mathbb{C}^{J\times L}$ is the sparse coefficient matrix. To achieve the sparse decomposition of the signal, a suitable dictionary needs to be constructed. The dictionary learning algorithm is employed to construct a suitable redundant dictionary to enhance signal reconstruction.

The K-SVD dictionary learning algorithm is a new dictionary learning algorithm proposed by Aharon and Elad et al (Aharon et al., 2006). The primary concept behind the K-SVD algorithm involves updating a set of atoms in the dictionary along with their

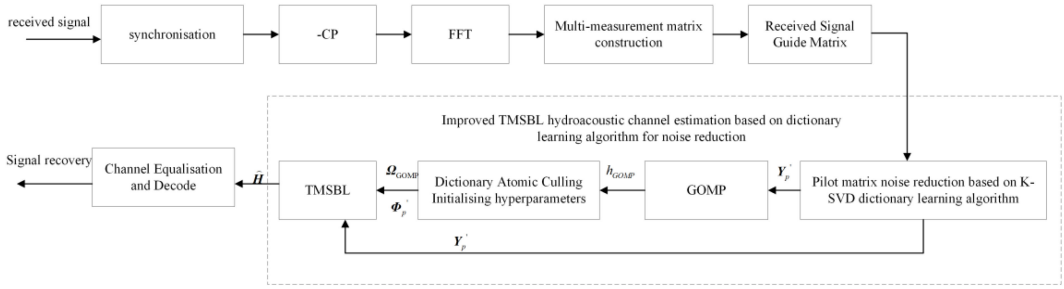


FIGURE 1
Block diagram of the receiver system based on the improved TMSBL channel estimation method.

sparse coefficients simultaneously. Through iterations of updating a set of atoms within the dictionary, if, when updating any atom, the remaining atoms remain unchanged, the dictionary is then updated with the sparse coefficients. $\boldsymbol{a}_i$ with no change in the remaining atoms, the $\boldsymbol{a}_i$ new sparse coefficients will be obtained after the update $\boldsymbol{X}_i$. When the error reaches the threshold, the whole dictionary post-sparse matrix is updated. Its solution model is:

$$\min_{A,\{x_i\}} \sum_{i=1}^{J} \| \boldsymbol{Y}_p - \boldsymbol{AX} \|_F^2 \quad s.t. \quad \| \boldsymbol{x}_i \|_0 \le k, \ 1 \le i \le J \tag{15}$$

where $\| \boldsymbol{Y}_p - \boldsymbol{AX} \|_F^2 = \| \boldsymbol{Y}_p - \sum_{i=1}^{J} \boldsymbol{a}_i \boldsymbol{x}_T^i \|_F^2 = \| (\boldsymbol{Y}_p - \sum_{i \ne K} \boldsymbol{a}_i \boldsymbol{x}_T^i) - \boldsymbol{a}_K \boldsymbol{x}_T^K \|_F^2$. $\boldsymbol{a}_i$ denotes the ith atom in the redundant dictionary $\boldsymbol{A}$, and x_T^i denotes the ith row vector of the sparse coefficient matrix $\boldsymbol{X}$. $\boldsymbol{a}_K$ is the updated atom, and $\boldsymbol{x}_T^K$ is the sparse solution corresponding to the updated atom. Then the error matrix of the signal is:

$$\boldsymbol{E}_K = \boldsymbol{Y}_p - \sum_{i \ne K} \boldsymbol{a}_i \boldsymbol{x}_T^i \tag{16}$$

where $\boldsymbol{E}_K$ is the error matrix of the signal. At this point the solution model can be described as:

$$\sum_{a_K, x_T^K} \left\| \boldsymbol{E}_K - \boldsymbol{a}_K \boldsymbol{x}_T^K \right\|_F^2 \tag{17}$$

To avoid the loss of sparsity in the sparse solution, the $\boldsymbol{E}_K$ in the corresponding x_T^K non-zero positions is extracted to obtain a new $\boldsymbol{E}_K^{'}$, the corresponding sparse coefficient vector is $\boldsymbol{x}_T^{K\prime}$, then (17) can be converted to:

$$\sum_{a_K, x_T^K} \| \boldsymbol{E}_K^{'} - \boldsymbol{a}_K \boldsymbol{x}_T^{K\prime} \|_F^2 \tag{18}$$

Then the singular value decomposition algorithm is used to $\boldsymbol{E}_K^{'}$ solving:

$$\boldsymbol{E}_K^{'} = \boldsymbol{U\Sigma V}^T \tag{19}$$

where $\boldsymbol{U}$ is the left singular matrices, take its first column as the update atom, i.e., $\boldsymbol{a}_K = U(\cdot, 1)$. V is the right singular matrix, take its first row with the first singular value as the $x_T^K\} = \Sigma(1,1)\boldsymbol{V}^T(1, \cdot)$. Then the corresponding update is obtained as $\boldsymbol{x}_T^K$, i.e., by updating each atom of the redundant dictionary in turn, the optimal sparse solution corresponding to each atom can be obtained. When all the atoms are updated, the updated dictionary and the optimal sparse coefficient matrix are obtained.

The GOMP algorithm is employed to achieve a sparse representation of the received pilot matrix, yielding the sparse coefficient matrix. Following this, the K-SVD dictionary learning algorithm is applied to mitigate the noise present in the received signal, resulting in the acquisition of the noise-reduced received pilot matrix $\boldsymbol{Y}_p^{'}$. The specific noise reduction process is shown in Figure 2.

3.2 *A priori* knowledge acquisition based on GOMP channel estimation

Obtaining the time-domain shock response of the underwater acoustic channel using GOMP channel estimation algorithm $\boldsymbol{h}_{\mathrm{GOMP}}$. To obtain the *a priori* knowledge of the TMSBL, the initial parameters of the EM algorithm are chosen based on the characteristics of the underwater acoustic channel, effectively reducing the algorithm's complexity.

Time-domain impact response of underwater acoustic channel obtained using GOMP channel estimation algorithm.

$$\boldsymbol{h}_{\mathrm{GOMP}} = \underset{h}{\mathrm{argmin}} \| \boldsymbol{Y}_p^{'} - \boldsymbol{\Phi}_{p,\Lambda_i} \boldsymbol{h}_i \| \tag{20}$$

where $\boldsymbol{\Phi}_{p,\Lambda_i}$ is the perceptual matrix corresponding to the set of indexes after the ith update of atoms, and $\boldsymbol{h}_i$ is the channel estimate after the ith iteration.

Set the average energy superposition function Q of the channel to be (Hong et al., 2022):

$$Q = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{h}_{\mathrm{GOMP}} i \tag{21}$$

Among them. $\boldsymbol{h}_{\mathrm{GOMP}} i$ is the ith column of $\boldsymbol{h}_{\mathrm{GOMP}}$, i.e., the channel time-domain impulse response of the ith OFDM symbol.

Set the threshold as $T_Q = \alpha \max(Q)$ where α is the energy coefficient, determined based on the characteristics of the underwater acoustic channel and considering the computational complexity. Compare the channel average energy superposition function Q and the threshold T_Q of the channel:

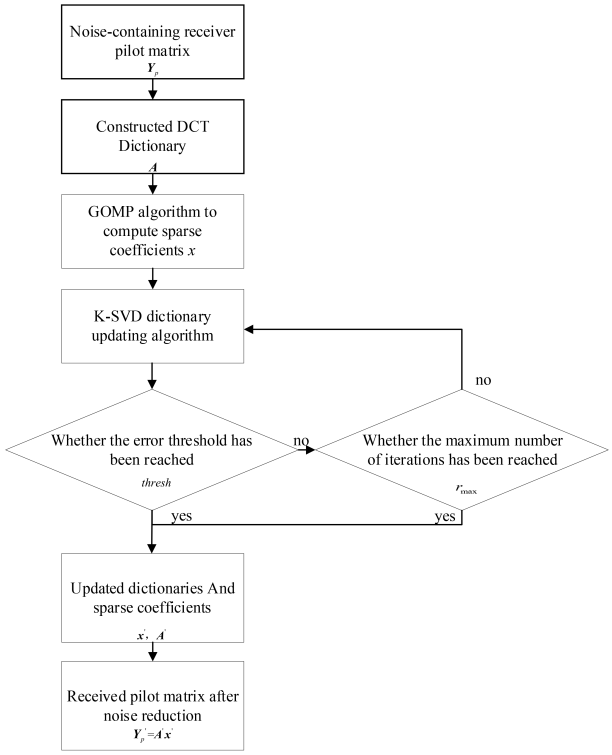


FIGURE 2
Flowchart of KSVD dictionary learning algorithm for noise reduction.

$$\boldsymbol{\Omega}_{\mathrm{GOMP}}(i)=\begin{cases}1, & Q(i)\geq T_Q\\ 0, & Q(i)<T_Q\end{cases}\qquad i=1,2,\ldots,M\,. \tag{22}$$

where $\boldsymbol{\Omega}_{\mathrm{GOMP}}$ is the hyperparameter vector of TMSBL. When the channel average energy superposition function Q is greater than the threshold, indicating that, at this time, the hyperparameter control channel impulse response is the channel tapping coefficient, $\boldsymbol{\Omega}_{\mathrm{GOMP}}$ = 1; on the contrary, it is considered that the hyperparameter is too small and its control channel impulse response probability is the noise, the $\boldsymbol{\Omega}_{\mathrm{GOMP}}$ = 0.

Define the initial hyperparameter matrix as $\boldsymbol{\Gamma}$ = diag($\boldsymbol{\Omega}_{\mathrm{GOMP}}$), whose diagonal elements are hyperparameter vectors $\boldsymbol{\Omega}_{\mathrm{GOMP}}$. Take the positions $\boldsymbol{\Gamma}$ whose diagonal elements are equal to zero as the indexed set $\boldsymbol{s}$ and eliminate the atoms corresponding to s in the dictionary matrix $\boldsymbol{\Phi}_p$ to obtain the initialized dictionary matrix $\boldsymbol{\Phi}_p{}'$ of the improved TMSBL algorithm.

Finally, the noise-canceled pilot receiver matrix $\boldsymbol{Y}_p{}'$, the initial hyperparameter matrix $\boldsymbol{\Gamma}$, and the initialized dictionary matrix $\boldsymbol{\Phi}_p{}'$ are substituted into the TMSBL channel estimation method for underwater acoustic channel estimation.

4 Simulation results and analysis

The simulation to validate the performance of the proposed algorithm is conducted using the implementation of an underwater acoustic OFDM communication system. An OFDM symbol comprises 1024 subcarriers, with 256 designated as frequency-conducting subcarriers (utilizing a comb-conducting structure), 30 as null subcarriers, and 738 as data subcarriers. Sixty-five OFDM symbols are transmitted in each frame, and the signal is modulated using 16QAM. The specific OFDM parameters are configured as presented in Table 1.

The simulated channel is generated using the BELLHOP underwater acoustic channel model to obtain the underwater acoustic channel impulse response. The sound velocity profile of the real marine environment, as experimented in the Yellow Sea in 2013, is depicted in Figure 3. This sound velocity profile is imported into BELLHOP, setting the sound source depth to 10m, the hydrophone depth to 9m, and the distance between the two to be 2000m. The seafloor is modeled as an elastic seafloor with seawater density of 1.5g/cm^3, seafloor absorption of 0.5dB, and the resulting channel impulse response is shown in Figure 4 as obtained through simulation. The seabed absorption is 0.5dB, and the sound line grazing angle is [-35°, 35°], leading to the channel impulse response displayed in Figure 4, obtained through simulation.

In order to measure the estimation accuracy of the proposed algorithm, the normalised mean square error (NMSE) of the channel estimation is defined as:

$$NMSE=\left(\sum_{i=0}^{L-1}\|\hat{\boldsymbol{h}}_i-\boldsymbol{h}\|_F^2\,/\,\|\boldsymbol{h}\|_F^2\right)/L \tag{23}$$

where $\hat{\boldsymbol{h}}_i$ denotes the channel estimate of the ith OFDM symbol, $\boldsymbol{h}$ is the true OFDM underwater acoustic channel impulse response, and L is the number of OFDM symbols.

10.3389/fmars.2024.1362416

TABLE 1 OFDM system parameter settings.

parameters	numerical
No. of subcarriers (number)	1024
Sampling frequency (kHz)	10
Number of comb pilots	256
Number of empty subcarriers (number)	30
OFDM symbol duration (ms)	102.4
Cyclic prefix duration (ms)	25.6
Number of symbols	65
mapping method	16QAM
Training symbolic numbers	1

The primary simulated comparison algorithms include LS, GOMP, TMSBL, and the enhanced TMSBL with the LS priori knowledge acquisition (LS-TMSBL) (Hong et al., 2022). The simulation is divided into two main aspects: firstly, the comparison of the normalized mean square error of the channel estimation to verify the performance of the channel estimation method; secondly, the comparison of the time used for the channel estimation to verify the complexity of the channel estimation method.

4.1 Simulation results and performance analysis

The proposed algorithm is described as KSVD-GOMP-TMSBL algorithm for simplicity of expression. The main comparison algorithms for the simulation are LS, GOMP, TMSBL, and LS-TMSBL methods, setting the maximum number of iterations r_{max} = 5000 and the error threshold is thresh = 1×10^{-6}.

Figure 5 displays the normalized mean square error plots for channel estimation using LS, GOMP, TMSBL, LS-TMSBL, and KSVD-GOMP-TMSBL algorithms. The received pilot matrix was denoised using the K-SVD dictionary learning method. The energy coefficient α is set to 0.05, and GOMP sparsity is fixed at 30. It can be observed from Figure 5 that the LS algorithm, devoid of channel sparsity utilization, exhibits poor estimation performance. The GOMP algorithm slightly outperforms the LS algorithm. The TMSBL algorithm, capitalizing on both the sparse nature of the channel and the temporal correlation between different symbols, demonstrates superior performance in channel estimation. The LS-TMSBL algorithm, incorporating the LS algorithm to acquire *a priori* knowledge of the TMSBL algorithm and influenced by the energy coefficient, exhibits performance slightly lower than the TMSBL algorithm. The KSVD-GOMP-TMSBL algorithm, leveraging the K-SVD dictionary learning algorithm for noise reduction, GOMP to obtain *a priori* knowledge of the TMSBL, and null subcarriers to determine noise variance, demonstrates improved performance compared to the comparison algorithms.

Figure 6 displays the normalized mean square error plots for channel estimation using LS, GOMP, TMSBL, LS-TMSBL, and KSVD-GOMP-TMSBL algorithms. No denoising was applied to the received pilot matrix. The energy coefficient α is set to 0.05, and GOMP sparsity is fixed at 30. It is evident from Figure 6 that, without noise reduction, the LS algorithm is more susceptible to noise, leading to a degradation in the performance of the LS-TMSBL algorithm. In contrast, the TMSBL algorithm and the KSVD-GOMP-TMSBL algorithm are relatively less affected by noise. The KSVD-GOMP-TMSBL algorithm improve its performance by obtaining the null subcarrier through noise variance.

Overall, the KSVD-GOMP-TMSBL algorithm exhibits improved channel estimation performance. The use of the K-SVD dictionary learning algorithm for noise reduction on the received pilot matrix enables the algorithm to achieve accurate estimation

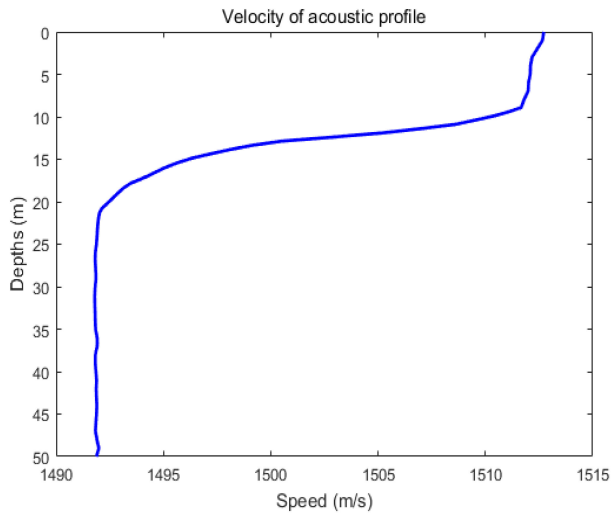


FIGURE 3
Velocity of acoustic profile.

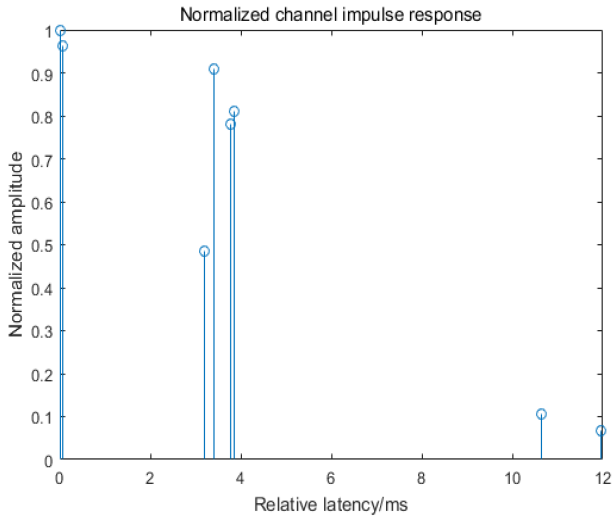


FIGURE 4
Channel impulse response.

results even at low SNR. The algorithm's performance is further enhanced by obtaining precise noise variance through null subcarriers instead of iteratively updating the noise variance. Additionally, the algorithm demonstrates some improvement in estimation performance even without noise reduction processing, showcasing its ability to mitigate certain noise interferences. The factors influencing the KSVD-GOMP-TMSBL algorithm are subsequently analyzed from various perspectives.

First, discuss the impact of the dictionary learning noise reduction method on the algorithm's performance. Figure 7 shows the lofar plot (left) of the received pilot matrix at an SNR of -10dB and the lofar plot (right) of the received pilot matrix after noise reduction using dictionary learning. It is clear that the lofar plots are significantly clearer after noise reduction. The SNR of the received pilot matrix before noise reduction is -10.28dB, and after noise reduction, it improves to 2.36dB. This demonstrates the superior noise reduction effect of the dictionary learning algorithm on the received matrix.

In Figure 8, compare the effectiveness of noise reduction between the dictionary learning algorithm and the singular value decomposition (SVD) noise reduction method. The figure demonstrates the impact of two noise reduction methods on the performance of both the TMSBL and GOMP-TMSBL algorithms. From Figure 8, the dictionary learning noise reduction method significantly improves channel estimation in both algorithms compared to the singular value decomposition method. For the GOMP-TMSBL algorithm, the NMSE of channel estimation under the dictionary learning noise reduction method is 0.0566, representing an almost tenfold decrease compared to the singular value decomposition noise reduction method. This demonstrates

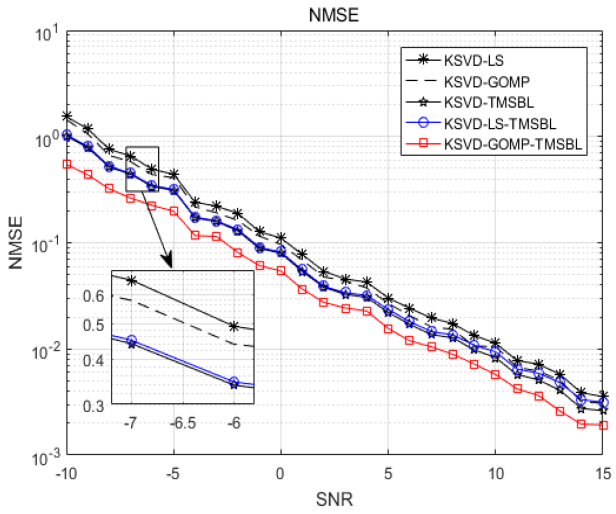


FIGURE 5
Channel estimation error with K-SVD dictionary learning noise reduction.

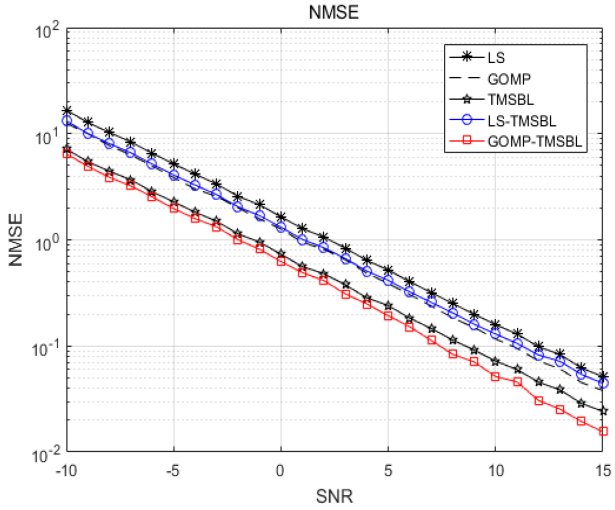


FIGURE 6
Channel estimation error without noise reduction processing.

that the dictionary learning noise reduction method is highly effective in noise reduction.

Figure 9 compares the channel estimation error of the KSVD-GOMP-TMSBL algorithm with other comparison algorithms, without noise reduction processing, to investigate the impact of dictionary learning noise reduction algorithms on the performance of channel estimation methods. As shown in Figure 9, the estimation error of the GOMP-TMSBL algorithm is 6.616, while the estimation error of the KSVD-GOMP-TMSBL algorithm is 0.5973 at a signal-to-noise ratio of -10 dB. The estimation error of the KSVD-GOMP-TMSBL algorithm is reduced by approximately 91.1% compared to the GOMP-TMSBL algorithm. This indicates that the performance of the channel estimation algorithm is improved by performing noise reduction on the received pilot matrix using the dictionary learning algorithm.

Next, discuss the impact of the energy coefficient (α) on the KSVD-GOMP-TMSBL algorithm. The KSVD-GOMP-TMSBL algorithm in the α value will affect the size of the threshold and,

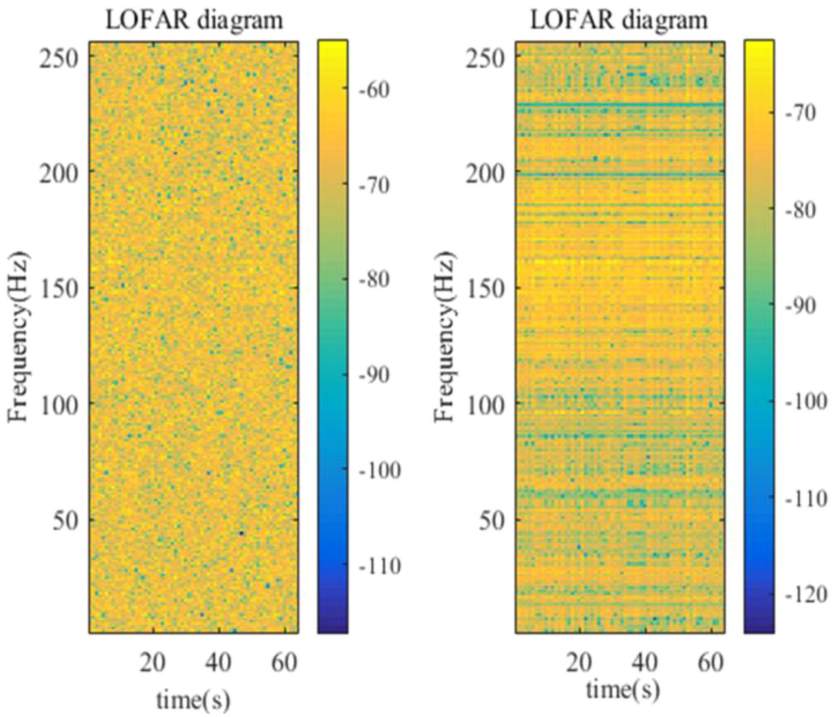


FIGURE 7
Lofar comparison figure of receive pilot matrix before(left) and after(right) noise reduction.

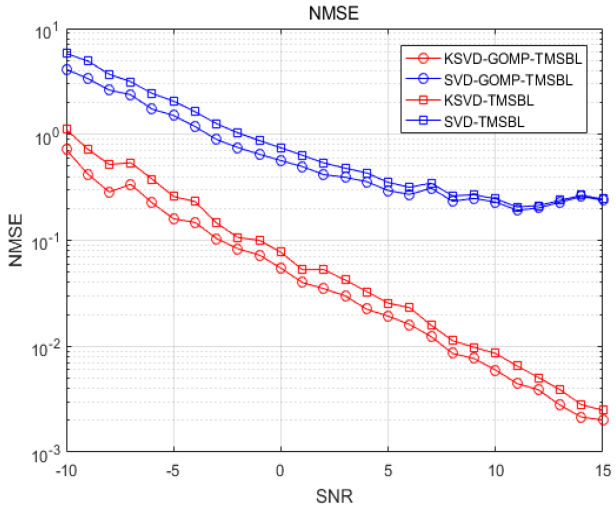


FIGURE 8
Channel estimation error with different noise reduction methods.

consequently, the hyperparameter matrix. When the α value is too small, the threshold becomes insufficient, leading to noise being misestimated as channel tapping coefficients, thereby affecting the accuracy of channel estimation. Conversely, when the α value is too large, the threshold becomes excessive, causing real but smaller channel tapping coefficients to be mistaken as noise, thus affecting the algorithm's performance. The chosen value of α significantly impacts the performance of the KSVD-GOMP-TMSBL algorithm. Figure 10 shows the energy coefficient α channel estimation error when different values are taken, where the GOMP algorithm sparsity is taken to be 30. From the figure, it can be observed that when the $0.01 \leq \alpha \leq 0.10$ the estimation errors of the KSVD-GOMP-TMSBL algorithms are relatively close, both achieve a better performance. When $\alpha = 0.15$, the estimation performance of the KSVD-GOMP-TMSBL algorithm is closer to that when α takes the value in the range [0.01 ~ 0.10]. However, as the threshold is critical at a = 0.15, confusion between channel tapping coefficients and noise can arise, which affects the algorithm's stability, exhibiting significant fluctuations when the signal-to-noise ratio is 11dB. For $\alpha > 0.15$, the estimation performance of the KSVD-GOMP-TMSBL algorithm is poorer, and it is slightly inferior to the TMSBL algorithm when the signal-to-noise ratio reaches a certain value. Therefore, to ensure optimal channel estimation performance and algorithm stability, the value of α should be in the range of [0.01 ~ 0.10].

The performance of the conventional GOMP algorithm is notably influenced by sparsity. Figure 11 illustrates the channel estimation error of the GOMP algorithm with varying sparsity values, with the energy coefficient (α) set to 0.05. It is evident from the figure that the subpar channel estimation performance when the

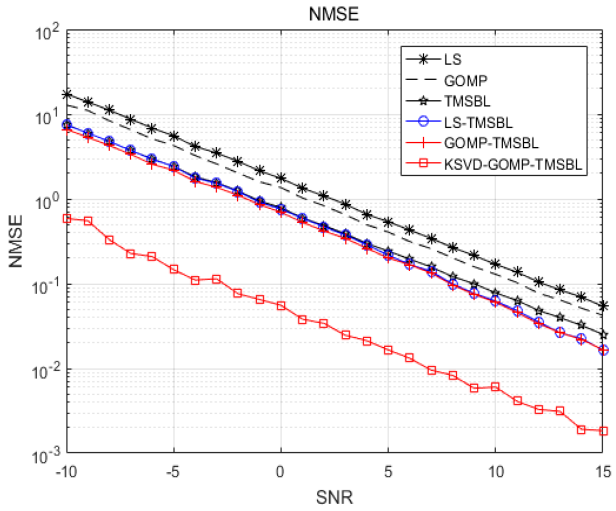


FIGURE 9
Channel estimation error with and without noise reduction.

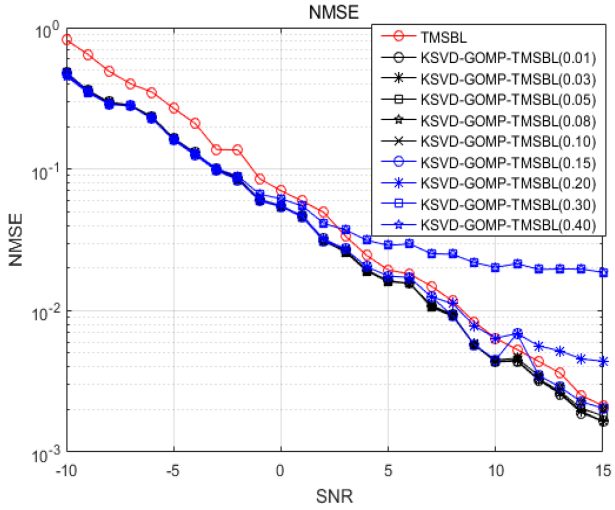


FIGURE 10
Channel estimation error for different energy coefficients.

sparsity is 5 is attributed to the multipath complexity of the shallow sea environment. In such an environment, where the number of propagated multipaths is higher, opting for a small sparsity value may erroneously set the channel's tapping coefficient to 0 when deriving hyperparameters from the *a priori* knowledge of the TMSBL. This leads to a biased channel estimation. The performance of the KSVD-GOMP-TMSBL algorithm is more consistent when the sparsity is set to the other five values, all of which surpass the TMSBL algorithm, enabling superior estimation performance. Thus, the KSVD-GOMP-TMSBL algorithm only needs to adopt a larger sparsity value, as the algorithm is minimally affected by the specific value of sparsity. Consequently, the performance of the KSVD-GOMP-TMSBL algorithm is not influenced by the sparsity chosen by the GOMP algorithm.

The influence of noise variance calculation methods on the performance of the TMSBL algorithm and the KSVD-GOMP-TMSBL algorithm is discussed below. Figure 12 illustrates the channel estimation errors of the KSVD-GOMP-TMSBL algorithm employing two noise variance calculation methods. At an SNR of -10 dB, the NMSE of the channel estimation for the KSVD-GOMP-TMSBL algorithm is 1.058 when the noise variance is computed through iterative updating using (Equation 13). Conversely, the NMSE is reduced to 0.5579 when the noise variance is determined using the null subcarrier in conjunction with (Equation 14). Compared to the previous method, employing null subcarriers to acquire the noise variance diminishes the NMSE of the channel estimation of the KSVD-GOMP-TMSBL algorithm by approximately 47.27%. It is evident that utilizing the null

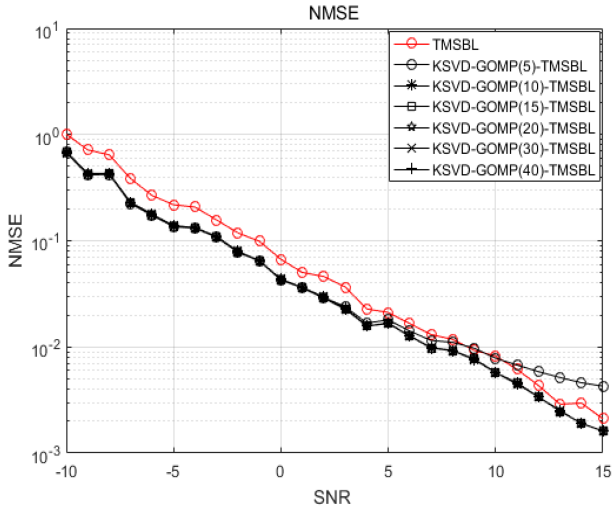


FIGURE 11
NMSE for different sparsities.

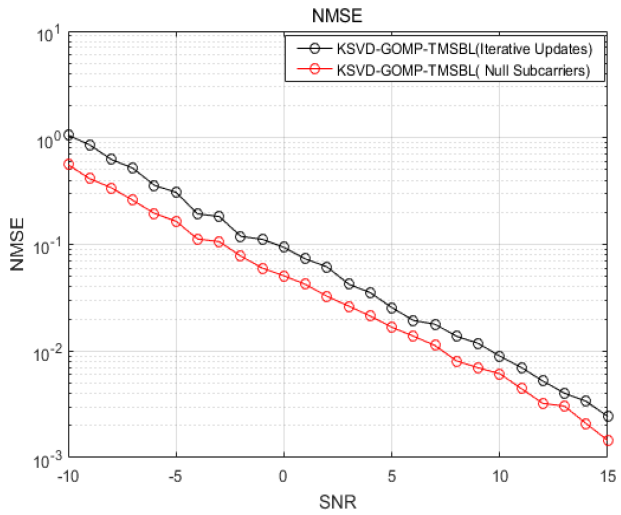


FIGURE 12
Channel estimation errors calculated with different noise variances.

subcarrier to determine noise variance in the OFDM system can enhance the performance of the KSVD-GOMP-TMSBL algorithm and improve the accuracy of channel estimation.

In the following sections, the proposed algorithm is compared with other methods. Figure 13 illustrates the NMSE of various channel estimation methods. The SBL method (Wipf and Rao, 2004) addresses the multi-measurement model by sequentially processing columns to estimate the channel. The TMSBL method (Qiao et al., 2018) leverages temporal correlation between channels to address the multi-measurement model. The SVD-LS-TMSBL method (Hong et al., 2022) utilizes SVD to reduce noise, LS to acquire *a priori* knowledge for the TMSBL algorithm, and TMSBL to achieve joint channel estimation. The figure illustrates that the SBL method exhibits the highest NMSE, which can be attributed to its failure to leverage the correlation between the channels. In contrast, TMSBL demonstrates superior channel estimation accuracy compared to the SBL algorithm due to its ability to capitalize on the temporal correlation between the underwater acoustic channels. The SVD-LS-TMSBL algorithm achieves lower NMSE in channel estimation than the TMSBL algorithm because it integrates singular value decomposition for noise reduction and employs the LS method to acquire the priori knowledge for the TMSBL algorithm. The KSVD-GOMP-TMSBL algorithm employs K-SVD dictionary learning for noise reduction, with *a priori* knowledge for the TMSBL algorithm obtained through the GOMP algorithm. This approach effectively addresses the limitations of the SVD-LS-TMSBL algorithm, where the LS channel estimation method is more sensitive to noise, and the noise reduction capability of SVD is suboptimal in low SNR environments, leading to reduced channel estimation accuracy.

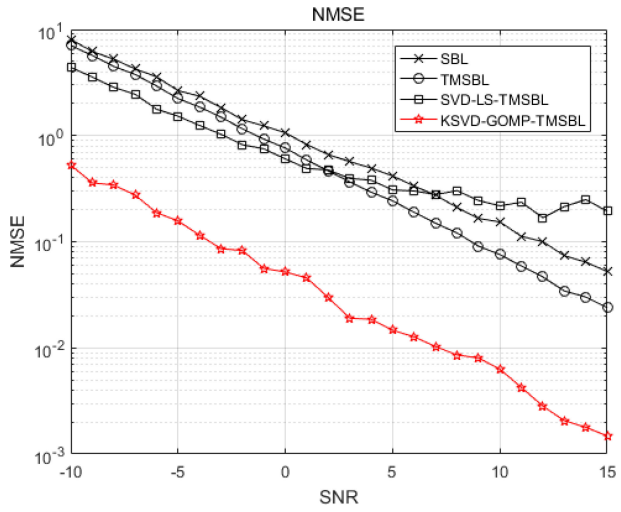


FIGURE 13
NMSE for different channel estimation methods.

These results demonstrate that the KSVD-GOMP-TMSBL method achieves superior channel estimation performance compared to the other algorithms. It can be observed that the proposed algorithm is capable of effectively reducing the impact of noise on the TMSBL algorithm, thereby enhancing the accuracy of channel estimation in OFDM communication systems.

The subsequent analysis centers on the influence of the KSVD-GOMP-TMSBL algorithm on the performance of the OFDM communication system. Figure 14 illustrates the bit error rate (BER) of various channel estimation methods, suggesting that the performance of the OFDM communication system is improved by the KSVD-GOMP-TMSBL algorithm. With an increase in SNR, all five channel estimation methods represented in the figure demonstrate a reduction in the BER of the OFDM communication system. At an SNR of 15 dB, the BERs for the LS, GOMP, TMSBL, LS-TMSBL, and KSVD-GOMP-TMSBL algorithms are 6.2%, 5.7%, 5.8%, 5.3%, and 4.5%, respectively. In comparison to the benchmark algorithms, the BER of the KSVD-GOMP-TMSBL algorithm decreases by at least approximately 15.1%. Based on the aforementioned analysis, the KSVD-GOMP-TMSBL algorithm contributes to the enhancement of the OFDM communication system's performance.

Table 2 presents the estimation errors of various channel estimation methods under identical signal-to-noise ratio conditions. The results clearly indicate that the estimation accuracy of the KSVD-GOMP-TMSBL algorithm is markedly improved compared to other methods. At a signal-to-noise ratio of -10 dB, the channel estimation error of the KSVD-GOMP-TMSBL algorithm decreases by around 96.6%, 95.4%, 92.2%, and 92.1% compared to the LS, GOMP, TMSBL, and LS-TMSBL channel estimation methods, respectively. The estimation error of the KSVD-GOMP-TMSBL algorithm similarly decreases across other signal-to-noise ratio values. This confirms the significant enhancement in channel estimation accuracy achieved by the KSVD-GOMP-TMSBL algorithm.

4.2 Algorithm complexity analysis

The following section will analyze the convergence and convergence rate of the KSVD-GOMP-TMSBL algorithm. Figure 15 illustrates the NMSE of the TMSBL and KSVD-GOMP-TMSBL algorithms for varying numbers of iterations at a signal-to-noise ratio (SNR) of -10 dB. It can be observed that both algorithms demonstrate convergence as the number of iterations increases. The KSVD-GOMP-TMSBL algorithm reaches convergence at 7 iterations, while the TMSBL algorithm reaches convergence at 26 iterations. The results demonstrate that the KSVD-GOMP-TMSBL algorithm exhibits a faster convergence rate than the TMSBL algorithm. It has been demonstrated that the integration of the KSVD-GOMP-TMSBL algorithm for noise reduction based on a KSVD dictionary and the utilization of GOMP to derive the prior knowledge of the TMSBL algorithm can effectively reduce the number of iterations of the TMSBL algorithm and accelerate the convergence speed.

The KSVD-GOMP-TMSBL algorithm is primarily composed of three key components: KSVD dictionary learning for noise reduction, *a priori* knowledge acquisition based on GOMP channel estimation, and TMSBL algorithm channel estimation. The KSVD dictionary learning technique for noise reduction incorporates GOMP sparse coding and dictionary updating. The computational complexity of GOMP sparse coding is $O(kpJL)$, and that of dictionary updating is $O(pJL(L+1))$. Consequently, the computational complexity of KSVD dictionary learning noise reduction is $O(pJL(k+L+1))$. The computational complexity of acquiring *a priori* knowledge through GOMP channel estimation is $O(kpJL+ML)$. The computational complexity of TMSBL channel estimation is $O(N(M^2p+L^4+L^2M))$, where N represents the number of iterations. In conclusion, the principal computational complexity of the KSVD-GOMP-TMSBL algorithm is $O(pJL(k+L+1)+ML+N(M^2p+L^4+L^2M))$. It can be observed that the complexity of the proposed algorithm is predominantly dictated

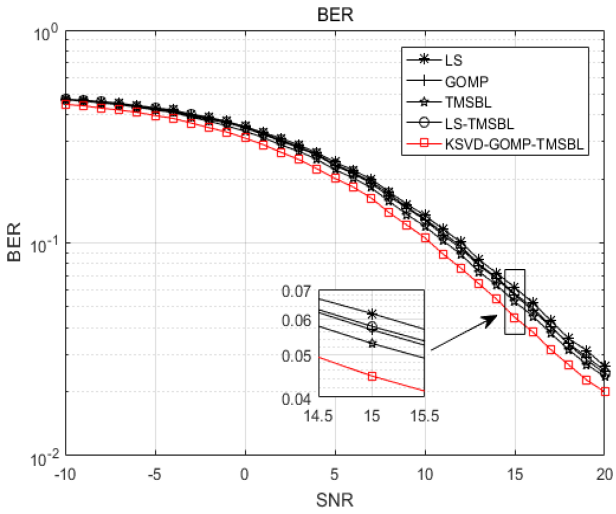


FIGURE 14
BER for different channel estimation methods.

TABLE 2 Estimation errors of different channel estimation methods.

Channel estimation methods	SNR (dB)	estimation error	Channel estimation methods	SNR (dB)	estimation error
LS	-10	17.2500	LS	5	0.5384
GOMP		12.7900	GOMP		0.4062
TMSBL		7.5060	TMSBL		0.2417
LS-TMSBL		7.4570	LS-TMSBL		0.2160
KSVD-GOMP-TMSBL		0.5873	KSVD-GOMP-TMSBL		0.0165
LS	-5	5.5200	LS	10	0.1687
GOMP		4.2530	GOMP		0.1292
TMSBL		2.4120	TMSBL		0.0779
LS-TMSBL		2.4120	LS-TMSBL		0.0623
KSVD-GOMP-TMSBL		0.1462	KSVD-GOMP-TMSBL		0.006
LS	0	1.7420	LS	15	0.054
GOMP		1.3590	GOMP		0.042
TMSBL		0.7897	TMSBL		0.0253
LS-TMSBL		0.7574	LS-TMSBL		0.0166
KSVD-GOMP-TMSBL		0.0556	KSVD-GOMP-TMSBL		0.0018

by the TMSBL channel estimation component. The computational complexity of the TMSBL algorithm is primarily driven by the value of N, especially when the number of iterations is substantial, which may result in a significant increase in the algorithm's overall complexity. The KSVD-GOMP-TMSBL algorithm employs GOMP to derive the initial hyperparameter matrix and the simplified dictionary matrix. This process is analogous to the iterative updating in TMSBL during the intermediate phase, leading to a reduction in the number of iterations and an acceleration of the convergence process. This is corroborated by the results shown in Figure 15. These findings demonstrate that the KSVD-GOMP-TMSBL algorithm effectively reduces the computational complexity of the TMSBL algorithm.

The KSVD-GOMP-TMSBL algorithm utilizes the GOMP channel estimation algorithm to derive the time-domain impulse response $\boldsymbol{h}_{\mathrm{GOMP}}$ of the underwater acoustic channel. It incorporates *a priori* knowledge from the TMSBL and leverages the characteristics of the underwater acoustic channel to determine the initial parameters of the EM algorithm. This approach can effectively reduce the complexity of the algorithm. Utilizing the null subcarrier in conjunction with (Equation 14) to calculate the noise variance, instead of iteratively updating with (Equation 13), can

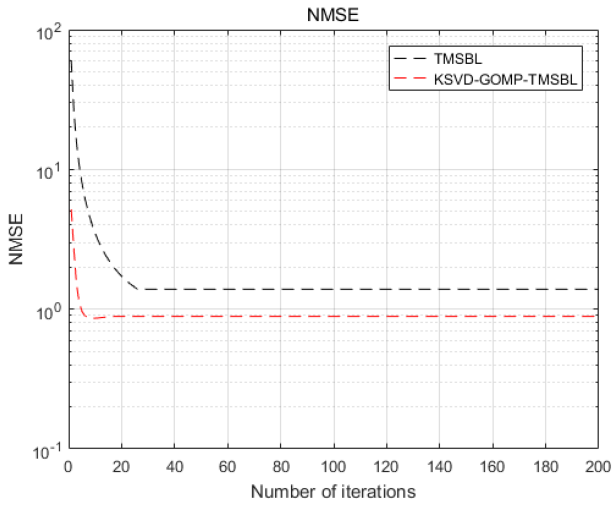


FIGURE 15
NMSE of TMSBL algorithm and KSVD-GOMP-TMSBL algorithm at different numbers of iterations.

theoretically decrease the number of iterations and thus simplify the complexity of the system. The computational complexity of the KSVD-GOMP-TMSBL algorithm will be subsequently analyzed by examining the algorithm's running time. The hardware configuration for the simulation comprises a Core i5 Intel processor at 2.3GHz with 12GB of RAM.

Figure 16 illustrates the comparison of the running time among various channel estimation methods. It is evident that the TMSBL algorithm exhibits the highest running time. The LS *a priori* TMSBL method demonstrates a lower running time compared to the TMSBL algorithm. This is attributed to the utilization of the LS algorithm to acquire *a priori* knowledge for the TMSBL algorithm, thereby reducing its overall complexity. The KSVD-GOMP-TMSBL Algorithm, employing K-SVD Dictionary Learning, the incorporation of a noise reduction algorithm, and utilizing the GOMP algorithm for acquiring *a priori* knowledge of the TMSBL algorithm, diminishes the number of iterations and decreases the running time in comparison to the TMSBL algorithm and the LS-TMSBL algorithm. At a signal-to-noise ratio of -10 dB, the running time of TMSBL is 52.71s, the LS *a priori* TMSBL method records a running time of 40.3s, and the KSVD-GOMP-TMSBL algorithm demonstrates a running time of 28.66s. In comparison to the preceding two methods, the running time diminishes by approximately 45.63% and 28.89%, respectively. Although the complexity of the KSVD-GOMP-TMSBL algorithm remains relatively high compared to channel estimation methods such as LS and GOMP, Figure 9 indicates that the KSVD-GOMP-TMSBL algorithm exhibits superior channel estimation accuracy. Overall, the computational complexity of the KSVD-GOMP-TMSBL algorithm is diminished in comparison to both the TMSBL algorithm and the LS-TMSBL algorithm.

Table 3 illustrates the running time of various channel estimation methods under identical SNR conditions. As evident from Table 3, the running time of the KSVD-GOMP-TMSBL algorithm decreases in comparison to both the TMSBL and LS-TMSBL algorithms. At a signal-to-noise ratio of 0 dB, the running time of the KSVD-GOMP-TMSBL algorithm decreases by approximately 50.3% compared to TMSBL and 33.9% compared to LS-TMSBL algorithms. The running time of the KSVD-GOMP-TMSBL algorithm also decreases at different signal-to-noise ratio values. This indicates a reduction in the computational complexity of the KSVD-GOMP-TMSBL.

5 Sea trial data validation

To validate the feasibility of the proposed algorithm, we utilized data obtained from sea trials in a specific maritime area for verification. For the offshore experiment, the sound source emission device UW350 was positioned at a depth of 5m. The pilot signal utilized was a 200-600Hz broadband long pulse signal with a sampling frequency of 10kHz, and the average in-band signal-to-noise ratio was -0.02dB. Based on GPS data, the distance between the transmitting ship and the receiving ship was calculated as 4672m. The receiving ship positioned the hydrophone in the seawater at a depth of 24m, and the depth of the experimental sea was measured at 25.5m. The depths of the mentioned equipment and seawater were measured by depth sensors.

Figure 17 depicts the channel estimation results obtained through the application of the KSVD-GOMP-TMSBL algorithm. The illustration reveals the relatively stable structure of the shallow-sea underwater acoustic channel, characterized by concentrated channel energy on a few paths, demonstrating sparse characteristics that manifest as a sparse multipath structure.

Figure 18 illustrates the received BER for various channel estimation methods. The sparsity of the GOMP algorithm is set to 20, the energy coefficient of the LS-TMSBL algorithm is set to 0.05, and the sparsity of the KSVD-GOMP-TMSBL algorithm is set to 20, with an energy coefficient of 0.05. As depicted in Figure 16, the BER of the LS algorithm and GOMP algorithm is the highest

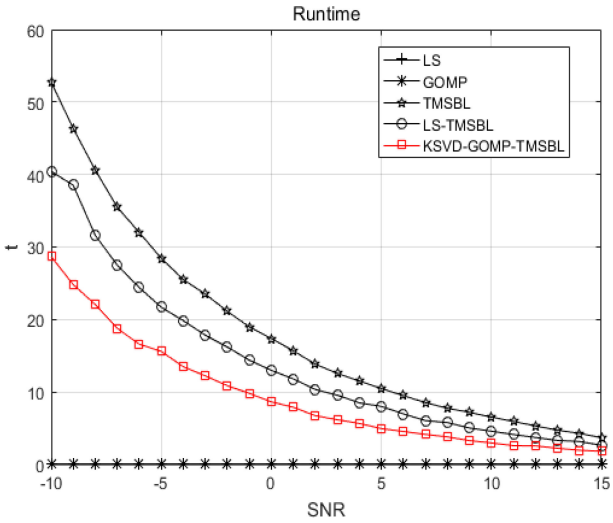


FIGURE 16
Running time of different channel estimation methods.

TABLE 3 Running time of different channel estimation methods.

Channel estimation methods	SNR (dB)	Running time (s)	Channel estimation methods	SNR (dB)	Running time (s)
TMSBL	-10	52.71	TMSBL	5	10.48
LS-TMSBL		40.42	LS-TMSBL		8.02
KSVD-GOMP-TMSBL		28.66	KSVD-GOMP-TMSBL		4.96
TMSBL	-5	28.45	TMSBL	10	6.54
LS-TMSBL		21.91	LS-TMSBL		4.58
KSVD-GOMP-TMSBL		15.6	KSVD-GOMP-TMSBL		2.98
TMSBL	0	17.37	TMSBL	15	3.69
LS-TMSBL		13.08	LS-TMSBL		2.65
KSVD-GOMP-TMSBL		8.64	KSVD-GOMP-TMSBL		1.83

among the considered algorithms. The LS algorithm is notably influenced by noise, leading to a high BER. Meanwhile, GOMP is extremely sensitive to sparsity, and the actual sparsity of the channel in the real environment is unknown, contributing to a high BER for the GOMP algorithm. The BER of both the TMSBL algorithm and the LS-TMSBL algorithm is lower than that of the LS algorithm and GOMP algorithm. In comparison to the aforementioned four algorithms, the BER of the KSVD-GOMP-TMSBL algorithm is lower, further validating its performance.

Figure 19 depicts the runtime of different channel estimation methods. It is evident from the figure that the TMSBL algorithm and the LS-TMSBL algorithm have the longest runtime, approximately 15 seconds. The KSVD-GOMP-TMSBL algorithm boasts a runtime of approximately 11 seconds, marking a reduction of about 26.7% compared to the two preceding algorithms. The LS algorithm and the GOMP algorithm demonstrate the shortest running time; however, as shown in Figure 16, the KSVD-GOMP-TMSBL algorithm outperforms both the LS algorithm and the GOMP algorithm in terms of BER. The superiority of the KSVD-GOMP-TMSBL algorithm in running time is evident.

Table 4 provides the average BER and average runtime for different channel estimation methods. The average BER of different channel estimation methods is higher due to a lower average in-band SNR, and no channel coding is applied. However, it is evident that the KSVD-GOMP-TMSBL algorithm exhibits a lower average BER compared to the comparison algorithms. The KSVD-GOMP-TMSBL algorithm demonstrates the shortest runtime among the three channel estimation methods with a closer BER. This indicates that the KSVD-GOMP-TMSBL algorithm performs effectively in reducing both system BER and system complexity.

6 Conclusion

The estimation of underwater acoustic channels using the TMSBL algorithm in shallow sea environments is challenged by

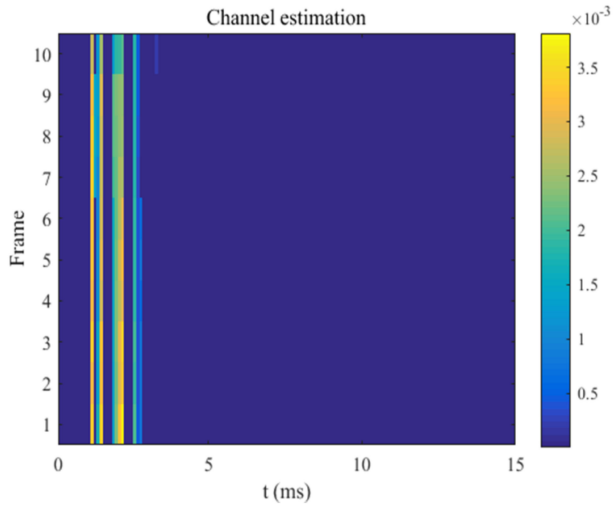


FIGURE 17
Channel estimation results.

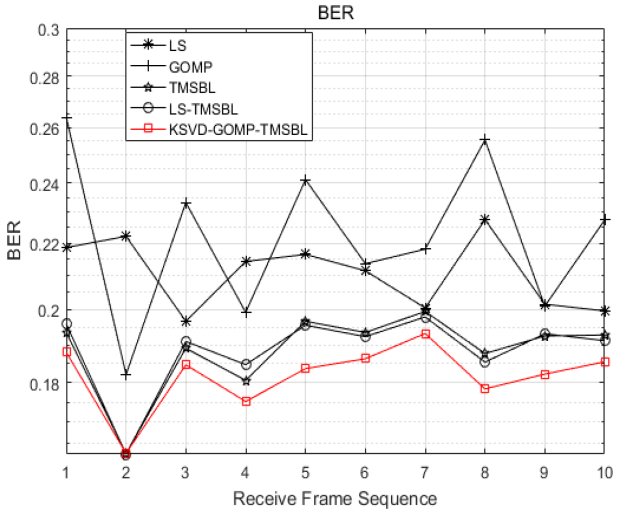


FIGURE 18
BER for different estimation methods.

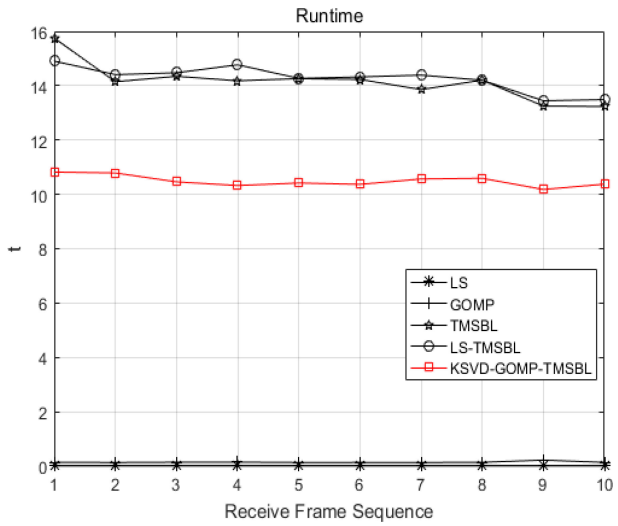


FIGURE 19
Running time of different channel estimation methods.

TABLE 4 Comparison of average BER and average runtime of different channel estimation methods.

Channel estimation methods	Average BER	Average running time (s)
LS	0.2109	0.0054
GOMP	0.2235	0.1672
TMSBL	0.1888	15.5599
LS-TMSBL	0.1889	15.3406
KSVD-GOMP-TMSBL	0.1820	11.5221

high computational complexity, and the algorithm's performance is significantly affected by the signal-to-noise ratio. The article proposes an improved channel estimation method for the temporal multiple sparse Bayesian learning OFDM underwater acoustic communication system. The K-SVD dictionary learning algorithm is employed to reduce noise in the received pilot matrix. Simultaneously, the null subcarrier is utilized to obtain a more accurate noise variance, thereby reducing the computational complexity of the algorithm and enhancing its noise immunity. The method employs the GOMP channel estimation algorithm to obtain the time-domain impulse response of the underwater acoustic channel. It acquires *a priori* knowledge of the TMSBL and selects the initial parameters of the EM algorithm based on the

characteristics of the underwater acoustic channel, thereby enhancing the accuracy of channel estimation. Simulations indicate that at a signal-to-noise ratio of -10 dB, the KSVD-GOMP-TMSBL algorithm reduces the NMSE of channel estimation by 92.2% compared to the TMSBL algorithm, significantly improving estimation accuracy. Furthermore, the running time is reduced by 45.6%, thereby accelerating the convergence of the TMSBL algorithm and reducing its computational complexity. The validation with experimental data from the sea trials demonstrate that the proposed algorithm has a lower impact on the signal-to-noise ratio compared to traditional channel estimation algorithms, and exhibits strong robustness. It achieves accurate estimates even in low signal-to-noise conditions in shallow water and operates at high speed.

The KSVD-GOMP-TMSBL algorithm effectively addresses the problems associated with low estimation accuracy and high computational complexity in the estimation of acoustic channels in shallow water under the conditions of low signal to noise ratio. This algorithm serves as a reference for estimating the acoustic channel in shallow water. Although the algorithm can effectively improve the performance of channel estimation, the reliance on pilot signals means that their quantity will influence the algorithm's performance. Therefore, the next step is to explore channel estimation methods that require fewer pilot signals to further enhance robustness.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

CX: Writing – review & editing, Writing – original draft. YR: Writing – original draft, Writing – review & editing. ML: Writing – review & editing. GT: Writing – review & editing. QM: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China under Grant (61761048), the Basic Research Special General project of Yunnan Province, China (202101AT070132) and Yunnan Minzu University Graduate Research Innovation Fund Project (2024SKY122).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199

Chen, P., Guo, Q., Li, P., and Cui, F. (2020). Joint sparse channel estimation and data detection based on bayesian learning in OFDM system. *Comput. Sci.* 47 (11A), 349–353. doi: 10.11896/jsjkx.191100090

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev.* 43, 129–159. doi: 10.1137/S003614450037906X

Cheng, H. K., and Wang, H. X. (2022). Time varying underwater acoustic channel estimation based on Kalman filter. *Tech. Acoustics* 41, 833–837. doi: 10.16300/j.cnki.1000-3630.2022.06.007

Cotter, S. F., and Rao, B. D. (2002). Sparse channel estimation via matching pursuit with application to equalisation. *IEEE Trans. Commun.* 50, 374–377. doi: 10.1109/26.990897

Giao, G., Song, Q. J., Ma, L., Liu, S. Z., Sun, Z. X., and Gan, S. W. (2018). Sparse Bayesian learning for channel estimation in time-varying underwater acoustic OFDM communication. *IEEE Access* 6, 56675–56684. doi: 10.1109/ACCESS.2018.2873406

Hong, D. Y., Wang, W., Yin, L., Pu, Z. Q., Zhou, C. Y., and Huang, H. N. (2022). An improved temporal multiple sparse Bayesian learning under-ice acoustic channel estimation method. *Acta Acustica* 47 (5), 591–602. doi: 10.15949/j.cnki.0371-0025.2022.05.013

Jia, S. Y., Zou, S. C., Zhang, X. C., and Tian, D. Y. (2022). Multi-block Sparse Bayesian learning channel estimation for OFDM underwater acoustic communication based on fractional Fourier transform. *Appl. Acoustics* 192, 108721. doi: 10.1016/j.apacoust.2022.108721

Jiang, W. H., Tong, F., Zhang, H. T., and Li, B. (2021). Dynamic discriminative compressed sensing estimation of hybrid sparse underwater acoustic channel. *Acta Acustica* 46 (6), 825–834. doi: 10.15949/j.cnki.0371-0025.2021.06.005

Lyu, X. R., Li, Y. M., and Guo, Q. (2021). Joint channel and impulsive noise estimation method for MIMO-OFDM systems. *J. Commun.* 42, 54–64. doi: 10.11959/j.issn.1000–436x.2021238

Meng, X. Y., and Liu, Z. L. (2023). TB-GOMP channel estimation algorithm for shallow underwater acoustic communication. *J. Ordnance Equip. Eng.* 44, 223–229. doi: 10.11809/bqzbgcxb2023.05.032

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B: Stat. Method.* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Tong, F., Wu, F. Y., and Zhou, Y. H. (2022). *Underwater acoustic channel estimation* (China: Science Press).

Wang, H., Sun, T. J., and Liu, T. (2020). Active sonar target classification based on dictionary learning. *Tech. Acoustics* 39 (5), 552–558. doi: 10.16300/j.cnki.1000-3630.2020.05.006

Wipf, D. P., and Rao, B. D. (2004). Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* 52, 2153–2164. doi: 10.1109/TSP.2004.831016

Wipf, D. P., and Rao, B. D. (2007). An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* 55, 3704–3716. doi: 10.1109/TSP.2007.894265

Wu, F. Y., and Tong, F. (2017). Improved compressed sensing estimation of block sparse underwater acoustic channel. *Acta Acustica* 42, 27–36. doi: 10.15949/j.cnki.0371-0025.2017.01.004

Xing, C., Dong, S., and Wan, Z. (2023). Direction-of-arrival estimation based on sparse representation of fourth-order cumulants. *IEEE Access* 11, 128736–128744. doi: 10.1109/ACCESS.2023.3332991

Xing, C. X., Wang, Z. L., Jiang, S. Y., and Yu, R. M. (2022). Direction of arrival estimation based on high-order cumulant by sparse reconstruction of underwater acoustic signals. *Acta Acustica* 47, 440–450. doi: 10.15949/j.cnki.0371-0025.2022.04.010

Xing, C. X., Wu, Y. W., Xie, L. X., and Zhang, D. Y. (2021a). A sparse dictionary learning-based denoising method for underwater acoustic sensors. *Appl. Acoustics* 180, 108140. doi: 10.1016/j.apacoust.2021.108140

Xing, C. X., Zhang, D. Y., Song, Y., Wu, Y. W., and Xie, L. X. (2021b). Research on inversion of sound speed profile using dictionary learning method. *Tech. Acoustics* 40 (6), 750–756. doi: 10.16300/j.cnki.1000-3630.2021.06.002

Xu, W., Yan, S. F., Ji, F., Chen, J. D., Zhang, J., Zhao, H. F., et al. (2016). Marine information gathering, transmission, processing, and fusion: current status and future trends. *Scientia Sinica(Informationis)* 46 (8), 1053–1085. doi: 10.1360/N112016-00064

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimed Tools Appl.* 83, 31957–31973. doi: 10.1007/s11042-023-16757-0

Yin, J. W., Gao, X. B., Han, X., Zhang, X., Wang, D. Y., and Zhang, J. C. (2021). Underwater acoustic channel estimation and impulsive noise mitigation based on sparse Bayesian learning. *Acta Acustica* 46 (6), 813–824. doi: 10.15949/j.cnki.0371-0025.2021.06.004

Zhang, X. (2023). An efficient method for the simulation of multireceiver SAS raw signal. *Multimed Tools Appl.* 83, 37351–37368. doi: 10.1007/s11042-023-16992-5

Zhang, X. B., Yang, P. X., Wang, Y. M., Shen, W. Y., Yang, J. C., Wang, J. F., et al. (2024). A novel multireceiver SAS RD processor. *IEEE Trans. Geosci. Remote Sens.* 62, 1–11. doi: 10.1109/TGRS.2024.3362886

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, Z., and Rao, B. D. (2011). Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Selected Topics Signal Process.* 5, 912–926. doi: 10.1109/JSTSP.2011.2159773

Zhu, L., Liu, S., Cao, S. N., and Liu, Y. Y. (2020). Nonparametric Bayesian dictionary learning in sparse gradient domain for image denoising. *Comput. Eng. Design* 41 (3), 802–807. doi: 10.16208/j.issn1000-7024.2020.03.032

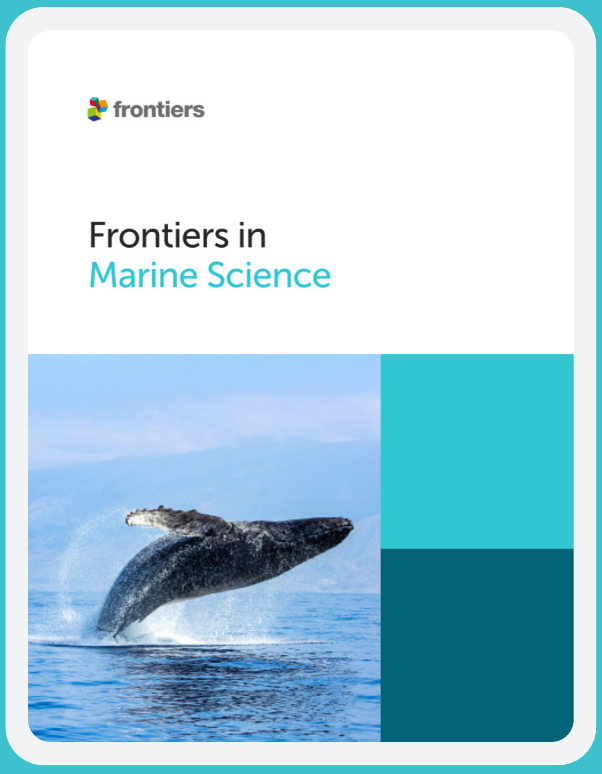
Frontiers in Marine Science

Explores ocean-based solutions for emerging global challenges

The third most-cited marine and freshwater biology journal, advancing our understanding of marine systems and addressing global challenges including overfishing, pollution, and climate change.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact