frontiers | Research Topics

Ocean observation based on underwater acoustic technology

Edited by
Xuebo Zhang, Haixin Sun and Arata Kaneko

Published in
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-2874-7
DOI 10.3389/978-2-8325-2874-7

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

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

Citation
Zhang, X., Sun, H., Kaneko, A., eds. (2023). *Ocean observation based on underwater acoustic technology*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2874-7

Table of contents

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 15 June 2023
DOI 10.3389/fmars.2023.1212840

OPEN ACCESS

EDITED AND REVIEWED BY
Mohammed Elamassie,
Özyeğin University, Türkiye

*CORRESPONDENCE
Xuebo Zhang
xuebo_zhang@sina.cn

RECEIVED 27 April 2023
ACCEPTED 05 June 2023
PUBLISHED 15 June 2023

CITATION
Zhang X, Sun H and Kaneko A (2023)
Editorial: Ocean observation based on underwater acoustic technology.
Front. Mar. Sci. 10:1212840.
doi: 10.3389/fmars.2023.1212840

Editorial: Ocean observation based on underwater acoustic technology

Xuebo Zhang[1]*, Haixin Sun[2] and Arata Kaneko[3]

[1]Research & Development Department, Whale Wave Technology Inc., Kunming, Yunnan Province, China, [2]School of Informatics, Xiamen University, Xiamen, China, [3]Hiroshima University, Hiroshima, Japan

KEYWORDS

underwater, detection, identification, communication, submarine

Editorial on the Research Topic
Ocean observation based on underwater acoustic technology

1 Introduction

The world's military powers have always attached great importance to the collection of Marine environment and hydroacoustic data. Marine monitoring technology plays an important role in promoting the development of economic, military and ecological protection. The performance of underwater communication and target detection algorithms is influenced by the nonlinear interference of underwater communication noise, speed of water surface ripples, weather changes, and geological structure of the seabed. To address the problems of environment interference, professionals and technical personnel propose corresponding technical solutions to address these issues. These solutions aim to improve and enhance the detection and identification capabilities when encountering such challenges. The use of sound waves as information carriers for long-distance transmission in the ocean is necessary, making the research and development of sound wave technology crucial in ocean exploration. Nevertheless, the extensive application of acoustic wave technology in ocean exploration relies on signal processing foundations and sensor support.

2 Underwater communication

Underwater communication is susceptible to various interfering factors. As a result, ensuring high efficiency and speed in underwater communication has become significantly important, leading to increased attention towards communication network protocols (Ahmed and Eltawil, 2015). On the other hand, due to realistic constraints, such as extreme spatiotemporal variations in the underwater acoustic channel, long transmission delays, and limited bandwidth, the application of mature network protocols in underwater communication often fails. Consequently, it is crucial to address the problem of improving system information throughput and network frequency spectrum efficiency in underwater acoustic communication with limited bandwidth (Lu et al). Addressing this challenge, a proposed method aims to overcome the limitations of previous self-interference cancellation (SIC) method, known as spatial self-interference cancellation (SSIC). This method effectively inhibits interference from array and direction-related aspects while also preventing distortion of desired signals over long distances. The implemented in-band full-duplex underwater acoustic communication (IBFD-UWAC) system utilizes improved constant beamwidth beamforming (ICBBF) (convex optimization theory), and a variable step-size least-squares algorithm is proposed for spatial noise threshold in the digtal self-interference cancellation (DSIC) method. Compared to the traditional variable step-size least mean square (VSS-LMS) method, this algorithm demonstrates a more effective elimination effect and achieves faster convergence (Wei et al).To address the hidden terminal problem caused by the unique channel characteristics of underwater sound, including long delay hidden terminals and multi-channel hidden terminals, a multichannel medium access control (MAC) solution is proposed. The channel allocation scheme is determined based on the physical transmission model, and graph coloring theory is utilized to improve channel utilization. Additionally, a channel selection rule method is proposed for graph coloring-based multichannel MAC protocol (GCMAC) data negotiation to enhance efficiency. Experimental results demonstrate that the designed multi-channel MAC protocol not only exhibits a higher success rate of receiving packets compared to single-channel protocols.

3 Underwater acoustic wave propagation

In shallow environments, several factors significantly interfere with the propagation characteristics of sound waves, including the unique structure of the seafloor, sound wave propagation, and attenuation in the ocean (Zhu et al). In their study on the inversion method of acoustic parameters in the marine environment, Bayesian theory was employed to verify the attenuation, sound velocity, seabed structure, and sound velocity density in the seabed. The focus was on matching the characteristic impedance of the vertical waveguide. The inversion results offer valuable insights for the application of existing underwater sensors. Accurate measurement of tidal currents in complex shallow sea environments is also crucial. Researchers have employed tools such as coastal acoustic tomography and mutual acoustic transmission for detecting tidal currents. Data assimilation techniques, specifically the use of the ensemble Kalman filter (EnKF), have been utilized to enhance measurement efficiency (Taniguchi et al). The experimental data used in their study were obtained from path average current results from Japanese researchers in a completed reverse acoustic transmission experiment in the Seto Inland Sea. By redesigning four acoustic station detection points, the EnKF was applied to observe temporal and spatial changes in tidal currents. Various combinations, including location values, the number of system members, and different inflation values, were employed during the implementation of EnKF. The study also identified the cause of high variation in the mean current of the path, attributing it to temporal and spatial variations produced by eddies associated with island wakes. Additionally, EnKF's mutual acoustic transmission capability allowed for the capture of short-period changes within long periods. These findings contribute to the understanding and development of coastal physical processes across different time scales.

4 Underwater imaging

Obtaining clear and accurate information about the natural distribution of underwater environments is particularly important in the study of the ocean, highlighting the significance of underwater imaging (Huang and Yang). Researchers have designed a technique that surpasses traditional imaging methods in terms of producing higher-resolution images. They employ the subblock processing method in the cross-trajectory dimension, leveraging the analysis of cross trajectory variance of super-resolution convolution (SRC). This approach yields similar high-resolution results to the back projection (BP) algorithm benchmark tests while significantly reducing processing time, resulting in improved efficiency (Li et al).Furthermore, a method based on a convolutional recursive neural network (CRNN) has been proposed to enhance underwater imaging performance. The underwater direction of arrival (DOA) estimation utilizes the acoustic array method. The results demonstrate that the proposed CRNN method achieves higher accuracy and faster processing times, making it suitable for broader applications.

5 A marine mobile device

In complex shallow water environments, a marine mobile device is essential for collecting experimental data for scientific research (Totland et al). Drawing on design concepts inspired by the appearance of a two-person kayak used for adventurous activities, researchers developed an unmanned surface vehicle (USV) known as the kayak drone. The primary objective was to construct the kayak drone using existing open-source software and hardware. Previous field experiments have demonstrated that the noise generated by the kayak drone is negligible, allowing it to record the data from fish echo-sounders in the surrounding waters without disturbing fish activities or the natural environment. Additionally, the kayak drone can be employed for baseline-free survey estimations of fish populations in shallow water, which is crucial for the management and assessment of various fish populations. The kayak platform finds applications beyond the aforementioned aspects and can be widely utilized in experiments involving different silent platforms.

6 Marine animal protection

With the passage of time and advancements in science and technology, various technologies utilized in marine research have become increasingly mature. Among these technologies, the protection of the marine environment and the well-being of sea-dwelling animals require dedicated attention and research (Rorstadbotnen et al).

In a study, researchers examined five hours of data from two fiber-optic telecommunication cables and successfully detected 1,808 fin whale vocalizations. This breakthrough opens up new possibilities for accurately mapping the presence and location of whales over large areas, spanning at least 60 km in length and 20 km in width, in near real-time. The simultaneous tracking of multiple whales also offers the potential for gaining new insights into their behavior and interactions within a corridor of approximately 10 km on either side of the cable. By utilizing two fiber cables separated by a few kilometers, the usual left-right ambiguity of a single straight-line array is resolved. The demonstrated capabilities present an opportunity for near-real-time whale tracking that can be implemented worldwide wherever whales are present. Furthermore, by incorporating fused data from other sources such as the Automatic Identification System (AIS), a real-time collision avoidance system, it becomes feasible to develop measures for reducing ship strikes. Successful examples of underwater acoustic systems include passive acoustic monitoring systems, which are not only stable, reliable and cost-effective, but have been widely used to estimate the location and seasonal presence of some animals.

7 Conclusion

Significant progress has been achieved in the study of coastal physical processes and underwater acoustic communication. The application of the GCMAC protocol and the IBFD-UWAC system with DSIC method has greatly enhanced the frequency and efficiency of underwater communication networks. In coastal surveillance and navigation, underwater acoustic array processing techniques, such as Synthetic Aperture Sonar and DOA estimation based on deep learning algorithms, have proven to be vital. These techniques contribute to improved capabilities for monitoring coastal regions and facilitating navigation tasks. Furthermore, advancements in underwater exploration technology are continuously improving. The development of innovative tools and methods enables more effective and efficient exploration of the underwater environment. Further research on underwater acoustic technology is needed in the future so as to be more familiar with the Marine environment.There is much room for future advances in underwater acoustic technology, such as enhancing the reliability and efficiency of underwater communication systems, and improving the accuracy of environmental monitoring so that the ocean can be observed and studied more comprehensively.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors would like to thank the support from the teachers and students in the Underwater Acoustic Communication Lab of Xiamen University.

Conflict of interest

XZ is employed by the R&D Department, Whale Wave Technology Inc., 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

Reference

Ahmed, E., and Eltawil, A. M. (2015). All-digital self-interference cancellation technique for full-duplex systems. *IEEE Trans. Wirel. Commun.* 14, 3519–3532. doi: 10.1109/TWC.2015.2407876

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 27 July 2022
DOI 10.3389/fmars.2022.973284

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Qingjian Liang,
South China Normal University, China
Pan Huang,
Weifang University, China

*CORRESPONDENCE
Keyu Chen
chenkeyu@xmu.edu.cn

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 20 June 2022
ACCEPTED 29 June 2022
PUBLISHED 27 July 2022

CITATION
Wei M, Chen K, Lin Y and Cheng E
(2022) Recognition of behavior state
of *Penaeus vannamei* based on
passive acoustic technology.
Front. Mar. Sci. 9:973284.
doi: 10.3389/fmars.2022.973284

Recognition of behavior state of *Penaeus vannamei* based on passive acoustic technology

Maochun Wei[1], Keyu Chen[2]*, Yating Lin[2] and En Cheng[2]

[1]Marine Electromechanical Department, Marine electromechanical Department, Xiamen Ocean Vocational College, Xiamen, China, [2]Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Xiamen, China

Underwater signal processing based on passive acoustic technology has carried out a lot of research on the behavioral sound of underwater creatures and the protection of marine resources, which proves the feasibility of passive acoustic technology for long-term and non-destructive monitoring of underwater biological sound production. However, at present, most relevant research focuses on fish but little research on shrimp. At the same time, as the main economic breeding industry, *Penaeus vannamei* has a backward industrial structure, in which the level of intelligence needs to be improved. In this paper, the acoustic signals generated by different physiological behaviors of *P. vannamei* are collected based on passive acoustic technology. Their different behaviors are finally classified and identified through feature extraction and analysis. Meanwhile, the characteristic non-parametric ANOVA is carried out to explore the relationship between the acoustic signals and the behavior state of *P. vannamei* to achieve the purpose of real-time monitoring of the behavior state of *P. vannamei*. The experimental results show that linear prediction cepstrum coefficient (LPCC) and Mel-frequency cepstrum coefficient (MFCC) characteristic coefficients are effective in the classification and recognition of different behavioral acoustic signals with interspecific acoustic signals of *P. vannamei*. Meanwhile, the SVM classifier based on OvR classification strategy can model the acoustic signal characteristics of different underwater biological behaviors more efficiently and has classification accuracy as high as 93%.

KEYWORDS

passive acoustic technology, acoustic signal of *Penaeus vannamei*, feature extraction, SVM, underwater

Introduction

For underwater organisms, specific psychological and physiological behaviors are difficult to observe visually. However, they can transmit information within and between species by making sound, which means that they could produce different signals according to specific behavior (Thomsen et al., 2020). Since the 1970s, acoustic methods have been used by fishery developed countries to investigate and evaluate fish resources. Among them, passive acoustic technology is the key technology of target detection, positioning, tracking, and recognition according to the radiated noise of the underwater target, which is gradually applied in the research of underwater organisms because of its low harmfulness and destructiveness to the research object. Meanwhile, the collected signal contains the characteristics of the sounding target itself to characterize the target behavior (Mann et al., 2016). Most of the organisms that can emit and rely on sound waves recorded in the study are fish. Now, it is known that there are 109 families, and more than 800 kinds of fish can make sound (Rountree et al., 2006), most of which are large mammals, such as whales and dolphins, and important economic fish, such as cod, grouper, and yellow croaker. Some invertebrates in important fisheries can also send out sound signals by behavioral movement, such as white shrimp (Berk, 1998), American lobster (Henninger and Watson, 2005), and squid (Iversen et al., 1963). Because of this, the difference between passive acoustic technology and other types of biological acoustics is using the underwater signal acquisition tools such as hydrophones, which are used to collect the acoustic signals of underwater marine organisms instead of the artificially generated sound. It can be used to find and monitor the organisms that emit the acoustic signals. As a non-invasive and non-destructive observation tool, it provides the ability of continuous long-term and remote monitoring. This long-term monitoring can provide important information about the daily and seasonal activity patterns of fish and other marine organisms. It can not only provide an effective acoustic signal database of aquatic organisms but also be of great significance for the development of marine resources and the protection of marine ecology (Noda et al., 2016).

Back in the 1950s, many scholars first studied the different vocal behaviors of aquatic species based on passive acoustic technology. In the field of fish biology and fishery, fishing (Fish et al., 1952) began to apply passive acoustic technology in the study of underwater sound signals generated by fish along the North Atlantic coast in 1952. Nordeide and Kjellsby (1999) recorded and collected the sound generated by the northeast Arctic cod (*Gadus morhua* L.) near Norway and suggested that passive acoustics can be used to study the spawning behavior. Scharer et al. (Schärer et al., 2014) conducted passive acoustic and synchronous video recording at two spawning gathering points in a spawning season on Mona Island, Puerto Rico, to study the sound related to the reproductive behavior of black grouper and quantify the correlation between the sound signal and the temporal and spatial distribution of reproductive behavior. Parks et al. (2011) recorded the vocal behavior of North Atlantic right whales (*Eubalaena glacialis*) in the Bay of Fundy, Canada, explored the differences in vocal call types and calling rates of North Atlantic right whales under different behavioral states, and proved that the behavioral state is the main factor affecting the calling rates. Soldevilla et al. (2010) used the passive acoustic records of Risso's dolphin at six locations in the Southern California Bight from 2005 to 2007 to explore the spatial and temporal trends of its echolocation behavior and motion patterns. The final results showed that Risso dolphins foraged at night, and the southern end of Santa Catalina Island was an important habitat for Risso dolphins throughout the year. In addition to other underwater organisms of fish, Au et al. (Au and Banks, 1998) studied the environmental noise in Kaneoh Bay, which generated when the claws of snapping shrimp are closed, so as to reduce its impact on other offshore organisms and underwater acoustic using for human. Silva et al. (2019) studied the vocal mechanism and main related acoustic variables of feeding signals produced by *Litopenaeus vannamei* with different body sizes under artificial feeding conditions and explored the relationship between feeding consumption rate and feeding sound signals. Daniel et al. (Smith and Tabrett, 2013) analyzed the feeding sound characteristics of tiger shrimp according to the feeding signal of tiger shrimp in a commercial pond with complex acoustics, which provided a reliable means to detect feeding activities.

Related works

At present, there are few studies on the acoustic signals produced by shrimp culture, especially the life behavior of *Penaeus vannamei*. Therefore, based on the passive acoustic technology, this paper collects the acoustic signals generated by different physiological behaviors of *P. vannamei* and analyzes the effective characteristics of the collected signals. Meantime, it cooperates with the efficient classifier and recognizes different behavior states to study the relationship between different behaviors and acoustic signal characteristics of *P. vannamei*. After receiving a large number of underwater acoustic signals by signal acquisition equipment such as hydrophone, underwater acoustic signal classification and recognition technology analyzes and transforms the acoustic signals accordingly to the automatic classification and recognition through various classification methods. In the practical application of underwater acoustic signal processing, due to the complexity of underwater environment and targets, underwater acoustic signal classification and recognition have always been a hot and difficult point. However, extracting effective signal features are important premises to improve the classification and recognition accuracy and the design of classification method. The difference between the combination method and features will also affect the final classification results.

So far, in the research on the classification and recognition of underwater biological vocal signals, the mainstream research tends to extract the characteristics of time, frequency, and time-frequency domains and classify them by using classical statistical methods, machine learning, and deep learning. The previous research work of this paper mainly focuses on the characteristics of acoustic variables in time and frequency domains (Wei et al., 2020). Mellinger et al. (Mellinger and Clark, 2000), based on the spectral correlation method, constructed a two-dimensional synthetic kernel according to the segmented sound signal of bowhead whale to include the shape of the spectrum diagram called by the target, cross-correlated with the spectrum diagram of the signal to be classified, so as to generate a recognition function and apply the threshold to this function to distinguish the sound signal of bowhead whale. Compared with the classification results of other classifiers such as matched filter, neural network, and hidden Markov model, it is the best method of classification and recognition. Gillespie et al. (2013) described the fully automatic detection and classification method of odontocete whistles. This method finds out that the connected data areas are higher than different pre-determined thresholds based on the spectrum of noise-eliminated sound data and then classifies different species. Finally, the correct classification rate of the four odontocete species has reached more than 94%. However, the classification rate depends largely on the number of species categories entered. When 12 species are included, the average correct classification rate decreases to 58.5%. Relatively, with the deepening of interdisciplinary research in recent years, more and more researchers use signal recognition methods in the field of speech signal processing based on human auditory perception mechanism to recognize underwater target acoustic signals, which opens up a new direction. Ibrahim et al. (2016) proposed a feature extraction method composed of discrete wavelet transform (DWT) and Mel-frequency cepstrum coefficient (MFCC), which is classified by support vector machine (SVM). Experimental studies show that the proposed detection scheme outperforms the spectrogram-based techniques in both detection accuracy and speed. Vieira et al. (2015)extracted cepstrum, MFCC, perceptual linear predictive (PLP), and other characteristic coefficients from different acoustic signals emitted by male toadfish, and classified them by hidden Markov model (HMM) classifier. The final cepstrum coefficient has the best classification effect. Pace et al. (2010) extracted MFCC feature coefficients from humpback whale vocal signals collected in Madagascar in August 2008 and 2009 and used K-means clustering algorithm for automatic classification, which has higher accuracy than manual classification. Taking the underwater acoustic signals of different behaviors of *P. vannamei* collected based on passive acoustic technology as the research object, and improving the passive recognition ability of *P. vannamei* underwater behavior as the research goal, this paper studies the feature extraction algorithm and classification recognition model of underwater acoustic signals of different behaviors of *P. vannamei*. The research results are expected to be applied to the field of underwater biological classification and recognition, including the analysis of underwater acoustic signals of *P. vannamei*, feature extraction, classification and recognition, and other related application fields.

The main work accomplished in this paper is as follows:

1. Extracting the MFCC and linear prediction cepstrum coefficient (LPCC) features of different behavioral acoustic signals of *P. vannamei*.
2. Performing non-parametric ANOVA on the extracted features to reduce feature dimension.
3. Using SVM classifier to classify and identify different behavioral signals of *P. vannamei* according to the characteristics.

The rest of the paper is organized as follows: in *s III*, the feature extraction algorithm of acoustic signals from different behaviors of *P. vannamei* is presented. In *Section IV*, the characteristic non-parametric ANOVA is done. In *Section V*, the experimental results are carried out. Finally, conclusion is provided in *Section VI*.

Acoustic signal feature extraction of *P. VANNAMEI*

Linear prediction cepstrum coefficient

As shown in Figure 1, the LPC coefficient is homomorphically processed to obtain the LPCC. Since the LPCC mainly reflects the channel frequency response, that is, the spectral envelope information of the signal, T can be approximately used as the short-time cepstrum of the original acoustic signal, which can improve the stability of the characteristic parameters (Gupta and Gupta, 2016).

The channel transfer function $H(z)$ is obtained by linear prediction analysis, and its impulse response is $h(n)$. The cepstrum of $\hat{h}(n)$ is obtained.

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{1}$$

p is the order, and a_i is the coefficient. Since $H(z)$ is resolved in the unit circle, take logarithms on both sides of the above formula at the same time to obtain:

$$\hat{H}(z) = \log H(z) = \sum_{n=1}^{+\infty} \hat{h}(n) z^{-n} \tag{2}$$

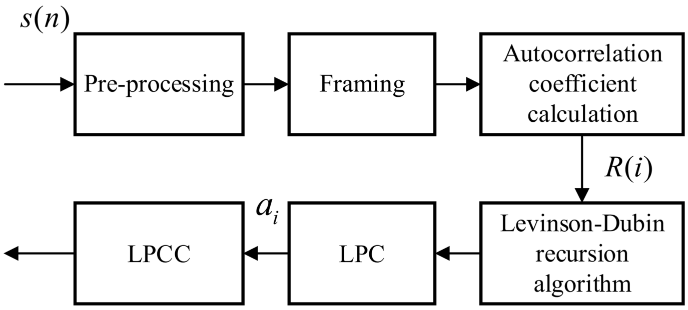


FIGURE 1
Flow chart of linear prediction cepstrum coefficient (LPCC) characteristic coefficient calculation.

If $\hat{h}(n) = 0$, the derivation of both sides of the above formula is obtained at the same time z^{-1}:

$$\frac{\partial}{\partial z^{-1}}\log\frac{1}{1-\sum_{i=1}^{p}a_iz^{-i}} = \frac{\partial}{\partial z^{-1}}\sum_{n=1}^{+\infty}\hat{h}(n)z^{-n} \tag{3}$$

After sorting, we get:

$$\frac{\sum_{i=1}^{p}ia_iz^{-i+1}}{1-\sum_{i=1}^{p}a_iz^{-i}} = \sum_{n=1}^{+\infty}n\hat{h}(n)z^{-n+1} \tag{4}$$

Namely:

$$(1-\sum_{i=1}^{p}a_iz^{-i})\sum_{n=1}^{+\infty}n\hat{h}(n)z^{-n+1} = \sum_{i=1}^{p}ia_iz^{-i+1} \tag{5}$$

Make the coefficients of z on both sides of **Equation (5)** equal to each power and obtain the recurrence relationship between $\hat{h}(n)$ and a_i:

$$\begin{cases} \hat{h}(1) = a_1 \\ \hat{h}(n) = a_n + \sum_{i=1}^{n-1}(1-\frac{i}{n})a_i\hat{h}(n-i) \quad 1 < n \le p \\ \hat{h}(n) = \sum_{i=1}^{p}(1-\frac{i}{n})a_i\hat{h}(n-i) \quad n > p \end{cases} \tag{6}$$

From the above formula, the cepstrum $\hat{h}(n)$ can be recursively obtained from the LPC prediction coefficient a_i. Finally, the LPCC characteristic coefficient a_i of behavioral acoustic signal can be obtained for subsequent behavior state classification. Two different behavioral acoustic signals with a time length of 5 s are intercepted, respectively, for 12-dimensional LPCC feature coefficient extraction. Setting the signal frame length to 20 ms, the frame shifts to 10% of the frame length, and the LPCC feature coefficient size to 2,487 frames in 12 dimensions, as shown in Figure 2.

Mel-frequency cepstrum coefficient

A large number of studies have shown that due to the non-linear auditory sensing mechanism of human ears, they have different auditory perception abilities for sounds with different frequencies. The specific performance is as follows: the perception ability for sounds below 1 kHz follows a linear relationship, while the perception ability for sounds above 1 kHz follows a logarithmic relationship. Therefore, the researchers introduced Mel frequency to describe the non-linear characteristics of human auditory system (Tong et al., 2020): high-frequency resolution at low frequency and low-frequency resolution at high frequency. Mel frequency is defined as follows:

$$f_{Mel} = 2595\,\lg(1 + f/700) \tag{7}$$

where f_{Mel} is Mel frequency, and the unit is Mel. f is the actual frequency, and the unit is Hz. The corresponding relationship curve between them is shown in Figure 3.

Therefore, according to the division method of critical bandwidth, the critical band-pass filter bank, namely, Mel filter bank, is set within the actual spectrum range of the signal, which can simulate the frequency perception characteristics of human ears, as shown in Figure 4.

As can be seen from the Figure 4, Mel filters are formed by superposition of several triangular band-pass filters, and the transfer function of each band-pass filter is:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \le k \le f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{8}$$

In the **Equation (8)**, $0 \le m < M$, and M is the number of filters.$f(m)$ is the center frequency. The spacing between the center frequencies increases with the increase m. $f(m)$ can be defined as:

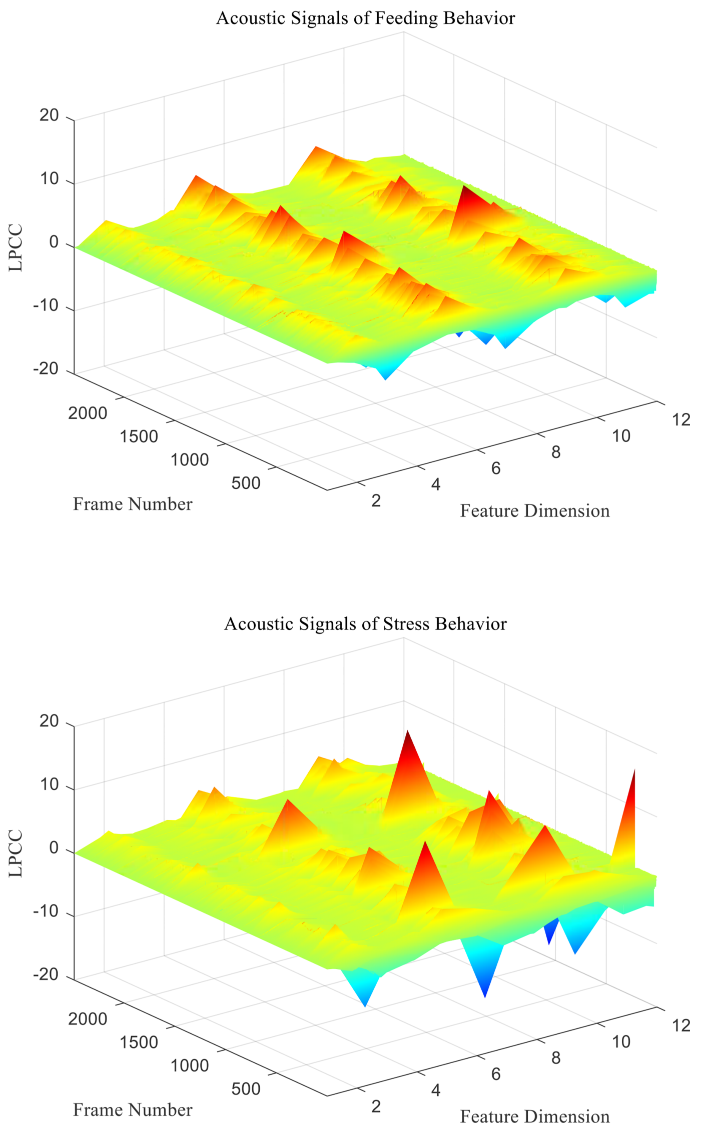


FIGURE 2
LPCC characteristic coefficients of acoustic signals from different behaviors of *P. vannamei*.

$$f(m) = (\frac{N}{f_s})f_{Mel}^{-1}(f_{Mel}(f_l) + m\frac{f_{Mel}(f_h) - f_{Mel}(f_l)}{M+1}) \qquad (9)$$

In **Equation (9)**, the lowest frequency of the filter range is f_l. The highest frequency is f_h. N is the DFT length, and f_s is the sampling frequency. The inverse function of f_{Mel} is:

$$f_{Mel}^{-1}(b) = 700(e^{b/1125} - 1) \qquad (10)$$

The key point of MFCC feature extraction is to transform the signal linear power spectrum into Mel scale power spectrum. The actual spectrum of the signal can be transformed into Mel scale spectrum through Mel filter bank, and then, the cepstrum coefficient can be calculated, which has good robustness. The main process of MFCC feature parameter extraction of behavioral acoustic signal of *P. vannamei* is shown in Figure 5.

The specific process is as follows:

1) Pre-processing

Pre-processing includes pre-denoising, framing, and windowing with Hamming window. The behavior signal after pre-processing is s_i (n), and the subscript i represents the ith frame after framing.

2) Fast Fourier Transform

The behavior signal of each frame after framing is transformed by FFT, and the signal is converted from time domain to frequency domain.

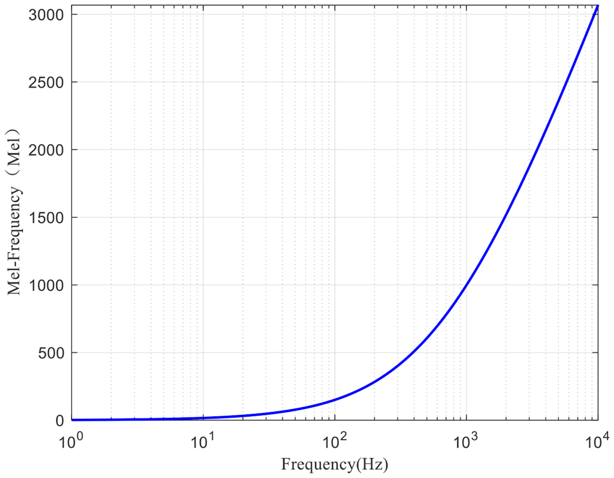


FIGURE 3
Corresponding curve between Mel frequency and actual frequency.

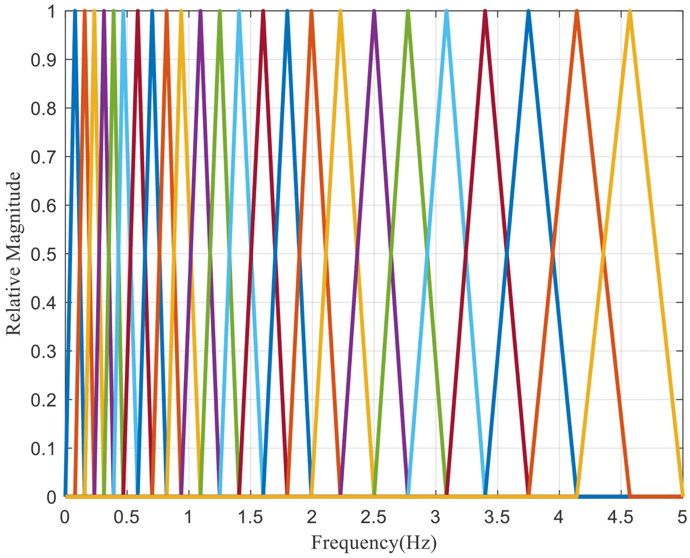


FIGURE 4
Frequency response curve of Mel filter bank.

$$X(i,k) = FFT[s_i(n)] \tag{11}$$

The above formula k is the kth spectral line in the frequency domain, and the spectral energy is calculated for the spectrum $X(i,k)$ to obtain the energy spectrum:

$$E(i,k) = |X(i,k)|^2 \tag{12}$$

3) Calculate the Mel spectral energy of the signal

The calculated signal energy spectrum $E(i,k)$ of each frame is passed through M Mel filter banks to calculate Mel spectrum energy:

$$Mel(i,m) = \sum_{k=0}^{N-1} E(i,k)H_m(k) \quad 0 \le m < M \tag{13}$$

4) Mel cepstrum coefficient is calculated by DCT inverse transformation

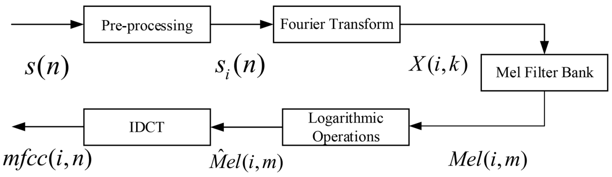


FIGURE 5
Mel-frequency cepstrum coefficient (MFCC) feature coefficient extraction principle block diagram.

Calculate the logarithmic energy of Mel spectrum energy:

$$\hat{M}el(i,m) = \ln[Mel(i,m)] \tag{14}$$

Calculate the inverse discrete cosine transform (DCT) of $\hat{M}el(i,m)$ to obtain the n-order MFCC of each frame signal:

$$mfcc(i,n) = \sqrt{\frac{2}{M}\sum_{m=0}^{M-1}}\hat{M}el(i,m)\cos\left[\frac{\pi n(2m-1)}{2M}\right] \quad n = 1,2,\ldots,p \tag{15}$$

The above formula M is the number of Mel filters, and m refers to the mth Mel filter. n is the order of MFCC.

In contrast, the standard MFCC only represents the static characteristics of the signal, while the dynamic characteristics can be characterized by the differential MFCC, which can further improve the classification and recognition rate. The calculation process of first- and second-order difference MFCCs is as follows:

$$D(i) = \frac{1}{\sqrt{\sum_{n=-k}^{k} n^2}} \sum_{n=-k}^{k} n \cdot mfcc(i+n) \tag{16}$$

$$D^{'}(i) = \frac{1}{\sqrt{\sum_{n=-k}^{k} n^2}} \sum_{n=-k}^{k} n \cdot D(i+n) \tag{17}$$

$mfcc\,(i)$ is the MFCC standard coefficient of the ith frame signal. $D\,(i)$ is the first-order differential MFCC of the ith frame signal. $D^{'}\,(i)$ is the second-order differential MFCC of the ith frame signal, and k is the number of differential signal frames, usually taken as 2. Two different behavioral acoustic signals with a time length of 5 s are intercepted, respectively, and the 12-dimensional standard MFCC characteristic coefficients are shown in Figure 6. Moreover, the first- and second-order differential MFCCs are extracted, as shown in Figures 7, 8. The signal frame length is set to 20 ms. The frame shift is 10% of the frame length, and the size of the three types of characteristic coefficients is 2,487 frames in 12 dimensions.

Since the channel impulse response in the low cepstrum period can more effectively characterize the signal itself, the main information of the acoustic signal is concentrated in the low dimension. By observing the MFCC characteristic amplitude and changing trend of each dimension of the acoustic signal in Figures 6–8, it can be found that the MFCC just emphasized the low order part.

There are differences in the low-order characteristic dimension. At the same time, for different behavioral acoustic signals, the first- and second-order differential MFCCs can reflect more differences between coefficients in different orders than the standard MFCCs, as shown in Figure 9.

Characteristic non-parametric ANOVA

All kinds of features of different behavioral acoustic signals obey the non-homogeneous distribution of normal variance. If the variance analysis is used, the analysis results will be wrong. Therefore, rank-based non-parametric ANOVA can be used for the data of such distribution. Non-parametric ANOVA does not require the data to obey the normal variance homogeneous distribution, which has wider applicability than variance analysis (Pardo-Fernandez et al., 2015).

In this paper, 1,000 acoustic signal samples of different behaviors were selected, and the sample duration was 100 ms. LPCC and MFCC features were extracted from feeding behavior and stress behavior signals. When the environment changes suddenly, the shrimps will be in a state of stress, such as hypoxia, salinity with different gradients, and water temperature. According to Kruskal–Wallis significance difference analysis, for a given significance level = 0.05, the original assumption is that there is no significant difference between the same type of characteristics of different behavioral acoustic signals. Under this hypothesis, if $p < 0.05$, the original hypothesis is rejected, and there are significant differences between the same characteristics of different behavioral acoustic signals. If $p > 0.05$, the original hypothesis is accepted, and there is no significant difference between the same characteristics of different behavioral acoustic signals.

According to the LPCC features, the 12th order LPCC feature coefficients are extracted and analyzed. Through the analysis, it can be found that there were significant differences in the LPCC characteristic coefficient values of each order of sound signals of feeding behavior and stress behavior. However, the L_2 coefficient value is $p>0.05$, and there is no significant

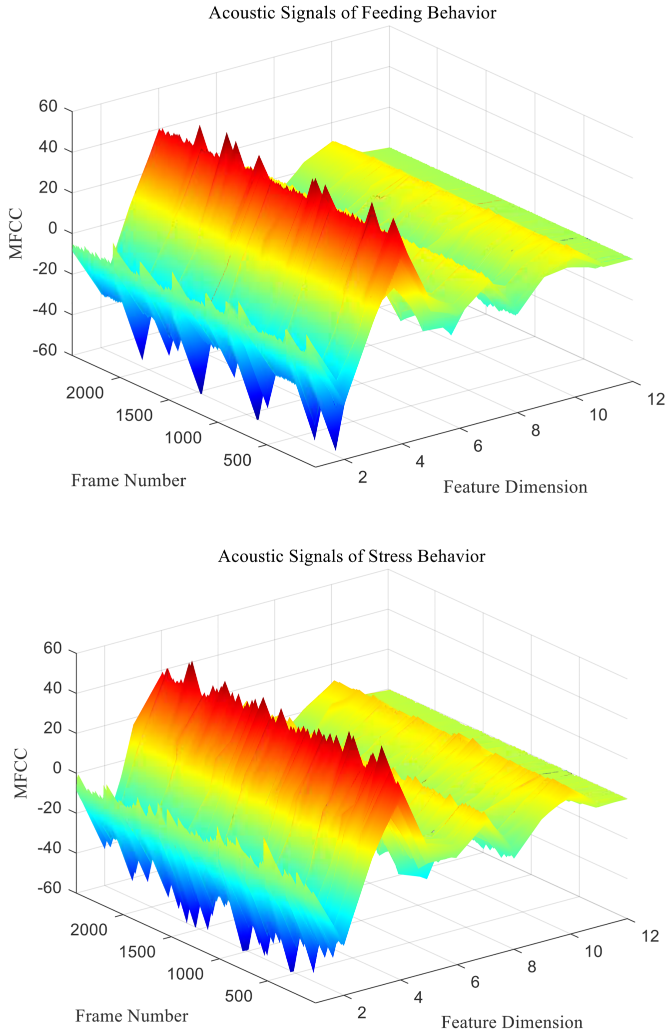


FIGURE 6
Standard MFCC characteristic coefficients of acoustic signals from different behaviors of *Penaeus vannamei*.

difference. Similarly, it can be discarded in the subsequent classification and recognition, as shown in Table 1.

For MFCC features, the total 35 orders of MFCC feature coefficients M_2 to M_{36} are extracted for significant difference analysis. Among them, $M_2$to M_{12} are the standard MFCCs, M_{13} to M_{24} are the first-order differential MFCCs, and M_{25} to M_{36} are the second-order differential MFCCs. Through analysis, it can be found that there are significant differences in the MFCC characteristic coefficient values of feeding and stress behavior signals, but the coefficient values of M_{13}, M_{14}, M_{17}, M_{18}, M_{20} to M_{23}, and M_{28} to M_{30} are $p>0.05$, so there was no significant difference in these coefficients, and it was difficult to distinguish different behavioral acoustic signals of *P. vannamei*. Therefore, in the subsequent classification and recognition, it can be discarded to achieve the effect of feature dimensional reduction, as shown in Table 2.

Therefore, based on non-parametric ANOVA of their significant differences, some features with redundant information are eliminated, and the dimension of the features is further reduced. It verifies the effectiveness of acoustic signal features of different behaviors and provides a certain implementation basis for subsequent classification and recognition research.

Experimental analysis

Based on the two kinds of characteristic coefficients of different behavioral acoustic signals of *P. vannamei*, SVM with different classification strategies is selected to classify and identify different behavioral states within the species. While exploring the classification effect of intraspecific behavior state of *P. vannamei*, this paper adds the vocal signals of other species of marine

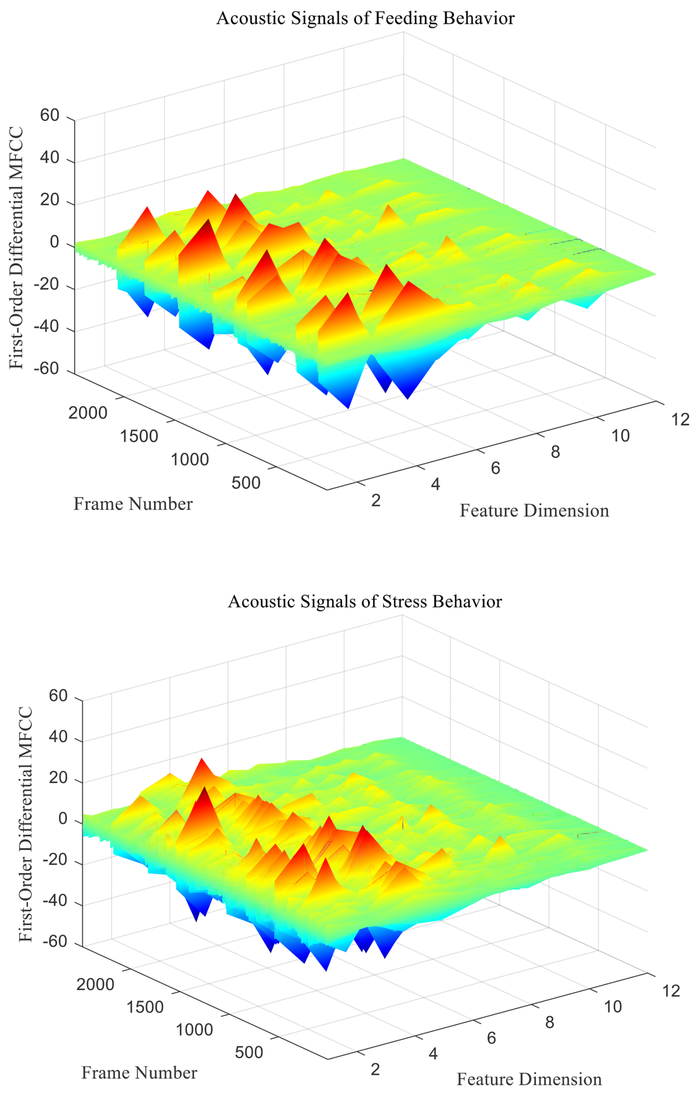


FIGURE 7
First-Order differential MFCC characteristic coefficients of acoustic signals from different behaviors of *Penaeus vannamei*.

organisms to study the differences in behavioral acoustic signals between *P. vannamei* and other biological species, further verify the effectiveness of feature extraction and classification model, and improve the generalization ability of classification system. This section will compare and analyze the acoustic signals of different behaviors within *P. vannamei* species and the classification and recognition results of interspecific signals.

Data set processing

The data set of the classification and recognition experiment in this paper consists of five categories of underwater biological signals. The shrimp sound signal experiment uses a WBT22-1107 hydrophone. The frequency range is 1–22 kHz. The receiving sensitivity response of the hydrophone is −193 ± 3dB @ 22 kHz (re 1 $V/\mu Pa$ @1 m, 20 m cable), and the experimental sampling rate is 100 kHz. After feeding, the shrimps mainly eat at the bottom of tank. Therefore, during the experiment, the hydrophone was placed at the center of the water tank down close to the shrimps, as shown in Figure 10 (Wei et al., 2020). In addition to the sound signals of feeding and stress behavior of *P. vannamei* collected according to the experimental methods introduced in *Section II*, the other three types of biological sound signals are humpback song signal (Mark), sperm whales "clicks" call signal, and bottlenose dolphin "clicks" call signal (Heimlich et al.).

First, the five types of acoustic signals are pre-processed in frames. A total of 1,500 signals are selected for each type of signal,

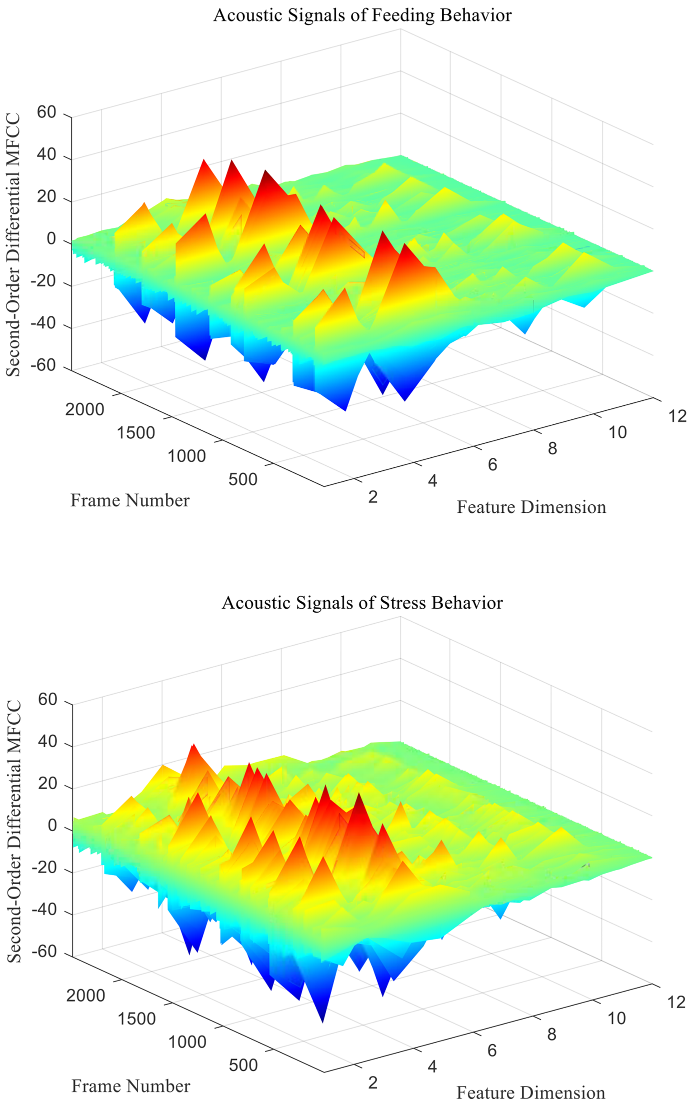


FIGURE 8
Second-Order differential MFCC characteristic coefficients of acoustic signals from different behaviors of *Penaeus vannamei*.

including 1,000 acoustic signals as the training set, 250 signals as the test set, and 250 signals as the verification set. The length of all signals is 0.1 s. The frame length is 0.02 s. The frame shift is 10%. According to the input form of the classifier and the previous feature significance analysis, the signal feature coefficient extraction of different classifiers is introduced:

1) MFCC characteristics

Calculating the 36-dimensional MFCC of each frame signal for each sample after framing includes the 12-dimensional standard and first- and second-order difference MFCCs. At the same time, after summing and averaging the coefficients corresponding to each order of each frame, the 12-dimensional coefficients of dimensions 1, 13, 14, 17, 18, 20–23, and 29–30 with small significant difference are discarded. The dimension of each sample signal characteristic matrix is 1 × 24.

2) LPCC characteristics

Calculating the 12-dimensional LPCC of each frame signal for each sample after framing, then summing and averaging the corresponding order coefficients of each frame, and discarding the first and second dimensional coefficients with small significant difference, then the dimension of each sample signal characteristic matrix is 1 × 10.

SVM classifier design

In the design and selection of SVM classifiers, based on the two classification strategies of OvR and OvO, this paper designs six SVM classifiers by using ECOC coding principle by selecting

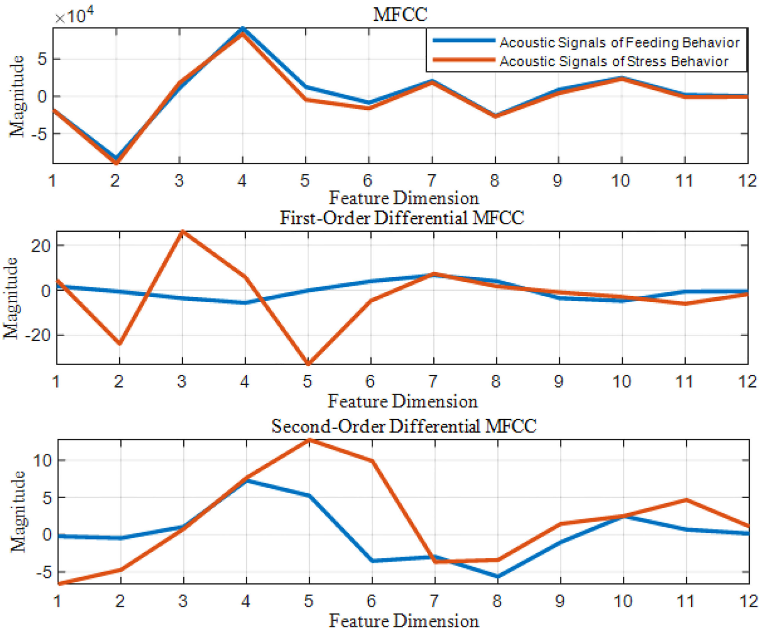


FIGURE 9
All dimensional MFCC and difference coefficient diagram of cumulative sum of all frames.

TABLE 1 Kruskal–Wallis non-parametric ANOVA for LPCC characteristics of acoustic signals with different behaviors.

LPCC	Mean of feeding state	Standard deviation of feeding state	Mean of stress state	Standard deviation of stress state	Kruskal–Wallis analysis	Significant difference
L_2	1.664	0.062	1.670	0.087	p=0.190	No

TABLE 2 Kruskal–Wallis Non-parametric ANOVA of MFCC order characteristics of acoustic signals with different behaviors.

MFCC	Mean of feeding state	Standard deviation of feeding state	Mean of stress state	Standard deviation of stress state	Kruskal–Wallis analysis	Significant difference
M_{13}	0.017	0.101	0.022	0.243	p=0.782	No
M_{14}	0.001	0.188	7e−05	0.436	p=0.091	No
M_{17}	−0.019	0.192	−0.035	0.345	p=0.272	No
M_{18}	−0.028	0.154	−0.013	0.349	p=0.095	No
M_{20}	−0.013	0.135	0.007	0.299	p=0.063	No
M_{21}	0.001	0.086	−0.001	0.158	p=0.411	No
M_{22}	0.004	0.081	−0.004	0.210	p=0.938	No
M_{23}	−0.004	0.055	−0.003	0.113	p=0.563	No
M_{28}	0.022	0.178	0.012	0.329	p=0.331	No
M_{29}	0.036	0.129	0.025	0.182	p=0.065	No
M_{30}	0.014	0.082	0.018	0.162	p=0.407	No

three kernel functions: linear, Gaussian, and polynomial kernel functions, as shown in Table 3.

Among them, the penalty factor of all SVM classifiers is C = 1. For the classification and recognition of five kinds of acoustic signals within and between *P. vannamei* species in this paper, different features based on SMO algorithm are used to quickly optimize the training SVM classifier. Finally, the test set is used to test the training classifier and analyze the results, as shown in Table 4.

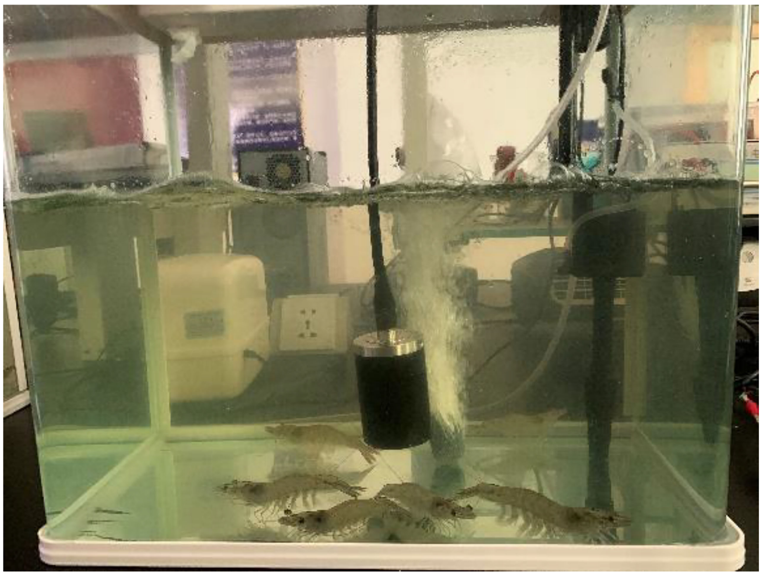


FIGURE 10
The experimental water tank.

In conclusion, LPCC and MFCC are effective in the classification of different behavioral acoustic signals and interspecific acoustic signals of *P. vannamei*. Meantime, classifiers based on different types and structures have different classification effects. Through experimental comparison and analysis, SVM classifier based on OvR classification strategy according to MFCC features has more stable and excellent classification results. It can model the acoustic signal characteristics of different underwater biological behaviors more efficiently and has the classification accuracy as high as 93%.

TABLE 3 SVM classifier based on different class partition strategies and kernel function selection.

Classification strategy	Kernel function	Number of subclassifiers	Number
OvR	Linear	5	SVM_1
	Gaussian	5	SVM_2
	Polynomial	5	SVM_3
OvO	Linear	10	SVM_4
	Gaussian	10	SVM_5
	Polynomial	10	SVM_6

Conclusion

Based on passive acoustic technology, this paper collects different behavioral acoustic signals produced by feeding and stress state of *P. vannamei* and establishes a behavioral acoustic signal database of *P. vannamei*. At the same time, the characteristics of acoustic signals are studied. After obtaining effective features, SVM classifiers with different structures are designed. According to different features, the classification of intra- and interspecific behavior states of *P. vannamei* is finally completed. Finally, the SVM classifier based on the OvR classification strategy based on MFCC features can roughly solve the five types of underwater biological acoustic signal classification problems proposed in this paper with the best performance and have a classification accuracy of 93%. Generally, underwater target behavior classification based on underwater acoustic signal is an important research direction of underwater acoustic signal processing, which involves many disciplines such as marine science, sonar technology, signal processing, feature engineering, pattern recognition, and computer technology. Therefore, as a comprehensive subject, the classification and recognition of different behavior states of

TABLE 4 Classification and recognition results of SVM classifiers of MFCC features.

Features	SVM_1	SVM_2	SVM_3	SVM_4	SVM_5	SVM_6
LPCC	90.4%	83.7%	20.6%	**92.7%**	89.8%	75.2%
MFCC	**93.0%**	80.9%	14.2%	80.0%	73.5%	40.3%

Bold values represents the highest value of each method.

P. vannamei based on passive acoustic technology not only overcomes the limitations of underwater visual observation research and application but also helps the intelligent development of related industries. It has important theoretical research and engineering practical application value.

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

MW performed the experiment. KC contributed to the conception of the study. YL performed the data analyses and wrote the manuscript. EC helped perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Xiamen marine and fishery development special fund project "Key technology research and industrialization demonstration of aquaculture automation equipment" (19CZB015HJ02); Fujian marine economic development subsidy fund project (ZHHY-2019-4); Fujian marine economic development special fund project (FJHJF-L-2021-12); Fujian Provincial Department of Education "Smart fishery Application Technology Collaborative Innovation Center" (XTZX-ZHYY-1914).

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

Au, W. W. L., and Banks, K. (1998). The acoustics of the snapping shrimp synalpheus parneomeris in kaneohe bay. *J. Acoustical Soc. America* 103 (1), 41–47. doi: 10.1121/1.423234

Berk, I. M. (1998). Sound production by white shrimp (penaeus setiferus), analysis of another crustacean-like sound from the gulf of Mexico, and applications for passive sonar in the shrimping industry. *J. Shellfish Res.* 17 (5), 1497–1500.

Fish, M. P., Kelsey, A. S., and Mowbray, W. H. (1952). *Studies on the production of underwater sound by north Atlantic coastal fishes* (University of Rhode Island: Narragansett Marine Laboratory).

Gillespie, D., Caillat, M., Gordon, J., and White, P.. (2013). Automatic detection and classification of odontocete whistles. *J. Acoustical Soc. America* 134 (3), 2427–2437. doi: 10.1121/1.4816555

Gupta, H., and Gupta, D. (2016). "LPC and LPCC method of feature extraction in speech recognition system," in *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence). IEEE.* Noida, India, 498–502.

Heimlich, S., Mellinger, D., and Klinck, H. *MobySound.org [EB/OL]*. Available at: http://www.mobysound.org/mobysound.html.

Henninger, H. P., and Watson, III W H. (2005). Mechanisms underlying the production of carapace vibrations and associated waterborne sounds in the American lobster, homarus americanus. *J. Exp. Biol.* 208 (17), 3421–3429. doi: 10.1242/jeb.01771

Ibrahim, A. K., Zhuang, H., Erdol, N., and Ali, AM. (2016). "A new approach for north atlantic right whale upcall detection," in *2016 International Symposium on Computer, Consumer and Control (IS3C). IEEE.* Xi'an, China 260–263.

Iversen, R. T. S., Perkins, P. J., and Dionne, R. D. (1963). An indication of underwater sound production by squid. *Nature* 199 (4890), 250–251. doi: 10.1038/199250a0

Mann, D., Locascio, J., and Wall, C. (2016). Listening in the ocean: new discoveries and insights on marine life from autonomous passive acoustic recorders. *Listening Ocean* 2016, 309–324. doi: 10.1007/978-1-4939-3176-7_12

Mark, A. M. *Whale acoustics [EB/OL]*. Available at: http://www.whaleacoustics.com.

Mellinger, D. K., and Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoustical Soc. America* 107 (6), 3518–3529. doi: 10.1121/1.429434

Noda, J. J., Travieso, C. M., and Sánchez-Rodríguez, D. (2016). Automatic taxonomic classification of fish based on their acoustic signals. *Appl. Sci.* 6 (12), 443. doi: 10.3390/app6120443

Nordeide, J. T., and Kjellsby, E. (1999). Sound from spawning cod at their spawning grounds. *ICES J. Mar. Sci.* 56 (3), 326–332. doi: 10.1006/jmsc.1999.0473

Pace, F., Benard, F., Glotin, H., Adam, O., and White, P. (2010). Subunit definition and analysis for humpback whale call classification. *Appl. Acoustics* 71 (11), 1107–1112. doi: 10.1016/j.apacoust.2010.05.016

Pardo-Fernandez, J. C., Jimenez-Gamero, M. D., and Ghouch, A. E. (2015). A non-parametric ANOVA-type test for regression curves based on characteristic functions. *Scandinavian J. Stat* 42 (1), 197–213. doi: 10.1111/sjos.12102

Parks, S. E., Searby, A., Celerier, A., Johnson, M. P., Nowacek, D. P., Tyack, P. L., et al. (2011). Sound production behavior of individual north Atlantic right whales: implications for passive acoustic monitoring. *Endangered Species Res.* 15 (1), 63–76. doi: 10.3354/esr00368

Rountree, R. A., Gilmore, R. G., Goudey, C. A., Hawkins, AD., Luczkovich, JJ., Mann, DA., et al. (2006). Listening to fish: applications of passive acoustics to fisheries science. *Fisheries* 31 (9), 433–446. doi: 10.1577/1548-8446(2006)31[433:LTF]2.0.CO;2

Schärer, M. T., Nemeth, M. I., Rowell, T. J., and Appeldoorn, RS. (2014). Sounds associated with the reproductive behavior of the black grouper (Mycteroperca bonaci)[J]. *Mar. Biol.* 161 (1), 141–147. doi: 10.1007/s00227-013-2324-3

Silva, J. F., Hamilton, S., Rocha, J. V., Borie, A., Travassos, P., Soares, R, et al. (2019). Acoustic characterization of feeding activity of litopenaeus vannamei in captivity. *Aquaculture* 501, 76–81. doi: 10.1016/j.aquaculture.2018.11.013

Smith, D. V., and Tabrett, S. (2013). The use of passive acoustics to measure feed consumption by penaeus monodon (giant tiger prawn) in cultured systems. *Aquac. Eng.* 57, 38–47. doi: 10.1016/j.aquaeng.2013.06.003

Soldevilla, M. S., Wiggins, S. M., and Hildebrand, J. A. (2010). Spatial and temporal patterns of risso's dolphin echolocation in the southern California bight. *J. Acoustical Soc. America* 127 (1), 124–132. doi: 10.1121/1.3257586

Thomsen, F., Erbe, C., Hawkins, A., Lepper, P., Popper, AN., Scholik-Schlomer, A., et al. (2020). Introduction to the special issue on the effects of sound on aquatic life. *J. Acoustical Soc. America* 148 (2), 934–938. doi: 10.1121/10.0001725

Tong, Y., Zhang, X., and Ge, Y. (2020). "Classification and recognition of underwater target based on MFCC feature extraction," in *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE* Macau, China. 1–4.

Vieira, M., Fonseca, P. J., Amorim, M. C.P., and Teixeira, CJC. (2015). Call recognition and individual identification of fish vocalizations based on automatic speech recognition: An example with the lusitanian toadfish. *J. Acoustical Soc. America* 138 (6), 3941–3950. doi: 10.1121/1.4936858

Wei, M., Lin, Y., Chen, K., Su, W., and Cheng, E. (2020). Study on feeding activity of litopenaeus vannamei based on passive acoustic detection. *IEEE Access* 2020 (99), 1–1. doi: 10.1109/ACCESS.2020.3019529

frontiers | Frontiers in Marine Science

TYPE Technology and Code
PUBLISHED 08 September 2022
DOI 10.3389/fmars.2022.986752

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Adwitiya Sinha,
Jaypee Institute of Information Technology, India
Xiao Feng,
Xiamen University, China

*CORRESPONDENCE
Atle Totland
atle.totland@hi.no

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 05 July 2022
ACCEPTED 18 August 2022
PUBLISHED 08 September 2022

CITATION
Totland A and Johnsen E (2022) Kayak Drone – a silent acoustic unmanned surface vehicle for marine research. *Front. Mar. Sci.* 9:986752. doi: 10.3389/fmars.2022.986752

Kayak Drone – a silent acoustic unmanned surface vehicle for marine research

Atle Totland* and Espen Johnsen

Research Group Ecosystem Acoustics, Institute of Marine Research, Bergen, Norway

Advancements in technologies have led to a rapid development of unmanned surface vehicles (USV) for marine ecosystem monitoring. The design, size, and scientific payload of the USVs differ as they are built for different purposes. Here, we present the design criteria and detailed technical solutions of a prototype USV which has been built to fulfill the following experimental and operational needs; the USV should be used for inshore and shallow water acoustic monitoring, offshore comparison of echo sounder recordings from the USV and research vessels, monitor natural fish schooling behavior and seabird-fish behavioral interactions. The prototype has been built over a period of 5 years with steadily quality improvements. As the hull is based on an expedition double kayak, the USV is named Kayak Drone, and we aimed at building the Kayak Drone using of-the-shelf hardware and existing open-source software. This allowed for the development of a modular and well-functioning USV at a relatively low cost. The Kayak Drone produces very little noise and *in situ* experiments show that the Kayak Drone can record echo sounder data of fish near the surface without disturbing their natural distribution and behavior. One *in situ* study shows that the Kayak Drone could navigate within a couple of meters from swimming puffin and other seabirds without triggering escape. These results demonstrate that the Kayak Drone can be utilized to produce unbiased survey estimates for fish distributed in shallow waters and near the surface, which is very important for many fish stock assessments and managements. Furthermore, it can also be used as a tool to observe the predation by seabirds on fish schools without interfering with their natural interspecific behavior, which traditionally has been very difficult. The use of the Kayak Drone is not restricted to these tasks, and we foresee that the Kayak Drone can be utilized in many different experiments where a silent platform is needed.

KEYWORDS

echo sounder, fish behavior, seabird - fish interactions, silent and nonintrusive vehicle, unmanned surface vehicle (USV), vessel avoidance, vertical fish distribution

1 Introduction

Worldwide, research vessel surveys have been crucial for monitoring marine ecosystems and for collection of information to assess the state of commercially important fish stocks (Gunderson, 1993). This traditional survey concept (Gunderson, 1993; Simmonds and MacLennan, 2008) has in recent years undergone a transition to also include additional platforms (Verfuss et al., 2019). Advancements in software, computers, sensors, and battery technologies have led to the development of a variety of unmanned surface vehicles (USV) for marine ecosystem monitoring. The USVs are partly replacing traditional tasks carried out by research vessels (De Robertis et al., 2019), but are more frequently used for other purposes (Shao et al., 2019; Chang et al., 2021) or as a supplement to traditional research vessels (Kum et al., 2020). With climate changes and other human induced stressors on the marine ecosystems, there is an increased need for more continuous and widespread monitoring, and it is highly likely that USVs will play a more important role in future data collection. The transition is ongoing, and several studies have tested USVs equipped with different sensors, and in a few cases the USVs are an integrated part of annual survey programs (De Robertis et al., 2019).

When USVs are used for abundance estimation and mapping of zooplankton and fish densities, they are usually equipped with scientific echo sounders. These echo sounders can be of the same type as installed on a research vessel, or of a more power efficient type (Benoit-Bird, 2018). Traditional use of echo sounders for fisheries research have been dependent on net-sampling for identification of the targets observed on the echograms. Such net-sampling requires relatively large, manned vessels, but more advanced multifrequency and broadband echo sounders together with deep learning acoustic classification may make trawling less important for identification of acoustic targets in the future (Johnsen et al., 2009; Brautaset et al., 2020; Handegard et al., 2021). Smaller unmanned vehicles can be used when regular biological sampling is of less importance or if samples can be obtained from fishing vessels or landing sites in the area. Nevertheless, the acoustic data format from the USVs is the same as recorded from a ship and makes comparison with data from research vessels feasible. This is important for the continuation of valuable survey time series that are repeated annually following standard protocols.

A long list of USVs for marine ecosystem monitoring have been developed in recent years and the range of purposes and designs are extensive (Othman, 2015). Some are designed for shallow waters such as the CatOne family (Romano and Duranti, 2012), whereas others are designed preliminary for ocean operations, e.g., Saildrone (De Robertis et al., 2019), Sailbouy (Gahni et al., 2014) and Kongsberg Sounder (Handegard et al., 2021b). The size ranges from the handheld USVs to the 22 m long Saildrone Surveyor USV (An et al., 2021). The prolusion may be sail, wave motion, underwater or air fan propeller, or a combination of these.

The scientific purposes affect the design of the USV, and our aim here is to present a new multipurpose prototype, the Kayak Drone USV, which was built to address a list of defined scientific challenges (Table 1). A detailed technical description of the Kayak Drone system is presented together with examples of research experiments carried out using the Kayak Drone. The paper is structured in scientific purposes, vehicle design, technical solutions, and examples of experimental field studies. We discuss advantages and limitation of the Kayak Drone and the future contribution of the Kayak Drone to marine research.

2 USV design and technical solutions

2.1 Scientific purposes and design criteria

The design of the Kayak Drone was defined in accordance with a list of research questions and purposes as presented in Table 1. The addressed research challenges required a silent acoustic platform to observe undisturbed densities and natural behavior of fish, marine organisms, and seabirds, without affecting the targets of interest. The main payload is broadband scientific echo sounders, and the acoustic signal to noise ratio was required to be low to enable recording of high-quality data from small organisms in the deep.

The Kayak Drone is primarily made for small scale experimental studies and not as a platform for large scale extensive surveys. To fulfill the different experimental and operational needs, two operation modes for the USV were included; autonomously following waypoints defined in a chart, and real-time remote-control operation. The hull should be small enough to enable high maneuverability close to the shore and easy transport on a boat trailer to various experimental sites. It was a criterium that the Kayak Drone could be launched both from the shore and from a research vessel during moderate sea states.

To study potential behavioral effects of a research vessel on the marine organisms, the USV had to have a maximum speed similar to the standard survey speed of a research vessel (10 knots), thus enabling it to run in front or behind the research vessel during standard operation. For other experiments, a speed of about 5 knots is acceptable and we aimed at having enough battery capacity for an operation time of around 10 hours at a speed of 5 knots. It was furthermore important to make a flexible and adaptive system design that allowed for easy alteration of the payload and implementation of additional or alternative sensors and their associated software.

TABLE 1 Overview of the scientific purposes and design criteria of the Kayak Drone.

Purpose	Design criteria
Inshore and shallow water acoustic monitoring	Silent propulsion to avoid disturbing the target organisms A slim hull to minimize the pressure field and thus prevent disturbing the target organisms Retractable keel set in a shallow position for near surface observation Good acoustic signal to noise level Modular design to allow easy alteration of sensors Launching from shore by boat trailer Semi-long endurance (8-10 hours) Accurate waypoint navigation for operation very close to the shore A reliable communication link for rapid operator intervention in trafficked waters
Offshore comparison of echo sounder recordings from the USV and a research vessel	Silent propulsion to avoid disturbing the target organisms A slim hull to minimize the pressure field and thus prevent disturbing the target organisms Retractable keel set in a deeper position to avoid the surface bobble layer and to stabilize the USV despite larger waves Good acoustic signal to noise level Modular design to allow easy alteration of sensors Speed similar to standard cruise speed of a research vessel (8-10 knots Launching from research vessel Sea-capability for offshore operation A reliable communication link over several nautical miles
Monitor natural fish schooling behavior	Silent propulsion to avoid disturbing the target organisms A slim hull to minimize the pressure field and thus prevent disturbing the target organisms Small hull (little noise and pressure wave, and small optical surface contrast) A reliable communication link
Seabird-fish behavioral interactions	Silent propulsion to avoid disturbing the target organisms A slim hull to minimize the pressure field and thus prevent disturbing the target organisms Modular design to allow easy alteration of sensors A reliable communication link Video capabilities to observe birds for simultaneous comparison with echo sounder recordings of prey fish

2.2 Technical solutions

During the building of the Kayak Drone, it was important to utilize available open-source software and off-the-shelf products as far as possible to reduce the costs, development time, and to increase the technical reliability. Reaching scientific goals was the main aim (Table 1), rather than the development of complex and novel technology.

2.2.1 The hull and keel

We purchased an existing and well-tested kayak hull after a review process using experts in kayaking. To satisfy the design criteria on speed, seaworthiness, and payload, we decided to modify a 6.7 m long, well-tested Triton expedition double kayak from Sea Kayaking UK as the basis for the USV (Figure 1). The total weight of the kayak, motor, batteries, and instrumentation is approximately 250 kg. The standard cockpits were converted to watertight chambers with lids. A well was made close to the stern where the propulsion motor is mounted, and a retractable keel was installed midships for acoustic transducers and to improve the sideway stability. The center board of the keel runs through a sheath in the hull and can be bolted in various draft positions (Figure 1). At the lower end of the keel there is a bulb, which houses the echo sounder transducers and lead weights (Figure 1). During survey operations, the adjustable keel is lowered to a bulb depth of 0.5 to 1.5 m below the sea surface. Air bubbles in the upper surface waters can attenuate or even block the acoustic signals, and lowering the transducers reduces this problem (Dalen and Løvik, 1981). During transport, the center board of the keel is fully retracted with only the bulb extending below the vehicle hull. A modified car trailer is used for transport, as a cradle onboard the research vessel and as a slipway during deployment from shore.

2.2.2 Propulsion, rudder, and battery system

A Torqeedo Cruise 2.0 RS electric outboard motor (Figure 1) is used for propulsion, which provide a maximum speed of almost 9 knots (Table 2). This is in line with the defined speed criteria of the USV. The engine is installed with a large propeller (30 cm diameter) mounted directly on the motor shaft, giving a high energy efficiency. The large propeller and high torque provide a low RPM (revolutions per minute) and a low noise level. In combination with the slim shape of the kayak hull, which produce a very low hydrodynamic pressure field, the critical design criterion of a silent vehicle within the audible and sensory frequency range of the target organisms was achieved. One marinized Torqeedo LiIon battery (Power 26-104) is the power source for the engine. The USV is designed to house two

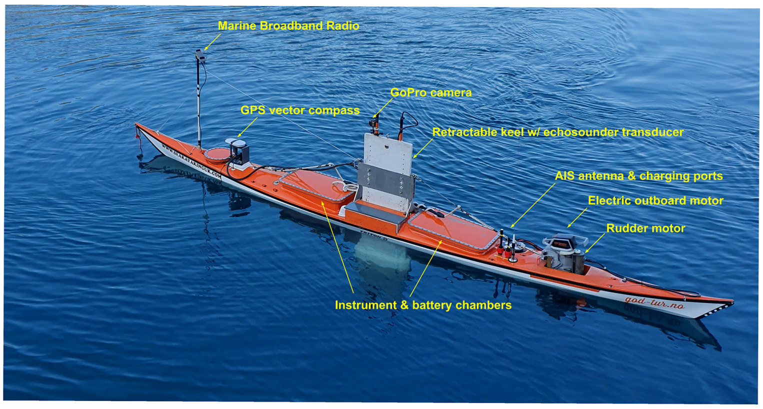


FIGURE 1
The Kayak Drone with the key external components.

such batteries, potentially doubling the current operation time of about 6 hours with an average speed of 5 knots (Table 2). In agreement with the Torqeedo company, we got access to the proprietary data protocols of the internal Torqeedo network, which allowed the USV software to alter the engine RPM and control the engine battery.

A Faulhaber brushless DC servo motor with position feedback (hall sensors) was marinized and attached to the propulsion engine to adjust rudder angle (Figure 1). Existing screw holes on the propulsion motor were used for attachment of the rudder motor and a cog. The corresponding cog on the rudder motor shaft transferred rudder positioning. Accurate control of rudder angle was achieved due to a Faulhaber 1:531 planetary gear attached to the servo motor. The rudder motor is powered by the Torqeedo battery. The charging time for fully discharged batteries is approximately 5 hours (Table 2).

2.2.3 Sensors

The main scientific payload of the Kayak Drone is currently a SIMRAD EK80 broad band echo sounder (Demer et al., 2017) with a SIMRAD ES200-7C 200 kHz split-beam transducer mounted on the retractable keel (Figure 1). The full echo sounder system consists of the transducer, a SIMRAD Wide Band Transceiver (WBT) and the EK80 SIMRAD software installed on a ruggedized industrial Windows computer placed in the cockpit chamber in front of the keel. There is free space in the bulb of the keel for an additional transducer. A scientific echo sounder can measure small organisms at great distances despite the week return echoes by amplifying the incoming signal. This makes the echo sounder vulnerable to noise in the frequency range of operation, and electric noise from other equipment may be detected as input signals. High-frequency signal noise is a well-known problem which can corrupt the acoustic recordings of an echo sounder (Mitson and Knudsen, 2003). Therefore, the echo sounder, computer and other sensors are powered by their own set of lead gel-batteries, separate from the power source of the propulsion and rudder motors. To further reduce the risk of electric noise from the motors, communication between the computer and motors goes through galvanically shielded adapters. Furthermore, the echo sounder and transducer cabling is situated in a compartment away from other electronics. The chassis of the echo sounder is grounded to sea and the transducer cable is shielded.

A small vector compass containing two GPS receivers is mounted on a bracket above deck (Seapilot Vector Compact GPS Compass) (Figure 1). In addition to standard GPS

TABLE 2 Summary of measured performance parameters of the Kayak Drone.

Performance parameter	Value	Comment
Maximum speed	9 knots	Tested on flat sea and no wind
Operation duration	6 hours	At a speed of 5 knots
Charging time	5 hours	From fully discharged
Range of radio link	> 1400 m	Not tested beyond this range. No interruptions of remote VNC control and data transfer.
Maximum wave height	2 meters	Limitation for safe retrieval back onboard research vessel and good quality of acoustic recordings.

Details on the performance of individual off-the-shelf hardware components can be found on the manufacturers' web sites.

navigation telegrams, the unit provides accurate heading information unaffected by magnetic fields, and information on roll, pitch and heave is also provided.

To increase navigation safety, a class B AIS transceiver is installed, which makes the Kayak Drone visible in other vessels' charts and in online ship maps (e.g. Marine Traffic). In the Kayak Drone chart program (OpenCPN) ships with AIS is visualized.

The compartments of the USV have leakage sensors, and alarms are visible in the control application. Independent GoPro video cameras are normally mounted on deck and on the keel bulb during experiments.

2.2.4 Wireless remote control

The Kayak Drone can operate autonomously based on a preplanned survey route without real-time connection. However, during most experiments the distance between the USV and a mother-vessel is within visual distance. As the vehicle has no anti-collision system installed, the operation requires continuous access control of the USV in coastal and heavily trafficked waters. This is less important during operations in high seas. Furthermore, several of the studies require a real time observation of the echogram. A VNC connection (open-source program TightVNC version 4.8.8) gives remote access to the Kayak Drone computer *via* a radio link. Traditionally, when two communicating units are moving independently from each other, omnidirectional antennas are used. This setup has challenges such as short range and unstable communication. Initially, we had a system with an omnidirectional antenna on the USV, and a four-sector directional and switching antenna on the mother vessel. Due to frequent communication breakdowns and a too short range, the original wireless link system was replaced by Kongsberg Marine Broadband Radios (MBR) (Ludvigsen et al., 2016). MBR radios are automatically steerable, phased array radios, that directs the antenna beam towards the node it is communicating with, both when transmitting and receiving. The phase steering or beam forming is done electronically without any moving parts. This technology has several advantages compared to an omnidirectional solution. Due to the narrow lobe, the system is robust towards unwanted external interference and noise. The transmit and receive gains are both high, thus giving a very long range. The electronic phase-controlled communication allows for smooth communication between nodes. Several nodes can be appended to an MBR network, but in the Kayak Drone system, we currently have only two nodes, the USV itself, and the mother vessel. Depending on the required communication range, three sizes of radios with integrated antennas can be chosen. For the USV and the mother vessel, we have so far used the smallest (Figure 1) and the intermediate radios, respectibly. The radio onboard the drone has an option to directly connect a camera and include live video streaming onto the MBR network. On the mother vessel or from a land station, the virtual computer screen on board the Kayak Drone can be seen, and keyboard and mouse signals are transmitted using a VNC solution (TightVNC). This makes the Kayak Drone very flexible and new sensors and software may be installed in the Kayak Drone without altering the overall system. The maximum range of the radio link has not yet been tested. The USV was 1400 m away from the mother vessel during one of the experiments, and the connection was still flawless.

2.2.5 Navigation and kayak operation

2.2.5.1 OpenCPN chart software and EchoDrone control software

Two internal software programs on the USV computer work in a joint effort to operate the drone (Figures 2, 3). These are the open-source chart plotter navigation program, OpenCPN (https://opencpn.org), and the self-made EchoDrone program developed in Labview version 2013.

The two programs interact and solve various tasks such as navigation, autonomous operation, sensor data reading and storage, handling error messages and navigational warnings, and control of the propulsion and rudder. The EchoDrone reads data from the GPS vector compass (latitude, longitude, UTC time, true heading, tilt, roll and heave, speed over ground, coarse over ground) and AIS before transmitting the data onto the USV intranet. OpenCPN reads from this Ethernet and displays information in the chart (see Figures 2, 4), calculates range and bearing to the next waypoint, cross track error and estimated time of arrival. EchoDrone reads these calculated navigation data sent from OpenCPN for the active waypoint. Based on this information, the rudder motor is commanded to set the rudder to an appropriate angle for the Kayak Drone to steer towards the next waypoint (Figure 3). In default operation mode, the user sets up a route with waypoints in the OpenCPN chart *via* the remote radio link and VNC. Once the route and the first waypoint are activated, the USV can be set to automatically follow it. Once the active waypoint is reached, OpenCPN will automatically switch and activate the next waypoint on the route and send out new relevant navigation data. The USV will automatically stop when it reaches the last waypoint on the route. It is however possible to make a closed route that forms an eternal loop. The throttle and thereby the vessel speed is set in the EchoDrone front panel by the remote operator.

The integration of OpenCPN into the drone system plays a central role in the overall design. Taking advantage of all the functionality in OpenCPN and integrate its navigational outputs, has significantly reduced the development time for the USV and consequently also the costs. It is highly unlikely that the Kayak Drone could have been developed to the current advanced operational level without OpenCPN, as advanced charts operations and visualization are vital, especially in coastal areas and narrow fjords. All this to the credit of the

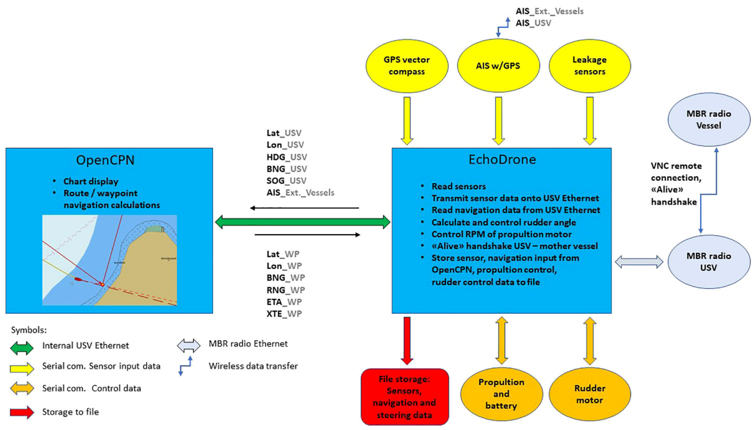


FIGURE 2
Overview of how the OpenCPN and EchoDrone programs share the task of controlling the Kayak Drone. The data flow between the two programs is shown as well as external sensors inputs and data outputs. Lat (latitude), Lon (longitude), HDG (heading), BNG (bearing), SOG (speed over ground), RNG (range), ETA (estimated time of arrival), XTE (cross track error), MBR (Maritime Broadband Radio), GPS (Global Positioning System) and AIS (Automatic Identification System), _USV (real time navigation data on the USV), _WP (navigation data on the active, next waypoint on the route), _Ext._Vessels (any vessel with AIS and withing range).

many developers of this open-source application. OpenCPN chart program can read and display various sea chart formats. The chart forms the basis for all OpenCPN operations. The project has access to high accuracy s-57 format charts (Figure 4) from the Norwegian Mapping Authorities. These high-resolution maps are a prerequisite for Kayak Drone experiments in coastal areas and narrow fjord systems as described in the design criteria for the USV. The Kayak Drone and its track is shown in the OpenCPN map together with the route and waypoints (Figure 4). At any given time, the operator onboard the mother vessel may modify or make a new route.

In manual operation mode, the user can control the rudder angle through buttons in the EchoDrone front panel. The battery status is also read and displayed in the EchoDrone program. The EchoDrone program continuously write time tagged sensor data and operational parameters to a log file. The program can be set in replay mode and will read data from a previous and actual survey at sea. This option has been useful for system development in the lab, as well as debugging and improvement of the navigation algorithms in hindsight, in the office. The behavior of the EchoDrone program is governed by close to 100 parameters. These are stored in an ASCII parameter file and can be edited by the user. The parameter file is read by EchoDrone during initiation.

2.2.5.2 Vessel software

A remote Labview program, "Vessel", may optionally be run on the mother vessel or at the command location on shore. In this program, a handshake option may be activated. Both the Vessel software operated by the controller and the EchoDrone software in the USV send and receive UDP ethernet telegrams *via* the MBR communication link at an interval of one second. If the remote communication link is lost, the USV will no longer receive handshake telegrams from the vessel, and propulsion is halted after a given time delay. The Vessel program can also automatically re-establish the VNC connection when the communication link is back and handshake telegram exchange resumes.

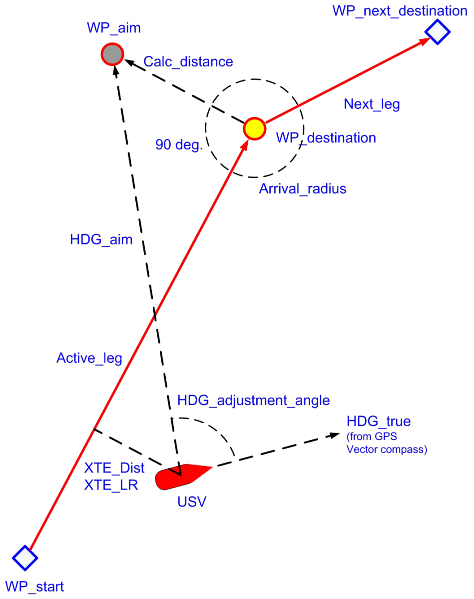
For specific experiments, the Vessel application can read GPS positions and transmit this as a moving target waypoint for the USV. This may typically be used to position the USV at a fixed distance and bearing relative to the mother ship to examine vessel avoidance reactions of fish to a large research vessel.

3 The Kayak Drone in experiments

There are several challenges with traditional vessel based acoustic surveys relating to vessel avoidance, near sea-surface acoustic blind zone, cost of operation and full coverage in shallow water and near the coastline. The Kayak Drone has been used in several scientific studies relating to these challenges, and here we briefly present preliminary results as illustrated by three different experiments carried out with the Kayak Drone.

3.1 Measuring the full vertical and horizontal density distribution of sprat in a Norwegian fjord

Sprat (Sprattus sprattus) is a small clupeid species with a wide distribution. It is an important key species in the



Step 1 (OpenCPN): Calculate bearing of Active_leg between WP_start and WP_destination, XTE_Dist (cross track error distance) and XTE_LR (cross track error left/right).

Step 2 (EchoDrone): Calculate bearing 90 degrees to Active_leg on the opposite side of XTE_LR.

Step 3 (EchoDrone): Calculate the Calc_distance from WP_destination to WP_aim, where the XTE_Dist is multiplied with a user defined Gain parameter.

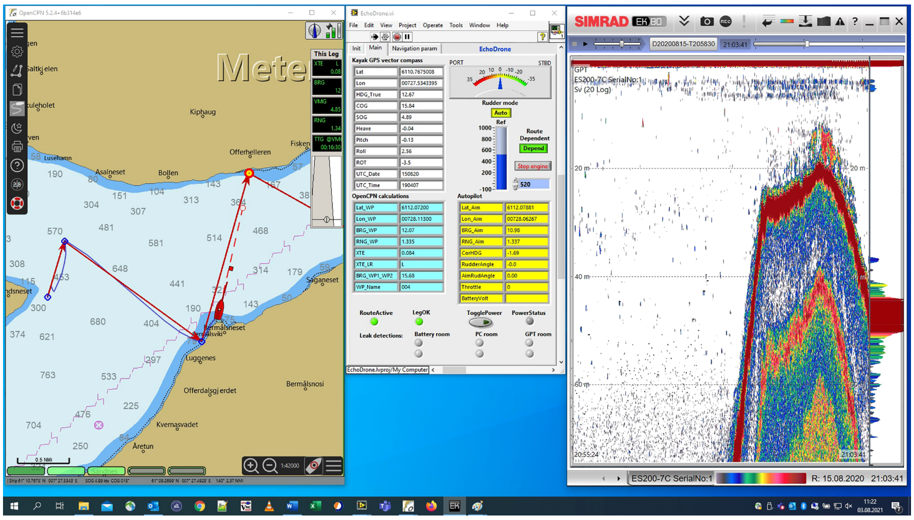
Step 4 (EchoDrone): Calculate Latitude and Longitude of WP_aim.

Step 5 (EchoDrone): Calculate the range and bearing from the current USV position to the WP_aim position.

Step 6 (EchoDrone): Calculate the HDG_aim to steer towards WP_aim. The HDG_adjustment_angle is also calculated and control the physical rudder angle. The rudder angle is adjusted in user defined rudder angle intervals to prevent continuous rudder adjustments for small changes in HDG_true.

Step 7 (OpenCPN): Calculate if the USV position is inside Arrival_radius. If inside the WP_next_destination will become the new active WP_destination and Next_leg will become Active_leg.

FIGURE 3
Illustration of the autonomous Kayak Drone navigation operation. A survey route with several legs (red arrows) are activated (similar to what is shown in the OpenCPN map). The Kayak Drone (red vessel) is sailing towards the next waypoint destination (WP_destination) marked as a red circle with yellow fill. Inactive waypoints are shown as blue diamonds. During operation, a set of calculations are continuously performed by OpenCPN and EchoDrone, and data are exchanged between the two programs. The calculation steps and resulting actions are described on the right side of the Figure. Note that the calculated WP_aim is used to compensate for drift due to wind and currents, which may result in a cross track error distance (XTE_Dist) to the left or right (XTE_LR) of the Active_leg. WP, way point; HDG, heading; XTE_Dist, cross track error (distance from active leg) and whether the USV is to the left or right side of the active leg (XTE_LR).



This Leg
XTE L 0.08
BRG 12
VMG 4.85
RNG 1.34
TTG @VMG 00:16:30

Kayak GPS vector compass

Lat	6110.7675008
Lon	00727.5343395
HDG_True	12.67
COG	15.84
SOG	4.89
Heave	-0.04
Pitch	-0.13
Roll	2.56
ROT	-3.5
UTC_Date	150820
UTC_Time	190407

OpenCPN calculations

Lat_WP	6112.07200
Lon_WP	00728.11300
BRG_WP	12.07
RNG_WP	1.335
XTE	0.084
XTE_LR	L
BRG_WP1_WP2	15.68
WP_Name	004

Autopilot

Lat_Aim	6112.07881
Lon_Aim	00728.06267
BRG_Aim	10.98
RNG_Aim	1.337
CorHDG	-1.69
RudderAngle	-0.0
AimRudAngle	0.00
Throttle	0
BatteryVolt	

FIGURE 4
The Kayak Drone PC screen as seen in real time by the controller *via* the VNC radio link. To the left is the OpenCPN chart program. The route is indicated by consecutive red arrows. Next waypoint is indicated by a red circle with yellow fill. The drone is shown as a red ship and the actual sailed track of the drone is shown as a blue curved line. In the middle of the screen is the EchoDrone USV control software. To the right is the echo sounder software showing a school of sprat in shallow waters close to the current position near the shore. .

ecosystems of the Norwegian fjords (Bakken, 1973). For almost five decades, changes in the biomass of sprat in these fjords have been monitored with acoustic trawl surveys carried out by research vessels. There has been a growing concern that a considerable and variable fraction of the sprat population have not been accessible for acoustic monitoring as the sprat may be distributed very close to the shore and the sea surface. To examine the full vertical and horizontal distribution of sprat in a fjord arm (Årdalsfjord) of Sognefjord, the Kayak Drone was used (Figure 5) in August 2020. The fjord arm was covered once at night and once during daytime, running zigzag transects (see survey report Johnsen et al., 2020). In each coverage, the Kayak Drone sailed in front of RV Kristine Bonnevie with a distance of at least 500 m. Echo sounder data were collected continuously from the two platforms and the data were analyzed using identical methods. The transducer depths were ~1.2 m and ~6 m, respectively, on the Kayak Drone and the RV Kristine Bonnevie.

The mean acoustic density measured from the Kayak Drone was more than three times higher the density measured with the vessel, and the Kayak Drone measured approximately the same density of sprat in both coverages. RV Kristine Bonnevie measured about 40% higher mean density during daytime than at nighttime. As expected, the vertical density distribution of sprat was considerably closer to the sea surface at night where about 65% of the density was found shallower than 8 m (Figure 6), which is approximately the upper depth from which the echo sounder on the large vessel can detect fish. During the day coverage, 30% of the fish density was found above 8 meters (Figure 6). In addition, during night many schools of sprat were found closer to the shore than the large vessel can navigate safely (Johnsen et al., 2020). Clearly, the Kayak Drone data shows that a high and variable fraction of the sprat density is not available during standard acoustic surveys using large research vessel. The large difference in diel vertical distributions and day-night density measurements made by the research vessel show that the standard survey estimates are biased as the fjord surveys are carried out day and night.

3.2 Generated noise and avoidance of herring from the kayak drone

It is well known that many fish species are very sensitive to approaching research vessels, and therefore may lead to biased survey estimates (De Robertis and Handegard, 2013). The vessel avoidance of the Norwegian spring spawning herring (Clupea harengus), which is one of the largest fish stocks in the North Atlantic (Devold, 1963), is well documented (Vabø et al., 2002; Ona et al., 2007). Despite large scientific effort to overcome the source of error caused by the blind zone and vessel avoidance, neither advance modelling of fish behavior (Vabø and Nøttestad, 1997) nor the development of 3D sonars for near surface monitoring (Holmin et al., 2016) have succeeded. In November 2020, we investigated the avoidance of the herring during their overwintering in fjords in northern Norway. The test site is very good as the herring is responsive to noise in this period due to large predation by whales (Vogel et al., 2021). The behavior is predictable and makes the experimental setup relatively simple as the herring emerges close to the surface at night, in a large dense layer (Ona et al., 2007).

In our experiment, we anchored an upward facing echo sounder below the herring layer and continuously monitored the behavior and backscattering strength of the herring layer at night without disturbing their natural behavior. We identified the exact position the upward facing echo sounder and repeatedly

FIGURE 5
The Kayak Drone in operation during the sprat acoustic survey in Årdalsfjorden, Norway, 2020.

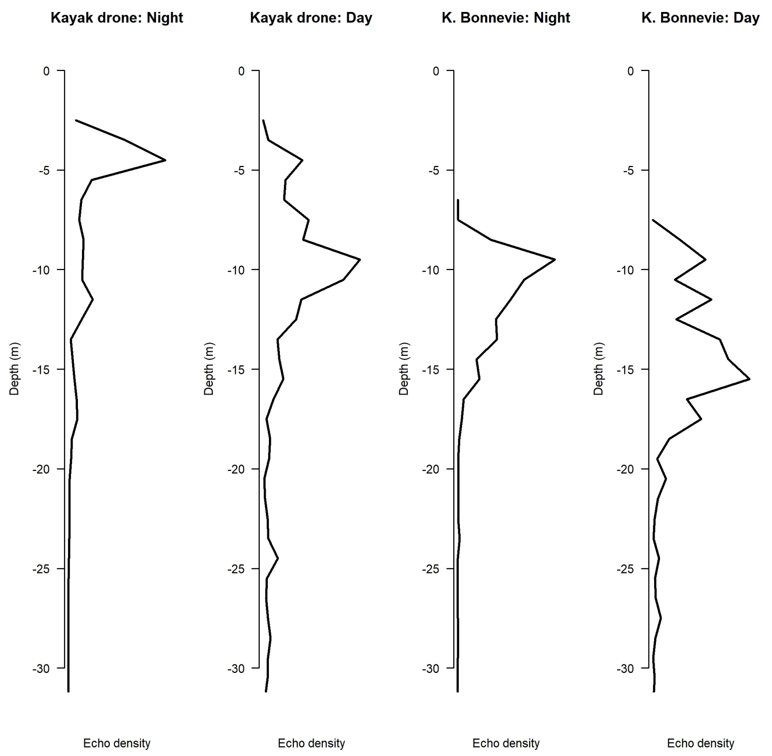


FIGURE 6
The echo density of sprat by depth, measured from the Kayak Drone and the research vessel Kristine Bonnevie (Årdalsfjorden, Norway, August 2020).

sailed over this position with the Kayak Drone, and later with a large research vessel G.O. Sars (Gross Registered Tonnage (GRT) 4067 tons, Length Overall (LOA) 77,5m). The upper part of the herring layer was about 20 m below the sea surface (Figure 7). The herring did not show any avoidance reaction when the Kayak Drone passed over the upward facing echo sounder (Figure 7). In contrast, the herring showed a strong diving response when the research vessel passed (Figure 7). The latter is in accordance with results previously presented by Ona et al. (2007). Clearly, surveys carried out by the Kayak Drone will lead to more accurate survey estimates as strong vessel avoidance of fish affects the survey estimate accuracy. The horizontal escapement reduces the number of individuals observed in the acoustic beam and the diving fish has a tilt angle that results in a reduced acoustic backscatter compared to fish with natural behavior (Vabø et al, 2002).

The noise produced by the USV, and the research vessel (with diesel electric propulsion) was measured in situ following ANSI S12.64-2009 (Grade C) (ANSI/ASA, 2009). A single hydrophone (Ocean Instruments Soundtrap HF) was deployed at 36.4 m depth on a bottom anchored mooring, which corresponds to a 20-degree angle from the hydrophone to the point of closest approach of the vessels (100 m from the surface above the hydrophone (see Figure 2 in ANSI S12.64-2009). Both the USV and the research vessel performed multiple measurement runs (port and starboard). Using the received sound pressure levels between 50 to 10 000 Hz, and assuming geometric sound propagation only, the average difference in source levels was estimated, with the USV source level being significantly lower than the source level of the research vessel (~ -37 dB re 1 µPa at 1m).

3.3 Using the kayak drone to observe seabirds and fish schools

It is difficult to observe the predation by seabirds on of fish schools without interfering with their natural interspecific behavior (Safina and Burger, 1988; Hentati-Sundberg et al., 2018), and there is a need to develop methods to study such fine-scale predator–prey patterns (Hentati-Sundberg et al., 2018). As presented above, several experiments shows that the Kayak Drone can observe fish near the sea surface using echo sounders without affecting their natural behavior. To test the avoidance reaction of swimming seabirds such as puffin (Fratercula arctica), common guillemot (Uria aalge) and

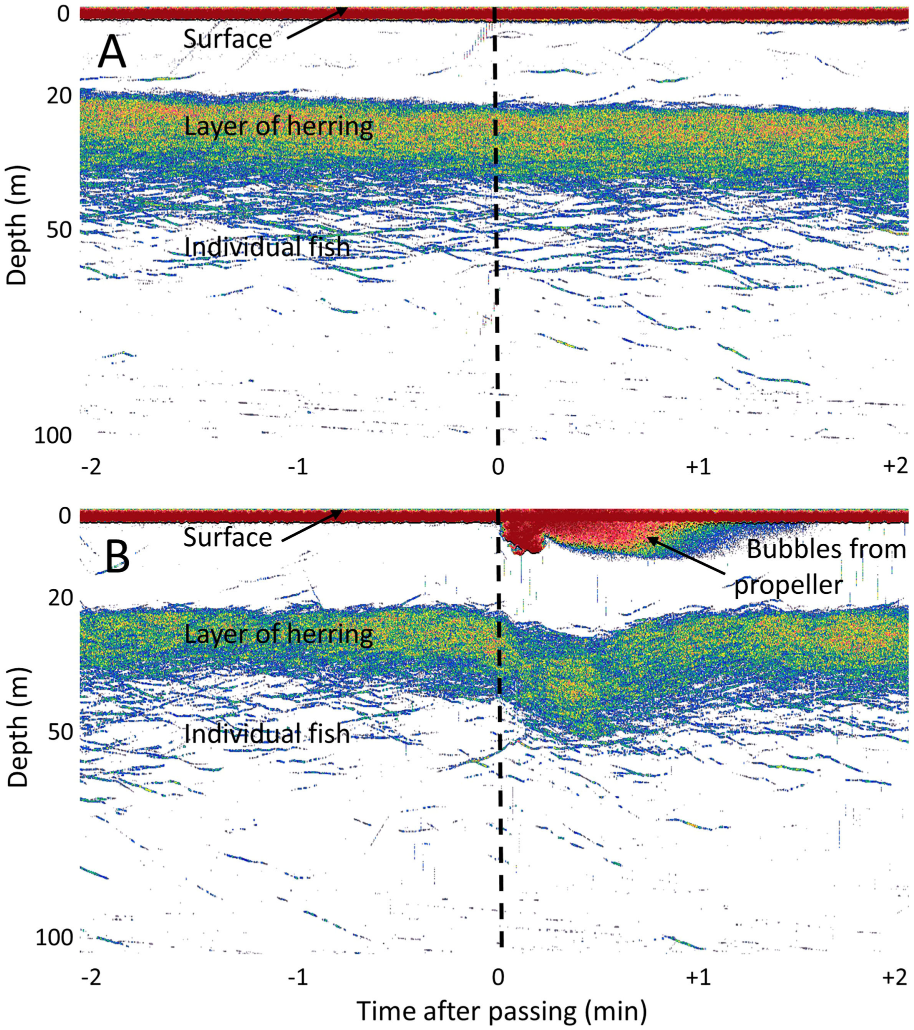


FIGURE 7
Echograms from an upward looking echo sounder mooring when the Kayak Drone **(A)** and a research vessel **(B)** pass over the mooring. A dense layer of herring can be observed from 20 to 40 meters below the surface. Below the herring layer, dispersed individual fish can be seen. Air bubbles from the propeller of the vessel **(B)** can be seen down to approx. 10 meters depth, while no bubbles are produced by the Kayak Drone **(A)**. The time of passing the mooring is shown as vertical black lines.

razorbills (Alca torda) to the Kayak Drone, an experiment was carried out near the bird cliffs at the island of Runde, Norway, May 2018.

A Gopro camera mounted on the antenna mast was used to observe the swimming seabirds, and the echo sounder was set to record fish aggregations and diving birds. By following the instructions of the local ornithologists, the Kayak Drone was steered slowly (~2 knots) into the aggregations of seabirds. No reaction was observed from the birds until the USV was about 2 meters away (Figure 8). Then the seabirds started swimming fast away from the approaching Kayak Drone. Observing diving seabirds targeting fish was also an objective in this experiment, but little fish and no diving birds were acoustically observed. Furthermore, the video recordings did not show any seabird with fish in their beaks. According to the experts, the non-feeding behavior was not very surprising as e.g. puffins typically aggregate near the cliffs in the afternoon before they return to their burrows at sunset. Nevertheless, the results are promising as the Kayak Drone seems to be applicable when observing seabird-fish interaction without affecting their natural behavior.

4 Discussion

The Kayak Drone project started with some scientific questions that could not easily be answered using traditional

FIGURE 8
The Kayak Drone approaching seabirds next to the Runde bird cliffs, Norway.

vessels or measurement platforms. To address these scientific objectives, at set of requirements and design criteria was listed (Table 1) and the Kayak Drone was built accordingly. The Kayak Drone is a very silent, fast-moving platform that supports high quality data collection from scientific broadband echo sounders. It can operate with high position accuracy in narrow fjords, lakes, shallow coastal waters and in open sea during favorable sea states. To do so, electric propulsion and battery power is required. Long-time survey operations covering large areas is therefore not within the scope. That USV niche is better covered by sail drones or petroleum powered USVs. The results from several experiments, shows that the Kayak Drone has answered the defined scientific needs. It is now in regular use and operates without any major technical problems.

The modular concept with regards to hardware and software design turned out to be important as is it difficult to foresee all future needs and challenges during the planning phase of such a project. Since the first sea trial in February 2018, the performance of the Kayak Drone has been considerably improved as a result of lessons learnt during sea operations. We gained practical experience with the USV during field experiments, leading to improvements in the control software, introduction of high accuracy maps and new compartment lids that does not leak. These upgrades have, together with a number of minor technical improvements, advanced the system to its current well-functioning level. One of the most significant upgrades was the installation of the MBR radio link which substantiality improved the stability of the remote communication. Thereafter, the USV has performed as originally intended, and without any significant challenges during surveys.

The Kayak Drone was not design as a prototype for future mass production, however, there is some lessons learnt that may be useful for others. The inclusion of the open-source program OpenCPN as a key component in the system, is probably the most consequential decision that we made. We regard this as a great success, and it has allowed safe USV operations in narrow fjords by using high resolution sea charts. The project would not have had the resources to develop a self-made, high quality chart system as an integrated part of the USV control program.

Despite the successful use of the Kayak Drone, it still has weaknesses and a potential for improvement. Upgrading and improvement of the current USV could include a more hydrodynamic keel/bulb to reduce drag, increase speed and make the USV more energy efficient. Some air bubbles also seem to originate from a sub-optimal keel design. These bubbles interfere negatively on the acoustic recordings. The GoPro cameras could be replaced by integrated MBR cameras to allow real time video streaming. The weight of the USV could be reduced and operation time increased if the batteries were shifted to more modern types. There is also space for more batteries than currently installed. For the analysis of the acoustic recordings, an additional echo sounder operating at a different frequency would enhance the scientific outcome. When exposed to large ocean waves, the roll becomes evident on the echograms as changes in density of school registrations. This could be mitigated by mechanical transducer gimbles, thus improving the quality of the acoustic recordings. It has been a challenge that the fiber glass hull is vulnerable to damage during retrieval in open sea and large waves. Cracks in the hull sometimes occur with consequential insignificant leakage. On the other hand, it is easy to repair, and the leakage alarms give a good real time overview of the situation. To mitigate this problem, a new hull might be built in a more rugged material than fiber glass.

The field experiments with the Kayak Drone have clearly shown that some types of marine science surveys cannot optimally be conducted by large research vessels. Their size may prevent adequate sampling of the target population as the acoustic sensor is mounted too deep, or because the vessel cannot navigate into the shallow habitats. In regions of the world with gradually sloping bottom from deep to shallow waters, this may negatively influence the research vessel's coverage of some target species and their distribution. For abundance estimation surveys it is a prerequisite that the survey brackets the target population (Gunderson, 1993), this is not always possible with large traditional research vessels. The field results have been very promising and support the assumption that USV technologies will become important new tools that will advance marine research in the future. The silent acoustic Kayak Drone has produced novel results within some fields of marine research. From the Kayak Drone experiments, new knowledge has been derived that possibly will lead to alteration of the current survey design for the Norwegian sprat monitoring.

Sometimes new tools provide unexpected observations. In the seabird-fish interaction experiment, several bird species showed almost no avoidance from the Kayak Drone, however, we observed that the northern gannets (Morus bassanus) circled the sky above the USV. It seems like the Kayak Drone acted as an aggregating device for the gannets. In another experiment on the sandeel grounds in the North Sea we carried an experiment in 2021 with the objective of measuring sandeel near the sea surface. From the GoPro video recordings on the keel, we observed a minke whale (Balaenoptera acutorostrata) that approached below the Kayak Drone three times over a period of 15 minutes. This may have been a random event, but such observations are nevertheless worth some reflection. A silent measuring platform may attract some species rather than scaring them away. In either case, a silent platform may be well suited to observe natural behavior of some species, but not necessarily others.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation.

Author contributions

EJ has been the project leader and responsible for defining the scientific requirements that the USV needed to meet the research goals. In addition, he is responsible for all the experimental survey designs. AT has been responsible for the technical design and construction of the USV. EJ and AT have authored the manuscript in a joint effort. All authors contributed to the article and approved the submitted version.

Funding

Norwegian Research Council, project number 243941/E40. Center for Research-based Innovation in Marine Acoustic Abundance Estimation and Backscatter Classification (CRIMAC; number 309512). Nordic Council research project, Hydroacoustic, project number 169 (2017). Internal research projects at Institute of Marine Research, Norway.

Acknowledgments

The authors want to thank Torqeedo GmbH, Germany for giving us access to the internal data protocols for their propulsion system. Nigel Dennis at Sea Kayaking UK and Ronny Riise at God Tur, Norway made the necessary rebuilding of the standard Triton kayak hull. Harald Fitje at IMR Norway made the housing and adaptation of the rudder motor. Kongsberg Seatex and Kongsberg Simrad have given us invaluable assistance throughout. Furthermore, Runde Environmental Centre and Cecilie Kvamme, IMR have invited us to participate in their field activities, and Geir Pedersen provided the noise measurement results from the herring experiments.

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

ANSI/ASA (2009). "ANSI S12.64-2009," in *Quantities and procedures for description and measurement of underwater sound from ships–part 1: General requirements*" (New York: American National Standards Institute/Acoustical Society of America).

An, Y., Yu, J., and Zhang, J. (2021). Autonomous sailboat design: A review from the performance perspective. *Ocean Engineering.* 238, 109753. doi: 10.1016/j.oceaneng.2021.109753

Bakken, E. (1973). Sprat in Norwegian waters, a short review of biology, fishery and current research. *Coun. Meet. Int. Coun. Explor Sea*, 1–13.

Benoit-Bird, K. J., Patrick Welch, T., Waluk, C. M., Barth, J. A., Wangen, I., McGill, P., et al. (2018). Equipping an underwater glider with a new echosounder to explore ocean ecosystems. *Limnol. Oceanog.: Methods* 16 (11), 734–749. doi: 10.1002/lom3.10278

Brautaset, O., Waldeland, A. U., Johnsen, E., Malde, K., Eikvil, L., Salberg, A. B., et al. (2020). Acoustic classification in multifrequency echosounder data using deep convolutional neural networks. *ICES J. Mar. Sci.* 77 (4), 1391–1400. doi: 10.1093/icesjms/fsz235

Chang, H. C., Hsu, Y. L., Hung, S. S., Ou, G. R., Wu, J. R., and Hsu, C. (2021). Autonomous water quality monitoring and water surface cleaning for unmanned surface vehicle. *Sensors* 21 (4), 1102. doi: 10.3390/s21041102

Dalen, J., and Løvik, A. (1981). The influence of wind-induced bubbles on echo integration surveys. *J. Acoustical Soc. America* 69 (6), 1653–1659. doi: 10.1121/1.385943

Demer, D. A., Andersen, L. N., Bassett, C., Berger, L., Chu, D., Condiotty, J., et al. (2017). evaluation of a wideband echosounder for fisheries and marine ecosystem science. In: *ICES cooperative research report*. Available at: https://archimer.ifremer.fr/doc/00585/69730/67619.pdf (Accessed May 23, 2022).

De Robertis, A., and Handegard, N. O. (2013). Fish avoidance of research vessels and the efficacy of noise-reduced vessels: a review. *ICES J. Mar. Sci.* 70 (1), 34–45. doi: 10.1093/icesjms/fss155

De Robertis, A., Lawrence-Slavas, N., Jenkins, R., Wangen, I., Mordy, C. W., Meinig, C., et al. (2019). Long-term measurements of fish backscatter from saildrone unmanned surface vehicles and comparison with observations from a noise-reduced research vessel. *ICES J. Mar. Sci.* 76 (7), 2459–2470. doi: 10.1093/icesjms/fsz124

Devold, F. (1963). The life history of the atlanto-scandian herring. *Rapp Proc. Verb Cons Int. Explor. Mer* 154, 98–108.

Ghani, M. H., Hole, L. R., Fer, I., Kourafalou, V. H., Wienders, N., Kang, H., et al. (2014). The SailBuoy remotely-controlled unmanned vessel: Measurements of near surface temperature, salinity and oxygen concentration in the northern gulf of mexico. *Methods Oceanography* 10, 104–121. doi: 10.1016/j.mio.2014.08.001

Gunderson, D. R.. (1993). *Surveys of fisheries resources* (John Wiley & Sons), 1–256.

Handegard, N. O., Algrøy, T., Eikvil, L., Hammersland, H., Tenningen, M., and Ona, E. (2021b)Smart fisheries in Norway: Partnership between science, technology, and the fishing sector (Accessed May 23, 2022).

Handegard, N. O., Tonje, F. N., Johnsen, E., Pedersen, G., Tenningen, M., Korneliussen, R., et al. (2021). Fisheries acoustics in Norway–wide band data, autonomous platforms and deep learning. *J. Acoustical Soc. America* 150 (4), A254–A254. doi: 10.1121/10.0008196

Hentati-Sundberg, J., Evan, T., Osterblom, H., Hjelm, J., Larson, N., Bakken, V., et al. (2018). "Fish and seabird spatial distribution and abundance at the largest seabird colony in the Baltic sea". *Mar. Ornithol.* 46, 61–68.

Holmin, A. J., Korneliussen, R. J., and Tjøstheim, D. (2016). Estimation and simulation of multi-beam sonar noise. *J. Acoustical Soc. America* 139 (2), 851–862. doi: 10.1121/1.4941913

Johnsen, E., Kvamme, C., and Totland, A. (2020). Measuring distribution and density of sprat in Årdalsfjorden with a kayak drone. *Rapport fra Havforskningen* 2020-28, 1893–4536.

Johnsen, E., Pedersen, R., and Ona, E. (2009). Size-dependent frequency response of sandeel schools. *ICES J. Mar. Sci.* 66 (6), 1100–1105. doi: 10.1093/icesjms/fsp091

Kum, B. C., Shin, D. H., Jang, S., Lee, S. Y., Lee, J. H., Moh, T., et al. (2020). Application of unmanned surface vehicles in coastal environments: Bathymetric survey using a multibeam echosounder. *J. Coast. Res.* 95 (SI), 1152–1156. doi: 10.2112/SI95-223.1

Ludvigsen, M., Albrektsen, S. M., Cisek, K., Johansen, T. A., Norgren, P., Skjetne, R., et al. (2016). Network of heterogeneous autonomous vehicles for marine research and management. *OCEANS 2016 MTS/IEEE Monterey*, 1–7. doi: 10.1109/OCEANS.2016.7761494

Mitson, R. B., and Knudsen, H. P. (2003)Causes and effects of underwater noise on fish abundance estimation. Available at: https://www.alr-journal.org/articles/alr/pdf/2003/03/alr3043.pdf (Accessed May 23, 2022).

Ona, E., Godø, O. R., Handegard, N. O., Hjellvik, V., Patel, R., and Pedersen, G. (2007). Silent research vessels are not quiet. *J. Acoustical Soc. America* 121 (4), EL145–EL150. doi: 10.1121/1.2710741

Othman, E. H. (2015)A review on current design of unmanned surface vehicles (usvs) (Accessed May 23, 2022).

Romano, A., and Duranti, P. (2012)Autonomous unmanned surface vessels for hydrographic measurement and environmental monitoring. In: *FIG working week 2012* (Rome Italy). Available at: https://www.fig.net/resources/proceedings/fig_proceedings/fig2012/papers/ts04d/TS04D_romano_duranti_6118.pdf (Accessed May 23, 2022).

Safina, C., and Burger, J. (1988). Use of sonar and a small boat for studying foraging ecology of seabirds. *Colonial Waterbirds* 11 (2), 234–244. doi: 10.2307/1521005

Shao, G., Ma, Y., Malekian, R., Yan, X., and Li, Z. (2019). A novel cooperative platform design for coupled USV–UAV systems. *IEEE Trans. Ind. Inf.* 15 (9), 4913–4922. doi: 10.1109/TII.2019.2912024

Simmonds, J., and MacLennan, D. N. (2008). *Fisheries acoustics: Theory and practice*. John Wiley & Sons, 1–472.

Vabø, R., and Nøttestad, L. (1997). An individual based model of fish school reactions: Predicting antipredator behaviour as observed in nature. *Fisheries oceanog.* 6 (3), 155–171. doi: 10.1046/j.1365-2419.1997.00037.x

Vabø, R., Olsen, K., and Huse, I. (2002). The effect of vessel avoidance of wintering Norwegian spring spawning herring. *Fisheries Res.* 58 (1), 59–77. doi: 10.1016/S0165-7836(01)00360-5

Verfuss, U. K., Aniceto, A. S., Harris, D. V., Gillespie, D., Fielding, S., Jiménez, G., et al. (2019). A review of unmanned vehicles for the detection and monitoring of marine fauna. *Mar. Pollut. Bull.* 140, 17–29. doi: 10.1016/j.marpolbul.2019.01.009

Vogel, E. F., Biuw, M., Blanchet, M., Jonsen, I. D., Mul, E., Johnsen, E., et al.(2021) Killer whale movements on the Norwegian shelf are associated with herring density. *Mar. Ecol. Prog. Ser* 665, 217–231. doi: 10.3354/meps13685

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fmars.2022.996493

*CORRESPONDENCE
Lei Xiao
xiaolei@wzu.edu.cn

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 17 July 2022
ACCEPTED 15 August 2022
PUBLISHED 30 September 2022

CITATION
Yu S, Xiao L and Sun W (2022)
Parametermeasurement of aircraft-radiated noise from a single
acoustic sensor node in three-dimensional space.
Front. Mar. Sci. 9:996493.
doi: 10.3389/fmars.2022.996493

Parametermeasurement of aircraft-radiated noise from a single acoustic sensor node in three-dimensional space

Shuwen Yu[1], Lei Xiao[2]* and Weitao Sun[3]

[1]Information Technology Center, Wenzhou University, Wenzhou, China, [2]Intelligent Information Systems Institute, Wenzhou University, Wenzhou, China, [3]The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang, China

A line spectrum presents the form of a narrow-band time-varying signal due to Doppler effect when the single hydrophone node observes flight-radiated noise. The modulation law of the time-varying signal contains a large number of feature information of moving targets, which can be used for detection and classification. This paper studies the possibility of using instantaneous frequency measurements from the hydrophone node to improve the precision of the flight parameter estimates when the source spectrum contains a harmonic line of constant frequency. First of all, we build up and analyze the underwater sound field excited by the aircraft using the ray theory model; then the Doppler shift in the two isospeed media, which is caused by the aircraft, is established; finally, a robust time–frequency transform describes the time–frequency distribution of the received signal, and a geometric approach solves the flying parameters.

KEYWORDS

underwater acoustic detection, Doppler effect, instantaneous frequency estimation, frequency estimation, white noise

1 Introduction

In most turboprop aircrafts, engines and propellers, as a power plant, produce periodic sounds (Chirico et al., 2018; Aygun et al., 2022) since they originate in mechanical rotation mechanisms. However, as these pressure waveforms are not purely sinusoidal, their spectrum consists of a fundamental frequency and higher-order Fourier components at multiples of the fundamental. The frequency of the first harmonic is between 15 and 100 Hz, and as many as a dozen higher harmonics may be detectable up to about 1 kHz. In terms of energy, the frequency of fundamental is a significant component of the source (Boashash, 2015; Zhang et al., 2022). The relative motion between the aircraft and hydrophone results in Doppler shift in the received signal. Under the condition of two isospeed sound propagation media, the Doppler shift depends on the altitude and velocity of the aircraft, the depth of the hydrophone, and many more. Therefore, obtaining the precise Doppler shift curve is the premise of solving the cruising parameters.

The modeling analysis and the theoretical research of underwater soundings by scholars are mainly in the two fields of a single hydrophone or an array of hydrophones. In the case of individual hydrophones, Ferguson used the ray theory to analyze the acoustic propagation model of air–water media and derived the over-the-top flight parameters of an air target based on the Doppler shift formula in the two-dimensional domain. With the continuous advancement of underwater acoustic detection equipment, vector hydrophones with "∞" type directivity have been applied to detect aerial targets. Compared with traditional scalar hydrophones, vector hydrophones can obtain not only the scalar field information of the signal of interest but also the vector field information. Using a reasonable mathematical model to process vector information, it is possible to determine the target orientation and other flight parameters. In terms of hydrophone arrays, Ferguson and Lo use wide-aperture hydrophone arrays to obtain the azimuth information of aerial target and perform parameter estimation and positioning of helicopters flying in the air, realizing the expansion of target motion information from a two-dimensional domain to three-dimensional domains. Peng Zhaohui *et al*. proposed a robust and reliable line spectral detection method based on hydrophone arrays in view of the interference problem of underwater detection of sound sources of aerial motion. In these studies, the underwater sounding algorithm based on the traditional single hydrophone can only estimate the sound source flight parameter problem in the two-dimensional plane, that is, assume that the helicopter just flew over the top of the hydrophone, without taking into account information such as yaw distance in the three-dimensional domain. Single-vector hydrophones can estimate azimuth information, but there are directivity blind spots which are more susceptible to background noise interference in practical applications. Although the hydrophone array can estimate the azimuth information, it has a high cost and is inconvenient to arrange and not suitable for small-volume underwater detection platforms.

The obvious directivity of acoustic ray presents in the seawater sound field excited by an aircraft. For the-same-level plane in the seawater, the maximum sound pressure lies in position directly below the aircraft, and the greater horizontal offset distance is, the less sound pressure is. For the same vertical line directly below the aircraft in the seawater, the greater the depth causes the less gradient of attenuation pressure. For the same vertical line non-directly below the aircraft in the seawater, with increasing depth, sound pressure increases until reaching a maximum at a certain depth, then with the further increasing of depth, the sound pressure decreases at its lowest level. Generally, time–frequency transform can project the time-varying signal in the time domain into the joint function of time and frequency in the two-dimensional plane. Currently, there are two methods of time–frequency analysis (Fang et al., 2016), and one is linear decomposition using kernel function, such as short-time Fourier transform, wavelet transform, and similar methods (Zhou et al., 2020; Aksenovich, 2020; Vandana and Arokiasamy, 2022). Similar to the main component of the signal, the linear combination of kernel function can represent it sparsely. The second one is quadratic time–frequency distribution, such as the Wigner–Ville distribution (Zcza et al., 2021; Kumar and Prasad, 2022). Without the interference of the window function, this method avoids constraints between time resolution and frequency resolution. The time–bandwidth product can reach the lower bound of the uncertainty principle, so the time–frequency representation can be improved to obtain accurate results (Zhang et al., 2021).

Compared with the incident rays, the energy of refracted rays has 30-dB energy attenuation; therefore, the time-varying signal that contained information about the parameters of movement is weak. Additionally, with the signal from background noise pollution, it is difficult to achieve the desired results by conventional time–frequency analysis. Therefore, in this paper, we propose a robust time–frequency analysis method to detect Doppler shift.

Furthermore, for the characteristics of the hydroacoustic signal excited by the airborne uniform flight motion target, a single hydrophone is used to solve the three-dimensional parameter estimation problem of the airborne fast-flight target state. Firstly, the discrete line spectrum of the helicopter is used as the characteristic sound source, and its three-dimensional Doppler propagation model in air–water medium is established. Then, based on the asymmetry of the Doppler frequency shift curve and its first- and second-order derivatives, the three-dimensional parameter estimation method of underwater sounding is derived. Finally, by analyzing the measured signals received by a single hydrophone, the paper verifies the estimation performance of the algorithm in the ocean noise scenario.

2 The model of a flight path along a straight line

The study only cares about the case that the direct refraction ray arrives at the hydrophone and considers that both aircraft altitude h and hydrophone depth d are far longer than the acoustic wavelength. Therefore, the aircraft-radiated noise is assumed to be a point source (Zhang et al., 2020). According to the ray theory, rays with the incident angle θ_I less than the critical angle θ_C can refract into the seawater, and only one of them can arrive at the hydrophone node.

2.1 Point source transmission model

The three-dimensional model of the aircraft flying over the hydrophone node in a straight line is depicted in Figure 1. A stationary hydrophone node is marked as H and is located at a constant depth d . An airborne source with a constant frequency f_0 is marked as S as it travels past the hydrophone node in a straight line l_s at a constant velocity v , a constant altitude h , and slant range $w(t)$. Additionally, the source velocity v is assumed to be subsonic, *i.e.*, $v<c_a<c_w$, where c_a and c_w are the velocity of sound in air and water. The moving source S is projected to the horizontal plane where the stationary point H is located and named as $S^{'}$, and the projection corresponding to its trajectory is named as l'_s . The projection of the straight-line distance $r(t)$ between S and H is $w(t)$. Specially, the projection of straight-line distance $r_{\min}\ (t)$ is $w_{\min}$ when the source is at the closest point of approach (CPA). S is in any position on the straight line l_s ; only one sound ray with the incident angle $\theta_I(t)$ refracts at the point T of the air–seawater interface and then arrives at H in the direction of the refraction angle θ_T . In this case, the velocity v of S can be decomposed into two mutually perpendicular components; among them, one component is in the horizontal plane with the straight line l_s , *i.e.*,

$$v_1(t) = v\cos\alpha(t)\,, \tag{1}$$

where $\alpha(t)$ is the slant angle between l_s and $\overline{S'H}$. Furthermore, $v_1(t)$ can be decomposed into two mutually perpendicular components as well, and one component is consistent with the direction of the incident sound ray arriving at the hydrophone node, *i.e.*,

$$v_{11}(t) = v_1(t)\cos\beta(t)\,, \tag{2}$$

where $\beta(t)$ is the line-of-sight angle. $\beta(t)$ and $\theta_I(t)$ are complementary angles to each other on the vertical plane as determined by S , H , and v_{11} .

The Doppler shift $f_d(t)$ is completely calculated by the four source transmission parameters $\{f_0,v,\alpha(t),\beta(t)\}$. Furthermore, the angles $\alpha(t)$ and $\beta(t)$ are solved from the geometric relations shown in Figures 2 and 3, respectively.

2.2 Doppler shift model

The relative position varies between S and H as presented in Figure 2. Note that the source velocity v and the moment t is either negative or positive, depending on whether H is on the left- or right-hand side of the CPA as S moves along its

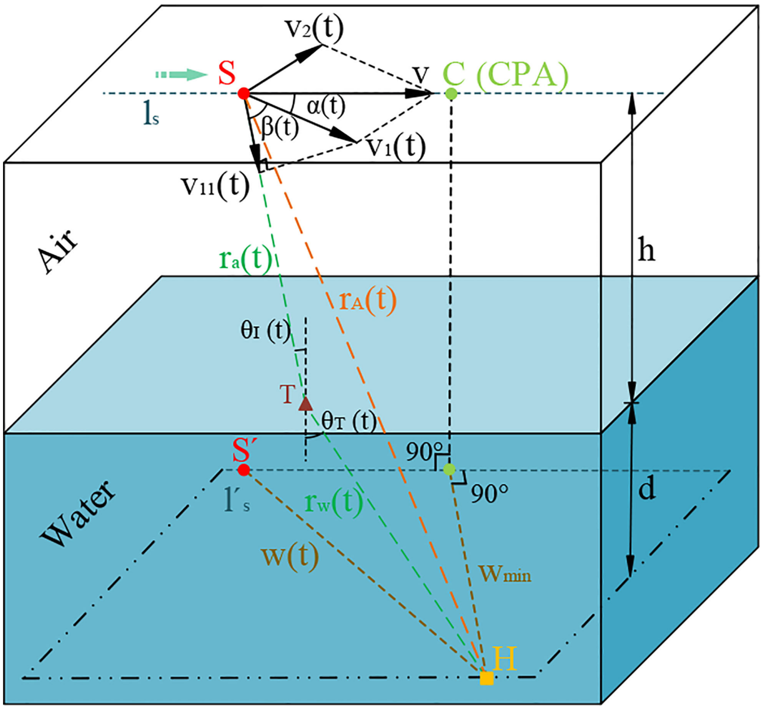


FIGURE 1
General geometrical configuration for a hydrophone node in the water and the linear trajectory of a source in the air as it transits past the hydrophone node in a straight line at constant velocity v , constant altitude h , and certain slant range w.

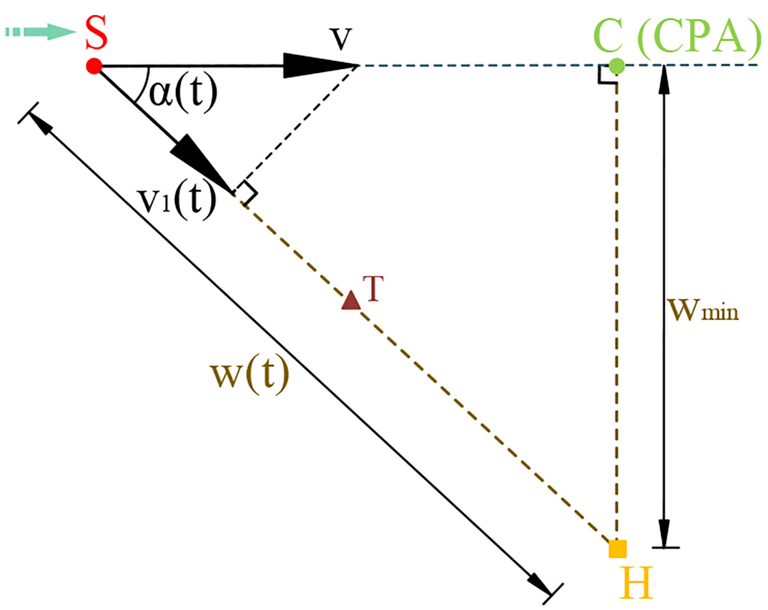


FIGURE 2
Top view of Figure 1. The slant range between the source and the hydrophone at time *t* is given by *w* . The distance from the hydrophone to the closest point of approach of the source is denoted by w_{min}.

trajectory. According to the Pythagorean theorem, the slant range $w(t)$ at time t can be derived as

$$w(t) = \sqrt{v^2t^2 + w_{\min}^2}\ , \tag{3}$$

Then, according to the trigonometric function, the slant angle $\alpha(t)$ is given by

$$\alpha(t) = \arcsin \frac{w_{min}}{\sqrt{v^2t^2 + w_{min}^2}}\ . \tag{4}$$

According to the geometric relations in Figure 3, there exists

$$\beta_I(t) = \frac{\pi}{2} - \theta_I(t)\ . \tag{5}$$

The propagation path of a sound ray with a specific incidence angle $\theta_I(t)$ is presented in Figure 3. The refraction angle $\theta_T(t)$ can be solved by the cosine theorem with the apex angle $\angle STH$ in $\triangle STH$ as follows:

$$\cos\{\pi - [\theta_I(t) - \theta_T(t)]\} = \frac{r_a^2(t) + r_w^2(t) - r^2(t)}{2r_a(t)r_w(t)}\ , \tag{6}$$

where the propagation distance $r_a(t)$ and $r_w(t)$ is given by

$$r_a(t) = \frac{h}{sin\ \ \theta_I(t)}\ ,\quad r_w(t) = \frac{d}{sin\ \ \theta_T(t)}\ , \tag{7}$$

and the straight-line distance between S and H is expressed as

$$r^2(t) = (d+h)^2 + v^2t^2 + w_{\min}^2\ . \tag{8}$$

Substituting Eq. (7) and Eq. (8) into Eq. (6), $\theta_T(t)$ with respect to t can be written in the following form:

$$\sin\ \theta_T(t)\sqrt{1 - n^2\cos^2\theta_T(t)}\left[(d+h)^2 + v^2t^2 + w_{\min}^2 - 2dh\right]$$
$$-2dhn\cos^2\theta_T(t) - \frac{d^2 sin\ \ \theta_T(t)}{\sqrt{1 - n^2\cos^2\theta_T(t)}} - \frac{h^2\sqrt{1 - n^2\cos^2\theta_T(t)}}{sin\ \ \theta_T(t)} = 0 \tag{9}$$

While S moves toward or away from the CPA, the relationship of various parameters with the moment t is shown in Table 1.

During the propagation of sound, it has traveled over time as

$$\tau(t) = \frac{h}{c_a sin\ \ \theta_I(t)} + \frac{d}{c_w sin\ \ \theta_T(t)}\ . \tag{10}$$

In particular, the propagation time traveled from the CPA to the hydrophone node is given by

$$\tau_{\min}(t) = \frac{h}{c_a sin\ \ \theta_{Ic}} + \frac{d}{c_w sin\ \ \theta_{Tc}}\ . \tag{11}$$

where

$$w_{\min} = \frac{h}{tan\ \ \theta_{Ic}} + \frac{d}{tan\ \ \theta_{Tc}}\ ,\quad \frac{c_a}{c_w} = \frac{sin\ \ \theta_{Ic}}{sin\ \ \theta_{Tc}}\ . \tag{12}$$

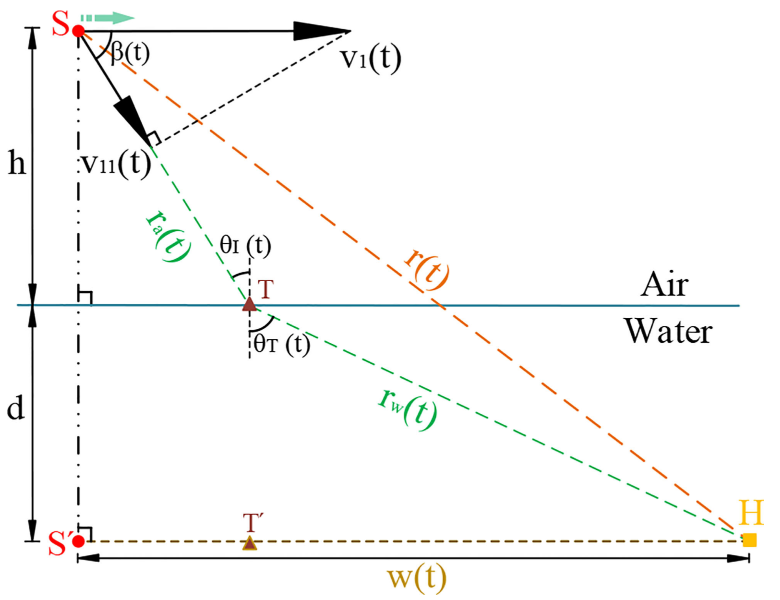


FIGURE 3
Vertical plane of a direct ray path from source to the hydrophone. Refraction of a ray with a specific incidence angle $\theta_I(t)$ through an air–water-layered environment at point T.

For the moving source with constant frequency f_0 , its Doppler frequency $f_d(t_r)$ received by the hydrophone node is given by

$$f_d(t_r) = \frac{f_0}{1 - \frac{v}{c_w} cos \enspace \gamma_T(t_r)} , \tag{13}$$

where t_r=t+$\tau(t)$, cos $\gamma_T(t_r)$=cos $\alpha(t_r)$cos $\beta_T(t_r)$, and $\gamma_T(t_r)$ is the incidence angle in three-dimensional space at the moment t . When the parameters $\{v,h,d,w_{\min}\}$ are known, $\alpha(t_r)$ and $\beta(t_r)$ can be obtained from Eq. (4) and Eq. (9). Therefore, for the straight-line flight, the Doppler frequency information $f_d(t_r)$ is completely specified by the five flight parameters, *i.e.*, $\{f_0,v,h,d, w_{\min}\}$.

In this case, the five flight parameters are changed respectively, as shown in Figures 4A–E. In order to observe the influence of the different parameters on the change trend of the Doppler frequency intuitively, the Doppler frequency $f_d(t)$ is depicted from the perspective of the source S . All five pictures demonstrate that there exists $f_d(t)>f_0$ or $f_d(t)<f_0$ during the process of the source S approaching the CPA or moving away from the CPA.

For each value of f_0 ranging from 40 to 80 Hz at an increment of 10 Hz, Figure 4A shows that the Doppler frequency $f_d(t)$ translates along the positive direction of the f_d axis with the increase of thez source frequency f_0 .

For each value of v ranging from 80 to 120 m/s at an increment of 10 m/s, Figure 4B shows that the maximum of the frequency shift $|f_d(t)–f_0|$ is relative with the velocity v . At the same time t , the greater the velocity is, the greater the frequency shift and its gradient are.

For each value of h ranging from 100 to 500 m at an increment of 100 m, this is illustrated in Figure 4C. For each value of d ranging from 20 to 80 m at an increment of 15 m, this is depicted in Figure 4D. For each value of $w_{\min}$ ranging from 0 to 100 m at an increment of 25 m, this is presented in Figure 4E. Figures 4C–E– show that the altitude h , the depth d , and the minimum slant $w_{\min}$ can affect the gradient of the Doppler frequency $f_d(t)$.

TABLE 1 Rule of parameter variation.

Position/parameters	From infinity to the CPA	CPA	From the CPA to infinity
t	$-\infty \rightarrow 0$	0	$0 \rightarrow +\infty$
α (t)	$0 \rightarrow \pi/2$	$\pi/2$	$\pi/2 \rightarrow 0$ $\pi/2 - arctan\ h/w_{a\ min} \rightarrow arcsin\ c^a/cw$
θ_I (t)	$arcsin\ c_a/cw \rightarrow \pi/2 - arctan\ h/w_{a\ min}$	$\pi/2 - arctan\ h/w_{a\ min}$	
θ_T (t)	$\pi/2 \rightarrow \pi/2 - arctan\ d/w_{w\ min}$	$\pi/2 - arctan\ d/w_{a\ min}$	$\pi/2 - arctan\ d/ww\ min \rightarrow \pi/2$

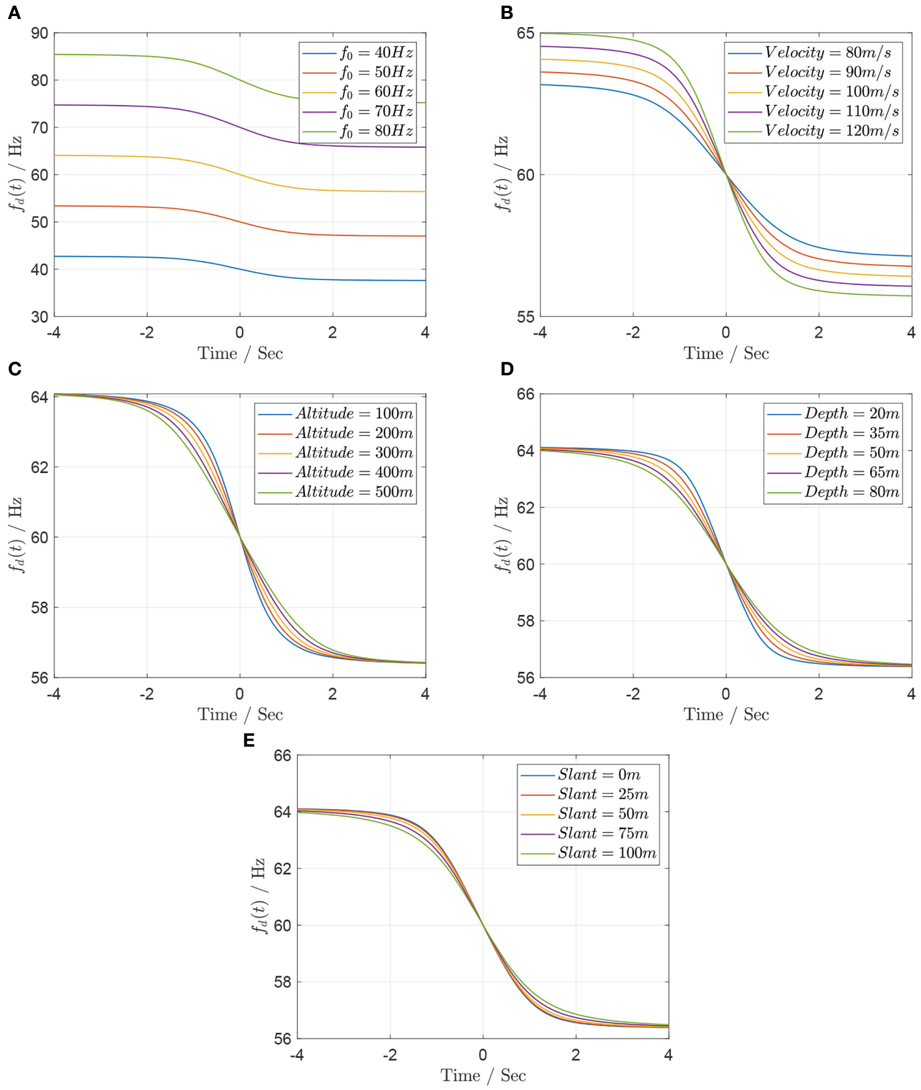


FIGURE 4
(A) Doppler shift curve f_d (*t*) *versus* frequency f_0 ∈ [30, 70] Hz every 10 Hz, assuming *v* = 100 m/s, *h* = 300 m, *d* = 50 m, and wmin = 50 m. **(B)** Doppler shift curve f_d (*t*) *versus* frequency *v* ∈ [80, 120] m/s every 10 m/s, assuming f_0 = 60 m/s, *h* = 300 m, *d* = 50 m, and wmin = 50 m. **(C)** Doppler shift curve f_d (*t*) *versus* frequency *h* ∈ [100, 500] m every 100 m, assuming f_0 = 60 m/s, *v* = 100 m/s, *d* = 50 m, and wmin = 50 m. **(D)** Doppler shift curve f_d (*t*) *versus* frequency *d* ∈ [20, 80] m every 15 m, assuming f_0 = 60 m/s, *v* = 100 m/s, *h* = 300 m, and wmin = 50 m. **(E)** Doppler shift curve f_d (*t*) *versus* frequency wmin ∈ [0, 100] m every 25 m, assuming f_0 = 60 m/s, *v* = 100 m/s, *h* = 300 m, and *d* = 50 m.

2.3 Estimation of flight parameters

Assuming that the sound velocity c_a and c_w in air and water, the hydrophone depth d , and its received Doppler frequency $f_d(t)$ are known, the section proposes a solution to estimate the sound source parameters, *i.e.*, f_0 , v , h and $w_{\min}$, using a single hydrophone.

Submitting Eq. (4) into Eq. (13), there exists

$$f_d(t) = f_0\left[1-\frac{v^2 t\cos\ \beta_T(t)}{c_w\sqrt{v^2t^2+w_{min}}}\right]^{-1}. \tag{14}$$

The velocity v and constant frequency f_0 of source can be obtained, respectively, as follows:

$$v = c_w\frac{f_d(-\infty)-f_d(+\infty)}{f_d(-\infty)+f_d(+\infty)},$$

$$f_0 = 2\frac{f_d(-\infty)f_d(+\infty)}{f_d(-\infty)+f_d(+\infty)}. \tag{15}$$

The first- and second-order derivative of $f_d(t)$ is derived in Appendix A. The four flight parameters $\{h, w_{\min}, \beta_T(t)\}\{h,$

$w_{\min}$,$\beta_T(t)$} are obtained by solving a set of equations as follows:

$$\frac{df_d(t)}{dt}\Big|_{t=\tau_{\min}} = f_d^{'}(f_0, v, h, w_{\min}) = f_d^{'}(\tau_{\min}) , \tag{16}$$

and

$$\frac{d^2 f_d(t)}{dt^2}\Big|_{t=\tau_{\pm 1}} = f_d^{''}(f_0, v, h, w_{\min}) . \tag{17}$$

Assume that the source with f_0=68*Hz* and *v*=123*m/s* flies along a straight line at *h*=200 m and the shortest projection $w_{\min}$ between *H* and $S^{'}$ is 30 m. The time–frequency curve of the signal received by the hydrophone with a depth of 50 m is shown in Figure 5.

Figure 5 shows that $f_d(-\infty)$ and $f_d(+\infty)$ can be approximated as f_d=73.8*Hz* at *t*=−4 s and f_d=63.4*Hz* at *t*=4 s, respectively. According to Eq. 15, *v* and f_0 , respectively, is given by

$$v = c_w \frac{f_d(-4) - f_d(+4)}{f_d(-4) + f_d(+4)} ,$$

$$f_0 = 2\frac{f_d(-4) f_d(+4)}{f_d(-4) + f_d(+4)} . \tag{18}$$

Next, we solve the following equations to obtain *h* , $w_{\min}$, and $\beta_T(t)$ at the moment *t*

$$f_d^{'}(0) = -6.921 ,$$

$$f_d^{''}(-0.5) = -7.614 ,$$

$$f_d^{''}(0.6) = 7.253 . \tag{19}$$

3 Time–frequency analysis method

Under strong noise conditions, most traditional methods are not able to ensure the robustness and accuracy of time–frequency analysis simultaneously. Therefore, this section

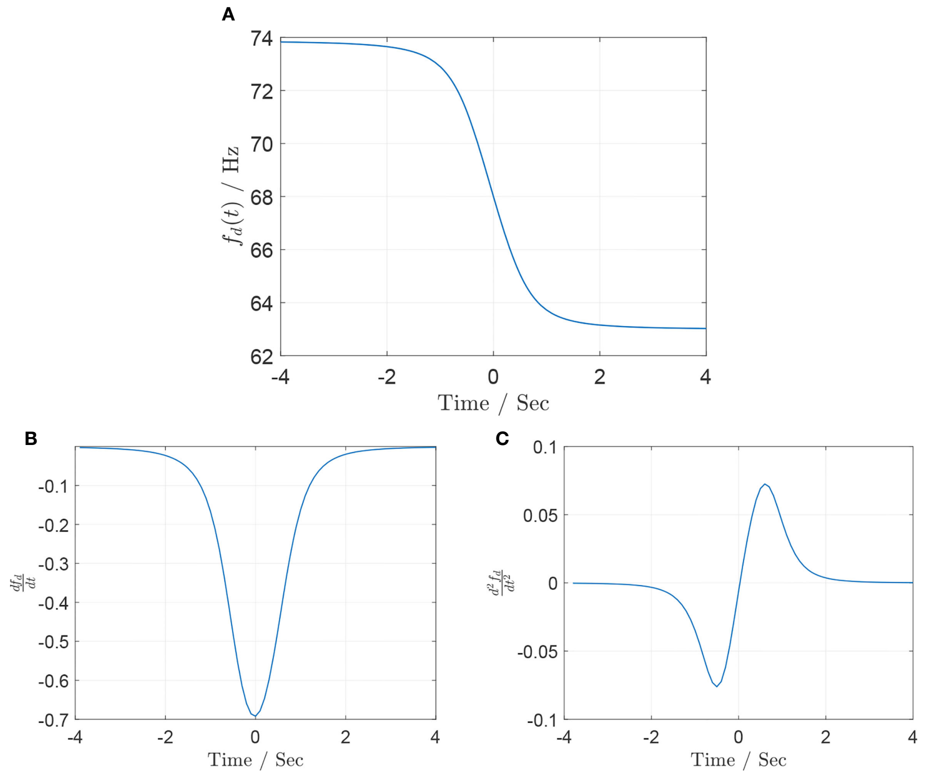


FIGURE 5
(A) Doppler shift curve f_d (*t*), assuming f_0 = 68 m/s, *v* = 123 m/s, *h* = 200 m, *d* = 50 m, and wmin = 30. **(B)** Doppler shift curve f_d (*t*), assuming f_0 = 68 m/s, *v* = 123 m/s, *h* = 150 m, *d* = 20 m, and wmin = 30. **(C)** Doppler shift curve f_d (*t*), assuming f_0 = 68 m/s, *v* = 123 m/s, *h* = 150 m, *d* = 20 m, and wmin = 30.

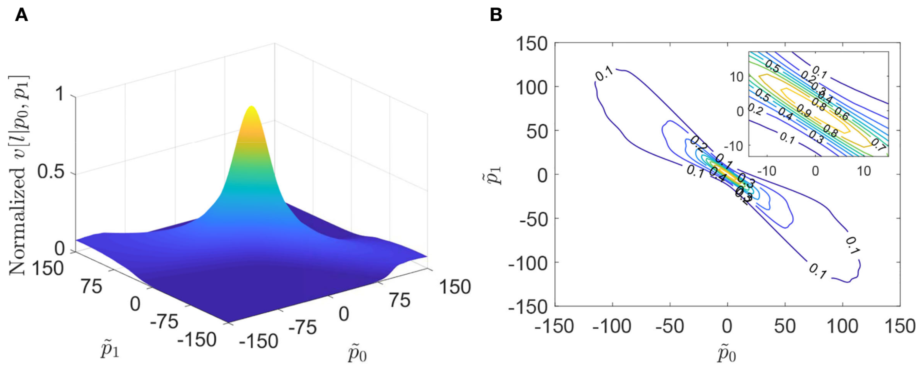


FIGURE 6
Accumulation power *v [l|p˜ n] versus p˜ n* with the two coefficients at *l* = 1,200. **(A)** Normalized surface and **(B)** contour map.

proposes a robust time–frequency analysis algorithm using a memory-dependent derivative.

3.1 The proposed algorithm

According to the theory of Ville, an analytic signal $x \in \mathbb{C}^L$ with the time-varying IF can be given by

$$x[l] = ae^{i\sum_{m=0}^{l}\varphi[m]}, \tag{20}$$

where $a \in \mathbb{R}_+$ and $\varphi[m]$ is the amplitude and IF, respectively, and $.i = \sqrt{-1}$. It should be noted that the FM signal may have a wide spectrum band since $\varphi[m]$ can vary in a wide range.

According to the Weierstrass approximation theorem, any continuous function in a closed and bounded interval can be uniformly approximated by a polynomial to any degree of accuracy. so $\varphi[m]$ can be expanded to U -order $(UłeqL{-}1)$ Taylor series at the moment $m{=}0$, *i.e.*,

$$\varphi[m] = \sum_{u=0}^{U} \frac{\varphi^{(u)}[0]}{u\,!}\left(\frac{m}{f_s}\right)^u + R_u\left[\left(\frac{m}{f_s}\right)^{u+1}\right], \tag{21}$$

where f_s is sample frequency, $\varphi^{(u)}[0]$ is the $u-$ th order difference coefficient about $\varphi[0]$, and $R_u(m)$ is the Maclaurin remainder. There exists $R_u(m){=}0$ when $U{=}L{-}1$. Additionally, the maximum order U is related to the approximation accuracy to FM function. Specifically, $U{=}0$ for the continuous wave and $U{=}1$ for the linear FM signal.

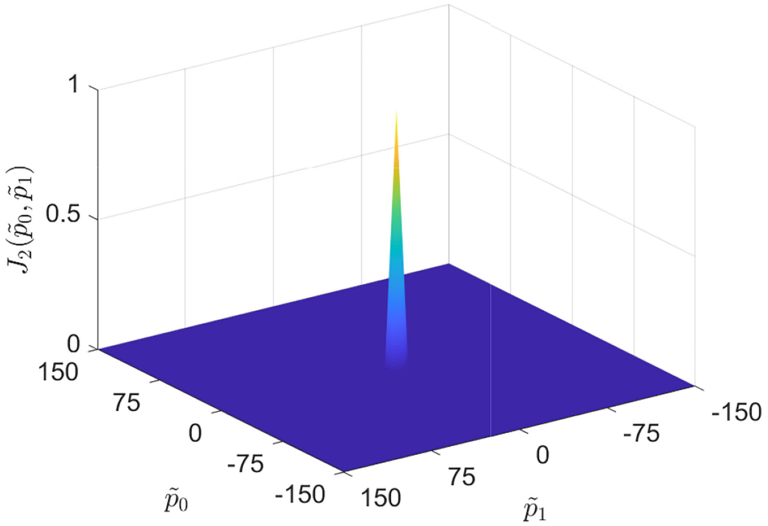


FIGURE 7
Normalized surface of J_2*[v(˜ pu)]* with *U = 1 versus* coefficients *p˜ u* ∈ [−150, 150].

Now, by inserting Eq. (21), omitting Maclaurin remainder $R_u(m)$, into Eq. (20), the reformulated formula is listed as follows:

$$x[l] = ae^{i\sum_{m=0}^{l}\sum_{u=0}^{U}p_u\left(\frac{m}{f_s}\right)^u}, \tag{22}$$

where $p_u = \frac{\varphi^{(u)}[0]}{u!}$ is the coefficient of the u -th power term. In Eq. (22), $e^{ip_0\frac{m}{f_s}}$ and $e^{i\sum_{m=0}^{l}\sum_{u=0}^{U}p_u(\frac{m}{f_s}ight)^u}$ is named as harmonic component and modulated element, respectively. Therefore, given a set of estimated parameters, a time–frequency estimation result is capable of producing a highly accurate approximation to a wide range of signals whose IF trajectories can be any continuous function of time. The goal of $\varphi[m]$ estimation in Eq. (20) is transformed to obtain p_u in Eq. (22) since IF can be recovered from its Taylor series.

In the signal estimation field, noise has always been considered as a serious cause of error. Although the most common noise can be modeled by well-known Gaussian distributions, some observed signals experience many types of i.i.d. noise with unknown distribution, for instance, underwater acoustic noise, low-frequency atmospheric noise, and other types of man-made noise (Liu et al., 2018). The signal x is corrupted by the i.i.d. noise with unknown distribution $z\in\mathbb{C}^L$; then, there is an observed signal y (Kozick and Sadler, 2000), *i.e.*,

$$y[l] = x[l] + z[l]\,. \tag{23}$$

Currently, there exists a class of i.i.d. noise, called symmetric α -stable ($S\alpha S$) noise, that can be used to model various types of real noise. The different noise statistical characteristics can be represented by changing parameters, where the characteristic exponent α controls the heaviness of the tails of the density function and the dispersion parameter γ determines the spread of the density around its location parameter. Thus, the probability density function of $S\alpha S$ gives a much better approximation to the i.i.d. noise with various distribution types.

These conventional methods can be considered as the maximum likelihood estimates resulting from corresponding minimization problems for Gaussian the noise environment. Thus, they are very sensitive to the i.i.d. noise with unknown distribution, and their robustness against a model mismatch is not ensured. In such cases, these methods under the Gaussian assumption, *e.g.*, parametric time–frequency analysis, may exhibit a serious performance degradation for the i.i.d. noise with arbitrary distribution. The choice of a more robust solution that can be used for various i.i.d. noises is thus salutary.

A fundamental issue discussed in this paper is the question of how to improve the resolution and robustness of the time–frequency estimation result in the i.i.d. noise with arbitrary distribution. The need for robustness and high resolution in signal analysis is self-evident, yet, at the present time, the traditional assessment of such properties remains largely *ad hoc*. The goal has been to develop a theory and techniques for a robust and high-resolution time–frequency estimation.

3.2 Power of phase difference accumulation

The phase difference is a common tool used in polynomial phase estimation, and it is performed by multiplying each moment with a kernel function element-wise (Vucijak and Saranovac, 2010). If FM is approximated by the U -th order Taylor series in Eq. (22) and the selected kernel function is also the U -th order Taylor series with the estimated $\hat{p}_u$,

$$h[l|\hat{p}_u] = e^{-i\sum_{m=0}^{l}\sum_{u=0}^{U}\hat{p}_u\left(\frac{m}{f_s}\right)^u}, \tag{24}$$

so the phase difference between x and h is given by

$$r[l|\tilde{p}_u] = x[l|p_u]h[l|\hat{p}_u] = ae^{ii\sum_{m=0}^{l}\sum_{u=0}^{U}\tilde{p}_u\left(\frac{m}{f_s}\right)^u}, \tag{25}$$

where $\tilde{p}_u = p_u - \hat{p}_u$ is the coefficient error, and $\phi[m|\tilde{p}_u] = \sum_{u=0}^{U}\tilde{p}_u(\frac{m}{f_s})^u$ is defined as the IF error. As p_u is a fixed unknown value, r has also been parameterized with $\tilde{p}_u$ to simplify the entire expression.

Coherence accumulation to phase difference can improve the SNR, which has obtained successful applications (Vucijak and Saranovac, 2010; Wang et al., 2017). For i.i.d. noise with arbitrary distribution, the averaging value of coherence accumulation will fluctuate slightly; however, the coherence accumulation of phase difference in Eq. (25) will keep increasing as time increases, especially when $\tilde{p}_u$ approximates to zero. Inspired by this fundamental rule, a new cost function is constructed to find the optimal solution for all coefficients $\tilde{p}_u$.

We define the accumulation power of phase difference accumulation (APP) by removing the average value as

$$v\,[l|\tilde{p}_u\,] = \frac{1}{l}\sum_{m=0}^{l}\{\,c[m|\tilde{p}_u] - c_M\,[m|\tilde{p}_u\,]\,\} \\ \{\,c[m|\tilde{p}_u\,] - c_M\,[m|\tilde{p}_u]\,\}!^{*}\,, \tag{26}$$

where * represents the complex conjugation, and

$$\begin{aligned} c\left[m|\tilde{p}_u\right] &= \sum_{n=0}^{m} \quad r\left[n|\tilde{p}_u\right], \\ c_M\left[m|\tilde{p}_u\right] &= \tfrac{1}{m} \quad \sum_{n=0}^{m} \quad c\left[n|\tilde{p}_u\right], \end{aligned} \tag{27}$$

are the phase accumulation of $r[m|\tilde{p}_u]$ and the average of $c[m|\tilde{p}_u]$ over the first $m\leq l$ sample numbers, respectively. Obviously, $v[l|\tilde{p}_u]$ is the accumulation power of phase accumulation $r[m|\tilde{p}_u]$ over the interval $[0,l]$.

THEOREM 1. The power $v[l|\tilde{p}_u]$ *versus* $|\tilde{p}_u|$ is the monotone function.

PROOF. We have provided the proof in the Appendix section. In Eq. (26), $\frac{dv[l|\tilde{p}_u]}{d\tilde{p}_u} \geq 0$ holds for $\tilde{p}_u \geq 0$ and $\frac{dv[l|\tilde{p}_u]}{d\tilde{p}_u} \leq 0$ holds for $\tilde{p}_u \leq 0$, so the maximal value of $v[l|\tilde{p}_u]$ will be obtained at $\tilde{p}_u = 0$.

An example will be exhibited with the sampling frequency f_s=1,200 H and the sample number L=1,200 , the order number U=1 , and the coefficient $\tilde{p}_u \in [-150, 150]$ for the power $v[l|\tilde{p}_u]$. Without loss of generality, the mesh plot of $v[l|\tilde{p}_u]$ *versus* coefficient $\tilde{p}_u$ at l=1,200 is demonstrated in Figure 6A, and the contour curve of $v[l|\tilde{p}_u]$ is shown in Figure 6B. It shows that $v[l|\tilde{p}_u]$ increases monotonically as coefficient $\tilde{p}_u$ tends to zero. Specifically, the maximum value of $v[l|\tilde{p}_u]$ lies at � $i\,ldep_u = 0$.

Theorems 1 and Figure 6 demonstrate that the power $v[l|\tilde{p}_u]$ at certain moment l is a monotonic function of $|\tilde{p}_u|$ and can obtain the maximum at $|\tilde{p}_u| = 0$ if the FM signal and h are coherent. Since $v[l|\tilde{p}_u]$ is a positive time series as l , different types of norm, like ℓ_1 -norm, ℓ_2 -norm, and ℓ_q -norm, can be used to measure the size of $v[l|\tilde{p}_u]$. From a statistical perspective, ℓ_2 -norm is suitable for Gaussian noise and ℓ_1 -norm is robust to outliers. Since $v[l|\tilde{p}_u]$ could be corrupted by $S\alpha S$ noise with impulses and the estimated IF can be determined by maximizing $v[l|\tilde{p}_u]$ at any moment, ℓ_1 - norm will be chosen. Therefore, the problem of coefficient $\tilde{p}_u$ estimation can be readily solved by ℓ_1 -norm maximization as follows:

$$\max_{\hat{p}_u} \| v[l|\tilde{p}_u] \|_1 \quad s.t.\ \tilde{p}_u = p_u - \hat{p}_u . \tag{28}$$

Moreover, Theorem 1 guarantees that the accurate coefficient. is available in the continuous domain, and the previous work in time–frequency estimation is fundamentally limited by the Heisenberg–Gabor inequality for a better resolution. The estimated IF can be approximated as $\hat{\phi}[m|\hat{p}_u] = \sum_{u=0}^{U} \hat{p}_u (\frac{m}{f_s})^u$.

3.3 IF estimation based on multiple log-sum

Under the condition of short window size L or high noise power, the estimated IF trajectory with ℓ_1 - norm shows poor accuracy and is easily interrupted by the accumulation of noises, although the maximal value in (28) is still obtained at the true p_u from the statistical perspective. In view of the nonlinear amplification for DOA estimation in Wang et al. (2008) and Wang and Kay (2010), the multiple logarithmic sum (MLS) is proposed to achieve the sharper peak for target IF and deeper null for non-target area, which also improves the accurate and robust performance for IF estimation.

Log-sum function has often been used to solve for maximization and is shown to be more convex and smoother than the ℓ_1 -norm (Zhang et al., 2019; Qiao et al., 2020), and this function can exhibit nonlinear amplification in some fixed intervals. Alternating convexity-promoting functions such as log-sum function can be used to replace the ℓ_1 -norm, and a surrogate function with a sharper peak for target frequency and deeper null for side lobes is achieved. Thus, we consider the use of the log-sum function to improve the convexity of the objective function. Replacing the ℓ_1 -norm in Eq. (28) with the multiple log-sum function over the entire sample number L will result in

$$\max_{\hat{p}_u} J_1 \quad (\tilde{p}_u) = \sum_{l=0}^{L-1} \log \{ v[l \,|\, \tilde{p}_u] + 1 \} \quad s.t.\ \tilde{p}_u = p_u - \hat{p}_u . \tag{29}$$

where “1” is added to ensure that $\log \{ v[l|\tilde{p}_u] + 1 \}$ is non-negative, and it does n’t change the location of the optimal solution. $J_1 \quad (\tilde{p}_u)$ *versus* $|\tilde{p}_u|$ is also monotonic since log is a monotonic function and $v[l|\tilde{p}_u]$ *versus* $|\tilde{p}_u|$ is a monotonic decreasing function as proved in Theorem 1. $J_1 \quad (\tilde{p}_u)$ in Eq. (29) is rewritten as follows:

$$J_1 \quad (\tilde{p}_u) = \log \coprod_{l=0}^{L-1} \{ \quad v[l \,|\, \tilde{p}_u] + 1 \quad \} = \log \nu(\tilde{p}_u), \tag{30}$$

so $J_1 \quad (\tilde{p}_u)$ can also be considered as the function of $\nu(\tilde{p}_u)$ and rewritten as $J_1[\nu(\tilde{p}_u)]$.

Hence, for $J_1[\nu(\tilde{p}_u)]$, the accumulation of the frequency domain is transformed into the integral of power in the logarithm space. Benefiting from the monotonicity of $\nu(\tilde{p}_u)$ *versus* $\nu(t|\tilde{p}_n)$as proved in Theorem 2, the sum of $J_1[\nu(\tilde{p}_u)]$ over $\nu(\tilde{p}_u)$ can significantly amplify the power $\nu(t|\tilde{p}_n)$ in a non-linear way. Without loss of generality, only the accumulation over the non-positive interval of $\tilde{p}_u$ is considered since $v[l|\tilde{p}_u]$ is an even function. According to Theorems 1 and 6, the power $v[l|\tilde{p}_u]$ tends to zero when $\tilde{p}_u \longrightarrow 0$. Thus, the lower integral limit can be obtained as follows:

$$\begin{aligned} &\lim_{\tilde{p}_u \to -\infty} \quad \log \nu(\tilde{p}_u) \\ &= \lim_{\tilde{p}_u \to -\infty} \log \coprod_{l=0}^{L-1} \{ v[l \,|\, \tilde{p}_u] + 1 \quad \} \quad = \quad C_0 . \end{aligned} \tag{31}$$

Consequently, the variable upper limit integral of $J_1 \quad [\nu(\tilde{p}_u)]$ over the continuous domain $\nu(\tilde{p}_u)$ is given by

$$\begin{aligned} J_2 \quad [\nu(\tilde{p}_u)] \quad &= \quad \int_{C_0}^{\nu(\tilde{p}_u)} \quad (\log \quad \kappa + 1) d \quad \kappa \\ &= \quad \nu(\tilde{p}_u) \log \nu(\tilde{p}_u) \quad + \quad C_1, \end{aligned} \tag{32}$$

where a fixed constant “1” added in the integral is added to obtain a simple analytical solution. Because the linear accumulation of $J_2 \quad [\nu(\tilde{p}_u)]$ with the constant term does not change the extreme value point, it does not affect the final solution coefficient $\tilde{p}_u$. Besides that, Eq. (32) with respect to $\nu(\tilde{p}_u)$ is convex according to Chapter 3 in Boyd et al. (2004).

THEOREM 2. $J_2 \quad [\nu(\tilde{p}_u)]$ *versus* $|\tilde{p}_u|$is monotone decreasing.

PROOF. According to the chain rule of derivation, the first-order derivative of $J_2[\nu(\tilde{p}_u)]$ about $|\tilde{p}_u|$ is given by

$$\frac{dJ_2[v(\tilde{p}_u)]}{d|\tilde{p}_u|} = \frac{dJ_2}{dv}\sum_{l=0}^{L-1}\left(\frac{dv}{dv[l]}\ \frac{dv[l|\tilde{p}_u]}{d|\tilde{p}_u|}\right).$$

Firstly, it follows from Theorem 1 that $v[l|\tilde{p}_u]$ is a monotonically decreasing function of $|\tilde{p}_u|$, and there exists $\frac{dv[l|\tilde{p}_u]}{d|\tilde{p}_u|} \leq 0$.

Secondly, the first-order derivative of the composite function $v\{v[l]\}$ about $v[l]$ from Eq. (30) is given by

$$\begin{aligned}\frac{dv\{v[l]\}}{dv[l]} &= \left\{\prod_{l=0}^{L-1} \{v[l|\tilde{p}_u] + 1\}\right\}^{\}}_{v[l]} \\ &= \sum_{l=0}^{L-1}\prod_{k} \{v[k|\tilde{p}_u] + 1\},\end{aligned}$$

where $k\in\{0,L-1\}\backslash\{l\}$. $\frac{dv\{v[l]\}}{dv[l]}$is non-negative since $v[l|\tilde{p}_u]$ is non-negative. Thus, it indicates that $v\ bracev[l]\}\ versus\ v[l]$ is monotonically increasing.

Thirdly, the first-order derivative of $J_2(v)$ is calculated as follows:

$$\frac{dJ_2[v]}{dv} = \frac{1}{v(\tilde{p}_u)} + 1,$$

which is positive since $v(\tilde{p}_u) \geq 0$. Thus, $J_2(v)$ is monotonically increasing.

To summarize the above-mentioned points, in the case of $\frac{d\phi}{dt} \lesseqgtr$, the first-order derivative of $J_2[v(\tilde{p}_u)]$ about $|\tilde{p}_u|$ is non-positive, *i.e.*,

$$\frac{dJ_2[v(\tilde{p}_u)]}{d|\tilde{p}_u|} \leq 0.$$

Therefore, it is proven.

After the nonlinear amplification with convexity-preserving operation and the accumulation in the frequency domain, $J_2[v(\tilde{p}_u)]\ versus\ \tilde{p}_u$ has a sharper peak and a deeper null level than $J_1[v(\tilde{p}_u)]$ in (29). $J_2[v(\tilde{p}_u)]$ with two coefficients is depicted in Figure 7, and the same simulation condition as in Figure 6 is kept. The deep null level of $J_2[v(\tilde{p}_u)]$ at the unmatched phase is beneficial to find the optimal coefficients and suppress the i.i.d. noise. Since $J_2[v(\tilde{p}_u)]$ is convex, any gradient algorithm can be used to solve this maximization problem.

In summary, the time–frequency estimation by the APP-MLS algorithm can be converted into the maximization problem as follows:

$$\begin{aligned}\max_{\hat{p}_u} J_3(\tilde{p}_u) &= v(\tilde{p}_u)\log v(\tilde{p}_u)\ s.t.\ \tilde{p}_u \\ &= p_u - \hat{p}_u. \end{aligned} \tag{33}$$

We propose the APP-MLS algorithm to find the optimal solution for (37), and the basic idea of the IF estimation by the APP-MLS algorithm is to determine the coefficients $\hat{p}_u$ of polynomial kernel to estimate the time–frequency trajectory. The details will be introduced in the next section. Due to the convexity of $J_3(\tilde{p}_u)$, any gradient algorithm can be used to solve this maximization problem. In this paper, the optimization solution $\hat{p}_n$ is obtained by maximizing Eq. (33) by the AdaBelief method, which is more efficient for the parameter tuning process. The intuition for AdaBelief is to adapt the step size according to the "belief" in the current gradient direction. This method uses the exponential moving average of the noisy gradient as the prediction of the gradient at the next time step; if the observed gradient greatly deviates from the prediction, we distrust the current observation and take a small step; if the observed gradient is close to the prediction, we trust it and take a large step. After localizing the $\hat{p}_u$ in one time interval, the estimated time–frequency trajectory $\hat{\varphi}[m|\hat{p}_u] = \sum_{u=0}^{U}\hat{p}_u(\frac{m}{f_s})^u$ can be obtained.

In practical application, the whole observed data will be partitioned into many overlapped slides with the window size τ and the overlap size S between two close windows; then, the number K of slides can be obtained. The polynomial order U and the initial coefficient $\hat{p}_u$ are given according to the rough character of the signal. In every slide, therein included are the following steps: firstly, the phase difference between the truncated signal and the kernel function is calculated, and its accumulated power can be obtained. Furthermore, by maximizing the multiple log-sum function with the AdaBelief algorithm, the estimated coefficient $\hat{p}_u$ can be obtained after several iterations. Finally, the estimated IF trajectory can be obtained within the overlap size S . The detailed procedure is illustrated in Figure 8.

During the application of the APP-MLS algorithm, the estimation accuracy of the IF trajectory depends on the order U , the window size W , and the overlap size S ($0 \leq S < W$). Specifically, as the order U increases, the Taylor series can approximate the estimated frequency function $\varphi[m]$ with a high accuracy within a fixed window size W . The longer window size W leads to larger deviations from the true IF trajectory at a fixed order U , especially at the window end. The window size W also affects the robustness of the algorithm, *i.e.*, the longer the window size, the better robustness to suppress noise; it should not be too short in practical application. Additionally, the higher-order U and longer window size W will increase the computation complexity, and their choice should be considered carefully.

4 Example

The 120-s aircraft noise data are used to validate the efficiency of the estimation model of flight parameters. These data were measured in an experiment where an aircraft traveled along a straight path at a constant speed. The speed of sound in air and seawater was 340 and 1,520 m/s, respectively. A hydrophone is placed at a depth of 90 m in the ocean, and its output is sampled at 44.1 kHz for 2 min. To reduce the computational burden, the data herein are downsampled to 1,200 Hz.

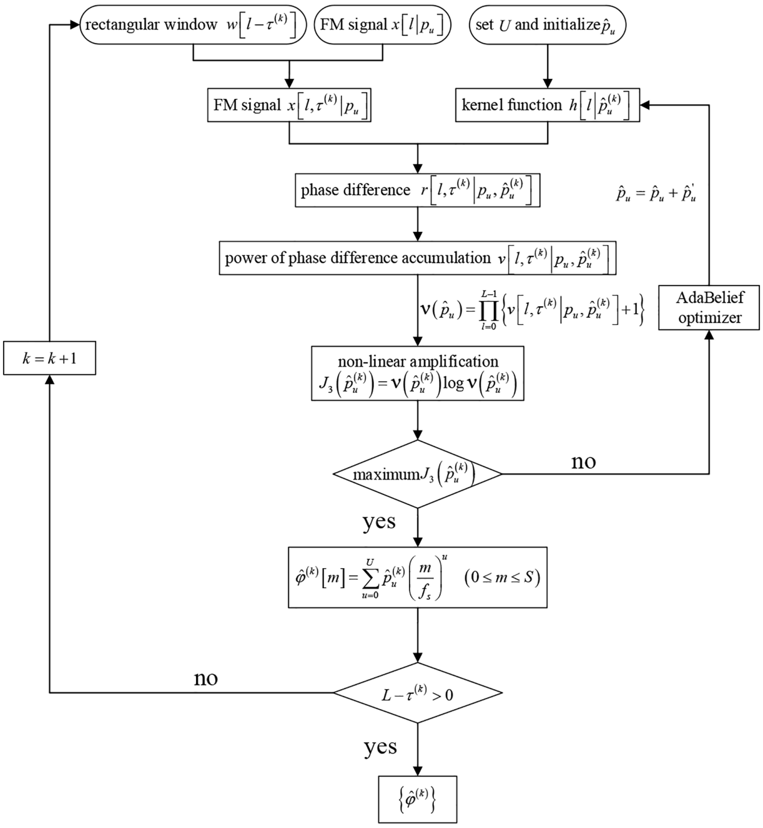


FIGURE 8
Time–frequency estimation based on APP-MLS.

The Acoustical Society of America provided a piece of audio data[1], where a hydrophone is placed at a depth of 90m in the ocean and records the underwater acoustic data from a turboprop aircraft with a constant frequency of 68 Hz flying straightly over the hydrophone at a fixed altitude. The sampling rate is 44.1 kHz, and the recorded time is 120 s, as shown in Figure 9. As the aircraft approximates and leaves the overhead of the hydrophone, the relative Doppler of the received acoustic signal has an apparent and quick change, and the IF trajectory can determine the flying parameters of the aircraft. The detailed analysis and principle can be referred to Urick R J (Urick et al, 1972) and the accurate estimation of the IF trajectory is very important. According to Tre DiPassio's 2019 Challenge Problem Solution, the original propeller frequency is estimated when the plane is directly overhead the hydrophone at a timestamp denoted as 54 s. To better show the estimated IF results for radiated noise, we intercepted the data from 49 to 59 s when the aircraft is nearly over the hydrophone.

The maximum frequency bin of each frame is used to estimate the fundamental frequency present in that frame. Figure 10 shows these estimates as well as the amplitude computed in each frame. The majority of the audio solely contain underwater noise. However, at around 54 s, a few seconds of a plane flying overhead can be heard when listening to the file. The presence of the plane is shown in the highlighted section of Figure 10. There is a spike in the amplitude above the noise floor, paired with a more defined instantaneous frequency track. Figure 11 shows only the frames where the plane flying overhead is audible commensurate to the amplitude spike in Figure 10.

In order to extract instantaneous frequencies from the provided file, a short-time Fourier transform (STFT) is utilized in Figure 11. The audio is split into frames with a length of 8,192 samples per frame, windowed with an 8,192-point Hamming window. Each frame is converted to the frequency domain using the fast Fourier transform (FFT) algorithm run on the frame data. In total, 75% overlap is used between the frames, and zero padding is added such that the frequency resolution (bin size) of the FFT in each frame is approximately 0.1 Hz. Consistent with the Nyquist Theorem, the maximal frequency present in the

---

[1] https://acousticstoday.org/iscpasp2019

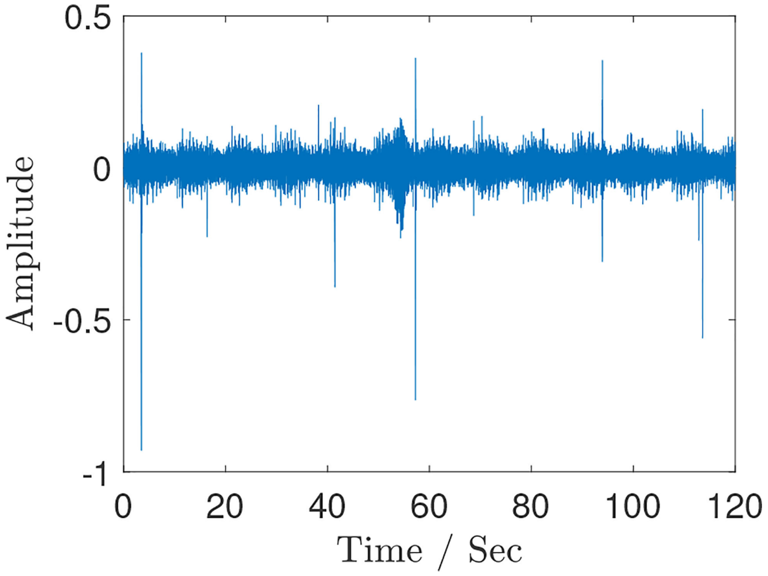


FIGURE 9
Helicopter waveform in each frame of the given audio file.

given audio is half the sampling frequency or 22,050 Hz. Therefore, the full-scale 22,050 Hz will be represented in the number of frequency bins equal to half of the FFT length due to the redundancy added when computing the FFT. The following equation thus shows how the desired resolution is achieved:

$$\text{Resolution} = \frac{f_s}{\text{FFTLEN}} \tag{34}$$

Solving Eq. (38) for the desired resolution of around 0.1 Hz in each frame shows that 53-times zero padding should be used.

The estimates from each frame in the region with audible plane motion are smoothed to create an instantaneous frequency track. This smoothing is done with a Savitzky–Golay filter. This type of filter is used to smooth sets of data points in the time domain using a least-squares fitting approach.

The estimated IF trajectory by the APP-MLS will be compared with STFT, PCT, and SST. Because of the high data sampling rate and the frequency of interest being located in the low-frequency range, the data were downsampled by 200 times. In order to estimate the IF better and compare fairly, some parameters of different methods will be set. The window size is set to W=128 , and the overlap rate is set as 75%. Moreover, the iteration number of PCT is seven, the synchronous squeeze factor of SST is 200, and the maximum order of APP-MLS is U=2 . Besides this, in Tre

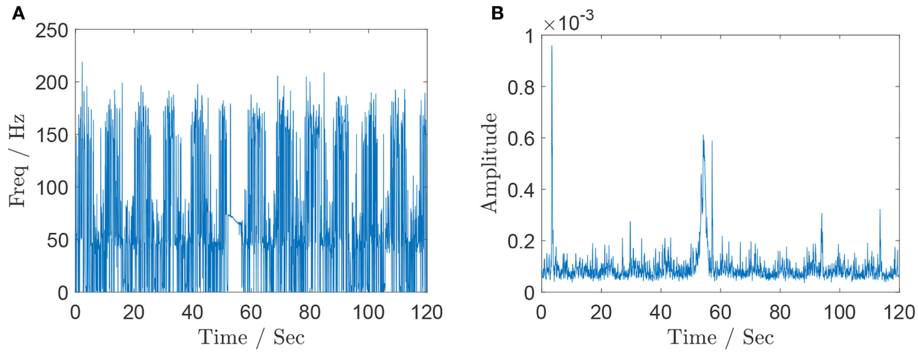


FIGURE 10
Instantaneous frequency estimation results for the aircraft acoustic data recorded by hydrophone with **(A)** the estimated fundamental frequency in each frame of the given audio file and **(B)** the estimated amplitude in each frame of the given audio file.

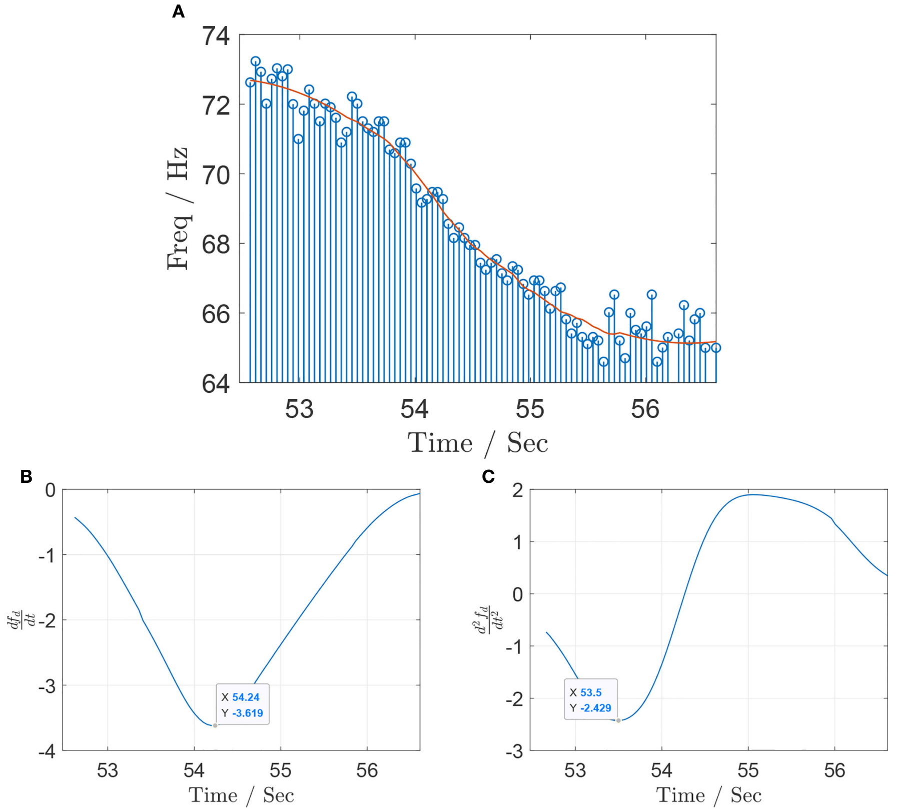


FIGURE 11
Instantaneous frequency estimation results for the aircraft acoustic data recorded by hydrophone with **(A)** raw and smoothed frequency track at the hydrophone position. **(B)** First-order derivative of the Doppler shift curve $f_d\ (t)$. **(C)** Second-order derivative of the Doppler shift curve $f_d\ (t)$.

DiPassio's 2019 Challenge Problem Solution[2], the author gives details about how to extract data from provided audio, which is given as follows: firstly, the audio is split into frames with a length of 128 samples per frame, windowed with a 128-point Hamming window. Secondly, each frame is converted to the frequency domain using the fast Fourier transform algorithm run on the frame data. Then, 75% overlap is used between the frames, and 16 times more zero padding is added such that the frequency resolution of these methods in each frame is approximately 0.1 Hz. Thirdly, the maximum frequency bin of each frame is used to estimate the fundamental frequency present in that frame. Finally, this smoothing is done with a Savitzky–Golay filter to smooth the IF curve.

Figure 12 presents the estimated results of the four methods. Although the changing trend of IF estimation by STFT can be observed as shown in Figure 12A, there exist severe fluctuations outside the interval 53–57 s when the aircraft deviates far from the head of hydrophone. Figures 12B, C show that the interference within the frequency range of interest by PCT and SST has been improved because the adaptive kernel function and the synchrosqueezing transform are used respectively. However, they are also distorted by the background noise with unknown probability distribution, and their estimated IF information is still not available effectively when the aircraft is far from the hydrophone. Benefitting from the gridless principle and nonlinear amplification, the IF trajectory yielded by APP-MLS can effectively suppress underwater acoustic noise even for low SNR. The IF result within the range of interest shown in Figure 12D is more accurate than the other three methods. The APP-MLS algorithm shows the robust IF estimation performance in the lower SNR environment, and its result is much closer to the true IF than those of the other methods.

The experimental frequency values from Figure 11 range from 64.17 to 73.45 Hz, for a total range of 9.28 Hz. Substituting these values into Eq. (15) will yield as follows:

$$\begin{cases} f_0 & = 2\frac{f_d(-\infty)f_d(+\infty)}{f_d(-\infty)+f_d(+\infty)} \approx 2\times\frac{73.45\times 64.17}{73.45+64.17} = 68.5 \quad \text{Hz} \\ v = c_w & \frac{f_d(-\infty)-f_d(+\infty)}{f_d(-\infty)+f_d(+\infty)} \approx 1520\times\frac{73.45-64.17}{73.45+64.17} = 101 \quad \text{m/s} \end{cases} \tag{35}$$

so the aircraft frequency is 68.5 Hz, and the speed of the aircraft is 101 m/s.

---

[2] https://www.tredipassio.com/asa2019?rq=asa

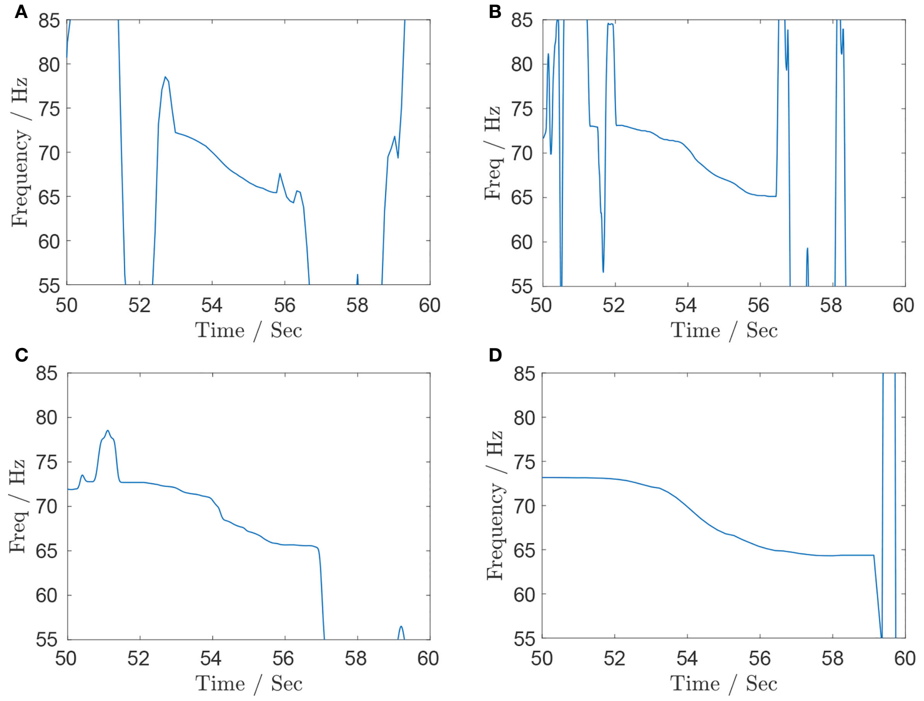


FIGURE 12
Instantaneous frequency estimation results for the aircraft acoustic data recorded by hydrophone with **(A)** STFT, **(B)** PCT, **(C)** SST, and **(D)** APP-MLS.

According to the first- and second-order derivative of $f_d(t)$, the rest of the two flight parameters $\{h,w_{\min}\}$ are obtained by solving a set of equations as follows:

$$\frac{df_d(t)}{dt} = f_d'(f_0, v, h, w_{\min}) = -3.619, \tag{36}$$

and

$$\frac{d^2 f_d(t)}{dt^2} = f_d^{\}}(f_0, v, h, w_{\min}) = -2.429. \tag{37}$$

Then, we obtain h=150.1m and $w_{\min}$ =1.4m .

5 Conclusion

This paper proposes the use of both time–frequency trajectory and three-dimensional space model to estimate the flight parameters, including $\{f_0,v,\alpha(t),\beta(t)\}$. This solution explores a computational approach using the geometric relationships of three-dimensional space and using the result of time–frequency estimation to create a predictive model for the instantaneous frequency observed at the hydrophone position by using Doppler effect calculations. The accuracy of this model will be evaluated *via* comparison with given experimental data. The model will then be used to extract the propeller frequency, speed, height, and slant of the aircraft source for a given hydrophone recording.

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

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding

This Project are supported by the Key Laboratory of Intelligent Image Processing and Analysis(Wenzhou, China, No.2021 HZSY0071) and the Key Project of Zhejiang Provincial Natural Science Foundation under Grant (No.LZ20F020004).

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

Aksenovich, T. V. (2020). "Comparison of the use of wavelet transform and short-time fourier transform for the study of geomagnetically induced current in the autotransformer neutral," in *2020 international multi-conference on industrial engineering and modern technologies* (FarEastCon). pp. 1–5. doi: 10.1109/FarEastCon50210.2020.9271210

Aygun, H., Kirmizi, M., and Turan, O. (2022). Propeller effects on energy, exergy and sustainability parameters of a small turboprop engine. *Energy* 249, 123759. doi: 10.1016/j.energy.2022.123759

Boashash, B. (2015). *Time-frequency signal analysis and processing: A comprehensive reference* (Massachussetts, Washington, DC: Academic Press).

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization* (University Printing House, Cambridge: Cambridge university press).

Chirico, G., George, N.Barakos and Nicholas Bown. (2018). Propeller installation effects on turboprop aircraft acoustics. 424, 238–62. doi:10.1016/j.jsv.2018.03.003

Fang, J., Wang, F., Shen, Y., Li, H., and Blum, R. S. (2016). Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach. *IEEE Trans. Signal Process.* 64, 4649–4662. doi: 10.1109/TSP.2016.2572041

Kozick, R. J., and Sadler, B. M. (2000). Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures. *IEEE Trans. Signal Process.* doi: 10.1109/78.887045

Kumar, A., and Prasad, A. (2022). Wigner-ville distribution function in the framework of linear canonical transform. *J. Pseudo Differential Operators Appl.* 13, 1–17. doi: 10.1007/s11868-022-00471-w

Liu, S., Zhang, Y. D., and Shan, T. (2018). Detection of weak astronomical signals with frequency-hopping interference suppression. *Digital Signal Process.* 72, 1–8. doi: 10.1016/j.dsp.2017.09.003

Qiao, C., Shi, Y., Diao, Y. X., Calhoun, V. D., and Wang, Y. P. (2020). Log-sum enhanced sparse deep neural network. *Neurocomputing* 407, 206–220. doi: 10.1016/j.neucom.2020.04.118

Urick, R. J. (1972). Noise Signature of an Aircraft in Level Flight over a Hydrophone in the Sea[J]. *J. Acoust. Soc. Am.* 52 (1A), 993–999.

Vandana, A. R., and Arokiasamy, M. (2022). Comparison of radar micro doppler signature analysis using short time fourier transform and discrete wavelet packet transform. doi: 10.1007/978-981-16-9012-9_35

Vucijak, N. M., and Saranovac, L. V. (2010). A simple algorithm for the estimation of phase difference between two sinusoidal voltages. *IEEE Trans. Instrum. Meas.* 59, 3152–3158. doi: 10.1109/TIM.2010.2047155

Wang, H., and Kay, S. (2010). Maximum likelihood angle-doppler estimator using importance sampling. *IEEE Trans. Aerospace Electronic Syst.* 46, 610–622. doi: 10.1109/TAES.2010.5461644

Wang, H., Kay, S., and Saha, S. (2008). An importance sampling maximum likelihood direction of arrival estimator. *IEEE Trans. Signal Process.* 56, 5082–5092. doi: 10.1109/TSP.2008.928504

Wang, P., Orlik, P. V., Sadamoto, K., Tsujita, W., and Gini, F. (2017). Parameter estimation of hybrid sinusoidal fm-polynomial phase signal. *IEEE Signal Process. Lett.* 24, 66–70. doi: 10.1109/LSP.2016.2638436

Zcza, B., Szq, A., Xian, J. A., Pyh, A., Xys, A., and Ayw, A. (2021). Linear canonical wigner distribution of noisy lfm signals *via* variance-snr based inequalities system analysis. *Optik* 237, 166712. doi: 10.1016/j.ijleo.2021.166712

Zhang, W., Fu, L., and Liu, Q. (2019). Nonconvex log-sum function-based majorization-minimization framework for seismic data reconstruction. *IEEE Geosci. Remote Sens. Lett. PP* 1–5, 1776–1780. doi: 10.1109/LGRS.2019.2909776

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., and Ying, W. (2022). Wide-bandwidth signal-based multireceiver sas imagery using extended chirp scaling algorithm. *IET Radar Sonar Navigation*. doi: 10.1049/rsn2.12200

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with class b model. *IET Radar Sonar Navigation*. doi: 10.1049/iet-rsn.2019.0477

Zhou, C., Zhang, G., Yang, Z., and Zhou, J. (2020). A novel image registration algorithm using wavelet transform and matrix-multiply discrete fourier transform. *IEEE Geosci. Remote Sens. Lett. PP* 1–5, 8002605. doi: 10.1109/LGRS.2020.3031335

Appendix

$$\frac{df_d}{dt} = \frac{f_0 c_w v^2}{\sqrt{v^2t^2+w_{min}^2}} \left\{ \frac{d\beta_T(t)}{dt} t\left(v^2t^2 + w_{\min}^2\right) \sin\ \beta_T \right. \tag{38}$$

$$\left. -\left(2v^2t^2 + w_{\min}^2\right) \cos\ \beta_T(t) + t \right\}\ .$$

$$\frac{d\beta}{dt} = (A_1 + A_2 - A_3 + A_4A_5 + A_3A_6A_7^{-1} - A_8A_9 - A_{10})^{-1}A_{11}\ . \tag{39}$$

where

$$\begin{aligned}
&A_1 = \cos\ \beta(t)\sqrt{1 - n^2\cos^2\beta(t)}\,,\\
&A_2 = n^2\left[1 - n^2\cos^2\beta(t)\right]^{-1/2}\sin^2\beta(t)\cos\ \beta(t)\,,\\
&A_3 = 2n\sin\ \beta(t)\cos\ \beta(t)\,,\\
&A_4 = \cos\ \beta(t)\sqrt{1 - n^2\cos^2\beta(t)}/(2dh)\,,\\
&A_5 = d^2/\left(1 - n^2\cos^2\beta(t)\right) + h^2/\sin^2\beta(t)\\
&-[\left(d+h\right)^2 + v^2t^2 + w_{\min}^2\,]\,, A_6 = d^2/\left(1 - n^2\cos^2\beta(t)\right) + h^2/\sin^2\beta(t)\\
&-\left[(d+h)^2+v^2t^2 + w_{\min}^2\,\right], \quad A_7 = 2dh\sqrt{1 - n^2\cos^2\beta(t)}\,, A_8 =\\
&\qquad \sin\ \beta(t)\sqrt{1 - n^2\cos^2\beta(t)}/(2dh)\,,\\
&A_9 = 2d^2h^2\sin^2\beta(t)\cos\ \beta(t)/\left(1 - n^2\cos^2\beta(t)\right)^2,\\
&A_{10} = h\sqrt{1 - n^2\cos^2\beta(t)}/[d\sin\ \beta(t)]\,,\\
&A_{11} = v^2t\sin\ \beta(t)\sqrt{1 - n^2\cos^2\beta(t)}/(2dh)\,,
\end{aligned} \tag{40}$$

$$\frac{d^2f_d}{dt^2} = \frac{c_wf_0v^2(G_1 + G_2 + G_3 - G_4)}{2G_5}\ . \tag{41}$$

$$\begin{aligned}
&G_1 = -2t\frac{d^2\beta}{dt^2}\sin\ \beta(t)(v^2t^2 + w_{\min}^2\,)^2\left[v^2t\cos\ \beta(t) - 1\right]\\
&G_2 = t\left(\frac{d^2\beta}{dt^2}\right)^2\left(v^2t^2 + w_{\min}^2\,\right)^2\left\{v^2t[\cos\ 2\beta(t) - 3] + 2\cos\ \beta(t)\right\}\\
&G_3 = 4\frac{d^2\beta}{dt^2}\sin\ \beta(t)\left(v^2t^2 + w_{\min}^2\,\right)\\
&\left(v^2tw_{\min}^2\ \cos\ \beta(t) + 2v^2t^2 + w_{\min}^2\,\right)\\
&G_4 = 4v^4t^3\cos\ \beta(t) + 3v^4t^2w_{\min}^2\ \cos\ 2\beta(t)\\
&+2v^2w_{\min}^4\ \cos\ 2\beta(t) + 3v^4t^2w_{\min}^4\ + 2w_{\min}^2\,G_5 =\\
&\qquad \left(v^2t^2 + w_{\min}^2\,\right)^{3/2}+\left[v^2t\cos\ \beta(t) - 1\right]^3
\end{aligned} \tag{42}$$

$$\frac{d^2\beta}{dt^2} = \left[\cos\ \beta(t)(B_{12} + B_{13})(B_{14} + nB_{15} - 2B_{16})\right]^{-1}$$

$$8\beta(t)\left\{B_1 - \beta(t)\begin{bmatrix} B_2 - B_3 + n(B_4 + B_5 + B_6) \\ \qquad + 2B_7B_8 - B_9 + B_{10} - B_{11}\end{bmatrix}\right\}\ . \tag{43}$$

Where

$$\begin{aligned}
&B_1 = 2v^2t\cos\ \beta(t)\left[n^2\cos\ 2\beta(t) - 1\right]\left[n^2\cos\ 2\beta(t) + n^2 - 2\right]^2/\\
&\qquad \left[\sin^2\beta(t)\right]^2,\\
&B_2 = n^6\cos^6\beta(t)\sin^{-1}\beta(t)\\
&\qquad \left\{\begin{matrix}[\cos\ 4\beta(t) - 4\cos\ 2\beta(t)]\left(d^2 + h^2 + v^2t^2 + w_{\min}^2\,\right) + \\ 3d^2 + 7h^2 + 3v^2t^2 + 3 + w_{\min}^2\end{matrix}\right\},\\
&B_3 = 2d^2n^2\left(11n^2 - 7\right)\sin^{-1}\beta(t)\,,\\
&B_4 = 5n^3\sin\ \beta(t)\left(d^2 + h^2 + v^2t^2 + w_{\min}\,\right),\\
&B_5 = n\sin\ \beta(t)\begin{bmatrix}9d^2\left(3n^2 - 2\right) + h^2\left(27n^2 - 14\right) + \\ \left(23n^2 - 14\right)\left(v^2t^2 + w_{\min}^2\,\right)\end{bmatrix},\\
&B_6 = 4dh\left(n^2 - 2\right)\sqrt{1 - n^2\cos^2\beta(t)}\,,\\
&B_7 = n^4\cos\ 4\beta(t) + 2n^2\left(n^2 - 2\right)\cos\ 2\beta(t) + n^4 - 2n^2 + 4\,,\\
&B_8 = 2dhn^3\cos^2\beta(t)\sin^{-2}\beta(t)\sqrt{1 - n^2\cos^2\beta(t)}\,,\\
&B_9 = 2dhn^3\sin^{-2}\beta(t)[\cos\ 4\beta(t) + 1]\sqrt{1 - n^2\cos^2\beta(t)}\,,\\
&B_{10} = 26h^2n^4\sin^{-3}\beta(t)\,,\\
&B_{11} = h^2\left[\left(1 - 8n^2\right)\cos\ 2\beta(t) + 12n^4 - 4n^2 + 3\right],\\
&B_{12} = \left[n^2\cos\ 2\beta(t)\sin^2\beta(t)\right]/2 + n^2 - 2\,,\\
&B_{13} = v^2\sin^3\beta(t)\left[n^2\cos\ 2\beta(t) + n^2 - 2\right],\\
&B_{14} = \cos\ 2\beta(t)[-d^2n^2\left(n^2 - 16\right) + h^2\left(-9n^4 + 16n^2 + 8\right)\\
&\qquad \left(v^2t^2 + w_{\min}^2\,\right)]\,,\\
&B_{15} = \left[n^3\cos\ 6(t) - 6n\cos\ 4\beta(t)\right]\left(d^2 + h^2 + v^2t^2 + w_{\min}^2\,\right) -\\
&\qquad 16\sqrt{2}dh\sin^3\beta(t)\left[-n^2\cos\ 2\beta(t) - n^2 + 2\right]^{3/2},\\
&B_{16} = 5d^2n^2 + h^2\left(4n^4 - 11n^2 + 12\right) + \left(n^2 + 4\right)\left(v^2t^2 + w_{\min}^2\,\right),
\end{aligned} \tag{44}$$

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fmars.2022.1015836

Spatial-digital joint self-interference cancellation method for in-band full-duplex underwater acoustic communication

Yinheng Lu[1,2,3], Xin Qing[1,2,3]*, Chenlu Yang[1,2,3], Yunjiang Zhao[4], Songwen Wu[1,2,3] and Gang Qiao[1,2,3]

[1]Acoustic Science and Technology Laboratory, 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]Yichang Testing Technique Research Institute, Yichang, China

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Zeyad Qasem,
Peking University, China
Xiaojun Mei,
Shanghai Maritime University, China
Hamada Esmaiel,
Aswan University, Egypt

*CORRESPONDENCE
Xin Qing
qingxin@hrbeu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 10 August 2022
ACCEPTED 14 September 2022
PUBLISHED 21 October 2022

CITATION
Lu Y, Qing X, Yang C, Zhao Y, Wu S and Qiao G (2022) Spatial-digital joint self-interference cancellation method for in-band full-duplex underwater acoustic communication.
Front. Mar. Sci. 9:1015836.
doi: 10.3389/fmars.2022.1015836

The in-band full-duplex underwater acoustic communication (IBFD-UWAC) mode has twice the information throughput of the traditional half-duplex communication mode, significantly increasing the communication efficiency. Extracting the weak desired signal from the high-power self-interference signal without distortion remains a challenging problem in implementing IBFD-UWAC systems. This paper proposes a spatial-digital joint self-interference cancellation (SDSIC) method for IBFD-UWAC. We first perform spatial self-interference cancellation (SSIC) and propose an improved wideband constant-beamwidth beamformer to overcome the problem of direction- and array-dependent interference in IBFD-UWAC systems. Convex optimization is used to maintain a constant beam response in the main flap and cancel the self-interference signal from a fixed direction, thus increasing the signal-to-interference ratio of the desired signal. Subsequently, we perform digital self-interference cancellation (DSIC) on the residual self-interference signal, and propose a variable-step-size least-mean-squares algorithm based on the spatial noise threshold. This algorithm modifies the least-mean-squares step-size adjustment criterion according to the noise level after SSIC and the desired signal, resulting in better DSIC. A series of simulations are implemented in a hardware-in-the-loop platform to verify the practicality and real-time performance of the proposed SDSIC method. The results show that the self-interference signal power can be reduced by 41.5 dB using the proposed method, an improvement of 13.5 dB over the conventional SIC method.

KEYWORDS

underwater acoustic, full duplex communication, self-interference (SI) cancellation (SIC), beamforming, VSS-LMS algorithm

1 Introduction

Recently, considerable attention has focused on underwater communication. Although there are numerous communication network protocols, mature land-based wireless network protocols often fail in underwater communication networks due to the long transmission delays, limited bandwidth, and the large spatial and temporal variability of the underwater acoustic channel (Yin, 2016; Song et al., 2019; Zhao et al., 2021). The scarce spectrum resources of the underwater acoustic channel also limit the frequency efficiency of underwater communication networks (Pan et al., 2012; Yan et al., 2021). Therefore, improving the spectrum efficiency and system information throughput of underwater acoustic communication (UWAC) networks under severe bandwidth limitations is an urgent problem. In-band full-duplex (IBFD) UWAC technology aims to achieve simultaneous transmission and reception of signals in the same frequency band from both sides of the communication. Compared with the traditional half-duplex mode, this increases the frequency efficiency by a factor of two, and increases the security of the communication system (Qiao et al., 2018a; Kolodziej et al., 2019). In engineering applications, IBFD-UWAC systems are typically used as relay nodes that receive data from one party while forwarding it to another party when the far-end desired signal is buried in high-power self-interference (SI) and communication cannot be achieved. To realize IBFD-UWAC, it is necessary to cancel the local high-power near-end SI signals (Sabharwal et al., 2014; Shen et al., 2020) and increase the signal-to-interference noise ratio (SINR) of the weak desired far-end signals to enable real-time communication between the two sides (Zhang et al., 2016; Zhao, 2021). Therefore, self-interference cancellation (SIC) is a critical technical issue. Spatial SIC (SSIC) is the first stage of the SIC process in IBFD-UWAC systems. The SSIC process does not attempt to perfectly cancel SI, but instead seeks to reduce SI by a sufficient amount to prevent the receiver's dynamic range from being swamped. After the SSIC process, a residual interference signal (RIS) remains. Digital SIC (DSIC) is needed to cancel the RIS (Everett et al., 2016; Satyanarayana et al., 2019).

For the SSIC method in full-duplex communication, a two-transmitter, three-receiver antenna system has been designed to provide 50 dB additional isolation, compared with a single-transmitter, single-receiver system, by adjusting the amplitude and phase of the transmitting and receiving antennas (Snow et al., 2011). Additionally, a compact switched-array reconfigurable antenna array has achieved a 65 dB passive SIC based on 10-cm transmitter and receiver antenna isolation (Ahmed and Eltawil, 2015). Vallese et al. (2017) proposed a joint transceiver–receiver beamforming algorithm for co-time, co-frequency full-duplex wireless communication systems, with SIC effects of 65.5 dB and 89.8 dB at transmitter signal-to-noise ratios (SNRs) of 30.0 dB and 60.0 dB, respectively. Everett et al. (2014) terminated the passive SIC effects in the radio carrier frequency range of 2.4–2.48 GHz at different distances. They obtained 27.9 dB and 24.5 dB interference cancellation effects at 50 cm and 35 cm, respectively, using an omnidirectional transmitting antenna. An additional 17 dB interference cancellation effect with a 90°-beamwidth transmitting antenna and 50 cm between the receiving and transmitting antennas was achieved by combining the directional antenna, absorbing medium, and antenna polarization. This resulted in interference cancellation of more than 70 dB and showed that passive cancellation performance is affected by the number of reflected paths (Everett et al., 2014). Choi et al. (2010) placed two transmit antennas half a wavelength away from a single receive antenna to achieve mutual cancellation. They conducted experiments at 2.4 GHz using a 7-inch antenna, and found that this configuration provided 26 dB of interference cancellation (5 MHz). The combination of noise cancellation and DSIC produced an overall SIC of 60 dB, but the results were not satisfactory for signal bandwidths over 100 MHz (Choi et al., 2010). Nawaz and Tekin (2016) designed and implemented a 2.4-GHz bipolar micro-bandwidth patch antenna, and evaluated its isolation performance. The results showed that the antenna could achieve 40 dB of interference cancellation. Cacciola et al. (2016) completely embedded the transmitting and receiving arrays in a metallic cavity and mounted them flush on the ground to reduce the interaction with nearby objects and antennas. This method obtained interference cancellation of more than 49 dB. Makar et al. (2017) proposed a toroidal hybrid T-shaped structure with a length of 35.6 mm that obtained an SIC effect of at least 35 dB in a multipath environment with near-omnidirectional transmitting and receiving antennas.

However, the above SSIC methods have been developed in the field of radio communication. In the engineering implementation of underwater communication, the equipment used for SSIC, such as acoustic baffles, has different apertures corresponding to different absorption frequencies, which attenuate the sound power by reflecting the sound within them. Qiao et al. (2013) used the vector hydrophone zero-point offset feature to reduce the received SI signal power. The full-duplex UWAC system developed by Li et al. (2015) uses a transmitting transducer with directivity (Zhao et al., 2021) to obtain an interference cancellation effect of about 25 dB. While acoustic baffles are inconvenient in engineering applications because of their weight and volume, directivity transducers are hardware-dependent. frequencies and attenuate the sound power by reflecting sound in them. They have limited cancellation capability and less flexibility, so the cancellation of local SI by array signal processing technology in wideband communication systems is worthy of study, but no SSIC scheme based on array signal processing has yet been proposed for

IBFD-UWAC systems. This paper proposes an improved constant beamwidth beamforming (ICBBF) algorithm.

For DSIC, Qiao et al. (2018b) proposed a maximum-likelihood algorithm with sparse constraints for estimating the sparse SI channels, achieving an SIC effect of 43 dB. This is a better cancellation effect than the least-squares algorithm. In the same year, Qiao et al. (2018a) considered the high probability of non-crossover regions for the transceiver signals during normal IBFD-UWAC, and proposed a DSIC algorithm for asynchronous full-duplex communication systems. This algorithm used non-crossover regions to collect the nonlinear distortion components combined with a sparse adaptive channel estimation algorithm for SIC. Their experimental results show that the method can effectively eliminate the nonlinear distortion caused by the power amplifier (PA). Shen et al. (2020) used the PA output channel signal as a reference and applied the dichotomous coordinate descent recursive least-squares (DCD-RLS) algorithm (Zakharov et al., 2008), which has a convergence rate similar to that of classical RLS. The complexity can be effectively reduced with the same numerical stability. Experiments show that the DCD-RLS algorithm achieves 46 dB cancellation, while an additional 23 dB SIC performance is obtained when the PA output channel signal is used as the reference signal. Note that nonlinear devices other than amplifiers have a limiting effect on this method, such as preamplifiers (Liu et al., 2009).

The literature discussed above is focused on the convergence speed of the algorithm and the cancellation effect of the DSIC process, with little consideration of the impact of algorithm complexity in real time. The real-time implementation of IBFD-UWAC systems would allow both parties to communicate without delay and with no need to wait for the algorithm process, which is of great significance. In addition, the signal-to-interference ratio (SIR) of the signal is still low after the SI signal is canceled by SSIC or the analog domain, leaving the RIS. In IBFD-UWAC systems, the performance of interference cancellation depends on the accuracy of interference propagation channel estimation, the desired signal can have an impact on the self-interference channel estimation process, resulting in a lower bit error rate (BER) performance (Qiao et al., 2018a). During the SI channel estimation process, previous studies do not consider the impact of the arrival of the desired signal, causing a reduction in BER performance.

Based on the above, this paper makes the following contributions.

- For SIC in IBFD-UWAC systems, this paper proposes a spatial-digital joint SIC (SDSIC) method that achieves better cancellation than the previous SIC.
- We propose an SSIC method using an ICBBF based on convex optimization theory for IBFD-UWAC systems. This SIC method prevents the distortion of the distant desired signal and suppresses the direction- and array-dependent interference.
- For DSIC, a variable-step-size least-mean-square algorithm based on the space domain noise threshold (SNT-VSS-LMS) is proposed. This method has a faster convergence speed and better cancellation effect than the general VSS-LMS method.

2 System model

The modulated signal $x[n]$ is transmitted by a digital-to-analog converter (DAC). After passing through a low-pass filter (LPF) and PA, the signal is transmitted by the transducer, and the transmitted signal $r_n(t)$ is received by the array after passing through the SI channel $h_{\mathrm{SI}}[n]$, while the remote desired signal $r_m(t)$ is received by the array. Thus, the array received signal $x_{\mathrm{m}}(t)$ is obtained as

$$x_{\mathrm{m}}(t) = r_{n\mathrm{m}}(t) + r_{f\mathrm{m}}(t) + n_{\mathrm{m}}(t) \tag{1}$$

Where, m=1,2,3,⋯,M , $r_{n\mathrm{m}}(t)$ is the local SI signal received by array element m, $r_{f\mathrm{m}}(t)$ is the desired signal received by array element m, and $n_{xtm}(t)$ is the ambient noise received by array element m. The purpose of SIC is to reconstruct the SI signal so that $r_n(t)$ can be subtracted from the received signal to obtain the pure desired signal. After being received by the array, the signals from different directional sources have different time delays. The purpose of SSIC is to use these time delays to increase the SINR. The received signal is acquired by an analog-to-digital converter (ADC) after a low-noise amplifier (LNA) and passed through the beamformer to obtain the RIS after the SSIC process (① in Figure 1). In the DSIC process (position ② in Figure 1), the RIS is compared with the reference beamformer. The RIS is passed through a linear filter together with the reference signal to complete the DSIC process. To consider the effect of the nonlinear distortion of the PA on the estimation process of the SI channel, $y'_{\mathrm{PA}}[n]$ is passed through an adaptive filter together with the RIS to obtain the estimated SI channel $\hat{h}_{\mathrm{SI}}$ and the reconstructed SI signal $y_{tASIC}[n]$, where $y_{\mathrm{ASIC}}[n]$ is related to $\hat{h}_{\mathrm{SI}}$ as follows. The error output of the filter $e(n)$ is the signal after DSIC, which is the desired signal to be demodulated. The error output can be written as

$$e(n) = r_{AB}(n) - y'_{\mathrm{PA}} \otimes \hat{h}_{\mathrm{SI}} \tag{2}$$

where ⊗ denotes the convolution operation. The reference signal vector y'_{PA} and the tap weights of the linear filter are

$$y'_{\mathrm{PA}} = \left[y'_{\mathrm{PA}}[n], y'_{\mathrm{PA}}[n-1], \ldots, y'_{\mathrm{PA}}[n-L+1]\right]$$

$$\hat{h}_{\mathrm{SI}} = \left[\tilde{h}_{\mathrm{SI}}[n], \tilde{h}_{\mathrm{SI}}[n-1], \ldots, \tilde{h}_{\mathrm{SI}}[n-L+1]\right] \tag{3}$$

where L is the number of tap weights of the linear filter. The DSIC performance in the IBFD-UWAC system depends on approximation with $y_{\mathrm{ASIC}}(t)$ and $y_{\mathrm{PA}}(t)$. The above DSIC process is position 173 in the system model of Figure 1.

3 Fundamentals

3.1 Spatial self-interference cancellation

Beamforming is a common interference cancellation technique that is widely used in radio and underwater acoustics. This paper focuses on the IBFD-UWAC system. After receiving the signal, beamforming is performed in the spatial domain for SIC and to increase the SIR of the received signal. For UWAC systems, the acoustic waves are spatially sampled by M array elements, and the output signal $y(t)$ from the system at time t is the instantaneous linear combination of the spatial samples $x_m(t)$, that is,

$$y(t) = \sum_{m=1}^{M} x_m(t) w_m^{\ *} \tag{4}$$

where w_m denotes the weight coefficient of array element m and * denotes the complex conjugate. Let the input signal be the impulsive complex plane wave $e^{j\omega t}$, which has angular frequency ω, and let the arrival angle be $\theta\in[-\pi/2,\pi/2t]\theta\theta\in[-\pi/2,\pi/2]$. Assume that the signal phase received by the first array element is 0, and the signal received by the first array element is $x_0(t)=e^{j\omega t}$. The signal received by array element m is $x_m(t)=e^{j\omega t-\tau_m}$, where the time delay in the wave from the first hydrophone to the m-th hydrophone is τ_m , which a function of the arrival angle θ. The beamformer output can be expressed by

$$y(t) = e^{j\omega t} \sum_{m=1}^{M} e^{(-\omega\tau_m)} w_m^{\ *} \tag{5}$$

where τ_0=0 and $j = \sqrt{-1}$. The response of the beamformer is

$$P\omega, \theta = \sum_{m=1}^{M} e^{(-\omega\tau_m)} w_m^{\ *} = w^H a(\omega, \theta) \tag{6}$$

where the vector $\boldsymbol{w}$ contains the complex conjugate coefficients of the M sensors and H denotes the conjugate transpose, i.e.,

$$w = [w_1, w_2, w_3 \cdots w_M]^T \tag{7}$$

in which T denotes the transpose. The vector $a(\omega,\theta)$ is the array response vector, also known as the guide vector, i.e.,

$$a(\omega, \theta) = [1, e^{-j\omega\tau_1}, e^{-j\omega\tau_2} \cdots e^{-j\omega\tau_{M-1}}] \tag{8}$$

The distance between each array element is $d=\lambda/2$, and so $\omega\tau_{\mathrm{m}}=(2\pi c/\lambda)(\mathrm{md}sin\ \theta/c)=\mathrm{m}\pi sin\ \theta$. The narrowband beamformer response is written as

$$\mathrm{P}(\omega, \theta) = \sum_{\mathrm{m}=1}^{M} e^{(-\mathrm{j}\mathrm{m}\pi \sin\theta)} w_m^{\ *} \tag{9}$$

Note that the SSIC method uses a frequency-domain wideband beamformer to convert the data from the time domain to multiple

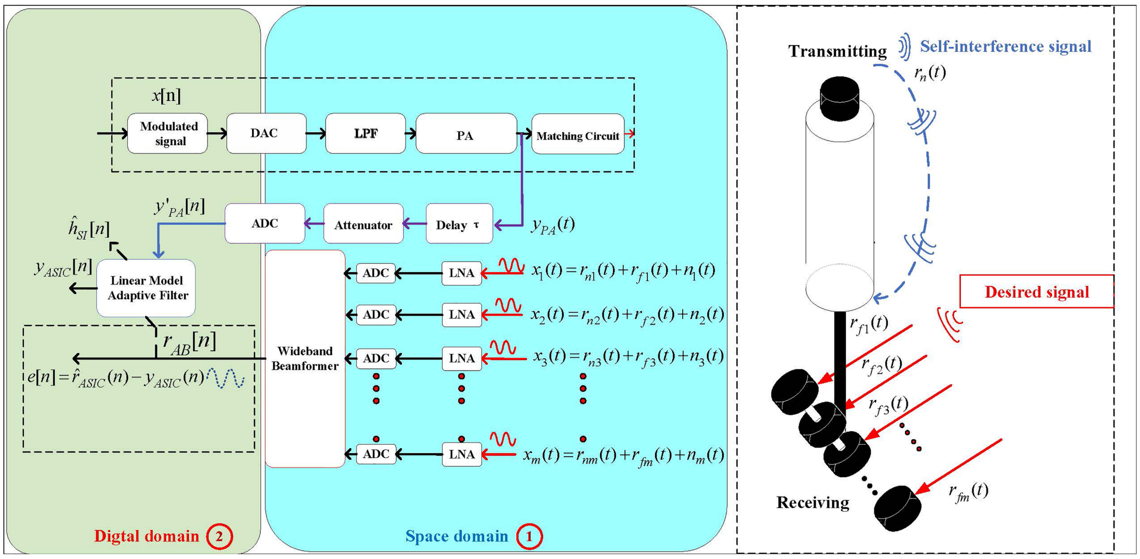


FIGURE 1
SDSIC system model in IBFD-UWAC.

sub-bands in the frequency domain through a Fourier transform. Each sub-band satisfies the narrowband condition to finish the beamforming process. Finally, the output of the beamformer is transformed to the time domain using an inverse discrete Fourier transform. For the problem of an inconsistent response at each frequency point, one method involves receiving the data of each frequency of the signal with different sizes of sub-arrays (Goodwin and Elko, 1993; Liu et al., 2021), but this requires a complex hardware structure and cannot achieve a constant beamwidth for the whole frequency band. More recently, researchers have designed different weighting vectors for the various sub-band frequencies, resulting in a constant beamwidth at these central frequencies. There are spatial re-sampling methods (Krolik and Swingler, 1990; Li et al., 2017), fitting methods based on Fourier transforms (Liu et al., 2007; Chan and Chen, 2007; Liu and Weiss, 2008), and Bessel function algorithms (Liu et al., 2005), among others. Furthermore, in the practical application of IBFD-UWAC systems, it is often necessary to design a special response beamformer according to the array shape (Roe et al., 1990; Fan and Liang, 2013). This process mainly keeps the main flap beam response consistent and suppresses direction- and array-dependent interference. In practical applications, if there are multiple desired signals at unknown angles in the far field, the beamformer should be designed such that the beam response of the side flaps can discriminate the multiple desired signals in different directions, which is critical for a full-duplex communication network. To solve the above problems, an ICBBF based on convex optimization theory is proposed in this paper. The proposed beamforming method constructs a "zero trap" for direction- and array-dependent interference source directions while maintaining the frequency invariance of the beam response of the main flap. We design the convex optimization process as follows. It is worth noting that the beamformer avoids significant amplitude cancellation of signals from other angles and retains the signal from other directions. The SSIC method designed in this paper uses a convex optimization method to cancel the SI signals of direction- and array-dependent interference in the IBFD-UWAC system. We design the specific convex optimization process as follows. Step 1: Apply Newton's iteration method to produce a beamformer according to the array structure, where the desired signal source angle is 0°, and divide the signal into several sub-bands. Perform the following steps at each sub-band frequency. Step 2: Take the main flap response at the center frequency of the beamformer designed above as the desired response $p_{\text{desired}}(\theta)$ for the convex optimization process, and minimize the two-parameter response of the optimized beamformer response $p_{MF}(\theta)$ in the main flap angle interval θ_{MF} with $p_{\text{desired}}(\theta)$. Step 3: Use convex optimization theory to constrain the beam response $p_{SF}(\theta)$ in the side flap angle interval θ_{SF} to be below the set value ξ_{SF} and the beam response $p_{ZF}(\theta)$ in the zero-trap interval θ_{ZF} below the set value ξ_{ZF} . Step 4: Use an inverse fast Fourier transform to restore the signal and obtain the signal after SSIC. According to the above steps, we construct the convex optimization equation as

$$
\begin{aligned}
&\min_{\omega_{HB}} \quad \|p_{MF}(\theta) - p_{desired}(\theta)\|_2 \\
&s.t. \|\omega_{HB}\|_2 \leq 1 \\
&\qquad \left|\omega_{HB}^T \vec{a}(f_k, \theta_{SF})\right| \leq \xi_{SF} \\
&\qquad \left|\omega_{HB}^T \vec{a}(f_k, \theta_{ZF})\right| \leq \xi_{ZF} \\
&\qquad \omega_{HB}^T \vec{a}(f_k, \theta_d) == 1
\end{aligned}
\tag{10}
$$

To facilitate the calculation using the SeDuMi toolbox, we transform Eq. (10) to

$$
\begin{aligned}
&\min \quad \epsilon_m \\
&s.t. \quad \omega_{HB}^T \vec{a}(f_k, \theta_d) == 1 \\
&\qquad \|p_{MF}(\theta_{MF}) - p_{desired}(\theta_{MF})\|_2 \leq \epsilon_m \\
&\qquad \left|\omega_{HB}^T \vec{a}(f_k, \theta_{SF})\right| \leq \xi_{SF} \\
&\qquad \left|\omega_{HB}^T \vec{a}(f_k, \theta_{ZF})\right| \leq \xi_{ZF} \\
&\qquad \|\omega_{HB}\|_2 \leq 1
\end{aligned}
\tag{11}
$$

where a new set of nonnegative variables is introduced: ϵ_m , m=1,2,⋯,M .

We set $b=[0,-1_{1\times M},0_{1\times N},]^{\mathrm{T}}$,$y=[1,\epsilon_1,\cdots,\epsilon_M,\omega_{HB1},\cdots,\omega_{HBN}]^{\mathrm{T}}$ and solve for the optimal weight vector at each frequency point, i.e., the weighting factor for each array element ω_{HB} . In SeDuMi, the standard convex cone problem is defined as

$$
\begin{aligned}
&\max \quad b^{\mathrm{T}} y \quad \text{subject} \quad \text{to} \\
&c - A^{\mathrm{T}} y \in K
\end{aligned}
\tag{12}
$$

where $\boldsymbol{y}$ contains the expected weights, $\boldsymbol{A}$ is an arbitrary coefficient matrix, $\boldsymbol{b}$ and $\boldsymbol{c}$ are arbitrary vectors, and $\boldsymbol{K}$ is a set of symmetric cones. The dimensions of $\boldsymbol{A}$, $\boldsymbol{b}$, and fc are matched and the q-dimensional second-order cone is defined as

$$
Q_c^q \underset{=}{\Delta} \left\{ \begin{bmatrix} x1 \\ x2 \end{bmatrix} \middle| x1 \in R, x2 \in C^{q-1}, x1 \geq ||x2|| \right\}
\tag{13}
$$

The equation constraint can be expressed as $\{0\} \underset{=}{\Delta} \quad \{x = C|x = 0\}$. The final problem then has the form

$$
\begin{aligned}
&\max \quad \boldsymbol{b}^{\mathrm{T}} \boldsymbol{y} \quad \text{subject} \quad \text{to} \\
&\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}} \boldsymbol{y} \in \{0\}^{q_1} \times R_+^{\ q_2} \times Q_c^{\ q_3}
\end{aligned}
\tag{14}
$$

This can be solved using the SeDuMi toolbox, where Eq. (14) has q_1 equation constraints, q_2 linear constraints, and q_3 second-order cones. The first equation in Eq. (11) can be expressed using a zero cone as

$$
\begin{aligned}
&\begin{bmatrix} 1 \\ a^{\mathrm{H}}(\theta_d)\omega_{HB} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0_{1\times M}^T & 0_{1\times N}^{\mathrm{T}} \\ 0 & 0_{1\times M}^{\mathrm{T}} & -a^{\mathrm{H}}(\theta_d) \end{bmatrix} y \\
&\underset{=}{\Delta} \mathrm{c}_1 - A_1^{\mathrm{T}} y \in \{0\}
\end{aligned}
\tag{15}
$$

The inequalities in Eq. (11) can be expressed using second-order cones as

$$\begin{bmatrix} \epsilon_m + 1 \\ 2p_{desired}(\theta_m) - 2\mathrm{a}^H(\theta_m)\omega_{HB} \\ \epsilon_m - 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2p_{desired}(\theta_m) \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & -\mathrm{t}^{\mathrm{T}}(m) & 0_{1\times N} \\ 0 & 0_{1\times M} & 2\mathrm{a}^{\mathrm{H}}(\theta_m) \\ 0 & -\mathrm{t}^{\mathrm{T}}(m) & 0_{1\times N} \end{bmatrix} y \underset{=}{\Delta} c_{1+m} - A_{1+m}^{\mathrm{T}}, y \in Q_{1+m}^3, m = 1, 2, \cdots, M \tag{16}$$

where $t(m)=[t_1,t_2,\cdots,t_i,\cdots,t_M]T$ and $t_i = \{ \begin{matrix} 0 & i \neq m \\ 1 & i = m \end{matrix} \}$.

$$\begin{bmatrix} \xi_{SF} \\ a^H(\theta_{iSF})\omega_{HB} \end{bmatrix} = \begin{bmatrix} \delta_S \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0_{1\times M} & 0_{1\times N} \\ 0 & 0_{1\times M} & -\mathrm{a}^H(\theta_{iSF}) \end{bmatrix} \mathrm{y} \underset{=}{\Delta} \mathrm{c}_{1+M+s} - A_{1+M+s}^{\mathrm{T}} \mathrm{y} \in Q_s^2, s = 1, 2, \cdots, S \tag{17}$$

$$\begin{bmatrix} \xi_{ZF} \\ \mathrm{a}^H(\theta_{iZF})\omega_{HB} \end{bmatrix} = \begin{bmatrix} n_z \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0_{1\times M} & 0_{1\times N} \\ 0 & 0_{1\times M} & -\mathrm{a}^H(\theta_{iSF}) \end{bmatrix} \mathrm{y} \underset{=}{\Delta} \mathrm{c}_{1+M+S+z} - A_{1+M+S+z}^{\mathrm{T}} \mathrm{y} \in Q_z^2, s = 1, 2, \cdots, Z \tag{18}$$

$$\begin{bmatrix} 1 \\ \boldsymbol{\omega}_{HB} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & \mathbf{0}_{1\times M} & \mathbf{0}_{1\times N} \\ \mathbf{0}_{N\times 1} & \mathbf{0}_{N\times M} & -I_N \end{bmatrix} \mathbf{y} \underset{=}{\Delta} \mathbf{c}_{M+S+Z+2} - \mathbf{A}_{M+S+Z+2}^{\mathrm{T}} \mathbf{y} \in Q_{M+S+Z+1}^{1+N} \tag{19}$$

where I_N is and N-dimensional unit vector. We set $c = [c_1^{\mathrm{T}}, c_2^{\mathrm{T}}, \cdots, c_{M+S+Z+2}^{\mathrm{T}}]$ and $A^{\mathrm{T}}=[A_1,A_2,\cdots,A_{M+S+Z+2}]^{\mathrm{T}}$, where c_i , $A_i(i=1,2,\cdots,M+S+Z+2)$ are given by Eqs. (15)–(19). The final conversion to the standard second-order cone problem is as follows.

$$\begin{aligned} &\max\ b^{\mathrm{T}} y \\ &\mathrm{s.t.}\ c - A^{\mathrm{T}} y \in \{0\} \times {Q_1}^3 \times \cdots \times {Q_M}^3 \times Q = {Q_{M+1}}^2 \\ &\quad \times \cdots \times {Q_{M+S}}^2 \times {Q_{M+S+1}}^2 \times \cdots \times {Q_{M+S+Z}}^2 \times {Q_{M+S+Z+1}}^{N+1} \end{aligned} \tag{20}$$

We obtain the final optimal solution of $\boldsymbol{y}$, where $M+2$ to $M+1+N$ are the optimal weights of the array ω_{HB} . After the above convex optimization process, ICBBF is able to suppress the signal in the specified direction according to the values of θ_{ZF} , θ_{MF} , and θ_{SF} in the convex optimization process. The 2-norm between $p_{MF}(\theta)$ and the desired beam response $p_{\mathrm{desired}}(\theta)$ is minimized, thus preventing distortion of the distant desired signal.

3.2 DSIC principle

The DSIC process is performed based on an adaptive filter. The amplifier signal collected locally $y'_{\mathrm{PA}}[n]]$ is reconstructed into a signal vector at time n by $L-1$ delay units, each of which has a corresponding tap weight coefficient $\omega[n]$. The overall weight coefficient is

$$W = [\omega_0[n], \omega_1[n], \omega_2[n], \cdots, \omega_{L-1}[n]]^T \tag{21}$$

The output signal of the filter $y_{\mathrm{ASIC}}[n]$ slowly approaches $r_{AB}[n]$, with the adaptive filter acting as a channel estimator. The offsetting effect depends on the magnitude of the error between the SI channel $h_{\mathrm{SI}}[n]$ and W, and the DSIC method is essentially an optimization of the adaptive filter algorithm to achieve the fastest convergence speed and the smoothest steady-state effect. The conventional LMS method is a stochastic gradient descent algorithm that includes filtering and adaptive processes based on the following procedure: (1) The input signal $x(n)$ is the input of the LMS adaptive filter. The output signal is $y(n)=w^T(n)\otimes x(n)$, where $w(n)$ is the weight vector. Let the expected output response of the LMS adaptive filter be $d(n)$. Then, the error between the expected output response of the filter and the output signal $y(n)$ is

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= d(n) - x^T(n) \otimes w(n) \\ &= d(n) - w^T(n) \otimes x(n) \end{aligned} \tag{22}$$

(2) From the above errors, the mean square error is obtained as

$$\begin{aligned} E\left[e^2(n)\right] &= E\left[d^2(n) - 2d(n)x^T(n)w(n) + w^T(n)x(n)x^T(n)w(n)\right] \\ &= E\left[d^2(n)\right] - 2E[d(n)x^r(n)w(n)] + E\left[w^T(n)x(n)x^T(n)w(n)\right] \end{aligned} \tag{23}$$

(3) We define

$$\begin{aligned} R_{xd}^T &= E\left[d(n)x^T(n)\right] \\ R_{xx} &= E\left[x(n)x^T(n)\right] \end{aligned} \tag{24}$$

and deduce that

$$E\left[e^2(n)\right] = E\left[d^2(n) - 2R_{xd}^T w(n) + w^T(n)R_{xx}w(n)\right] \tag{25}$$

(4) The LMS algorithm uses the gradient of the mean square error

$E[e^2(n)]$ to find the minimum value:

$$\nabla(n) = \nabla E\left[e^2(n)\right] = \left[\frac{\partial E\left[e^2(n)\right]}{\partial w_1(n)}, \frac{\partial E\left[e^2(n)\right]}{\partial w_2(n)}, \ldots, \frac{\partial E\left[e^2(n)\right]}{\partial w_M(n)}\right]^T \tag{26}$$

(5) The weights of the LMS algorithm are updated according to

$$w(n+1) = w(n) - \mu\hat{\nabla}(n) = w(n) + 2\mu e(n)x(n) \tag{27}$$

The conventional LMS algorithm causes the steady-state effect and convergence speed to be mutually constrained because of the invariance of its step size. In contrast, the core of VSS-LMS is the use of different step size adjustment criteria. The filter is assigned a more significant step size when the power of the interference in the filter is significant, resulting in a faster convergence speed, and the step size is adaptively reduced to ensure that the filter has a better steady-state effect when the system is close to the steady state. In contrast, in the IBFD-UWAC system, SSIC ensures that most SI signals are canceled. Suppose there is an overlap between the desired signal and the SI signal. In this case, an error is introduced to the filter when estimating the SI channel, which reduces the estimation accuracy. Thus, this paper proposes an SNT-VSS-LMS algorithm and sets an expectation threshold based on SSIC and the real-time state error. The step adjustment criterion is designed to minimize the influence of the desired signal in the SI channel estimation and reduce the influence of the desired signal on the filter performance under low-SINR conditions. In contrast, the step size in the filter changes according to parabolic constraints. The step adjustment criterion is designed for the RIS of the IBFD-UWAC system, and can be described as follows:

a. If the power of the SI signal after SSIC is sufficiently high that the BER performance cannot be improved, the filter applies the maximum step size in DSIC.
b. If the SI signal after SSIC no longer affects the BER performance, the filter uses the minimum step size in DSIC to ensure optimal steady-state results.
c. Between these two cases, the filter step size is constrained by the parabolic function to achieve the optimal SIC effect. In the IBFD-UWAC system, the energy range for the arrival of the desired signal is predicted because the desired signal after SIC must have a certain SINR that guarantees the performance conditions of the communication BER. The SNT-VSS-LMS algorithm with the step-size adjustment criterion based on a parabolic function allows the step size to be adjusted using the relationship between the instantaneous state error and the threshold value, as described in Eq. (28).

$$\mu(n) = \begin{cases} \mu_{\min} & 0 < \quad |\sigma_e(n)| \quad < \sigma_v \\ \left(\frac{|\sigma_e|-\sigma_v}{|\gamma_v\sigma_d|-\sigma_v}\right)^2 \times \mu_{\max} + \left(1 - \left(\frac{|\sigma_e|-\sigma_v}{|\gamma_v\sigma_d|-\sigma_v}\right)^2\right) \times \mu_{\min} & \sigma_v < \quad |\sigma_e(n)| \quad < \gamma_v\sigma_d \\ \mu_{\max} & \gamma_v\sigma_d < \quad |\sigma_e(n)| \end{cases} \tag{28}$$

where $\sigma_e(n)$ is the square root of the error variance in the instantaneous update state at time n and σ_v is the SINR of the desired signal that guarantees the BER performance. This is calculated as follows:

$$\sigma_v = \sqrt{\sigma_n^2 \cdot 10^{\frac{SINR_{known}}{10}}} \tag{29}$$

where σ_n^2 is the noise floor of the SI signal after SSIC, $SINR_{known}$ is the SINR of the desired signal that guarantees the BER performance condition, σ_d is the square root of the

TABLE 1 Specific flow of the SNT-VSS-LMS algorithm.

1) Setting of initial values:

Initial filter weight value $w(0)$, initial output signal value $y(0)$, initial step size $\mu(0)$, instantaneous state error $\sigma_e(0)$, far-end desired signal arrival threshold σ_v (obtained by Eq. (29)), and desired signal variance $\sigma_d(0)$

2) Parameters of the proposed algorithm are set according to the actual situation: Determine α, β, compensation factor β_v, maximum step length μ_{max}, and minimum step length μ_{min}

3) Iterative update process: a) Output signal $y(n) = u(n)^* w(n)$ of adaptive filter, in which input signal $u(n)$ is the RIS after spatial-domain SI suppression

b) Calculate estimation error $e(n) = d(n) - y(n)$

c) Calculate error variance in instantaneous state and variance of desired signal in instantaneous state

d) Collect the background noise under the experimental conditions, using the noise level σ_n obtained by SSIC to obtain σ_v, and set the compensation factor β_v according to the experimental conditions

e) Adjust step size according to the above step-size adjustment criterion

f) Update equation of weight vector

$w(n+1) = w(n) + \tilde{\mu}(n)e(n)u(n)$

4) Calculate average error M-NMSE

instantaneous desired signal variance, γ_v is a compensation factor, and α, β are parameters controlling the shape of the modified sigmoid function. The specific flow of the SNT-VSS-LMS algorithm is described in Table 1.

Unlike the traditional VSS-LMS algorithm, we consider the influence of the arrival of the desired signal on the RIS channel estimation process. The SNT-VSS-LMS algorithm adaptively adjusts the step size by examining the relationship between $\sigma_e(n)$ and σ_v,$\gamma_v\sigma_d$ in the convergence process according to Eq. (28), i.e., when $\sigma_e(n)$ is less than σ_v , the main error of the SI channel estimation process depends on the desired signal, so the step size is reduced to μ_{min} ; when $\sigma_e(n)$ is greater than $\gamma_v\sigma_d$, the main error of the SI channel estimation process depends on the RIS and the step size is adjusted to μ_{max} for faster convergence. When $\sigma_e(n)$ is in the intermediate state, the step is constrained using the parabolic function. In this way, the SNT-VSS-LMS algorithm obtains better steady-state results and faster convergence. Compared with the traditional VSSLMS algorithm, our proposed algorithm determines the step value referring to the SI signal after SSIC, which can minimize the impact of the desired signal while ensuring the convergence speed and steady-state effect.

Figure 2 shows a flowchart of the proposed SDSIC method. After SSIC, the RIS is subjected to DSIC to complete the SIC process, resulting in the distant desired signal to be demodulated.

To evaluate the performance of our proposed algorithm, we examine the results of simulations and experiments in the next section.

4 Simulation analysis

4.1 SSIC simulation

To illustrate the SIC effect and the signal restoration effect, the simulation experiments use a high-power single-frequency SI signal. The desired signal is an LFM signal. To make the experimental scenario closer to that of engineering applications, the SI signal and the desired signal overlap. We collected simulation data from an anechoic pool. The frequency of the SI signal is 4 kHz, and the LFM frequency range is 3–5 kHz. The direction of the signal is 70°, the number of array elements is $M=12$, and the distance between each array element is $d=0.185$ m.

Figure 3 shows the beam response of the beamformer with the minimum variance distortion less response (MVDR)-MAILLOUX algorithm (Mailloux, 1995). This algorithm uses the MVDR with some virtual interference sources in a preset angle range to accomplish a more extensive range of zero trapping. The MVDR-MAILLOUX beamformer achieves excellent cancellation from 65–75°, but the beam responses are significantly different at each frequency point around the desired direction of 0°, and the lowest frequency point of the main flap widths is 19.5°. The maximum frequency distortion reaches 4.2 dB in the range [-5°, 5°]. In actual applications, the array will drift due to waves and other environmental factors, and the position is not fixed. Thus, the guiding vector will be

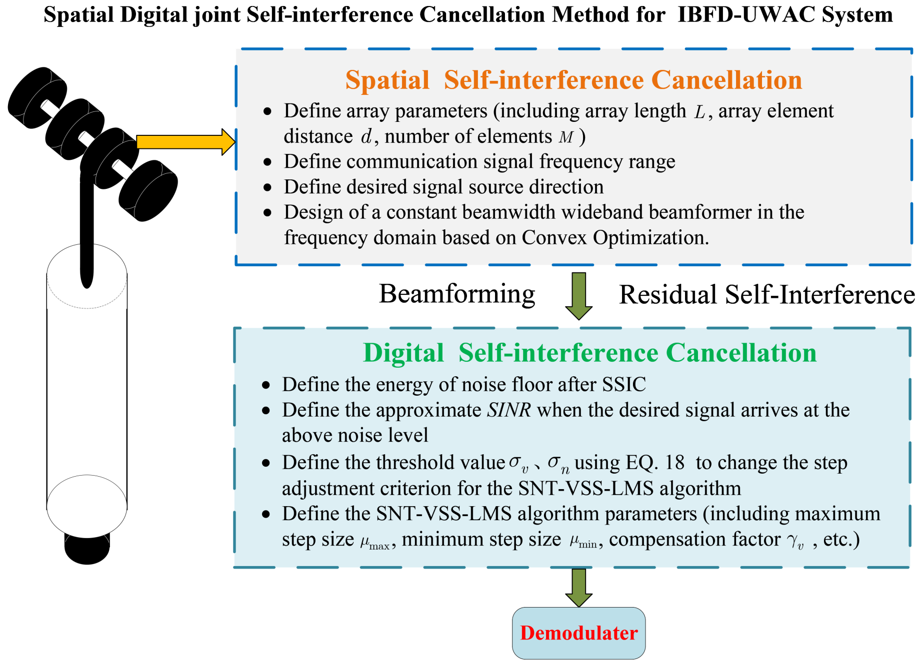


FIGURE 2
Flowchart of the SDSIC method.

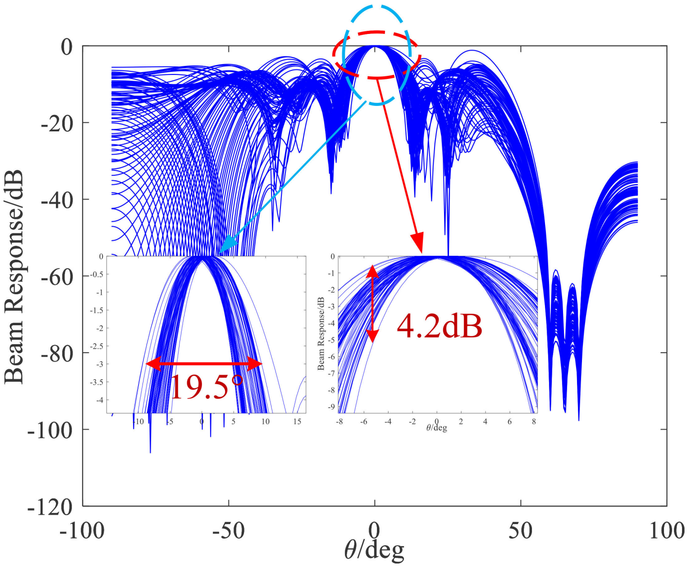


FIGURE 3
Beam response of the MVDR-MAILLOUX beamformer.

mismatched, and there will be frequency distortion when the signal is restored. This will result in a loss of frequency characteristics from the far-end desired signal after restoration, which will have a specific impact on the BER performance of the communication system. To solve this problem, it is necessary to make the spectrum of the base array signal distortion less after passing through the beamformer, and the beam response of each frequency point should be the same. As the beamformer should have approximately the same primary flap beam response at each frequency point, the "constant-beamwidth" beamformer is of considerable interest (Krolik and Swingler, 1990).

Figure 4 shows the beam response of the adaptive beamformer based on the LMS algorithm. We can see that the main flap width of the lowest frequency point response of the

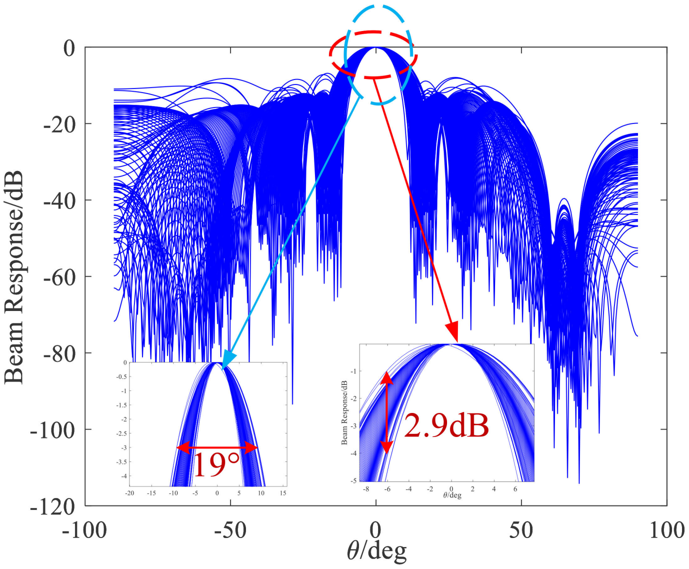


FIGURE 4
Beam response of the adaptive beamformer based on LMS algorithm.

beamformer optimized by the LMS adaptive algorithm is 19°. The maximum value of the primary flap frequency distortion is 2.9 dB in the range [-5°, 5°]. The constant-beamwidth beamformer exhibits an excellent ability to restore the signal. However, the amplitude of the array drift is large, and the beam response needs to be kept consistent within a specific angle range.

Figure 5 shows the beam response of the constant-beamwidth beamformer based on Newton's iteration method. The maximum frequency distortion is 0.4 dB in the range [-5°, 5°] and 2 dB in the range [-10°, 10°]. The beam response of this method controls the main flap within a specific range, which solves the problem of array drift, i.e., the mismatch of the guiding vector.

However, the beam response of this beamformer is below -40 dB for the rest of the side flaps beside the interference source, and the width of the main flap is too large. Additionally, the beamformer's angle resolution decreases. The simulation results of the proposed SSIC method are given below. According to the experimental data, the parameters of the convex optimization process described in Section 3.1 are set as follows: $\theta_{MF}=[-5°,5°]$, $\theta_{SF}=[-90^{,irc},-8°]\cup[8°,90°]$, $\theta_{ZF}=[65°,75°]$. In the simulation, $\xi_{SF}=10^{-15/20}$, $\xi_{ZF}=10^{-60/20}$, and $\theta_d=0°$ is the desired signal source direction. After solving the convex optimization problem, the obtained beam response is shown in Figure 6. The main flap width of the lowest frequency point response of the proposed method is 10°, and the maximum frequency distortion in the range [-5°, 5°] is 1.9 dB, and so the main flap width has been reduced by achieving a consistent main flap beam response and the beamformer's resolution has been enhanced. Moreover, the beamformer cancels signals around the 70° direction and preserves signals from other directions, which is of great significance for constructing a full-duplex UWAC network. In subsequent simulations, we performed the SSIC process with the ICBBF based on convex optimization.

In the SSIC simulation experiments, the distance between each array element was set to 0.185 m, the number of array elements was fixed at 12, and the sampling frequency was 50 kHz. The desired signal had a direction of 0° and the SI signal had a direction of 70°. Note that, because the beamformer weights the data of each array element to obtain the new data, we compared the far-end desired signal with the SI signal after the beamformer. To determine the SIC effect, the non-overlapping part of the signal was used to calculate the cancellation effect:

$$\zeta_{BF} = 10 \times \lg\left(\sigma_1^2/\sigma_2^2\right) - 10 \times \lg\left(\sigma_3^2/\sigma_4^2\right) \tag{30}$$

where σ_1 is the standard deviation of the non-overlapping part of the desired signal after beamforming, σ_2 is the standard deviation of the non-overlapping part of the desired signal from the reference array, σ_3 is the standard deviation of the non-overlapping part of the SI signal after beamforming, and σ_4 is the standard deviation of the non-overlapping part of the SI signal from the reference array.

The purpose of the above equation is to keep the amplitude of the non-overlapping part of the desired signal consistent before and after beamforming. Figure 7 shows the normalized spectrum of the signal after SSIC and the signal received by the

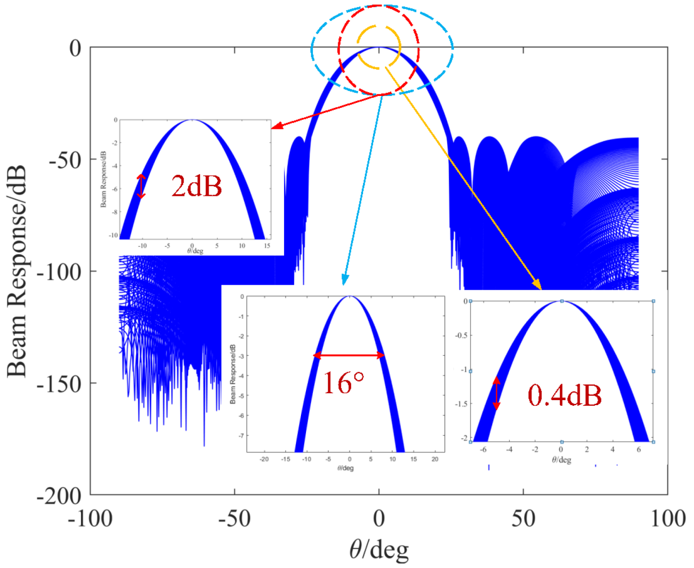


FIGURE 5
Beam response of the constant-beamwidth beamformer based on Newtonian iteration.

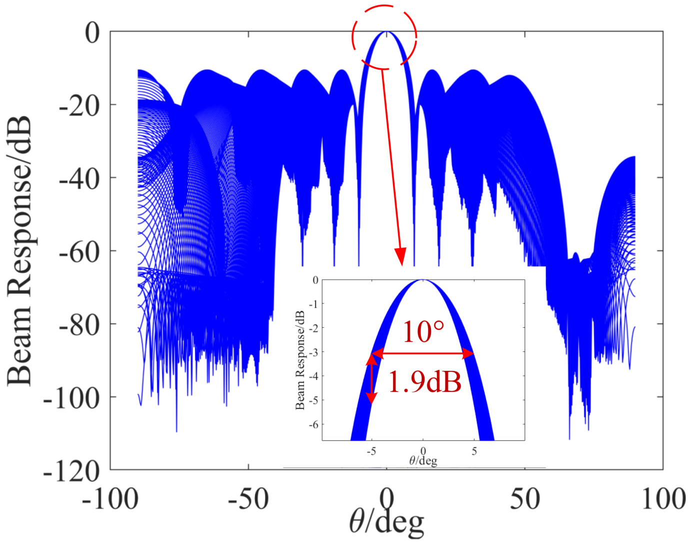


FIGURE 6
Spectrum of the signal after the constant-beamwidth beamformer based on Newtonian iteration.

reference array.

The simulation results show that the SI signal is cancelled by 31.5 dB after SSIC. Next, the RIS was subjected to DSIC.

4.2 DSIC for RIS

In the DSIC simulation experiments, the local SI signal and the desired signal were wideband LFM signals. For the VSS-LMS algorithm, we set c=0.003 ; for the SNT-VSS-LMS algorithm, γ_v=0.8 , μ_{max} =0.003 , μ_{min} =0.0005 . The SNT-VSS-LMS algorithm obtains $\sigma_v = \sqrt{\sigma_n^2 \cdot 10^{\frac{20}{10}}}$ with $SINR_{known}$=20 dB, and the SNR of the SI signal is SNR_{SI}=30 dB. Therefore, the step size starts to decrease, and when the NMSE reaches -10 dB, the step size is adjusted to the minimum μ_{min} =0.0005 . As can be seen, the SNT-VSS-LMS algorithm starts to reduce the step size before the cancellation is complete, preventing “over-cancellation” of the SI signal and reducing the influence of the desired

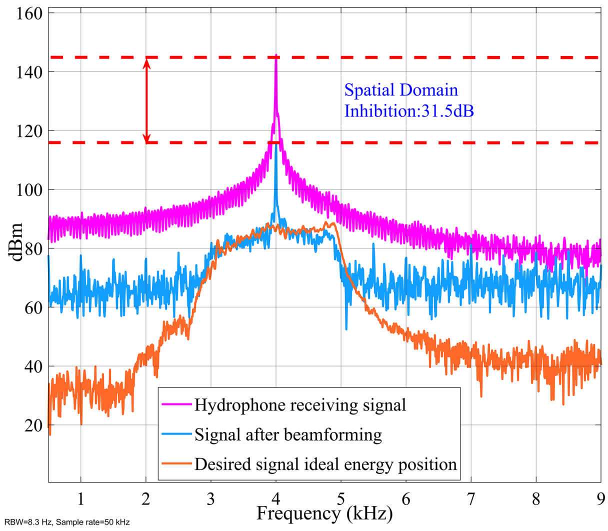


FIGURE 7
Spectrum of the signal after SSIC (after normalization).

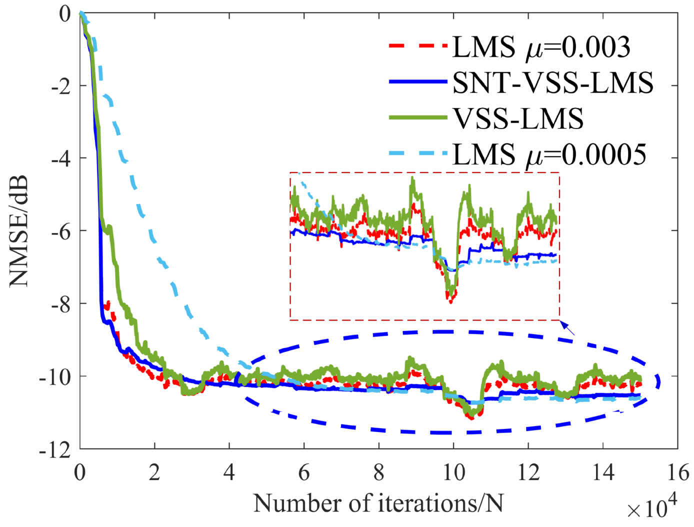


FIGURE 8
NMSE curves of DSIC using various algorithms.

signal on the filter steady state. To better demonstrate the superiority of the SNT-VSS-LMS algorithm, we simulated the case in which the desired signal completely overlaps with the SI signal. The NMSE curves are shown in Figure 8.

Figure 8 shows the NMSE curves for DSIC using the SNT-VSS-LMS, VSS-LMS, and LMS algorithms in the case of an overlap between the desired signal and the SI signal. The SNT-VSS-LMS algorithm produces the best results when the filter cancellation is close to the steady-state and the fastest convergence rate for the cancellation iterations in the initial stages when the SI signal energy is high. We also show the BER under different interference-to-signal ratios (ISRs) in Figure 9 to illustrate the effect of SIC. The simulation analysis of the BER under a 10-dB SNR is carried out below. The ISR varies in the range [12, 22], and the BER is obtained at 2-dB intervals.

Figure 9 compares the BER performance of the three schemes under an SNR of 10 dB. As most of the RIS in the iterative process overlaps with the desired signal in the simulation considered in this paper, the LMS algorithm has a slow convergence speed. Indeed, the convergence time is longer than the desired signal arrival time, so demodulation without complete cancellation will significantly reduce the BER performance. The VSS-LMS and SNT-VSS-LMS algorithms achieve good BER performance because of the faster convergence speed. Although VSS-LMS has a reasonable convergence speed, it cannot adaptively judge the relationship between the RIS energy and the desired signal, resulting in poor steady-state performance and high BER. When the ISR is above a certain threshold, the BER increases with increasing ISR. This is because the energy of the SI signal increases, and when the ISR is less than a certain threshold, the BER increases instead. When the ISR is small, the desired signal worsens the filter steady state, this decreasing the BER.

5. Anechoic pool experiment

To prevent multipath effects from creating uncertainty in the direction of the signal source, the experiments were performed in an anechoic pool. The results are shown in Figure 10.

The simulations were implemented in the Simulink ® platform based on a hardware-in-the-loop framework, where the real-time implementation of the IBFD-UWAC system has more practical application value. In the experiments, the array element spacing was 0.185 m, the number of array elements was fixed to 12, and the sampling frequency was 50 kHz. The desired signal source had a direction of 0° and the SI signal had a direction of 70°.

The locally transmitted signal was an LFM signal plus Gaussian white noise, and the far-end desired signal at a distance of 20 m was an LFM signal. The aim of this experiment was to verify the cancellation effect of the proposed SDSIC method. The ICBBF described in Section 3.1 was used to cancel the SI signal at 70° degrees and enhance the desired signal at 0°.

Figure 11 compares the cancellation effects of the SDSIC and traditional SIC methods. The right side of Figure 11 shows a time-frequency diagram of the signal after the SDSIC method and the signal without SIC. In general, SSIC reduces the burden of SIC in the analog and digital domains and improves the upper limit of SIC under the same hardware conditions. This figure shows that using the traditional SIC scheme can cancel up to

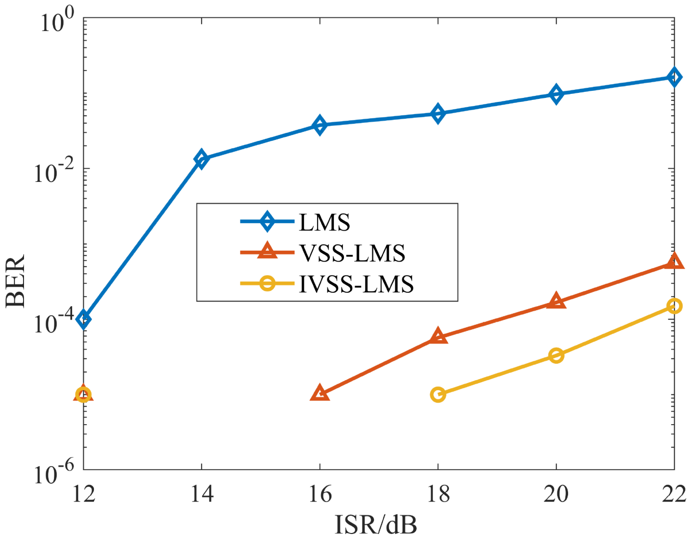


FIGURE 9
BER in simulations.

36.2 dB of the SI signal, whereas the proposed SDSIC method can cancel up to 56.4 dB. The scheme and algorithm proposed in this paper are therefore highly effective. In a second experiment, the local transmitting signal was a 3–5 kHz LFM signal, and the far-end desired signal at a distance of 20 m was a direct-sequence spread spectrum signal in the same frequency band. This experiment was conducted to verify the cancellation effect of the proposed SDSIC method. Note that, in the IBFD-UWAC system, the SIC process takes a certain amount of time, and the two sides can communicate stably only after the SI signal has

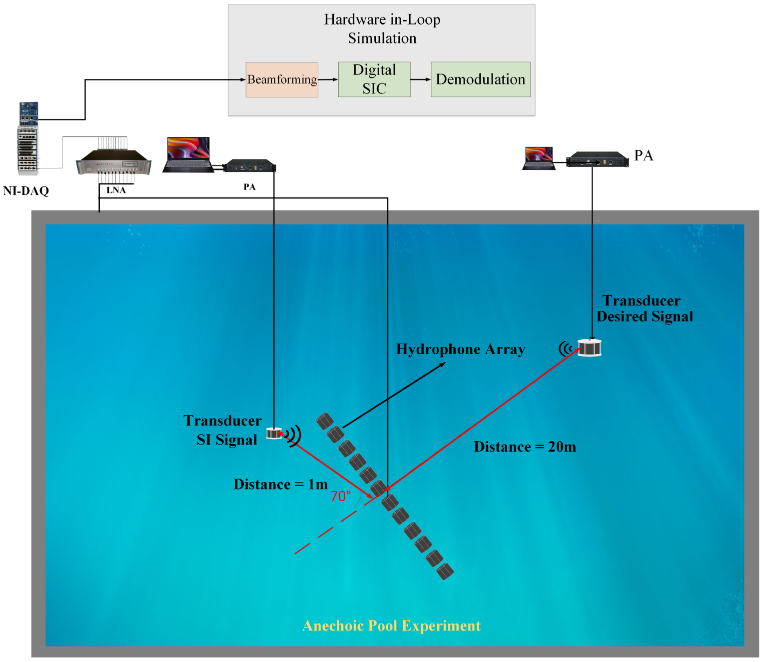


FIGURE 10
Schematic diagram of the experimental scenario.

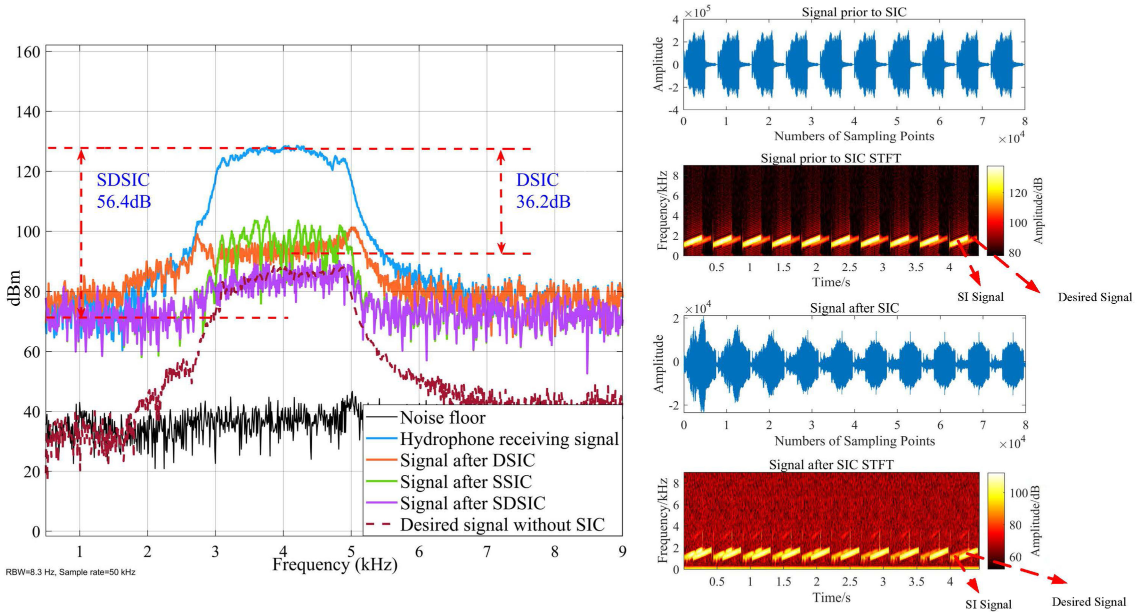


FIGURE 11
Comparison of cancellation effects between SDSIC method and traditional DSIC method.

been canceled to a certain degree. Compared with the DSIC method, the SDSIC method first filters the received signal in the spatial domain, which reduces the width of the main flap while achieving a consistent main flap beam response, and accurately cancels the signal in the set direction (θ_{ZF} = [65°,75°]). This improves the resolution of the beamformer. The parameters of the signals in this experiment are listed in Table 2.

As shown in Figure 12, both SDSIC and DSIC reach a steady state, although the SDSIC method has a better cancellation effect. The right side of Figure 12 shows a time-frequency diagram of the signal after the SDSIC method and the signal without SIC. The SSIC effect reaches 20 dB, and the SNR of the canceled signal SNR_{RIS} = 27 dB. The noise energy of the current signal σ_n^2 is obtained from the RIS, and the expected SINR of the desired signal $SINR_{known}$ = 18 dB. The DSIC of the RIS was completed by the SNT-VSS-LMS algorithm using the step-size adjustment criterion of Eq. (28), and the expected threshold value based on the spatial domain noise σ_v was obtained from Eq. (29). We set c = 0.005 in the VSS-LMS algorithm and γ_v = 0.8, $\mu_{\max}$ = 0.005, $\mu_{\min}$ = 0.0001 in the SNT-VSS-LMS algorithm. The NMSE curves obtained from the RIS after applying the LMS, VSS-LMS, and SNT-VSS-LMS algorithms are shown in Figure 13.

The SNT-VSS-LMS algorithm proposed in this paper has the fastest convergence speed, and the best steady-state effect after applying the DSIC process to the RIS. This algorithm can adjust the step size in real time according to the variance of the instantaneous state error. At first, the step size is set to 0.005 to achieve the fastest convergence speed when the SI signal is canceled by 24 dB (i.e., the NMSE result of the signal is close to SNR_{SI}–$SINR_{known}$ = 9 dB), and then the step size is adjusted to prevent the SI signal from being over-canceled, which ensures the best steady-state effect. Figure 13 verifies the effectiveness of the algorithm and its superior performance compared with the VSS-LMS algorithm. The VSS-LMS technique gives a poorer steady-state effect because of its inability to determine adaptively where the desired signal is coming from and lack of optimization of the convergence speed. In summary, the proposed SNT-VSS-LMS algorithm produces an excellent DSIC effect for the RIS.

Next, we consider the BER performance of the IBFD-UWAC system. Note that, during the anechoic pool trial, the PA of the far-end transmitting ship was kept at its maximum value. When the two sides communicate, the arrival time of the far-end

TABLE 2 Parameters of the spread spectrum communication signal.

Chip duration	0.8 s
M sequence spread spectrum order	6
Single spread spectrum symbol duration	50.4 ms
Communication band	3–5 kHz
Sampling frequency	50 kHz
Communication rate	19.8 bps (synchronous head is not included)
Channel coding	None

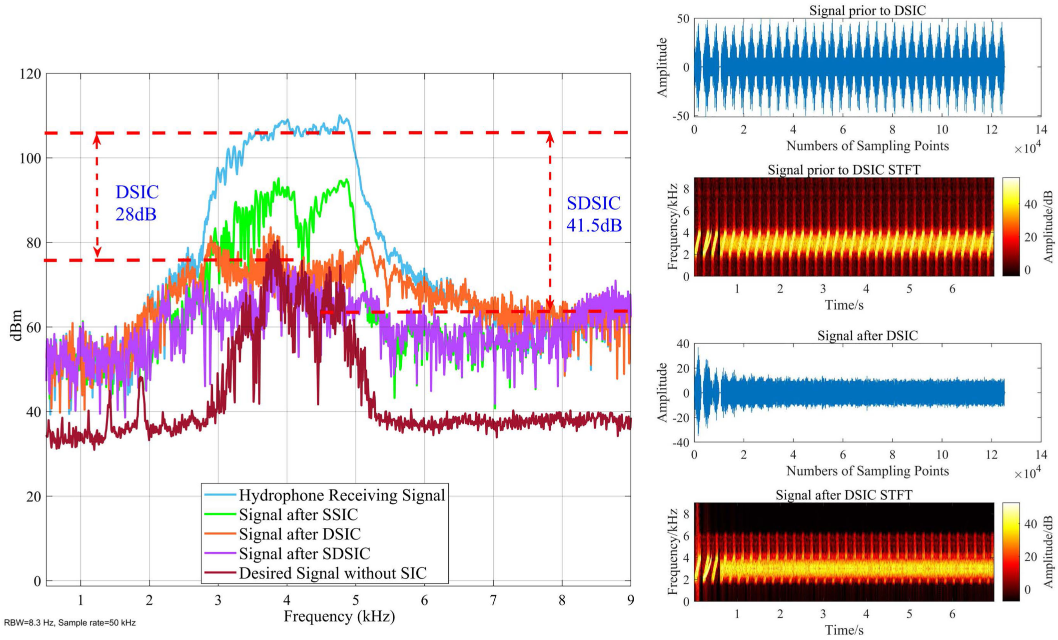


FIGURE 12
Cancellation effect of SDSIC method and traditional DSIC method.

desired signal is not known, so the SIC algorithm requires a fast offset speed to ensure that the communication information is not lost. We now illustrate the offset effect of each SIC method under the different arrival times of the desired signal. In the communication experiment, we transmitted the SI signal locally without interruption, the desired signal was transmitted randomly, and we selected the data with different desired signal arrival times for subsequent processing to obtain the BER curve

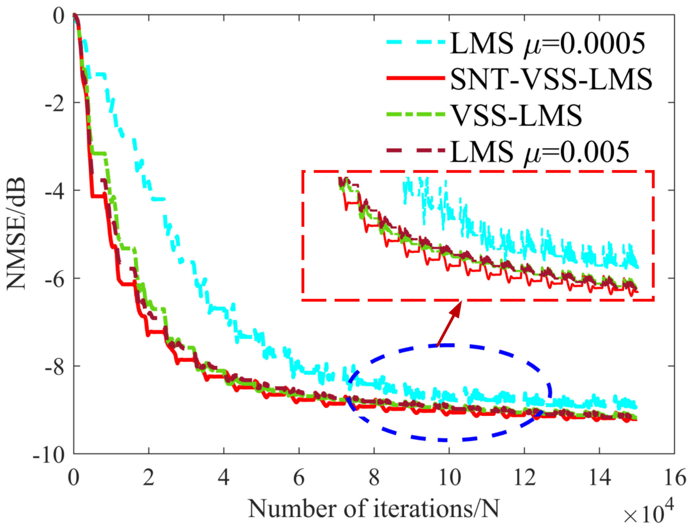


FIGURE 13
Comparison of cancellation performance between LMS, VSS-LMS, and SNT-VSS-LMS.

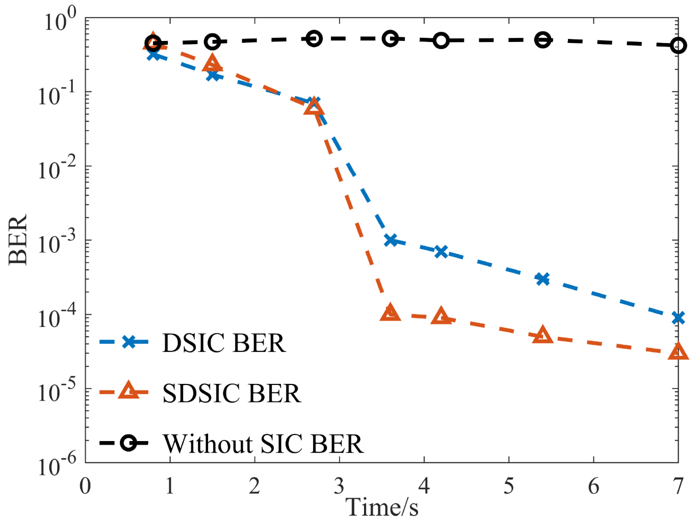


FIGURE 14
BER versus number of iterations for DSIC, SDSIC, and signal without SIC.

with respect to time. The results are shown in Figure 14, where *t* in the BER curve is the time from the start of the SIC process.

From Figure 14, a slower arrival of the desired signal results in a lower BER, because the SIC process takes some time. The proposed SDSIC significantly improves the BER performance after 2 s, but we cannot complete the communication before 2 s because the SSIC process also takes some time. When the desired signal arrives slowly, both SDSIC and DSIC have BERs of less than 10^{-3} after offset, but SDSIC achieves better BER performance. Moreover, SDSIC has a faster SIC speed when the desired signal arrives early.

6. Conclusion

This paper has described a new SIC method for the unique problems of IBFD-UWAC systems. The proposed SDSIC method, including an SSIC method that is applicable to the IBFD-UWAC, uses a beamformer based on convex optimization theory to cancel the SI signal accurately, and can ensure perfect restoration of the distant desired signal. In addition, for the RIS after SSIC, we have proposed an SNT-VSS-LMS algorithm that modifies the step-size adjustment criterion of the traditional VSS-LMS algorithm. This algorithm initially cancels most of the RIS with a rapid convergence rate, and then sets the expected threshold based on the noise floor of the RIS and the SINR under which the far-end desired signal satisfies the performance condition of the communication BER. The expected threshold value of the DSIC process is intended to minimize the impact of the desired signal for the SI channel estimation process, resulting in a smaller mean square error and the optimal steady state. An experimental verification of the proposed SDSIC method was conducted in an anechoic pool using an IBFD-UWAC system. The traditional SIC method can cancel up to 28 dB of the SI signal, while the SDSIC method proposed in this paper cancels up to 41.5 dB, significantly improving the upper limit of SIC in IBFD-UWAC systems. In addition, the proposed SNT-VSS-LMS algorithm performs optimally in terms of both convergence speed and steady-state effect. The anechoic pool experiments verify the real-time effectiveness of the proposed method. The algorithm proposed in this paper is easy to implement and has low complexity, allowing it to be implemented in engineering applications. The proposed algorithm could be directly ported to a field-programmable gate array, with toolboxes such as HDL Coder in MATLAB® used to realize the production of engineering prototypes.

Future work will investigate analog-domain SIC. Previous studies suggest that analog-domain SIC will be limited by the parameters of hardware circuits, e.g., PA, ADC sampling accuracy. Further research should investigate strong analog-domain SIC methods and attempt to combine them with SSIC methods. In this paper, DSIC was applied to the RIS signal after SSIC, and the experiments were performed in an anechoic pool. For environments with complex channels (e.g., marine environment), the published literature shows that the convergence process of LMS filters under fast time-varying channels can have a serious impact on the demodulation of information corresponding to the time variation. Thus, it would be worthwhile investigating an efficient DSIC scheme under fast time-varying channels.

Finally, the balanced relationship between analog-domain SIC and SSIC/DSIC should be considered, and the channel characteristics in the context of engineering applications should be further investigated to obtain the best SIC performance.

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

YL, CY, and YZ contributed to the conception and design of the study. GQ organized the database. YL and XQ performed the statistical analysis. YL and CY wrote the first draft of the manuscript. YL, CY, SW, and XQ wrote sections of the manuscript. All authors contributed to manuscript revision, and have read and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Heilongjiang (Grant No. LH2021F010), Natural Science Foundation (Grant No. U1806201), Open Foundation of the Key Laboratory of Underwater Information and Control (Grant No. J2322048), and Open Foundation of Key Laboratory of Underwater Acoustic Countermeasure Technology (Grant No. JCKY2022207CH01).

Acknowledgments

The authors would like to thank all those who participated in the preparation of this paper, particularly the editors and reviewers.

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, E., and Eltawil, A. M. (2015). All-digital self-interference cancellation technique for full-duplex systems. *IEEE Trans. Wirel. Commun.* 14, 3519–3532. doi: 10.1109/TWC.2015.2407876

Cacciola, R., Holzman, E., Carpenter, L., and Gagnon, S. (2016). "Impact of transmit interference on receive sensitivity in a bi-static active array system," in *2016 IEEE International Symposium on Phased Array Systems and Technology (PAST, IEEE)*, IEEE, Waltham, Massachusetts, USA Vol. 1–5. doi: 10.1109/ARRAY.2016.7832623

Chan, S. C., and Chen, H. H. (2007). Uniform concentric circular arrays with frequency-invariant characteristics–theory, design, adaptive beamforming and doa estimation. *IEEE Trans. Signal Process.* 55, 165–177. doi: 10.1109/TSP.2006.882109

Choi, J., Jain, M., Srinivasan, K., Levis, P., and Katti, S. (2010). "Achieving single channel, full duplex wireless communication," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking.* Association for Computing Machinery, New York, United States 1–12. doi: 10.1145/1859995.1859997

Everett, E., Sahai, A., and Sabharwal, A. (2014). Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Trans. Wirel. Commun.* 13, 680–694. doi: 10.1109/TWC.2013.010214.130226

Everett, E., Shepard, C., Zhong, L., and Sabharwal, A. (2016). Softnull: Many-antenna full-duplex wireless *via* digital beamforming. *IEEE Trans. Wirel. Commun.* 15, 8077–8092. doi: 10.1109/TWC.2016.2612625

Fan, Z., and Liang, G.-l. (2013). Broadband beamforming with minimum sidelobe and constant beamwidth based on convex optimization. *Acta Electon. Sin.* 41, 943. doi: 10.3969/j.issn.0372-2112.2013.05.018

Goodwin, M., and Elko, G. (1993). "Constant beamwidth beamforming," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP, IEEE, Minneapolis, MN, USA), Vol. 1. 169–172. doi: 10.1109/ICASSP.1993.319082

Kolodziej, K. E., Perry, B. T., and Herd, J. S. (2019). In-band full-duplex technology: Techniques and systems survey. *IEEE Trans. Microw. Theory Tech.* 67, 3025–3041. doi: 10.1109/TMTT.2019.2896561

Krolik, J., and Swingler, D. (1990). Focused wide-band array processing by spatial resampling. *IEEE Trans. Acoust. Speech Signal Process.* 38, 356–360. doi: 10.1109/29.103073

Li, L., Song, A., Cimini, L. J., Xia, X.-g., and Shen, C.-C. (2015). "Interference cancellation in in-band full-duplex underwater acoustic systems," in *OCEANS 2015 - MTS/IEEE Washington*. (IEEE, Washington, DC, USA) 1–6. doi: 10.23919/OCEANS.2015.7404411

Liu, R., Cui, G., Lu, Q., Yu, X., Feng, L., and Zhu, J. (2021). "Constant beamwidth receiving beamforming based on template matching," in *2021 IEEE Radar Conference (RadarConf21)*. (IEEE, Atlanta, GA, USA) 1–5. doi: 10.1109/RadarConf2147009.2021.9455192

Liu, J., Lee, J., Li, L., Luo, Z.-Q., and Wong, K. (2005). Online clustering algorithms for radar emitter classification. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1185–1196. doi: 10.1109/TPAMI.2005.166

Liu, W., and Weiss, S. (2008). Design of frequency invariant beamformers for broadband arrays. *IEEE Trans. Signal Process.* 56, 855–860. doi: 10.1109/TSP.2007.907872

Liu, W., Weiss, S., Mcwhirter, J., and Proudler, I. (2007). Frequency invariant beamforming for two-dimensional and three-dimensional arrays. *Signal Process.* 87, 2535–2543. doi: 10.1016/j.sigpro.2007.03.018

Liu, J., Zakharov, Y. V., and Weaver, B. (2009). Architecture and fpga design of dichotomous coordinate descent algorithms. *IEEE Trans. Circ. Syst. I: Reg. Pap.* 56, 2425–2438. doi: 10.1109/TCSI.2009.2015725

Li, S., Yang, X., Ning, L., Long, T., and Sarkar, T. K. (2017). "Broadband constant beamwidth beamforming for suppressing mainlobe and sidelobe interferences," in *2017 IEEE Radar Conference (RadarConf)*. (RadarConf, IEEE, Seattle, USA) 1041–1045. doi: 10.1109/RADAR.2017.7944358

Mailloux, R. (1995). "Covariance matrix augmentation to produce adaptive array pattern troughs," in *IEEE Antennas and Propagation Society International Symposium*, (IEEE, Newport Beach, CA, USA) Vol. 1. 102–105 (1995 Digest (IEEE). doi: 10.1109/APS.1995.529973

Makar, G., Tran, N., and Karacolak, T. (2017). A high-isolation monopole array with ring hybrid feeding structure for in-band full-duplex systems. *IEEE Antennas Wirel. Propag. Lett.* 16, 356–359. doi: 10.1109/LAWP.2016.2577003

Nawaz, H., and Tekin, I. (2016). "Three dual polarized 2.4ghz microstrip patch antennas for active antenna and in-band full duplex applications," in *2016 16th Mediterranean Microwave Symposium (MMS)*. (IEEE, Abu Dhabi, United Arab Emirates) 1–4. doi: 10.1109/MMS.2016.7803854

Pan, C., Jia, L., Cai, R., and Ding, Y. (2012). Modeling and simulation of channel for underwater communication network. *Int. J. innov. comput. Inf. control: IJICIC*, 2149–2156.

Qiao, G., Gan, S., Liu, S., Ma, L., and Sun, Z. (2018a). Digital self-interference cancellation for asynchronous in-band full-duplex underwater acoustic communication. *Sensors* 18, 1700. doi: 10.3390/s18061700

Qiao, G., Gan, S., Liu, S., and Song, Q. (2018b). Self-interference channel estimation algorithm based on maximum-likelihood estimator in in-band full-duplex underwater acoustic communication system. *IEEE Access* 6, 62324–62334. doi: 10.1109/ACCESS.2018.2875916

Qiao, G., Liu, S., Sun, Z., and Zhou, F. (2013). "Full-duplex, multi-user and parameter reconfigurable underwater acoustic communication modem," in *2013 OCEANS - San Diego (IEEE)*. (IEEE, San Diego, CA, USA) 1-8. doi: 10.23919/OCEANS.2013.6741096

Roe, J., Cussons, S., and Feltham, A. (1990). Knowledge-based signal processing for radar esm systems. *IEEE Proc. Part F. Commun. radar Signal Process.* 137, 293–301. doi: 10.1049/ip-f-2.1990.0045

Sabharwal, A., Schniter, P., Guo, D., Bliss, D. W., Rangarajan, S., and Wichman, R. (2014). In-band full-duplex wireless: Challenges and opportunities. *IEEE J. Sel. Area. Commun.* 32, 1637–1652. doi: 10.1109/JSAC.2014.2330193

Satyanarayana, K., El-Hajjar, M., Kuo, P.-H., Mourad, A., and Hanzo, L. (2019). Hybrid beamforming design for full-duplex millimeter wave communication. *IEEE Trans. Veh. Technol.* 68, 1394–1404. doi: 10.1109/TVT.2018.2884049

Shen, L., Henson, B., Zakharov, Y., and Mitchell, P. (2020). Digital self-interference cancellation for full-duplex underwater acoustic systems. *IEEE Trans. Circ. Syst. II: Express Briefs* 67, 192–196. doi: 10.1109/TCSII.2019.2904391

Snow, T., Fulton, C., and Chappell, W. J. (2011). Transmit–receive duplexing using digital beamforming system to cancel self-interference. *IEEE Trans. Microw. Theory Tech.* 59, 3494–3503. doi: 10.1109/TMTT.2011.2172625

Song, A., Stojanovic, M., and Chitre, M. (2019). Editorial underwater acoustic communications: where we stand and what is next? *IEEE J. Ocean. Eng.* 44, 1–6. doi: 10.1109/JOE.2018.2883872

Vallese, P., Varanese, N., and Spagnolini, U. (2017). "Self-interference cancellation for multi-antenna full duplex radio systems," in *WSA 2017; 21th International ITG Workshop on Smart Antennas*. (IEEE, Berlin, Germany) 1–6.

Yan, H., Ma, T., Pan, C., Liu, Y., and Liu, S. (2021). "Statistical analysis of time-varying channel for underwater acoustic communication and network," in *2021 International Conference on Frontiers of Information Technology*. (FIT, IEEE, Greenville, SC, USA) 55–60.

Yin, Y. L. (2016). *Multi-carrier communication and cross-layer design of underwater acoustic communication network* (Harbin Engineering University, Harbin).

Zakharov, Y. V., White, G. P., and Liu, J. (2008). Low-complexity rls algorithms using dichotomous coordinate descent iterations. *IEEE Trans. Signal Process.* 56, 3150–3161. doi: 10.1109/TSP.2008.917874

Zhang, Z., Long, K., Vasilakos, A. V., and Hanzo, L. (2016). Full-duplex wireless communications: Challenges, solutions, and future research directions. *Proc. IEEE* 104, 1369–1409. doi: 10.1109/JPROC.2015.2497203

Zhao, Y. (2021). *Research on key technology of self-interference cancellation in full duplex underwater acoustic communication* (Harbin Engineering University, Harbin).

Zhao, Y., Qiao, G., Liu, S., Zakharov, Y., and Ahmed, N. (2021). Self-interference channel modeling for in-band full-duplex underwater acoustic modem. *Appl. Acoust.* 175, 107687. doi: 10.1016/j.apacoust.2020.107687

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fmars.2022.1047332

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Lei Zhufeng,
Xi'an Aeronautical Institute, China
Qi Liu,
Southeast University, China
Shizhe Wang,
Naval University of Engineering, China

*CORRESPONDENCE
Yuxing Li
liyuxing@xaut.edu.cn

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 18 September 2022
ACCEPTED 18 October 2022
PUBLISHED 28 October 2022

CITATION
Yi Y, Li Y and Wu J (2022) Multi-scale permutation Lempel-Ziv complexity and its application in feature extraction for Ship-radiated noise.
Front. Mar. Sci. 9:1047332.
doi: 10.3389/fmars.2022.1047332

Multi-scale permutation Lempel-Ziv complexity and its application in feature extraction for Ship-radiated noise

Yingmin Yi[1,2], Yuxing Li[1,2]* and Junxian Wu[1]

[1]School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, China, [2]Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an, China

Permutation Lempel-Ziv complexity (PLZC) is a recently proposed method for analyzing signal complexity. However, PLZC only characterizes the signal complexity from single scale and has certain limitations. In order to overcome these shortcomings and improve the performance of feature extraction for underwater acoustic signal, this paper introduced coarse graining operation, proposed the multi-scale permutation Lempel-Ziv complexity (MPLZC), and proposed an automatic hybrid multi-feature extraction method for ship-radiated noise signal (S-S) based on multi-scale Lempel-Ziv complexity (MLZC), multi-scale permutation entropy (MPE) and MPLZC. The results of simulation and realistic experiments show that MPLZC can better reflect the change of signal complexity in detecting the dynamic change of signals, and more effectively distinguish white noise, pink noise and blue noise than MPE and MLZC; compared with the three feature extraction methods based on MLZC, MPE and MPLZC respectively, the proposed method has the highest recognition rates of six S-Ss under the same number of features, and the recognition rate reaches 100% when the number of features is 5; the introduction of MPLZC significantly improves the performance for ship-radiated noise signal of the automatic hybrid multi-feature extraction method. It is indicated that the proposed method, as a new underwater acoustic technology, is valid in other underwater acoustic signals.

KEYWORDS

multi-scale permutation Lempel-Ziv complexity, multi-scale permutation entropy, multi-scale Lempel-Ziv complexity, ship-radiated noise, feature extraction

Introduction

Nonlinear dynamics index is a powerful metric for analyzing signals with non-stationary, non-Gaussian and non-linear characteristics. Among them, the nonlinear dynamic indexes based on entropy and Lempel-Ziv complexity (LZC) are effective, which have been widely used in underwater acoustic signal processing, fault diagnosis and biomedicine in recent years, and have reached some achievement (Zhang et al., 2016; Li et al., 2021; Zhang et al., 2021; Shi et al., 2022a; Li et al., 2022).

The nonlinear dynamic index based on entropy can reflect the complexity of the signal. The common nonlinear dynamic indexes based on entropy include approximate entropy (AE), sample entropy (SE), and permutation entropy (PE). Pincus et al. proposed AE (Pincus, 1991), and AE does not need a large amount of data to measure the complexity of time sequences. Richman et al. proposed the SE as an improved algorithm of AE (Richman and Moorman, 2000), the calculation of SE does not depend on the length of time series, and the calculation error is small. Bandt and Pompe proposed PE (Bandt and Pompe, 2002), compared with AE and SE, PE has stronger anti-noise ability and higher calculation speed.

The nonlinear dynamic index based on LZC can reflect the irregular degree of signal (Cui et al., 2016; Li et al., 2020). The frequently used nonlinear dynamic indexes based on LZC include: LZC, weight Lempel-Ziv complexity (WLZC) and permutation Lempel-Ziv complexity (PLZC). Lempel et al. proposed LZC (Lempel and Ziv, 1976), and LZC has the advantages of easy calculation and no parameter setting (Xie et al., 2012). introduced the concept of weight into LZC and proposed the WLZC, WLZC can better detect the changes of time sequences through weighted calculation. Bai proposed the PLZC the by combining the PE and LZC (Bai et al., 2015), and the proposed PLZC can more simply, stably and effectively quantize the dynamic changes of signals.

However, for the nonlinear dynamic indexes based on entropy and LZC, both of which are difficult to fully reflect the signal effective information from a single scale, many scholars introduced the coarse graining operation into these indexes (Anne, 2020), including multi-scale permutation entropy (MPE) and multi-scale Lempel-Ziv complexity (MLZC). Previous studies have shown that the MLZC achieved satisfactory results in the fields of biomedicine and mechanical fault diagnosis (Yeh and Shi, 2018; Yan et al., 2021; Shi et al., 2022b), and MPE has been applied in the fields of underwater acoustic signal with excellent performance (Choi et al., 2016; Liu et al., 2017; Chen et al., 2019).

Ship-radiated noise signal (S-S) as a kind of underwater acoustic signal (Li et al., 2019; Zhang et al., 2020; Esmaiel et al., 2022), which can indicate the physical characteristics of ships (Wang et al., 2017; Yang et al., 2022). The extraction of nonlinear dynamic indexes from S-Ss are helpful to the classification and recognition of different ships (Bao et al., 2010). It has been found that dispersion entropy-based Lempel-Ziv complexity (DELZC) is introduced into the feature extraction method for S-Ss, which effectively improves the recognition rate of different S-Ss (Li and Geng, 2022). In the past decade, PE and PE-based improved multi-scale algorithms have been applied successively in the field of feature extraction for S-Ss, such as MPE and multi-scale weight permutation entropy (MWPE), and the separability of extracted features for S-Ss is improved (Li and Li Y, 2016; Li and Li Y, 2017; Xie et al., 2021). As the results shows that the introduction of multi-scale in entropy improved the feature extraction performance of S-Ss. However, multi-scale also has not been introduced into PLZC at present.

In order to comprehensively reflect the complexity of PLZC and improve the feature extraction performance of S-Ss, we proposed the multi-scale permutation Lempel-Ziv complexity (MPLZC) by introducing coarse graining operation, in addition, we combined with MLZC, MPE and MPLZC to propose the automatic hybrid multi-feature extraction method for feature extraction of S-Ss. The overall structure of this paper is as follows: Section 2 introduces the principle of MPLZC and the automatic hybrid multi-feature extraction method; in Section 3, the effectiveness of MPLZC is verified by simulation signal experiment; Section 4 performs single feature extraction experiments and multi-feature extraction experiments of six S-Ss and classification experiments based on K-Nearest Neighbor (KNN) classifier; finally, Section 5 gives the main conclusions of the full paper.

Theory and method

Multi-scale permutation Lempel-Ziv complexity

MPLZC is used to detect the dynamic changes of nonlinear systems, which combines the coarse graining operation, the permutation pattern of PE and LZC. The specific calculation steps of MPLZC are as follows:

(1) For a given series $X=\{x_1,x_2,\ldots,x_N\}$, the series $Y^{(g)}=\{y_1^{(g)},y_2^{(g)},\ldots,y_j^{(g)},\ldots,y_l^{(g)}\}$ is obtained by coarse graining operation. $y_j^{(g)}$ can be expressed as:

$$y_j^{(g)}=\frac{1}{g}\sum_{i=(j-1)g+1}^{jg}x_i,\quad 1\le j\le l \tag{1}$$

where g represents the scale factor, l is equal to $[\frac{N}{g}]$, and $[\frac{N}{g}]$ represents the largest integer not greater than $\frac{N}{g}$. If $g=1$, $Y^{(g)}$ is the same as the original sequence X.

(2) For series $Y^{(g)}$, phase space reconstruction is performed to obtain the following reconstructed time series matrix:

$$\begin{bmatrix} y^{(g)}(1) & y^{(g)}(1+\tau) & \ldots & y^{(g)}(1+(m-1)\tau) \\ y^{(g)}(2) & y^{(g)}(2+\tau) & \ldots & y^{(g)}(2+(m-1)\tau) \\ y^{(g)}(3) & y^{(g)}(3+\tau) & \ldots & y^{(g)}(3+(m-1)\tau) \\ \vdots & \vdots & \ldots & \vdots \\ y^{(g)}(k) & y^{(g)}(k+\tau) & \ldots & y^{(g)}(k+(m-1)\tau) \end{bmatrix} \tag{2}$$

where m is the embedding dimension; τ is the delay time, and k represents the number of reconstructed row vectors which value of k is $N - (m-1)\tau$.

(3) Each reconstructed row vector is rearranged in ascending order as follows:

$$y^{(g)}(i+(j_1-1)\tau) \le y^{(g)}(i+(j_2-1)\tau) \le \ldots$$
$$\le y^{(g)}(i+(j_m-1)\tau), \quad 1 \le i \le k \tag{3}$$

if the values of two elements are equal in the same row vector as:

$$y^{(g)}(i+(j_1-1)\tau) = y^{(g)}(i+(j_2-1)\tau), \quad (j_1 < j_2) \tag{4}$$

their order can be arranged as follows:

$$y^{(g)}(i+(j_1-1)\tau) \le y^{(g)}(i+(j_2-1)\tau) \tag{5}$$

any reconstructed line vector corresponds to a symbol sequence $s(i)$:

$$s(i) = [j_1, j_2, \ldots, j_m] \tag{6}$$

where $i = 1,2,\ldots,k$, and $k \le m!$. The matrix composed of these symbol sequences is defined as:

$$S = \begin{bmatrix} s(1) \\ \vdots \\ s(k) \end{bmatrix} \tag{7}$$

(4) Matrix S has k symbol sequences, each symbol sequence is numbered according to the class of corresponding patterns, and there are $m!$ classes of patterns in total. For example, for a given finite symbol sequence with $m = 3$, there are a total of $m! = 6$ patterns, the corresponding relationships are [1, 2, 3]→1, [1, 3, 2]→2, [2, 1, 3]→3, [2, 3, 1]→4, [3,1, 2]→5, [3, 2, 1]→6, and there are similar corresponding relationships when m is set to other values.

(5) After the conversion in step (4), the numbers corresponding to each $s(i)$ are arranged in rows. Then we get a pattern sequence G with length k, and all the elements are integers between 1 to $m!$.

(6) Take G as a new sequence and perform LZC calculation on it, more specifically, its normalized complexity is as follows:

$$LZC(G) = \frac{c(k)}{\lim_{k\to\infty} c(k)} \approx \frac{c(k)\log_{m!}k}{k} \tag{8}$$

where $c(k)$ represents the complexity of sequence G by process of LZC, and the specific steps of LZC are described in reference (Lempel and Ziv, 1976)

(7) The LZC of the series Gunder different g are calculated, and the MPLZC of the original series X is obtained, which can be expressed as:

$$MPLZC(X,g) = LZC(G) \tag{9}$$

Automatic hybrid multi-feature extraction method

Flow chart of the automatic hybrid multi-feature extraction for S-Ss is shown in Figure 1, the specific steps are as follows:

(1) Input the different types of S-Ss.
(2) Calculate the MLZC, MPE and MPLZC of various S-Ss under different scale factors (SFs) and mix them to obtain hybrid features (HF).
(3) Initialize the number of extracted features $i = 3$.
(4) Judge whether i is greater than 10, if it is greater than 10, the process ends; otherwise, execute the next step.
(5) Select i features of various S-Ss, and KNN classifier is used to classify S-Ss.
(6) Output the highest average recognition rate (ARR) under the current i, judge whether ARR is equal to 100%, if the ARR reaches 100%, the process ends; otherwise, the i is increased by 1 and return to step (4).

Simulation signal complexity analysis

MIX signal

To verify the effectiveness of PLZC in reflecting the complexity change when the random sequence becomes a periodic deterministic time series, Mixed signal is used in this experiment, which can be defined as:

$$\begin{cases} MIX(t) = (1-z) \times x(t) + z \times y \\ x(t) = \sqrt{2}\sin(2\pi t/12) \end{cases} \tag{10}$$

where z is a random variable that decreases from 0.99 to 0.01, $x(t)$ is a periodic function, and y is a variable evenly distributed in $[-\sqrt{3}, \sqrt{3}]$. In the process of signal generation, the greater the z value, the stronger the randomness; the smaller the z value, the stronger the periodicity.

The sampling frequency of the Mix signal is 1000Hz, and the total sampling time is 15s. The sliding window with 1000 sampling points is used to move sampling at the overlap rate

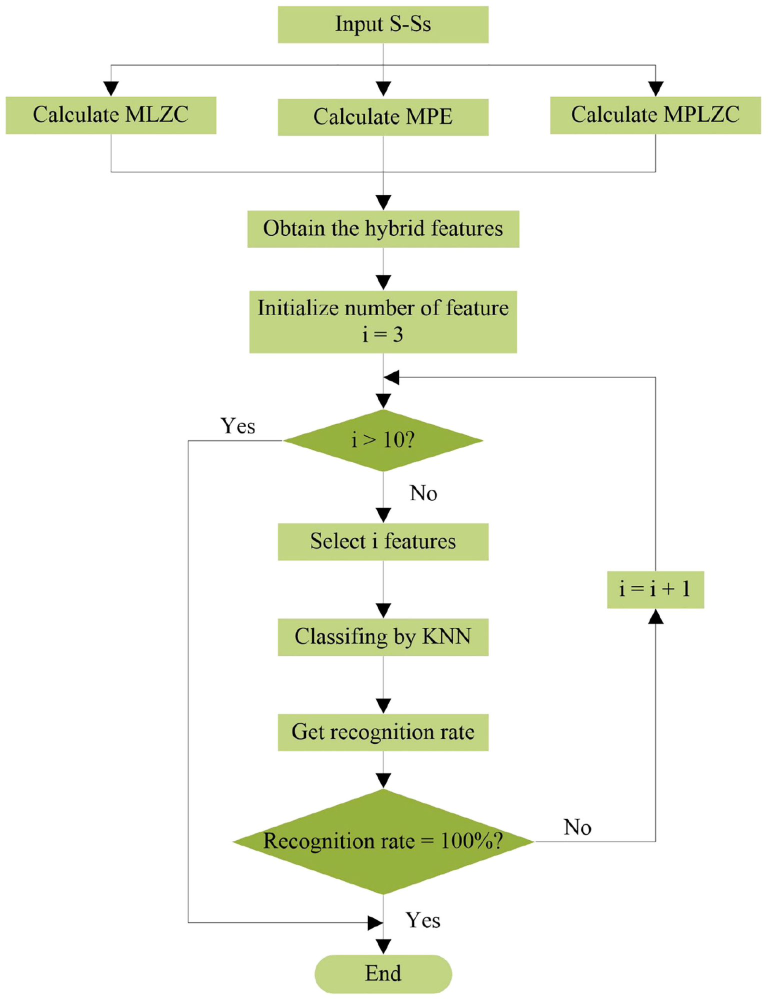


FIGURE 1
Flow chart of the automatic hybrid multi-feature extraction for S-Ss.

of 90%, a sliding window is a sample, a total of 140 samples. Three kinds of complexity of each sample are calculated respectively, in which the embedding dimension m of PE and PLZC are set to 4, and the delay time τ are set to 1. Figure 2 shows Waveform and complexity curves Mix signal: (a) Waveform; (b) Complexity curves. It can be seen from Figure 2B that the increase of periodicity for Mix signal with the time increase, and the complexity of Mix signal should also reduce.

It can be seen from Figure 2B that PE, PLZC and LZC all have a downward trend, which can reflect the complexity change of Mix signal; the PLZC curve is smoother and has less fluctuation than LZC and PE when falling. The experimental results indicate that PLZC can better reflect the complexity change of Mix signal.

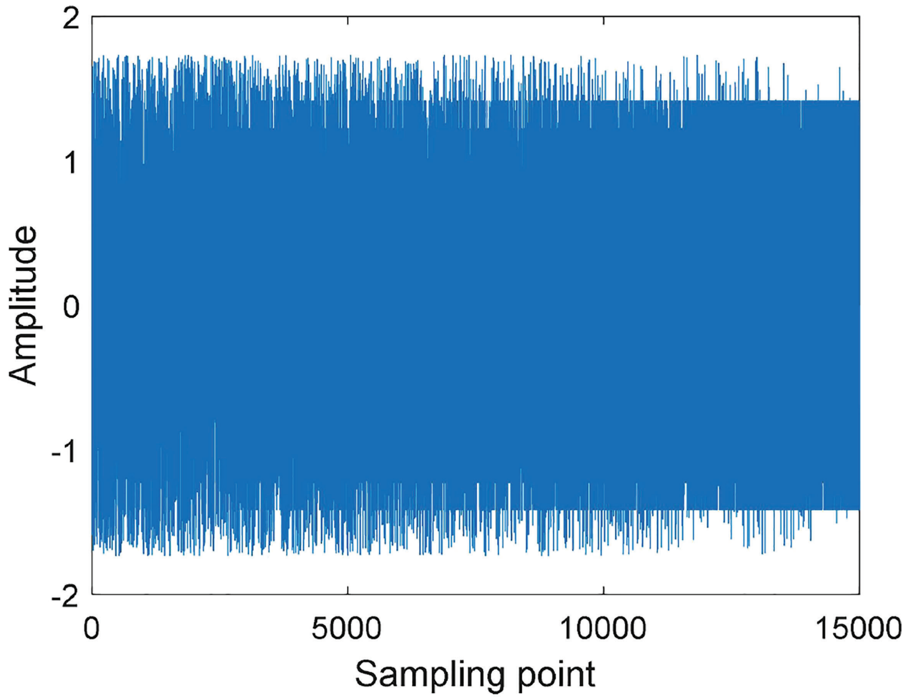
Logistic model

To verify the effectiveness of PLZC, a comparative experiment was performed to detect the dynamic changes of logistic model. Logistic model can be defined as:

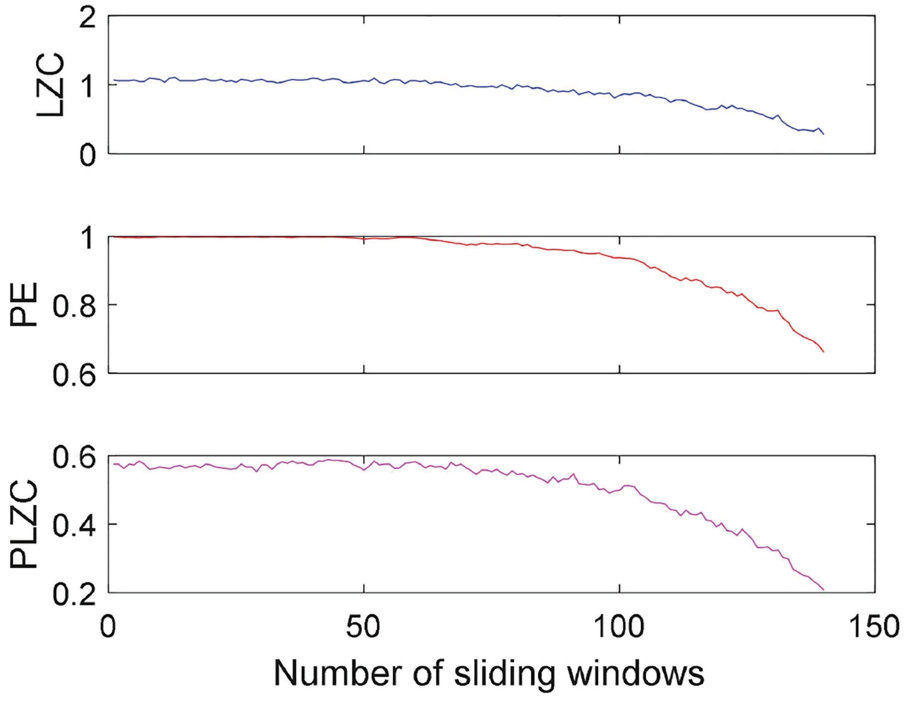
$$x_{i+1} = u \times x_i \times (1 - x_i) \tag{11}$$

where $\boldsymbol{x}_1$ is 0.65 $\boldsymbol{u}$ is the control parameter with the range of [3.5,4]. LZC, PE and PLZC are calculated respectively, the parameters of PE and PLZC are set to m = 4 and τ = 1. Figure 3 depicts waveform and complexity curves of logistic model: (a) Waveform; (b) Complexity curves.

It can be seen from Figure 3A that the randomness of logistic model becomes generally higher with the increase of u. It can be concluded from Figure 3B that the curves of LZC, PE and PLZC

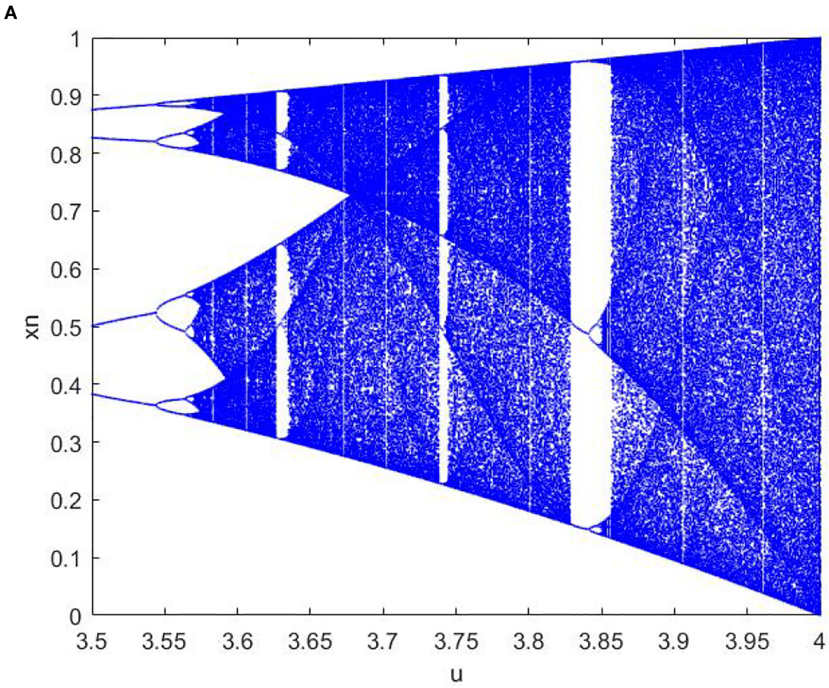


A Waveform

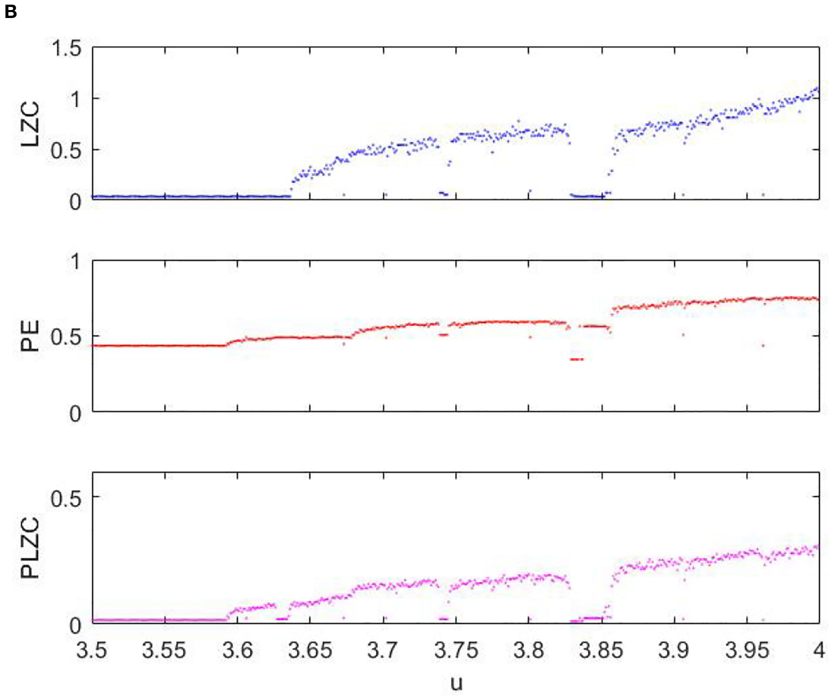


B Complexity curves

FIGURE 2
Waveform and complexity curves Mix signal: **(A)** Waveform; **(B)** Complexity curves.



Waveform



Complexity curves

FIGURE 3
Waveform and complexity curves of logistic model: **(A)** Waveform; **(B)** Complexity curves.

all have an upward trend in general; LZC cannot detect the complexity upward trend of logistic model when u is in the range of 3.57 to 3.62, only PLZC can accurately reflect the complexity downward trend of logistic model when u near 3.63. Experimental results show that PLZC is more sensitive to complexity changes of logistic model than PE and LZC, and can better describe dynamic changes for logistic model.

Different noise signals

To confirm the ability of MPLZC to discriminate different noise signals, a comparison experiment was conducted for pink noise, white Gaussian noise (WGn) and blue noise. Figure 4 shows the normalized results of three types for noise signal.

20 samples are taken for the three noise signals separately, and the number of sampling points of each sample is 5000. MLZC, MPE and MPLZC of each sample are calculated respectively. The SF of MLZC, MPE and MPLZC changes from 1 to 20, the embedding dimension m of MPE and MPLZC are set to 2, and the delay time τ are set to 1. Figure 5 is means and standard deviations of complexity for three noise signals under different SFs.

From Figure 5, the mean and standard deviation curves of MPE are overlapped for pink noise, white noise and blue noise; for MLZC and MPLZC, mean and standard deviation curves of three noise signals all have significant difference. The experimental results suggest that compared with MPE, MLZC and MPLZC have better ability to distinguish three noise signals, especially when SF is less than 5.

Feature extraction of S-Ss

Six types of S-S

Six types of S-Ss are used in this feature extraction experiment with the sampling frequency of 44.1kHz and sampling interval is [1, 200000] which are from the online website (National Park Service) and called as Ship-①, Ship-②, Ship-③, Ship-④, Ship-⑤ and Ship-⑥ respectively. Figure 6 shows the normalized results of six types of S-Ss.

Single feature extraction and classification

100 samples are selected for each S-S, and the sampling points of each sample is 2000. MLZC, MPE and MPLZC of the six kinds of S-Ss are extracted with the SF from 1 to 10. For comparison and analysis, the parameters of MPE and MPLZC

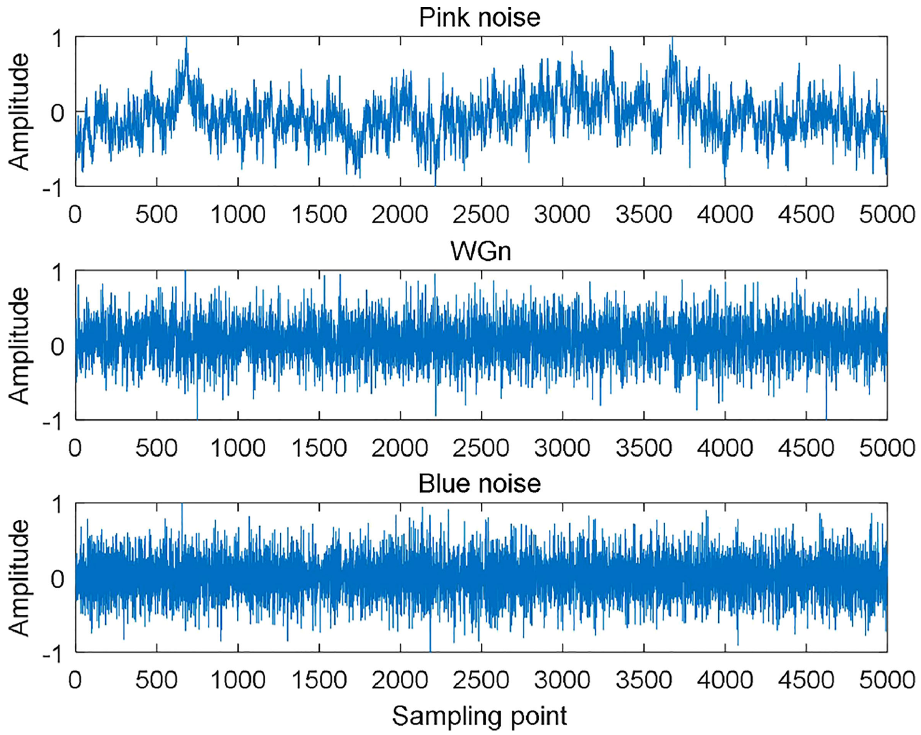
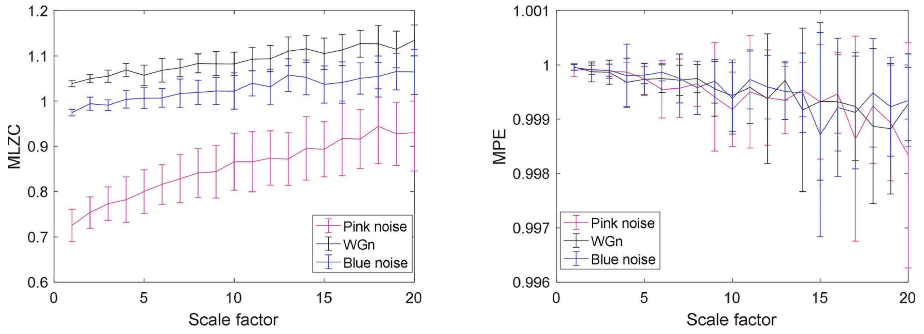
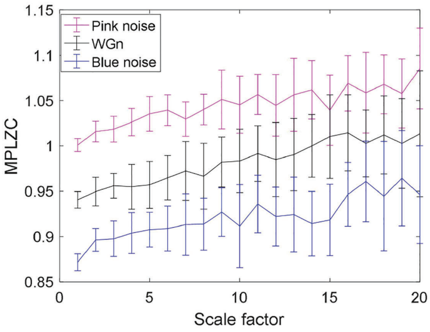


FIGURE 4
The normalized results of three types for noise signal.



A MLZC **B** MPE



C MPLZC

FIGURE 5
Means and standard deviations of complexity for three noise signals under different SFs: **(A)** MLZC, **(B)** MPE, **(C)** MPLZC.

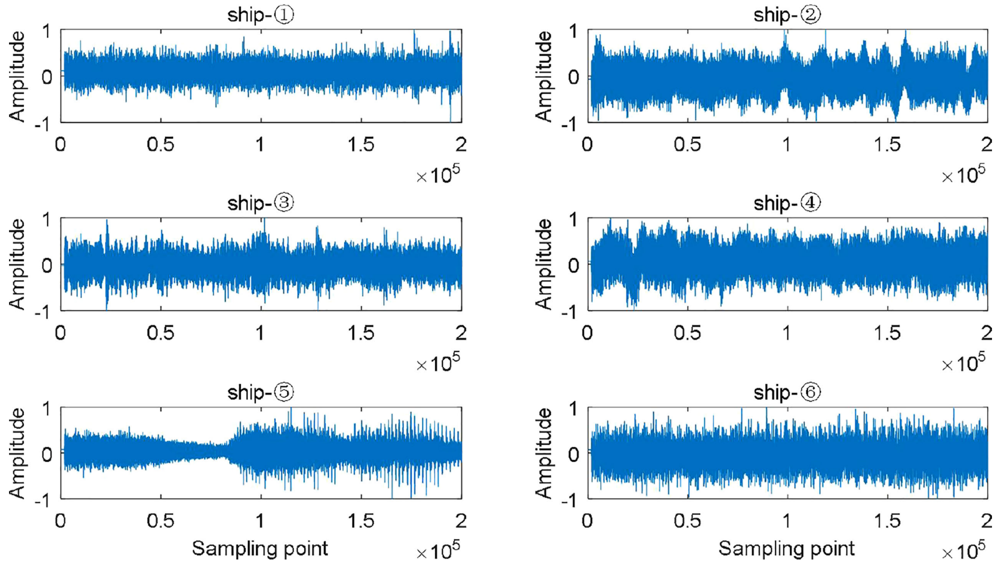
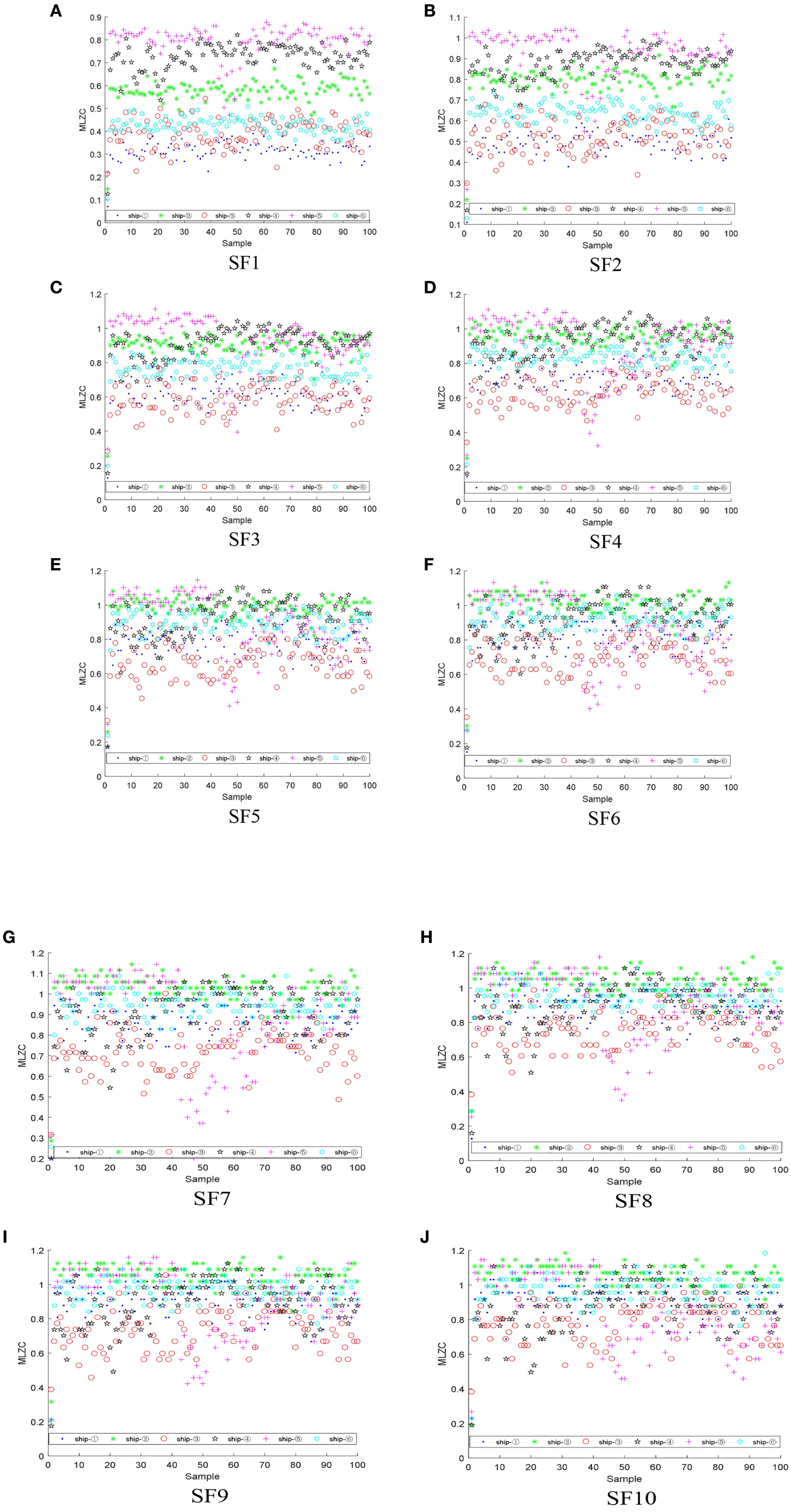


FIGURE 6
The normalized results of six types of S-Ss.



SF1 SF2 SF3 SF4 SF5 SF6 SF7 SF8 SF9 SF10

FIGURE 7
The feature distributions of MLZC for six S-Ss under each SF:**(A)** SF1, **(B)** SF2, **(C)** SF3, **(D)** SF4, **(E)** SF5, **(F)** SF6, **(G)** SF7, **(H)** SF8, **(I)** SF9, **(J)** SF10.

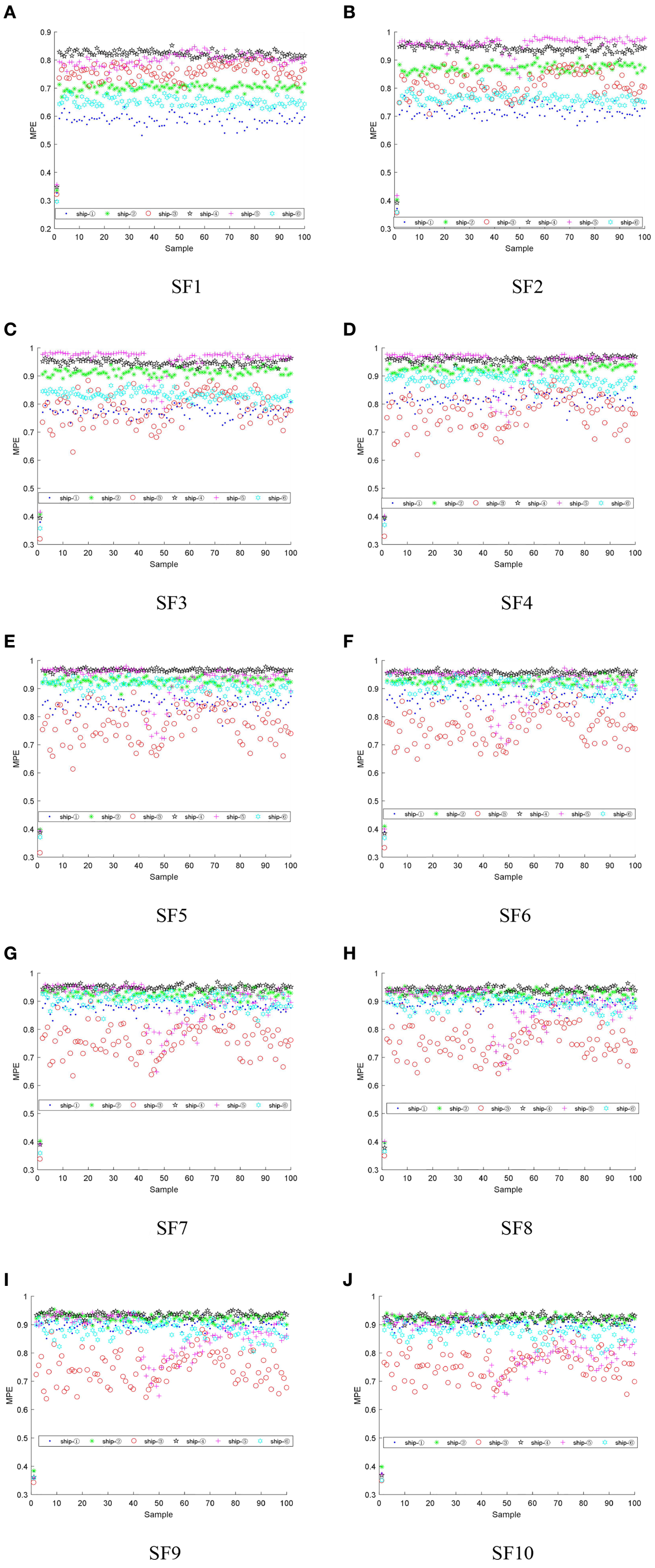
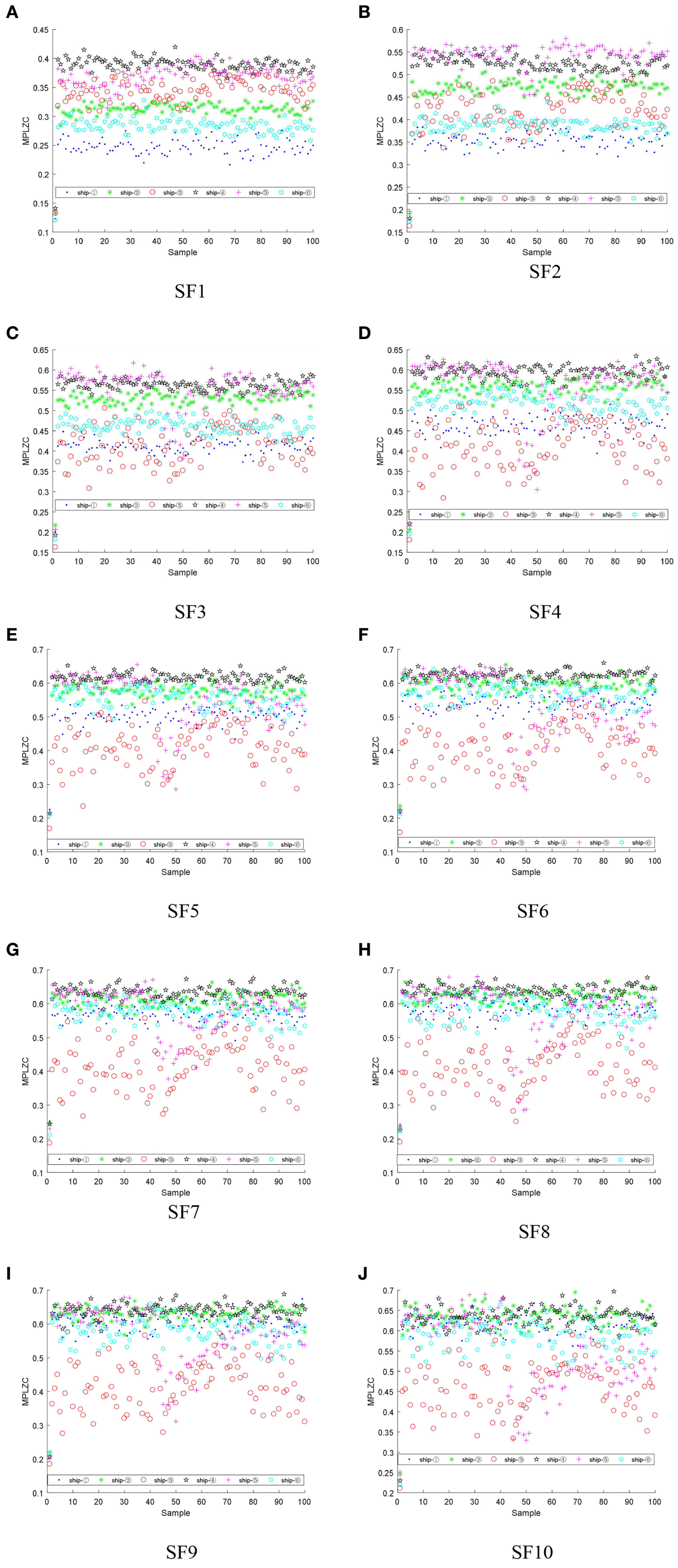


FIGURE 8
The feature distributions of MPE for six S-Ss under each SF:**(A)** SF1, **(B)** SF2, **(C)** SF3, **(D)** SF4, **(E)** SF5, **(F)** SF6, **(G)** SF7, **(H)** SF8, **(I)** SF9, **(J)** SF10.



SF1 SF2 SF3 SF4 SF5 SF6 SF7 SF8 SF9 SF10

FIGURE 9
The feature distributions of MPLZC for six S-Ss under each SF:**(A)** SF1, **(B)** SF2, **(C)** SF3, **(D)** SF4, **(E)** SF5, **(F)** SF6, **(G)** SF7, **(H)** SF8, **(I)** SF9, **(J)** SF10.

are set to $m = 5$ and $\tau = 1$. Figures 7–9 demonstrate the feature distributions of MLZC, MPE and MPLZC for six S-Ss under each SF, where SF1 represents scale factor 1, SF2 represents scale factor 2 and so on.

From Figures 7–9, for MLZC, MPE and MPLZC, the overlapping phenomenon of the feature distributions for the six S-Ss gradually became serious with the increase of SF; for MPE and MPLZC, compared with MLZC, the feature dstributions of the six S-Ss have more significant differences under each SF; compared with other SFs, the feature distributions of MPE for six S-Ss have less overlapping parts under SF1, SF2 and SF3, especially ship-② and ship-⑥; from SF4 to SF8, the MPE of ship-④ and ship-⑤ are approximate, which is difficult to distinguish these two S-Ss; for MPLZC, the distinguishing effect between the feature distributions of ship-② and ship-④ under SF2 is better than that under SF1; the difference between the MPLZC of ship-② and ship-③ under SF1 is more obvious than that in SF2; from SF5 to SF10, the MPLZC distributions of ship-② and ship-⑤ are seriously overlapped, which is difficult to distinguish these two S-Ss. In summary, under different SF, MLZC, MPE, and MPLZC also have different ability to distinguish different S-Ss, only the single feature is adopted, which is difficult to distinguish the six S-Ss.

To more clearly compare the ARRs of six S-Ss, six S-Ss are classified by KNN classifier (Venkatesan et al., 2018). 100 samples of each S-S are selected, of which the 50 samples are used as training samples and the other 50 samples are test samples. Table 1 shows the ARRs of six S-Ss under each SF.

From Table 1, with the increase of SF, the ARR of single feature extraction methods based on MLZC, MPE and MPLZC is decreased generally; for the three feature extraction methods based on MLZC, MPE and MPLZC, the highest ARR is 63.7% under SF1, 75.3% under SF2 and 68% under SF1. In conclusion, the ARRs of the six S-Ss are lower than 80%, and single feature extraction method is difficult to accurately identify various S-Ss.

Automatic hybrid multi-feature extraction and classification

The automatic hybrid multi-feature extraction and recognition method is adopted to further improve the recognition rates of six S-S, and three multi-feature extraction methods based on MLZC, MPE and MPLZC are used as comparative experiments. Figure 10 shows that the triple feature distributions corresponding to the highest ARRs of six S-Ss, MLZC1 is the MLZC under SF1, MLZC2 is the MLZC2 under SF2 and so on, the same is true for MLZC and MPLZC. Table 2 shows that the ARRs of the triple feature extraction methods for six S-Ss.

As can be seen from Figure 10 and Table 2, compared with the single feature distributions, the triple feature distributions of six S-Ss have stronger separability; for ship-①, the recognition rate of the multi-feature extraction method based on MLZC is the highest; from ship-② to ship-⑥, the recognition rates of multi-feature extraction method based on MLZC are the lowest, and the ARRs of the automatic hybrid multi-feature extraction method are the highest; compared with the other three multi-feature extraction methods, the automatic hybrid multi-feature extraction method based on HF has the highest average ARR for six S-Ss. The analysis indicate that the automatic hybrid multi-feature extraction method can better identify the six S-Ss.

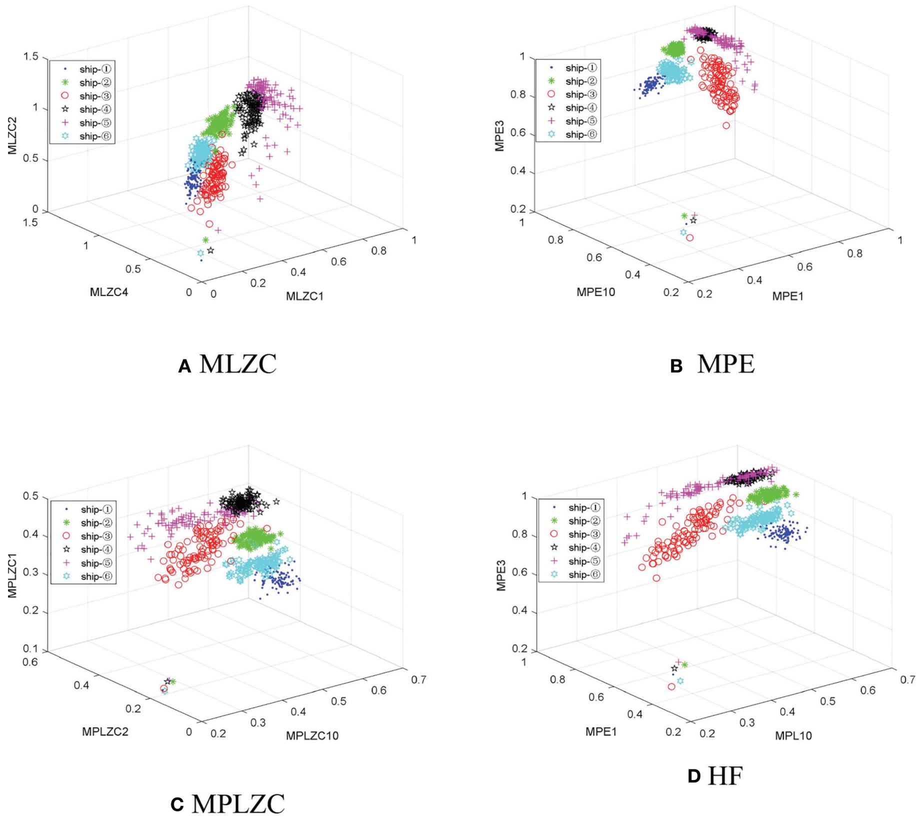
Since the ARR of the six S-Ss do not reach 100% and the number of extracted features is also less than 10, the number of extracted features is increased and the ARRs are calculated. Table 3 shows the highest ARRs of multi-features for six S-Ss.

According to Table 3, under the same number of features, the ARRs of multi-feature extraction method based on MLZC are the lowest for six S-Ss, and the ARRs of the automatic hybrid multi-feature extraction method are the highest; for proposed method and multi-feature extraction method based on MPE, the highest ARR increases with the increase of the number of extracted features; when the number of extracted features is 5, the ARR of the proposed method reaches 100%. Hence, the automatic hybrid multi-feature extraction method proposed in this paper has the best feature extraction performance.

Table 4 shows the feature combinations corresponding to the multi-feature extraction methods. It can be concluded from Table 4, for the feature extraction methods based on MLZC, MPE and MPLZC, the features of the same SF can be selected under different number of extracted features, such as MLZC1; for the automatic hybrid multi-feature extraction method based on HF, MLZC is not selected when the number of extracted features is 3, and MPE is not selected when the number of

TABLE 1 The ARRs of six S-Ss under each SF.

Type	**Average recognition rate (%)**									
	SF1	**SF2**	**SF3**	**SF4**	**SF5**	**SF6**	**SF7**	**SF8**	**SF9**	**SF10**
MLZC	63.7	50.7	36.7	34.7	35.7	31.0	31.3	24.3	28.0	19.3
MPE	72.0	75.3	68.3	58.0	49.3	49.3	49.3	38.3	39.0	38.0
MPLZC	68.0	63.3	54.7	50.3	48.7	45.7	37.3	40.7	36.3	37.0



A MLZC **B** MPE **C** MPLZC **D** HF

FIGURE 10
The triple feature distributions under the highest ARRs of six S-Ss: **(A)** MLZC, **(B)** MPE, **(C)** MPLZC, **(D)** HF.

TABLE 2 The ARRs of the triple feature extraction methods for six S-Ss.

Triple feature	Recognition rate						Averagerecognition rate
	ship-①	ship-②	ship-③	ship-④	ship-⑤	ship-⑥	
MLZC	98.0%	94.0%	90.0%	82.0%	56.0%	82.0%	83.7%
MPE	94.0%	100%	100%	98.0%	98.0%	94.0%	97.3%
MPLZC	92.0%	100%	98.0%	100%	90.0%	94.0%	95.7%
HF	94.0%	100%	100%	100%	100%	96.0%	98.3%

TABLE 3 The highest ARRs of multi-features for six S-Ss.

Multi-feature	Number of extracted features		
	3	4	5
MLZC	83.7%	81.7%	80.0%
MPE	97.3%	98.3%	98.7%
MPLZC	95.7%	97.0%	96.3%
HF	98.3%	99.3%	100%

TABLE 4 The feature combinations corresponding to the multi-feature extraction methods.

Multi-feature	**Number of extracted features**		
	3	**4**	**5**
MLZC	MLZC1, MLZC2, MLZC4	MLZC1, MLZC2, MLZC4, MLZC8	MLZC1, MLZC2, MLZC4, MLZC6, MLZC8
MPE	MPE1, MPE3, MPE10	MPE1, MPE2, MPE4, MPE10	MPE1, MPE2, MPE3, MPE5, MPE8
MPLZC	MPLZC1, MPLZC2, MPLZC10	MPLZC1, MPLCL2, MPLZC4, MPLZC10	MPLZC1, MPLZC2, MPLZC3, MPLZC4, MPLZC10
HF	MPE1, MPE3, MPLZC10	MLZC1, MPLZC1, MPLZC2, MPLZC9	MLZC1, MPE3, MPE10, MPLZC1, MPLZC4

extracted features is 4; MPLZC is selected in the automatic hybrid multi-feature extraction method under the different number of extracted features. The results indicate that MPLZC plays a significant role in automatic hybrid multi-feature extraction, and the introduction of MPLZC improves the S-S feature extraction performance.

Conclusions

This paper raised the MPLZC as well as employed it in the field of underwater acoustic, and presented an automatic hybrid multi-feature extraction method for S-Ss. The superiority of MPLZC in feature extraction for S-Ss and feasibility of the proposed method are verified by simulation and realistic experiments, and the main conclusions are as follows:

(1) PLZC is a complexity metric that combines LZC and PE, which has certain advantages in detecting dynamic changes for nonlinear sequences. Experiments show that PLZC can more accurately detect the complexity changes of Mix signal and logistic model than LZC and PE.

(2) This paper presented MPLZC as a complexity feature, which is on the basis of PLZC and combines with coarse graining operation. Compared with MLZC and MPE, MPLZC can more effectively distinguish pink noise, white Gaussian noise and blue noise.

(3) Combining with MLZC, MPE and MPLZC, a new automatic multi-feature extraction method for S-Ss is proposed. The highest ARR of proposed method for six S-Ss is higher than other three methods under the same number of features, and the ARR reaches 100% when five features are extracted.

(4) Under each feature number, the feature combinations corresponding to the highest ARRs of automatic hybrid multi-feature extraction method for S-Ss all include MPLZC. Therefore, the introduction of MPLZC improves the S-S feature extraction performance.

In the future research, we will apply the automatic hybrid multi-feature extraction method to other underwater acoustic signals, such as marine background noise, marine animals, etc.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.nps.gov/glba/learn/nature/soundclips.

Author contributions

YY provided the research object and original data, YL determined the experimental method and operation of the experiment, and JW completed the writing of this paper. All authors contributed to the article and approved the submitted version.

Funding

Key Research and Development Plan of Shaanxi Province (2020ZDLGY06-01) Key Scientific Research Project of Education Department of Shaanxi Province (21JY033) Science and Technology Plan of University Service Enterprise of Xi'an (2020KJRC0087) Natural Science Foundation of Shaanxi Province (grant no. 2022JM-337, 2021JQ-487, and 2020JQ-644).

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

Anne, H. (2020). Multiscale entropy approaches and their applications. *Entropy* 22 (6), 644. doi: 10.3390/e22060644

Bai, Y., Liang, Z., and Li, X. (2015). Permutation lempel-ziv complexity measure of electroencephalogram in GABAergic an aesthetics. *Physiol. Meas.* 36, 2483–2501. doi: 10.1088/0967-3334/36/12/2483

Bandt, C., and Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 88, 174102. doi: 10.1103/PhysRevLett.88.174102

Bao, F., Li, C., Wang, X., Wang, Q., and Du, S. (2010). Ship classification using nonlinear features of radiated sound: An approach based on empirical mode decomposition. *J. Acoust. Soc. America* 128, 206–214. doi: 10.1121/1.3436543

Chen, Z., Li, Y., Cao, R., Ali, W., Yu, J., and Liang, H. (2019). A new feature extraction method for ship-radiated noise based on improved CEEMDAN, normalized mutual information and multiscale improved permutation entropy. *Entropy* 21 (6), 624. doi: 10.3390/e21060624

Choi, Y., Hyun, K., and Choi, J. (2016). Assessing multiscale permutation entropy for short electroencephalogram recordings. *Cluster Comput.* 19, 2305–2314. doi: 10.1007/s10586-016-0648-8

Cui, L., Gong, X., and Zhang, J. (2016). Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the lempel-ziv complexity. *J. Sound Vib.* 385, 372–388. doi: 10.1016/j.jsv.2016.09.008

Esmaiel, H., Xie, D., Qasem, Z., Sun, H., Qi, J., and Wang, J. (2022). Multi-stage feature extraction and classification for ship-radiated noise. *Sensors* 22 (1), 112. doi: 10.3390/s22010112

Lempel, A., and Ziv, J. (1976). On the complexity of finite sequences, inform. *Theory IEEE Trans.* 22, 75–81. doi: 10.1109/TIT.1976.1055501

Li, Y., Gao, P., Tang, B., Yi, Y., and Zhang, J. (2021). Double feature extraction method of ship-radiated noise signal based on slope entropy and permutation entropy. *Entropy* 24 (1), 22. doi: 10.3390/e24010022

Li, Y., and Geng, B. (2022). 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, Y., and Li Y, C. X. (2016). Feature extraction of ship-radiated noise based on permutation entropy of the intrinsic mode function with the highest energy. *Entropy* 18 (11), 393. doi: 10.3390/e18110393

Li, Y., and Li Y, C. X. A. (2017). Novel feature extraction method for ship-radiated noise based on variational mode decomposition and multi-scale permutation entropy. *Entropy* 19 (7), 342. doi: 10.3390/e19070342

Li, Z., Li, Y., Zhang, K., and Guo, J. (2019). A novel improved feature extraction technique for ship-radiated noise based on IITD and MDE. *Entropy* 21 (12), 1215. doi: 10.3390/e21121215

Li, Y., Mu, L., and Gao, P. (2022). Particle swarm optimization fractional slope entropy: A new time series complexity indicator for bearing fault diagnosis. *Fractal Fract.* 6 (7), 345. doi: 10.3390/fractalfract6070345

Liu, T., Yao, W., Wu, M., Shi, Z., Wang, J., and Ning, X. (2017). Multiscale permutation entropy analysis of electrocardiogram. *Phys. A.: Stat. Mech. its Appl.* 471, 492–498. doi: 10.1016/j.physa.2016.11.102

Li, Y., Wang, S., and Deng, Z. (2020). Intelligent fault identification of rotary machinery using refined composite multi-scale lempel-ziv complexity. *JManufSyst* 61, 725–735. doi: 10.1016/j.jmsy.2020.05.004

National park service. Available at: https://www.nps.gov/glba/learn/nature/soundclips.htm.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U. States America* 88 (6), 2297–2301. doi: 10.1073/pnas.88.6.2297

Richman, J., and Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circulatory Physiol.* 278 (6), H2039–H2049. doi: 10.1152/ajpheart.2000.278.6.H2039

Shi, J., Su, Z., Qin, H., Shen, C., Huang, W., and Zhu, Z. (2022a). Generalized variable-step multiscale lempel-ziv complexity: A feature extraction tool for bearing fault diagnosis. *IEEE Sens. J.* 22 (15), 15296–15305. doi: 10.1109/JSEN.2022.3187763

Shi, J., Su, Z., Qin, H., Shen, C., Huang, W., and Zhu, Z. (2022b). Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Sens. J.* 22 (15), 15296–15305. doi: 10.1109/JSEN.2022.3187763

Venkatesan, C., Karthigaikumar, P., and Varatharajan, R. (2018). A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection. *Multimed. Tools Appl.* 77, 10365–10374. doi: 10.1007/s11042-018-5762-6

Wang, Q., Zeng, X., Wang, L., Wang, H., and Cai, H. (2017). Passive moving target classification *via* spectra multiplication method. *IEEE Signal Process. Lett.* 24 (4), 451–455. doi: 10.1109/LSP.2017.2672601

Xie, D., Hong, S., and Yao, C. (2021). Optimized variational mode decomposition and permutation entropy with their application in feature extraction of ship-radiated noise. *Entropy* 23 (5), 503. doi: 10.3390/e23050503

Xie, F., Yang, R., and Zhang, B. (2012). Analysis of weight lempel-ziv complexity in piecewise smooth systems of DC-DC switching converters. *Acta Phys. Sin.* 61, 110504. doi: 10.7498/aps.61.110504

Yang, H., Li, L., Li, G., and Guan, Q. (2022). A novel feature extraction method for ship-radiated noise. *Def. Technol.* 18 (4), 604–617. doi: 10.1016/j.dt.2021.03.012

Yan, X., She, D., Xu, Y., and Jia, M. (2021). Application of generalized composite multiscale lempel–ziv complexity in identifying wind turbine gearbox faults. *Entropy* 23 (11), 1372. doi: 10.3390/e23111372

Yeh, C., and Shi, W. (2018). Generalized multiscale lempel–ziv complexity of cyclic alternating pattern during sleep. *Nonlinear Dyn.* 93, 1899–1910. doi: 10.1007/s11071-018-4296-9

Zhang, Y., Wei, S., Di Maria, C., and Liu, C. (2016). Using lempel–ziv complexity to assess ECG signal quality. *J. Med. Biol. Eng.* 36, 625–634. doi: 10.1007/s40846-016-0165-5

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

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Nav.* 14 (7), 1055–1060. doi: 10.1049/iet-rsn.2019.0477

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 02 November 2022
DOI 10.3389/fmars.2022.1019385

Motion parameter estimation of AUV based on underwater acoustic Doppler frequency measured by single hydrophone

Shaowei Rong[1] and Yifeng Xu[2]*

[1]Northwestern Polytechnical University, Xi'an, China, [2]College of Information Engineering (Wintec International College), Jinhua Polytechnic, Jinhua, China

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Sheng Miao,
Southwest Forestry University, China
Ruiping Song,
Xiamen University, China

*CORRESPONDENCE
Yifeng Xu
xuyifeng123@mail.nwpu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 15 August 2022
ACCEPTED 01 September 2022
PUBLISHED 02 November 2022

CITATION
Rong S and Xu Y (2022) Motion parameter estimation of AUV based on underwater acoustic Doppler frequency measured by single hydrophone. *Front. Mar. Sci.* 9:1019385. doi: 10.3389/fmars.2022.1019385

This paper describes the use of a single hydrophone to estimate the motion parameters of an autonomous underwater vehicle (AUV) from the underwater acoustic signal excited by its propulsion motor. First, the frequency range of the hydroacoustic signal radiated by the AUV motor is determined, and a detection and recognition model is designed. In the case of uniform linear motion of the AUV, the geometric relationship between the Doppler frequency shift curve of the sound source is derived and the motion model of the sound source and sound line propagation is established. An estimation algorithm for the motion parameters of multiple AUVs based on data from a single hydrophone is derived. Then, for Doppler underwater acoustic signals disturbed by independent identically distributed noise with an arbitrary probability distribution, a cumulative phase difference power amplification instantaneous frequency estimation method is proposed. This method is based on the sum of multiple logarithmic functions. Finally, the effectiveness and accuracy of the algorithm in estimating the motion parameters of multiple AUVs are verified through simulations and experiments.

KEYWORDS

AUV motion parameter estimation, underwater acoustic measurements, single hydrophone Doppler measurement, accumulated logarithmic product sum ratio, multiple AUVs

1 Introduction

The cooperative use of autonomous underwater vehicles (AUVs) can expand the application range of underwater ocean unmanned systems (Nielsen et al., 2018). AUVs are usually propelled underwater by a motor. This means that the navigation noise of AUVs is low (Zhang et al., 2020a), which increases the difficulty of underwater acoustic AUV detection. However, the vast majority of AUVs use pulse width modulation (PWM) to control their motors, which will produce unique frequency characteristics. Lo et al.

(2000); Acarnley and Watson (2006); Le Besnerais et al. (2009); Hara et al. (2018); Railey et al. (2020), and De Viaene et al. (2018) hhave studied the analytical characteristics of the Maxwell radial vibration caused by PWM power supplies in the motor, focusing on the spatial mode, frequency, voltage, reference update time, the interaction of PWM harmonics, and the noise radiated by PWM converters when feeding. Railey et al. (2020) and Lindberg et al. (2022) investigated the unique high-frequency underwater acoustic characteristics of brushless DC motor propulsion systems for the passive measurement of unmanned underwater vehicles, and determined the main characteristics of the noise to be the PWM frequency of the motor and the strong tone at its second harmonic, and the sideband spacing at the propeller rotation frequency multiplied by the number of poles in the motor.

If a traditional array is used to detect AUVs, the detection distance means that a large underwater acoustic array would be required, making the detection cost very high. Coraluppi et al. (2018); Lexa et al. (2020); Zhu et al. (2021), and Brinkmann and Hurka (2009) described the use of distributed multi-target tracking and passive sonar tracking with fixed and mobile sensors. Single-sensor narrowband and wideband measurements can be used for spatial tracking, and a distributed underwater monitoring network composed of a single hydrophone can reduce the detection costs considerably.

In recent years, the characteristics of active Doppler sonar have been applied to estimate the velocity by accurately estimating the Doppler frequency (Yang and Fang, 2021; Saffari et al., 2022). Further, hydrophones have been used to estimate the motion information of aerial sound sources. Buckingham et al. (2002) measured helicopter flight data and obtained audio signals in air, seawater, and seabed sediments. The measured data show that the signal has time-varying characteristics, which verifies the Doppler frequency shift signal generated by the sound source and illustrates the possibility for underwater acoustic detection. For a single hydrophone, Ferguson used ray theory to analyze a sound propagation model in an air–water medium, and deduced a method of determining the flying parameters on the top of air targets according to the Doppler frequency shift in a two-dimensional domain (Ferguson, 1992). Li et al. (2021a) developed a probabilistic multi-hypothesis tracker algorithm based on a batch recursive extended Rauch–Tung–Striebel–Smooth algorithm to process a large number of passive measurements including clutter. Wong and Chu (2002) reported a positioning method using a single vector hydrophone combined with a correlation spatial matched filter beam pattern and a minimum variance distortion free response. Underwater detection algorithms based on a single hydrophone can only estimate the flight parameters of the sound source in a two-dimensional plane, that is, it is assumed that the helicopter only flies in a single plane above the hydrophone, without considering the yaw distance and other information in the three-dimensional domain. Single-vector hydrophones can estimate the azimuth information, but there is a directional blind spot that is vulnerable to the influence of background noise in practical applications. Although a hydrophone array can estimate azimuthal information, it is not suitable for small underwater detection platforms because of its high cost and inconvenient layout.

Detecting the relative motion between a radiation noise source e (Buckingham et al., 2002; Qiao et al., 2021; Gusland et al., 2021; Hanif et al., 2022), the water surface, and an underwater vehicle containing the receiving hydrophone will produce the Doppler effect, and the signal frequency collected by the hydrophone will be different from that of the original radiation sound source. This time-varying frequency difference is the Doppler frequency shift, and contains state information of the speed, distance, depth, direction of motion, and trajectory of the moving noise source (Lo, 2017). Accurate Doppler instantaneous frequency (IF) estimations can obtain accurate motion information. To obtain accurate IF information, many classical methods have been proposed. The short-time Fourier transform (STFT) and continuous wavelet transform (CWT) are essentially linear transforms, characterized by static resolution in the time–frequency plane. However, due to the limitation of the Heisenberg–Gabor inequality, neither STFT nor CWT can achieve good resolution in both the time domain and frequency domain, which means that good time resolution means poor frequency resolution (Liu et al., 2016; Weber and Oehrn, 2022). In recent years, polynomial chirplet transforms, based on linear frequency chirplet transforms, have become a generalized IF parameter estimation method that can be used to describe the IF characteristics of a variety of analytical signals (Zhou et al., 2018). The synchronous compression transform only compresses the IF coefficients to the frequency modulated (FM) track in the frequency direction, and can redistribute or compress the IF curve coefficients to the FM track through STFT and CWT, similar to the ideal IF estimation (Yu et al., 2017; Li et al., 2021b). However, this method is limited by the low IF resolution of the window function. Stankovic and Katkovink proposed a Wigner–Ville distribution (WVD) with an adaptive window length and a WVD with variable and data-driven window lengths (Stankovic and Katkovnik, 1999). An algorithm based on the TFD peak is also applicable to other WVD representations, including polynomial WVD, which is unbiased for nonlinear FM trajectories, and pseudo WVD and smooth pseudo WVD, which can suppress the crossover terms caused by noise in the time–frequency plane.

In this paper, an underwater acoustic Doppler model of AUV motion is established. The motion situation of the AUV is estimated by using the Doppler frequency shift measured by a single hydrophone. An algorithm based on the phase difference power cumulative logarithmic product sum ratio is proposed. The cumulative phase difference between the FM signal and the kernel function removes the average value, and it is proved that its power is a monotonic function of the difference between the estimated frequency value and the actual frequency value. The cumulative multiple logarithm algorithm of cumulative phase difference power plays the role of nonlinear amplification and maintains monotonicity. For the differential frequency characteristics (Fang et al., 2016; Qiao et al., 2020), the Adabellie algorithm is used to maximize the sum of multiple logarithms, and the estimated coefficients can be fitted with kernel functions.

The remainder of this paper is organized as follows. In section 2, the motor model of the AUV is established, and the hydroacoustic Doppler frequency shift characteristics of the system are analyzed. Additionally, the motion positions of the AUV are retrieved according to the Doppler frequency shift model. In section 3, the intermediate frequency estimation model is established, and the idea and principle of the accumulate logarithmic product sum ratio of power (PAL-PSR) algorithm are derived. Sections 4and 5 verify the effectiveness of the algorithm through simulations and experiments, respectively. Finally, some concluding remarks are given in Section 6.

2 Doppler shift model of multi-AUV motion

In the process of underwater AUV detection and identification, a single array hydrophone is used to monitor and identify AUVs in an area of water. The use of single array hydrophone reduces energy consumption and processing units, and is superior for concealed detection and identification of AUVs. However, AUVs are now mainly propelled by brushless DC (BLDC) motors and permanent magnet synchronous motors (PMSM), which enable low-speed, low-noise and long-term underwater operation. Therefore, it is difficult for a single element hydrophone sensor to passively search for AUVs. Because AUVs are mainly driven by BLDC motors, however, the motor itself has many signal characteristics. In most cases, AUVs perform straight-line navigation or a serpentine search pattern in which a long route is traversed at a constant speed. Based on these conditions, the following assumptions can be stated:

Assumption 1: The AUV uses BLDC or PMSM as the main propulsion motor.

Assumption 2: The position of the AUV is random at the initial search time.

Assumption 3: The AUV moves in a straight line at a uniform speed during the search process.

2.1 AUV motor characteristics

The AUV system has two modules with significant size and power, both of which generate mechanical noise: the driving motor for stabilizing the control surface and the motor for propulsion. The noise of an AUVs rudder drive motor is mainly broadband transient noise, which is concentrated in the low-frequency region from 10 Hz–10 kHz. The motor speed of BLDC motors and PMSMs is controlled by PWM, which is an effective method for this purpose. The harmonic of the PWM switching frequency in the phase current will produce radial vibration. The vibration caused by radial electromagnetic force under the PWM current harmonic produces strong noise in BLDC motors, PMSMs, and induction motors. The PWM output frequency is a constant, usually above 10 kHz and below 30 kHz.

In BLDC motors and PMSMs, electronic commutation is carried out through quasi square wave six step commutation. The coils are energized in turn and interact with the permanent magnets on the rotor, resulting in the rotation of the motor. Since the phase current in the quasi square wave commutation of the motor cannot rise and fall instantaneously, pulsating torque will be generated during rotation.

In Railey et al. (2020), the radial electromagnetic forces created by the air-gap flux density in motors are a function of the pole combination, current harmonics, and back EMF harmonics. These radial forces F_{rmnl} lead to deformation in the stator core, causingvibration. The radial force harmonics F_{rm} can be written as

$$F_{rm} = \sum_{n}\sum_{l} F_{rmnl} \cos\left[np\omega_r t - \left(np \pm lN_s\right)\theta\right] \tag{1}$$

where p is the number of poles, is the slot number, $u = 2n + 1$, $k = n$, $(n = 0,1,2\ldots)$ are the harmonics, and ωr is the rotation frequency of the motor in radians, $f1 = \omega / 2\pi$. Therefore, the frequency and mode number of the radial force harmonics are

$$\left[npf_1, -np \pm lN_s\right] \tag{2}$$

with $n,l = 0,1,2\ldots$. Thus, when the motor rotates, the excitation frequency f_s of the radial force on each stator tooth is

$$f_p = pf_s \tag{3}$$

where p is the number of poles in the motor and f_s is the shaft rotation frequency. The motor speeds for both BLDC motors and PMSMs are controlled using PWM. Although PWM is an efficient method for controlling the motor speed, the harmonics of the PWM switching frequency in the phase current produce radial vibrations.

Remark 1. According to Eq. (3), when the velocity of the AUV is constant, the frequency of the propeller and motor is also constant. At the same time, the pulse frequency generated by torque ripple and pole pairs will be modulated in the vibration frequency generated by PWM. Through the modulation frequency and PWM signal, the motion characteristics of the AUV can be determined, and different AUVs can also be identified. Figure 1 shows the PWM signal of a BLDC motor and the FM signal generated by the propeller and motor cogging torque.

Figure 1 indicates that a large part of the AUV noise source vibrates below the motor tone. We can determine the PWM switching frequency, PWM, and its multiples. In these propulsion systems, the PWM driving the modulated voltage signal of the BLDC motor is the main noise source. In addition, there are spacing sidebands, f_p and f_s , both around f_PWM . We have experimentally verified that f_p is equal to f_s multiplied by the number of permanent magnet poles, p , in the motor. When the sideband spacing f_s also appears in the motor noise signal, p can be determined.

The AUV is passively tracked by a single element hydrophone. When the AUV is driven by a BLDC motor, we can detect and classify the PWM through its unique high-frequency harmonic acoustic characteristics. Combined with the Doppler signal of PWM generated by AUV movement, the AUV speed can be estimated. Extending these two detection protocols to any AUV scene can realize multi-source tracking and classification.

2.2 Design of AUV detection model for multiple distributed single element hydrophones

In Figure 2, the underwater monitoring network composed of multiple single hydrophones can track AUVs in a large range, thus realizing inexpensive and wide coverage underwater monitoring.

A single hydrophone can calculate the AUV velocity and the distance to the hydrophone from the AUV Doppler frequency. Therefore, according to the condition of Assumption 3, monitoring an AUV using two distributed hydrophones will enable the course of the AUV to be estimated from the tangent of the circle formed by the shortest distance between the AUV and the hydrophone. With three hydrophones, it becomes possible to completely determine the route of the AUV, except in one special case. When the AUV course is parallel to the straight line formed by the three hydrophones, only the course can be judged, and two routes will be formed. Figure 3 shows a schematic diagram of the AUV trajectory calculated by the hydrophone network measurement system.

Lemma 1. According to the common tangent calculation method of (Haiying and Zufeng, 2014), when the position coordinates of R1, R2 which are the yaw distance between the AUV and the hydrophone estimated by the Doppler acoustic signal of the AUV motor of a single hydrophone in Figure 3, and the two hydrophones are known, the parameter equations of the four common tangents L1, L2, L3, and L4 can be obtained. In Figure 3A, L1 and L2 are the outer tangents of two circles, and L3 and L4 are the inner tangents of two circles. The outer tangent length is $l_{\text{out}} = \sqrt{d^2_{H_1H_2} - (R_1 - R_2)^2}$. The inner tangent length is $l_{inner} = \sqrt{d^2_{H_1H_2} - (R_1 + R_2)^2}$. According to the time sequence and time difference of the AUV detected and identified by the hydrophone, combined with the velocity of the AUV calculated from the Doppler frequency, the direction of the AUV can be calculated, and two routes can be excluded, that is, L1 and L2 or L3 and L4 are retained. When three hydrophones detect the AUV, the entire AUV trajectory can be calculated, except for the situation shown in Figure 2C.

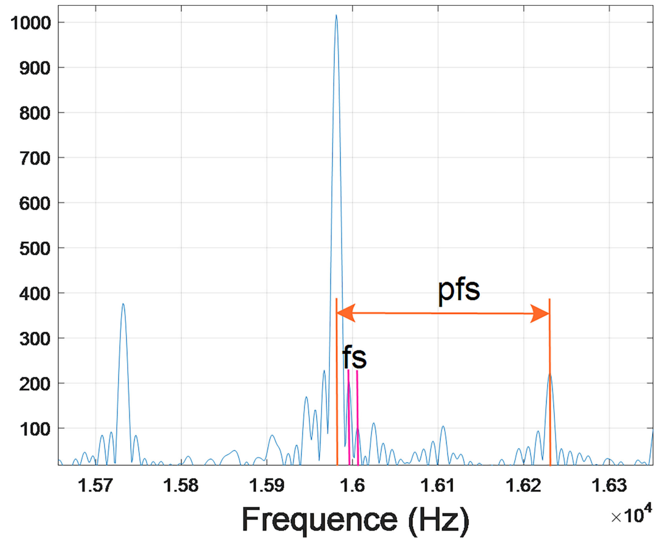


FIGURE 1
AUV propulsion motor PWM frequency and harmonic frequency.

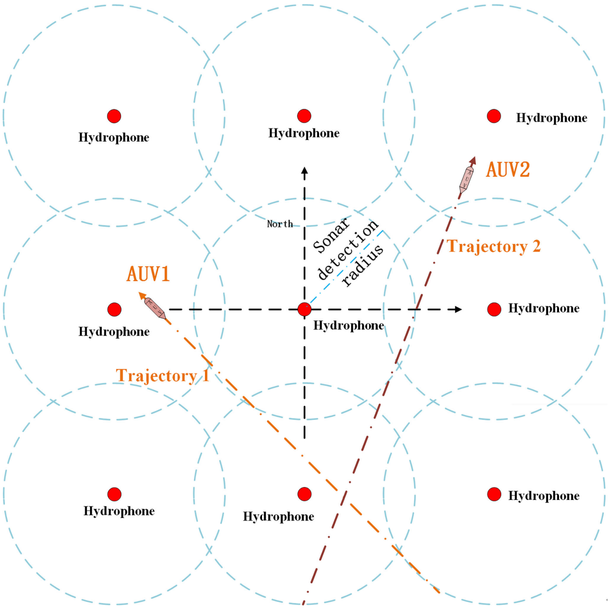


FIGURE 2
Multiple AUV trajectories and hydrophone position.

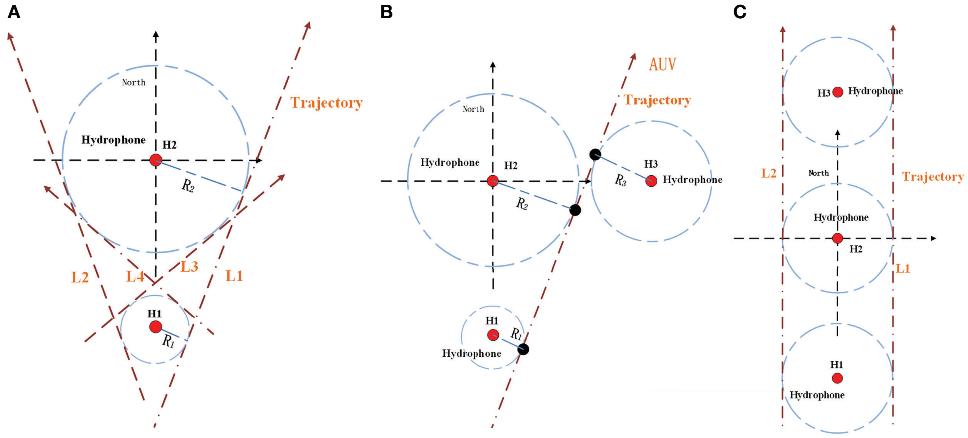


FIGURE 3
(A) Four possible trajectories of AUV; **(B)** AUV has only one possible trajectory; **(C)** Two possible trajectories of AUV.

2.3 Doppler model for single hydrophone detecting multiple AUV motions

According to Assumptions 1–3, the AUV trajectories form straight lines and the velocities remain constant.

The AUV travels in a straight line at a constant speed, and the position of the hydrophone is the coordinate origin. Figure 4 shows the model of a single AUV detected by a single hydrophone. A single hydrophone can only estimate the velocity of the AUV and the yaw position of the shortest-range hydrophone for a linear moving target through the Doppler frequency. The AUV moves in a straight line with a velocity of v_a . Define the vertical distance between the AUV sailing direction and the hydrophone as the yaw distance R_{min} between the AUV and the hydrophone, i.e., segment HA3. When displayed in the AUV velocity coordinate system, the AUV Doppler model can be simplified in Figure 4B.

According to Figure 4A, the location of the AUV is (X_A, Y_A) and the location of the hydrophone is (0, 0) with

$$\begin{cases} X_A(t) = X_A(0) + v_A t \cos\theta_A \\ Y_A(t) = Y_A(0) + v_A t \sin\theta_A \end{cases} \tag{4}$$

The AUV velocity v_A is constant and the direction in which the AUV is sailing θ_A is constant. The AUV motion coordinates are $(X_A(t), Y_A(t))^{\mathrm{T}}$. When the coordinates of the AUVchange with time, the line of sight (LOS) angle of the AUV relative to the hydrophone also changes. The LOS angle is expressed as

$$\lambda_A(t) = \arctan\left(\frac{Y_A(t)}{X_A(t)}\right) \tag{5}$$

Here, the LOS angle between the AUV and the hydrophone is $\lambda_A(t)$. The projection angle α_A of the velocity vector in the LOS direction between the AUV and the hydrophone is

$$\cos\alpha_A = \frac{\vec{v}_a \cdot (-X_A(t), -Y_A(t))^T}{\left|(X_A(t), Y_A(t))^T\right| |\vec{v}_{ui}|} \tag{6}$$

In addition, according to the position relationship between the moving sound source and the hydrophone node in the top view shown in Figure 4A, the deflection angle can be determined. When the moving point sound source is on the left or right side of $A_1A_2A_3$, the velocity and time are considered to be negative or positive, respectively. The position relationship between the moving point sound source and the stationary node is shown in Figure 4A. Using simple trigonometry, the deflection angle $\alpha(t)$ as a function of time is

$$\alpha(t) = \arccos\frac{vt}{\sqrt{v^2t^2 + R_{min}^2}} \tag{7}$$

The instantaneous frequency $f_d(t)$ of the AUV sound source received by the hydrophone is

$$f_d(t) = \frac{f_0}{1 - \frac{\vec{v}_a \cos\alpha_A}{c_w}} \tag{8}$$

where c_w is the underwater sound velocity.

As the AUV moves along a straight line, the sound source deviation angle and LOS angle observed by the hydrophone will change as the relative position between the sound source and the hydrophone varies. In the coordinate system shown in Figure 4, the distance between the sound source and the hydrophone is relatively small, the relative velocity is positive, and the Doppler frequency shift is positive. As the distance between the sound source and the hydrophone increases, the relative velocity and the Doppler shift become negative. When the hydrophone is perpendicular to the course of the AUV, the relative velocity is zero and the Doppler frequency shift is zero. At this time, the distance between the AUV and the hydrophone reaches its minimum.

Figure 5 shows the theoretical values and first and second derivatives of the Doppler frequency shift curve received by the hydrophone before and after a sound source with a resonant frequency of 16 kHz passes through the nearest point (located at a distance of 200 m). The positive and negative values at time t_s cause the asymmetry of the frequency offset. When the AUV is at the same distance from the acoustic receiver h , the frequency offset when it is close to the state is greater than that when it is away from the state.

From the relationship between Eq. (4) to Eq. (8), it can be seen that the Doppler frequency shift of the AUV motor underwater acoustic signal is related to the velocity and yaw distance of the AUV. The AUV heading information in Figure 4A cannot be calculated. The actual calculation is completed using the simplified model in Figure 4B.

The natural frequency f_0 and velocity v_A of a point sound source can be calculated from the asymmetry of the curve with respect to time t_s , i.e., positive and negative, and the Doppler frequency shift equation when the sound source is infinite.

$$\begin{cases} f_0 = 2\frac{f_d(-\infty) f_d(+\infty)}{f_d(-\infty) + f_d(+\infty)} \\ v_A = c_w \frac{f_d(-\infty) - f_d(+\infty)}{f_d(-\infty) + f_d(+\infty)} \end{cases} \tag{9}$$

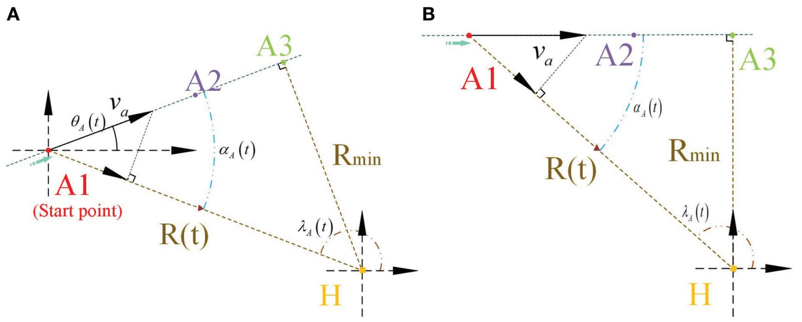


FIGURE 4
Top view of an AUV's radiated noise propagation and hydrophone nodes. **(A)** AUV radiated noise propagation and hydrophone node in Fixed-frame. **(B)** AUV radiated noise propagation and hydrophone node in AUV velocity frame.

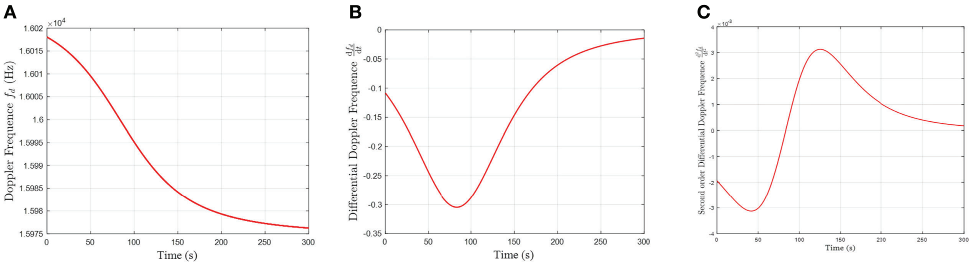


FIGURE 5
(A) Doppler frequency and its **(B)** first and **(C)** second derivatives.

Set the time $t = s$ corresponding to the natural frequency f_0 . Then, the first derivative and the second derivative of the Doppler frequency shift $f_d\,(t)$ can be calculated, and the time at which the pole value is zero is substituted into

$$\begin{cases} \frac{df_d}{dt}\,|_{t=\tau_0} = f_d'\,(\tau_0 \mid f_0,\, v,\, h,\, w_{\min}\,) \\ \frac{d^2 f_d}{dt^2}\,|_{t=\tau_{-1}} = f_d''\,(\tau_{-1} \mid f_0,\, v,\, h,\, w_{\min}\,) \end{cases} \tag{10}$$

Here, τ_0 is the time corresponding to the minimum of $f_d^{\}}\,(t)$ and τ_{-1} is the time corresponding to the minimum value of $f_d^{\}}\,(t)$. Thus, the shortest horizontal projection of the point sound source can be obtained. In Eq. (14), the first derivative of the Doppler frequency shift $f_d\,(t)$ is

$$\frac{df_d}{dt} = -f_0 \frac{\frac{\vec{v}_A \sin\alpha_A}{c_w}\,\dot{\alpha}_A}{\left(1 - \frac{\vec{v}_A \cos\alpha_A}{c_w}\right)^2} \tag{11}$$

where c_w is the underwater sound velocity, $\vec{v}_A$ is velocity of AUV.The second derivative of the Doppler shift $f_d\,(t)$ is

$$\frac{d^2 f_d}{dt^2} = -f_0 \left[\frac{c_w \vec{v}_A\left(\ddot{\alpha}_A \sin\alpha_A + \dot{\alpha}_A^2 \cos\alpha_A\right)}{\left(c_w - \vec{v}_a \cos\alpha_A\right)^2} - 2\,\frac{c_w \vec{v}_A^2 \dot{o} t \alpha_A^2 \sin^2\alpha_A}{\left(c_w - \vec{v}_A \cos\alpha_A\right)^3}\right] \tag{12}$$

3 Multi-frequency estimation design

When a single hydrophone receives navigation noise from multiple AUVs, it can be considered as a multi-component signal. The multi-component signal can be expressed as the linear superposition of multiple single-component signals. By decomposing a multi-component signal into multiple single-component signals, the advantages of the time–frequency method can be effectively used to realize the time–frequency analysis of the multi-component signals. This allows multiple AUVs to be distinguished and the motion parameter information of each AUV to be obtained.

According to the Doppler frequency shift model presented in the previous section and the design principles of AUV motors, combined with Assumptions 1 to 3, when the motor speed of an AUV is constant, Eq. (3) implies that the harmonic frequency difference between f_{pwm} and f_p is constant. Therefore, an AUV can be identified by a constant frequency difference, and the change trend of the harmonic Doppler frequency of f_{pwm} and f_p from the same AUV is also consistent. Thus, the Doppler frequency estimation of the harmonics of f_{pwm} and f_p AUV becomes particularly important.

Multi-component signals are a linear superposition of multiple single-component signals, that is,

$$s\,(t) = \sum_{k=1}^{N} A_k \exp\,[j\phi_k\,(t)] = \sum_{k=1}^{N} A_k \exp\left[j2\pi \int \omega\,(t)\,dt\right] \tag{13}$$

The multi-component signal can be reduced to the superposition of multiple signals, and each single signal is a variable that needs to be accurately estimated. As the single Doppler signals are time-varying, the dispersed Doppler signal can be represented by an FM signal. According to [34], the FM signal can be expanded as an n -order Taylor series ($n \le l$) at the point i=0, i.e.,

$$f\,(l) = ae^{j\sum_{i=0}^{l}\omega[i]} = ae^{j\sum_{i=0}^{l}\sum_{n=0}^{N}\frac{\omega^{(n)}[0]}{n\,!}\left(\frac{i}{f_s}\right)^n}, \tag{14}$$

where $a \in R^+$ is the amplitude and. is the transient angular frequency. $\omega^{(n)}$ [0] is the n -th derivative of $\omega\,[i]$. As the value of n increases, the signal expanded by this Taylor series becomes closer to the real FM signal. f_s is the sample frequency. The existence of noise has always been considered as a serious effect on the estimation accuracy of IF trajectories. Underwater acoustic noise has an unknown distribution in practical applications, and prior information of the noise distribution is often difficult to obtain. The signal $f\,(l)$ is corrupted by underwater acoustic noise with an unknown distribution.

3.1 Power of phase-difference accumulation

Sun et al. (2021b) designed a polynomial to approximate the original signal. The IF estimation function obtained by the n

-order Taylor series expansion is designed as

$$h[i|\hat{\omega}_n] = e^{-j\sum_{i=0}^{l}\sum_{n=0}^{N}\frac{\hat{\omega}^{(n)}[0]}{n!}\left(\frac{i}{f_s}\right)^n}. \tag{15}$$

The error between the estimator and the true frequency is

$$r\left[i \mid \tilde{\omega}_u\right] = f\left[i \mid \omega_n\right] h\left[i \mid \hat{\omega}_n\right] = ae^{j\sum_{i=0}^{l}\sum_{n=0}^{N}\frac{\hat{\omega}^{(n)}[0]}{n!}\left(\frac{i}{f_s}\right)^n}, \tag{16}$$

where $\tilde{\omega}_u = \omega_u - \hat{\omega}_u$ is the frequency expansion error and $\phi\left[i \mid \tilde{\omega}_u\right] = \sum_{n=0}^{N}\frac{\tilde{\omega}^{(n)}[0]}{n!}(\frac{i}{f_s})^n$ is the phase error. When the phase error $\phi\left[i \mid \tilde{\omega}_u\right] = 0$, the FM signal frequency is $\hat{\omega}_u = \omega_u$.

The coherent accumulation of phase difference can improve the signal-to-noise ratio (SNR) (Du et al., 2020). For underwater acoustic noise with an arbitrary distribution, the average value of coherence accumulation will fluctuate slightly; however, the coherence accumulation of phase difference in Eq. (16) will increase with time, especially when $\phi\left[i \mid \tilde{\omega}_u\right]$ is approximately zero [42]. Inspired by this fundamental rule, a cost function is constructed to find the optimal solution for all coefficients $\phi\left[i \mid \tilde{\omega}_u\right]$:

$$p\left[l \mid \tilde{\omega}_u\right] = \frac{1}{l}\sum_{i=0}^{l}\left\{\sum_{k=0}^{i} r\left[k \mid \tilde{\omega}_u\right] - r_m\left[i \mid \tilde{\omega}_u\right]\right\}\left\{\sum_{k=0}^{i} r\left[k \mid \tilde{\omega}_u\right] - r_m\left[i \mid \tilde{\omega}_u\right]\right\}^{*} \tag{17}$$

with $r_m\left[i \mid \tilde{\omega}_u\right] = \frac{1}{i}\sum_{k=0}^{i} r\left[k \mid \tilde{\omega}_u\right]$. In the above cost function, $\sum_{k=0}^{i} r\left[k \mid \tilde{a}_u\right]$ is the sum of the accumulated phase difference of i points in $r\left[k \mid \tilde{\omega}_u\right]$ and * denotes the complex conjugate. The power $p\left[l \mid \tilde{\omega}_u ght\right]$ and its mean value over the period l improve the anti-interference ability of the spectrum to a certain extent.

Lemma 2. The cumulative phase difference power $p\left[l \mid \tilde{\omega}_u\right]$ minus the mean is a monotonically decreasing function of the coefficient modulus $\left|\tilde{\omega}_u\right|$ (Sun et al., 2021b; Sun et al., 2021a).

3.2 Instantaneous frequency estimation based on nonlinear accumulated phase power logarithm

In view of the nonlinear amplification for DOA estimation and IF estimation described by Filippini et al. (2019); Sun et al. (2021b), and Zhang et al. (2018), a multiple logarithmic sum is proposed to achieve a sharper peak for the true IF and a deeper trough for the area that is not of interest. This will also improve the accuracy and robustness of the IF estimation. On this basis, a log-sum-multiple function is used for the maximization, as this is more convex and smoother than a log-sum. This function also exhibits nonlinear amplification in some fixed intervals.

$$\max_{\hat{p}_u} J_1\left(\tilde{\omega}_u\right) = \log\frac{\prod_{l=0}^{L-1}\left\{p\left[l \mid \tilde{\omega}_u\right] + 1\right\}}{\sum_{l=0}^{L-1} t\left\{p\left[l \mid \tilde{\omega}_u\right] + 1 = \sum_{l=0}^{L-1}\log\left\{p\left[l \mid \tilde{\omega}_u\right] + 1\right\} - \log\sum_{l=0}^{L-1}\left\{p\left[l \mid \tilde{\omega}_u\right] + 1\right\} \mathrm{s.t.}\, \tilde{\omega}_u = \omega_u - \hat{\omega}_u} \tag{18}$$

where “1” is introduced into the integrand to ensure a simple analytical solution.

Under the condition of short-time window function or strong noise, the L1 norm method has poor estimation accuracy or is susceptible to noise interference. The log product sum ratio is used to increase the peak value of the target instantaneous frequency in the time spectrum and suppress the side lobe of the non target region. Under the condition of small sample data or low signal-to-noise ratio, this method can effectively improve the accuracy and stability of the time-frequency estimation curve.

Remark 2. Eq. (18) can be understood as the logarithmic mapping of the cumulative power at each time minus the logarithmic mapping of the cumulative power at all times.

The lower limit of $J_1\left(\tilde{\omega}_u\right)$ is

$$\lim_{\tilde{p}_u \to -\infty}\left\{\sum_{l=0}^{L-1}\log\left\{p\left[l \mid \tilde{\omega}_u\right] + 1\right\} - \log\sum_{l=0}^{L-1}\mathrm{p}\left[\mathrm{lt} \mid \mathrm{\omega_u}\right] + 1\right\} = 0 \tag{19}$$

If the upper limit of $J_1\left(\tilde{\omega}_u\right)$ is a variable $v\left(\tilde{\omega}_u\right)$, the variable upper limit integral of $J_1\left(\tilde{\omega}_u\right)$ can be expressed as

$$\begin{aligned} J_2\left[v\left(\tilde{\omega}_u\right)\right] &= \int_1^{v\left(\tilde{\omega}_u\right)}\left(\log\kappa + 1\right)d\kappa \\ &= v\left(\tilde{\omega}_u\right)\log v\left(\tilde{\omega}_u\right) \end{aligned} \tag{20}$$

where “1” is introduced into the integrand to ensure a simple analytical solution. Sun et al., 2021a showed that $J_2\left[v\left(\tilde{p}_u\right)\right]$ is a convex function about $v\left(\tilde{p}_u\right)$.

Lemma 3. $J_2\left(\tilde{\omega}_u\right)$ is a convex function (Boyd et al. (2004)).

Theorem 1. $J_2\left(\tilde{\omega}_u\right)$ is a monotonically decreasing function of $\tilde{\omega}_u$. According to the chain rule, the first derivative of $J_2\left[v\left(\tilde{\omega}_u\right)\right]$ with respect to $\left|\tilde{\omega}_u\right|$ is

$$\frac{dJ_2\left[v\left(\tilde{p}_u\right)\right]}{d\left|\tilde{\omega}_u\right|} = \frac{dJ_2}{dv}\sum_{l=0}^{L-1}\left(\frac{dv}{dp\left[l\right]}\;\frac{dp\left[l \mid \tilde{\omega}_u right\,.\right]}{d\left|\tilde{\omega}_u\right|}\right) \tag{21}$$

To determine whether Eq. (21) is positive or negative, we analyze the positive and negative nature of the first-order functions $\frac{dJ_2}{dv}$, $\frac{dv}{dp\left[l\right]}$, and $\frac{dp\left[l \mid \tilde{\omega}_u\right]}{d\left|\tilde{\omega}_u\right|}$.

According to Lemma 1, $p\left[l \mid \tilde{\omega}_u\right]$ is a monotonically decreasing function of $\left|\tilde{\omega}_u\right|$ and $\frac{dp\left[l \mid \tilde{\omega}_u\right]}{d\left|\tilde{o}mega_u\right|} \leq 0$. Then, in Eq. (20), $v\left\{p[l]\right\} = \frac{\prod_{l=0}^{L-1}\left\{p\left[l \mid \tilde{\omega}_u\right] + 1\right\}}{\sum_{l=0}^{L-1}\left\{p\left[l \mid \hat{\imath} lde\omega_u\right] + 1\right\}}$, $\frac{dv}{dp\left[l\right]}$ can be expressed as

$$\begin{aligned}\frac{dv\{p[l]\}}{dp[l]} &= \left(\frac{\prod_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}}{\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}}\right)_{p[]} \\ &= \frac{\left(\prod_{i=0}^{L-1}\{p[l\mid lde\omega_u]+1\}\right)_{p[l]}}{\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}} \\ &\quad - \frac{\prod_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}}{\left(\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}\right)^2} \\ &\quad \left(\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}t\right)_{p[l]} \\ &= \frac{\prod_L\{p[k\mid\tilde{\omega}_u]+1\}}{\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}} \\ &\quad - \frac{\prod_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}}{\left(\sum_{l=0}^{L-1}\{p[l\mid\tilde{\omega}_u]+1\}\right)^2}L \qquad (22)\end{aligned}$$

where $k \in \{0, L-1\} \setminus \{l\}$. In Eq. (22), because the power $p\,[l\mid\tilde{\omega}_u] \geq 0$, we have that $\frac{L}{\sum_{l=0}^{L-1}\{p\,[l\mid\tilde{\omega}_u]+1\}} \leq 1$ and $\frac{dv\,\{p\,[l]\}}{dp\,[l]} \geq 0$. $v\,\{p\,[l]\}$ is a monotonically increasing function of $p\,[l\mid de\omega_u]$.

Finally, the first derivative of $J_2\,[v\,(\tilde{\omega}_u)]$ with respect to $v\,(\tilde{\omega}_u)$ is

$$\frac{dJ_2\,[v\,(\tilde{\omega}_u)]}{dv} = \log\,v\,(\tilde{\omega}_u) + 1 \qquad (23)$$

Because $v\,(\tilde{\omega}_u) \geq 1$, Eq. (23) is positive. $J_2\,[v\,(\tilde{\omega}_u)]$ is a monotonically increasing function of $v\,(\tilde{\omega}_u)$. In summary, the first derivative of $J_2\,[v\,(\tilde{\omega}_u)]$ with respect to $|\tilde{\omega}_u|$ is non-positive:

$$\frac{dJ_2\,[v\,(\tilde{\omega}_u)]}{d\,|\tilde{\omega}_u|} \leq 0 \qquad (24)$$

Therefore, $J_2\,[v\,(\tilde{\omega}_u)]$ is a monotonically decreasing function of $|\tilde{\omega}_u|$.

The estimation coefficient $\hat{\omega}$ can be obtained by solving the following optimization problem:

$$\max_{\hat{\omega}_u} J_2\,(\tilde{\omega}_u) = v\,(\tilde{\omega}_u)\,\log\,v\,(\tilde{\omega}_u)\,\mathrm{s.t.}\,\tilde{\omega}_u = \omega_u - \hat{e}ga_u \qquad (25)$$

When $\hat{\omega}_u$ is known, the estimated IF can be obtained as $\hat{\phi}\,[i\mid\hat{\omega}_u] = \sum_{u=0}^{U}\hat{\omega}_u\,(\frac{i}{f_s})^u$. According to Theorem 1, $J_2 łeft\,(\tilde{\omega}_u)$ is a convex function, so the optimal value can be determined using a gradient method.

The algorithm based on the log-sum multiple function has the characteristics of a deep zero sidelobe and extremely sharp main lobe. In Figure 6, compared with the APP-MLS algorithm (Sun et al., 2021b), the proposed algorithm gives a lower main lobe value than the APP-MLS algorithm. This reduces the computational complexity and hardware requirements caused by possible large data values in the calculation. At the same time, because PAL-PSR algorithm has sufficiently high IF resolution, there is no side sidelobe leakage of Fourier transform in Figure 6.

In the process of using this algorithm, the estimation accuracy of the time–frequency curve will be affected by the order U , the length of the rectangular window W , and the sliding length L . Specifically, if the length of the rectangular window W is constant, an increase in the order U means that the fitting polynomial function will provide a better approximation to the real frequency function. Thus, the final estimated time–frequency $\varphi\,[m]$ will be of higher accuracy. If the order is fixed, an increase in the length of the rectangular window W will increase the error between the fitting polynomial function and the real frequency function as the number of sample points increases, so the accuracy of the final time–frequency curve will become worse. The error will be greatest at the last point in the window. However, W also determines the stability of the algorithm. A longer window length produces a stronger ability to suppress different distributed noise. Therefore, the value of W should not be too small in practical applications. A longer sliding length L gives a larger deviation between the estimated IF time–frequency $\varphi\,[m]$ and the actual curve. In addition, larger values of the order U , rectangular window length W , and sliding length L will increase the computational complexity of the algorithm. In practice, appropriate parameter values should be selected according to accuracy, stability, and computational complexity requirements.

3.3 Multi-AUV frequency identification

In Figure 6, the PAL-PSR algorithm produces a sharper peak and lower sidelobe region, so it can accurately distinguish multiple frequencies.

$$\hat{\omega}_i\,(t) - \hat{\omega}_j\,(t) = a \qquad (26)$$

where a is a constant and i , j are nonadjacent spectral lines. According to Eq. (3), such a set of constant frequencies can be considered as the PWM frequency of a BLDC motor, propeller modulation frequency, and electrode modulation frequencyfrom the same AUV.

When a single hydrophone receives AUV motor noise, if the motor frequency f_{pwm} used by the AUV is the same or close, some frequency crossover will occur after time–frequency estimation of multiple frequency bands, as shown in Eq. (3).

The energy of the same frequency signal increases at the frequency intersection, so it can be judged that the frequency has crossed over. The crossover frequency is used as the Doppler information of the motion of two AUVs at the same time, combined with the first-order derivatives f_{pwm} and f_p and the second-order derivatives f_{pwm} and f_p of AUV motion. This modulation signal assists in distinguishing AUV motion information.

According to the above analysis, this paper proposes an AUV signal detection algorithm. The specific steps are as follows:

Step 1: Calculate the STFT of the signal and judge the signal interval according to the stable energy value of the STFT in the

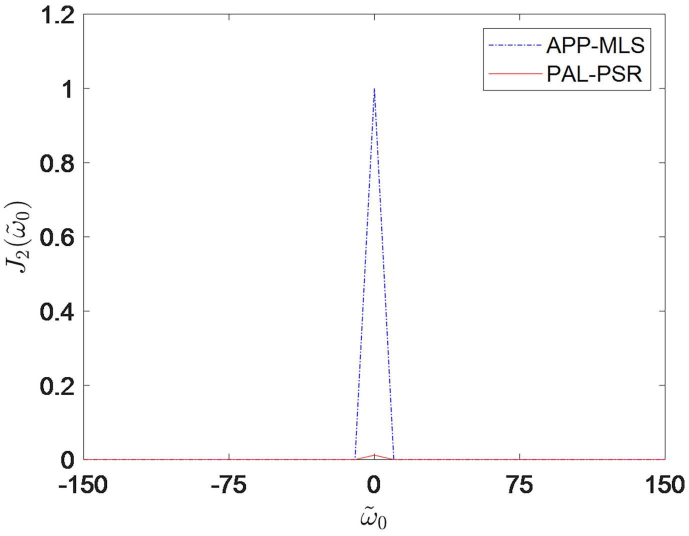


FIGURE 6
$J_2(\tilde{\omega}_u)$ comparison between PAL-PSR and APP-MLS.

frequency domain. Specifically, search for a stable value that exceeds the threshold in the frequency domain within each time window. According to the theory of signal detection and estimation, for a given false alarm probability, it is difficult to give an accurate detection threshold. In practice, the detection threshold is generally determined by the method of numerical statistics (Li et al., 2011; Zhu et al., 2021). According to whether the selected amplitude frequency continuously exceeds the threshold in continuous time, the false alarm probability can then be reduced.

Step 2: Among the frequencies screened in step 1, the PAL-PSR algorithm is used for frequency estimation. The peak of the deep zero sidelobe is then formed after accurate frequency estimation, allowing the frequencies that the STFT algorithm cannot distinguish to be identified.

Step 3: Multiple AUVs are distinguished according to the frequency estimated in step 2, combined with the BLDC frequency characteristics of the AUVs and the Doppler frequency shift.

Remark 3. According to Eq. (8), if the speed difference of multiple AUVs equipped with similar BLDC motors is large, the modulation frequency generated by f_p at $f_p wm$ will be relatively different, and the frequency may not crossover. When the sailing speeds are similar, the modulation frequency generated by f_p at $f_p wm$ may also cross frequency. When frequency crossover occurs, as shown in Eq. (9), the Doppler frequency is a continuous differentiable function. Thus, different AUVs can be distinguished and AUV motion information can be extracted by judging the continuity of $f_p wm$ and the fixed frequency difference between $f_p wm$ and the harmonic signal.

4 Simulations

This section compares the performance of the PAL-PSR algorithm with several classical IF estimation algorithms, including the STFT, WVD, polynomial chirplet transform, and synchrosqueezing transform (SST). The algorithms are assessed using a sinusoidal FM signal (SFM). The sample time is $T = 1$ s and the sample frequency is $f_s = 50$ kHz. In Figure 7A, B, the window size W and the overlap size S are set to 200 and 20, respectively, considering the tradeoff between computation time and the ability to suppress noise. Symmetric α-stable (SaS) noise (Zhang et al., 2022) is often used to model underwater acoustic noise, as it represents real-world noise with changing statistical characteristics (Jobst et al., 2020). The generalized SNR (GSNR) (Ferrari et al., 2020) isused, which is defined as the ratio of the signal's average power to the noise dispersion in the finite interval of interest:

$$\mathrm{GSNR} = 10\log_{10}\frac{a^2}{\gamma} \tag{27}$$

An SFM signal is often used as the detection signal for sonar and radar because it offers increased bandwidth and can maintain a large pulse interval:

$$x_{GSFM}[l] = e^{i\sum_{m=0}^{l}\varphi_{SFM}[m]}, \tag{28}$$

where the IF is set to

$$\varphi_{SFM}[m] = 15000 + 10.*\cos(a2\pi m + a2\pi) \tag{29}$$

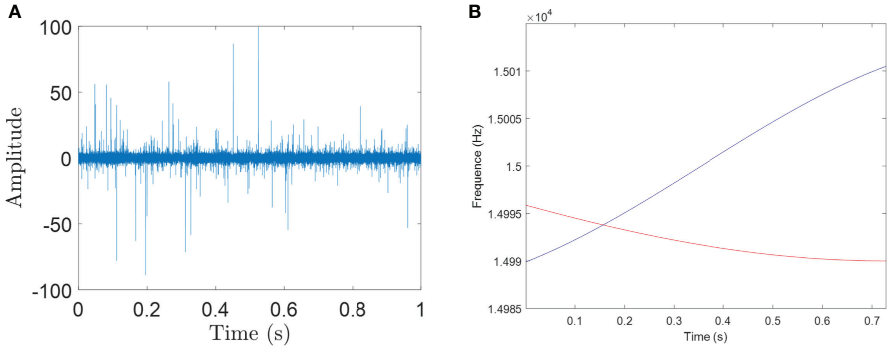


FIGURE 7
(A) SFM signal with noise, **(B)** designed frequency curve of cross SFM signal.

Although most noise can be assumed to follow a Gaussian distribution, there exist other types of underwater acoustic noise with unknown distributions, and prior information of the noise distribution is often difficult to obtain. SaS noise has been used to model various types of real underwater acoustic noise (Gogineni et al., 2020; Zhang et al., 2021). Under the condition of a heavy tailing effect, that is, adding SaS noise with a characteristic index of $\alpha = 1.6$ to the signal, and setting GSNR=−10dB , we compare the performance of the proposed time–frequency estimation algorithm with four classical methods. In the simulations, we set the window length $W = 200$ and sliding length $L = 50$. The SFM signal polluted by strong noise is shown in Figure 11A.

Figure 8 shows the estimation results of STFT, WVD, and SST. Due to the intersection of signal frequencies, the energy aggregation of the time–frequency estimation curves is scattered, and the STFT is more vulnerable to noise interference, resulting in a fuzzy time–frequency distribution. Figure 8B shows that WVD is highly ambiguous around the signal frequency intersection. Figure 8C shows the results given by the SST. The large number of pulses distorts the time–frequency trajectories to varying degrees, making it difficult to suppress high-energy noise. As can be seen from Figure 8D, compared with the four existing methods, the proposed time–frequency estimation algorithm not only maintains high energy aggregation, but also obtains high time–frequency trajectories, which verifies the accuracy and stability of our approach. Figure 8E shows the relative errors of the five time–frequency estimation curves. The error results show that the proposed time–frequency estimation algorithm effectively suppresses high-energy noise.

5 Experiments

5.1 Experiments

According to Coraluppi et al. (2018) and Zhu et al. (2021), the thruster operates at a fixed PWM frequency of 15–20 kHz, and it has been verified that there is a strong signal at the PWM frequency of the motor. An experiment was carried out using an electrically propelled unmanned surface vessel (USV). During the movement of the USV, a strong single-frequency signal at the PWM frequency of the USV propeller motor can be observed in the hydrophone measurement data (Figures 9A, B). When the propeller state remains unchanged, the PWM frequency will not vary. According to the conditions specified in Assumption 1 to Assumption 3, the USV was set to sail in a straight line at a uniform speed in the due north direction.

Experiments were carried out on a lake, with the position of the hydrophone taken as the coordinate origin. The hydrophone was fixed at a depth of 0.5 m. Two ships with the same motor were used to conduct experiments, and the data obtained from the motor acoustic signals of the two ships were processed.

In the experiment, f_{pwm} = 16005 Hz under the static condition of the USVs. The USVs move according to Assumption 1 to Assumption 3. USV1 has a velocity of 2.9 m/s and USV2 has a velocity of 3.8 m/s. The USVs sail in a straight line. Figures 19–21 compare the estimated time–frequency curves of the hydrophone after receiving the experimental data through PAL-PSR and the Doppler time–frequency curve calculated according to the coordinates of the USV positions and the receiving hydrophone. According to the motion parameters of the USVs calculated from the time–frequency curve, combined with Assumptions 1–3, the estimated navigation parameters of the USVs were obtained. When the Doppler frequency of USV motion is used in Eq. (8), it is possible to solve Eq. (8). The movement time of the USVs is 15 s. Through comparative experiments, the maximum deviation between f_{pwm} measured at rest and f_{pwm} obtained by calculating the derivative of the Doppler signal of the USVs is 0.8 Hz. The deviation in f_p is 0.9 Hz. The results of the lake experiment are presented in Table 1. The relative error between

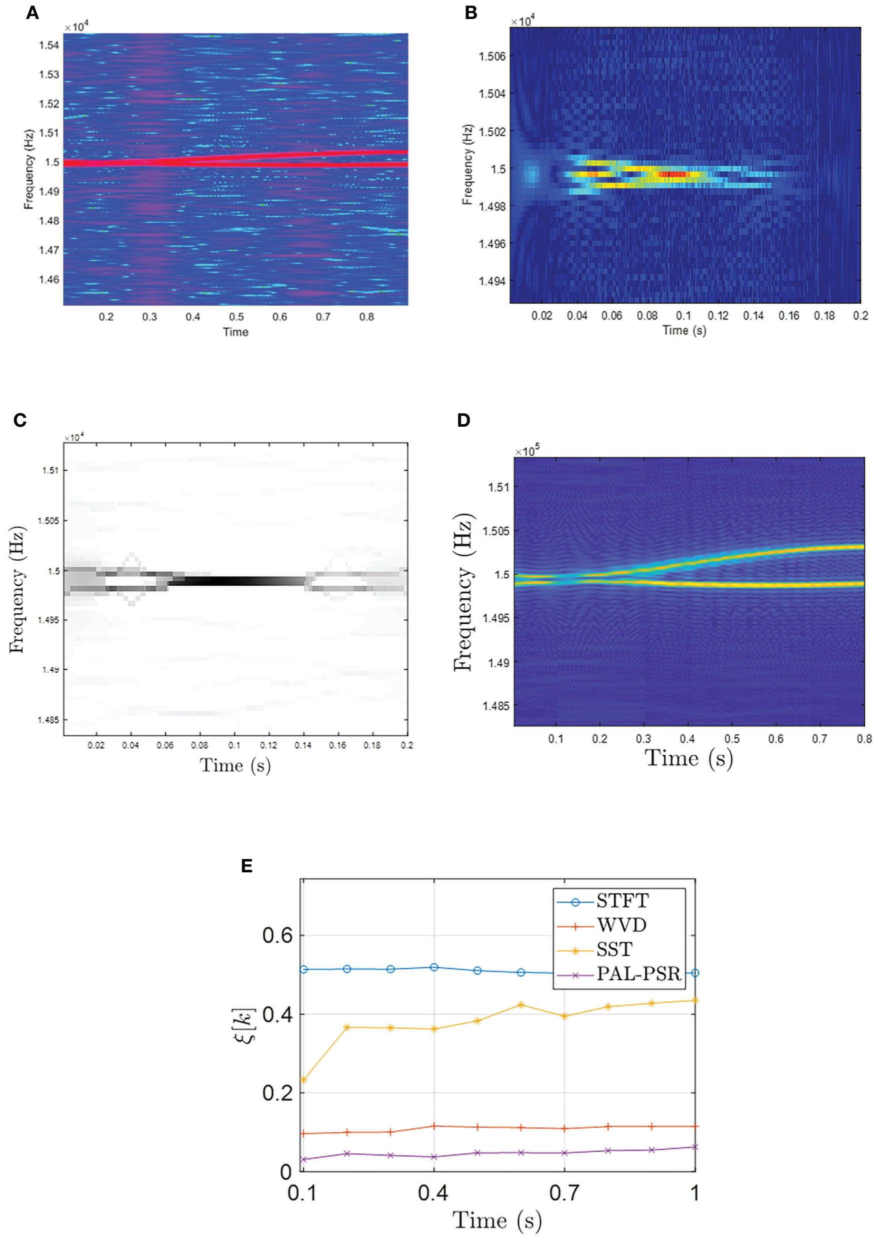


FIGURE 8
(A) STFT curve of cross SFM signal, **(B)** WVD curve of cross SFM signal, **(C)** SST curve of cross SFM signal, **(D)** PAL- PSR curve of cross SFM signal, **(E)** eelative error curves for STFT, WVD, SST, and PAL-PSR. GSNR2=-10 dB.

the USV and hydrophone positions retrieved by Eq. (8) and the GPS locations is not more than 7%, and the USV velocity estimation error is not more than 4%.

Comparing the transformation results of the experimental data by the two algorithms in Figures 10B, C, it can be seen that the spectrum obtained by the STFT in Figure10A has obvious frequency broadening, and it is unable to accurately judge the frequency intersection generated by multiple AUVs, and it is unable to accurately estimate the frequency. Only the trend of frequency transformation can be roughly seen. However, STFT can determine the approximate frequency range, and then use PAL-PSR algorithm to accurately estimate the frequency and distinguish the targets.

From Figure 10C, combined with the Doppler frequency of Eq. (8) and the motor modulation frequency in Eq. (3), when the sailing velocity of the USVs is different, the motor speeds are also different. Thus, the position where the frequency crossover occurs will change. However, because f_p is modulated in f_{pwm} , the frequency difference between f_{pwm} and f_p remains unchanged, as shown in Eq. (26).

The first and second derivatives of f_{pwm} for the different USVs are shown in Figures 11A, B. According to Eq. (7) and Eq. (8), the minimum value of the first derivative and the zero value of the second derivative indicate the shortest distance between the USV and the hydrophone, that is, the yaw position. According to the time at which the yaw position occurs, the USV velocity, f_{pwm} , f_p , and yaw distance of the USV can be calculated.

5.2 Experimental data error analysis

During the experiment, there was a deviation between the GPS locations and measured Doppler signal, as shown in Table 1. The algorithm was used to estimate the motion state parameters of the USV. The estimated results and errors are shown in Table 2. In the estimation process, external interference, including measurement errors associated with GPS and the USV inertial navigation system, created some deviation between the propeller and GPS measurement points, resulting in errors in the nonlinear calculation process. In the experiment, the hydrophone monitored the PWM signal of the propulsion motor. The distance between the propulsion motor and the GPS/INS position was 2.5 m. In addition, due to the influence of noise, the estimated Doppler frequency shift will have some degree of error. Thus, using only the starting time and the shortest distance time to estimate the yaw position and navigation speed of the USV will introduce a certain deviation. If multiple points were used as the start time (e.g., 0.5 s, 1 s) and the shortest distance time, and the data were averaged to reduce the error, the results presented in Table 3 would be obtained.

6 Conclusion

In this paper, we investigated the signal characteristics of AUV propulsion motors and designed a multi-AUV linear motion model. According to the motion model and the Doppler information generated by the motor signal, the relative position between the AUV and a hydrophone was calculated. The proposed PAL-PSR algorithm can accurately estimate the underwater acoustic frequency curve, even in the presence of impulse signal noise. As the cost function is a monotonic function of frequency, the estimation result is stable. Accurate estimations are possible because the nonlinear amplification mapping allows the signal to grow exponentially in energy, whereas the noise grows linearly at

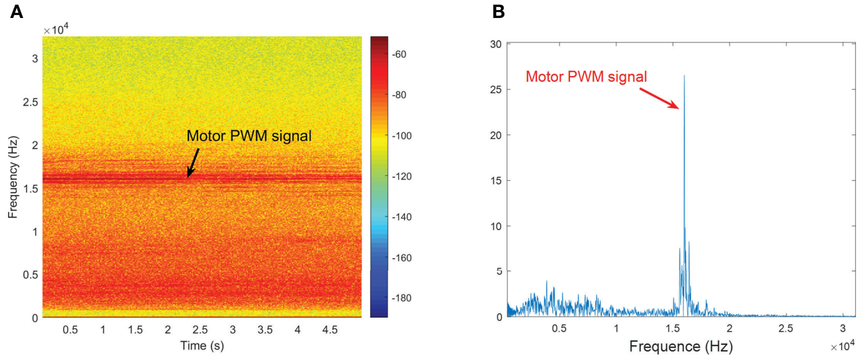


FIGURE 9
(A) Time–frequency diagram of USV motor, **(B)** frequency diagram of USV motor.

TABLE 1 USV navigation conditions.

	USV velocity (m/s)	USV initial position	Yaw distance (m)	USV $f_{p}wm$	Motor pole	Rad (rad/min)	USV f_p
USV 1	2.9	(35.5, -33.4)	35.5	16005Hz	12	600	120Hz
USV 2	3.8	(17.4,-19.1)	17.4	16005 Hz	12	780	156Hz

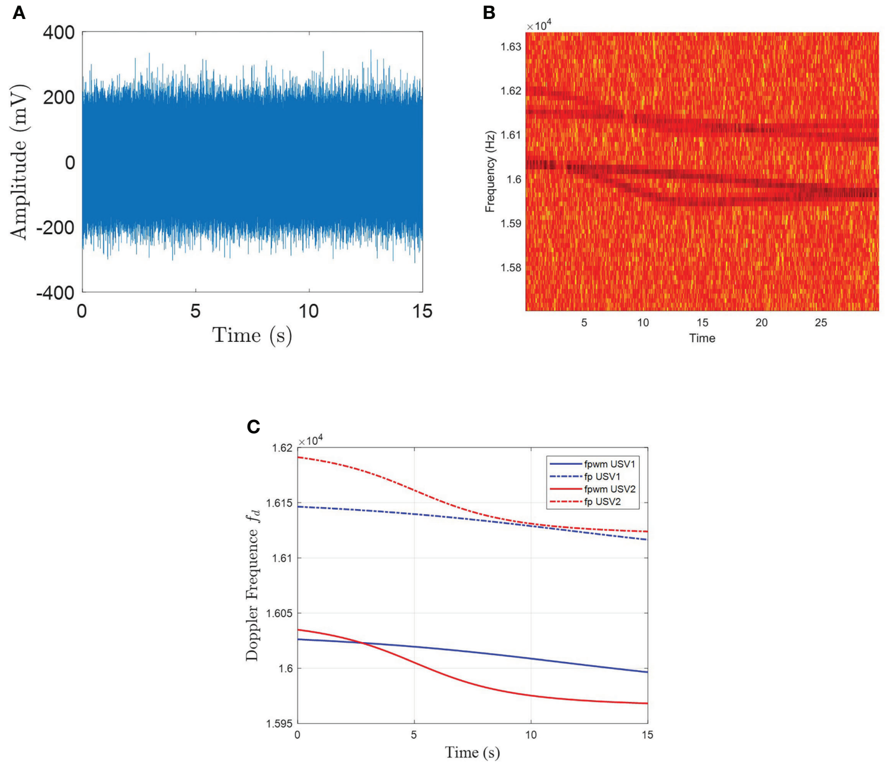


FIGURE 10
(A) Time-domain signal collected by the hydrophone in the experiment, **(B)** STFT curve of f_{pwm} and f_p in the experiment, **(C)** PAL-PSR curve of f_{pwm} and f_p in the experiment.

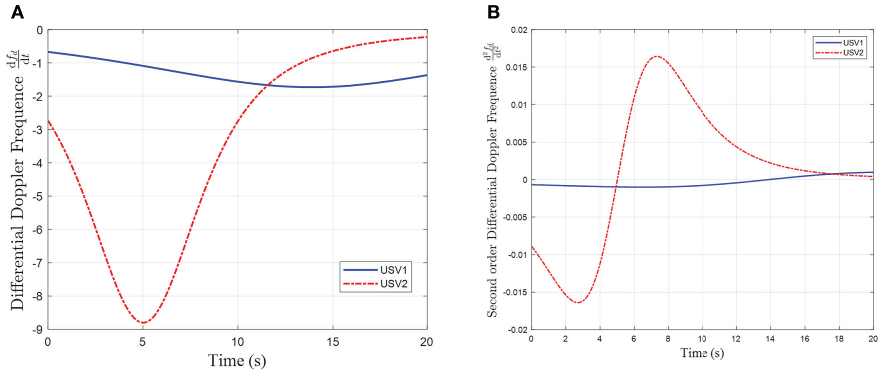


FIGURE 11
(A) First-order derivative of the estimated frequency f_{pwm}, **(B)** second-order derivative of the estimated frequency f_{pwm}..

most. Through computer simulations, the effectiveness of the PAL-PSR method was verified under different GSNRs and window sizes. Compared with some classical methods, PAL-PSR exhibits excellent performance and robustness in the presence of an unknown noise distribution. In addition, by examining the characteristics of AUV motors, an AUV judgment mechanism based on the PAL-PSR frequency estimation method was designed, enabling a single hydrophone to distinguish multiple AUVs through their signal characteristics. A lake experiment was conducted using

TABLE 2 Inversion calculation of USV circle navigation conditions and errors.

	Estimation of AUV velocity (m/s)	Error of AUV initial velocity estimation	Estimation of yaw distance (m)	Error of yaw distance	Estimation of $f_p wm$	Estimation of f_p
USV 1	2.79	3.79%	33.1	6.76%	16005.5Hz	120.6 Hz
USV 2	3.92	3.16%	18.3	5.17%	16005.8Hz	156.9 Hz

TABLE 3 Motion trajectory at multiple times after initial calculation time.

	Estimation of AUV velocity (m/s)	Error of AUV initial velocity estimation	Estimation of yaw distance (m)	Error of yaw distance (m)	Estimation of $f_p wm$	Estimation of f_p
USV 1	2.84	2.07%	34.1	3.94%	16005.5 Hz	120.3 Hz
USV 2	3.88	2.11	17.9	2.87%	16005.8 Hz	156.4 Hz

USVs and a single hydrophone. The PAL-PSR algorithm was used to estimate the FM signal curve received by a single hydrophone. According to the frequency modulation curve, the positions of the USVs were retrieved and compared with the ground-truth GPS coordinate positions.

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

SR: Conceptualization, Methodology, Software, Writing - Original Draft. YX: Writing - Review and Editing, Supervision, Software, Funding acquisition, Project administration. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No.LY21F010001.

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

Acarnley, P. P., and Watson, J. F. (2006). Review of position-sensorless operation of brushless permanent-magnet machines. *IEEE Trans. Ind. Electron.* 53, 352–362. doi: 10.1109/TIE.2006.870868

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization* (Cambridge:Cambridge University Press, British).

Brinkmann, K., and Hurka, J. (2009). "Broadband passive sonar tracking," in *Informatik 2009–im focus das leben* (Germany:Gesellschaft f{\"u}r Informatik e. V).

Buckingham, M. J., Giddens, E. M., Simonet, F., and Hahn, T. R. (2002). Propeller noise from a light aircraft for low-frequency measurements of the speed of sound in a marine sediment. *J. Comput. Acoust.* 10, 445–464. doi: 10.1142/S0218396X02001760

Coraluppi, S., Carthel, C., and Coon, A. (2018). "An MHT approach to multi-sensor passive sonar tracking," In *2018 21st International Conference on Information Fusion (FUSION)*. (Cambridge, UK:IEEE) 480–487.

De Viaene, J., Verbelen, F., Derammelaere, S., and Stockman, K. (2018). Energy-efficient sensorless load angle control of a BLDC motor using sinusoidal currents. *IET Electr. Power Appl.* 12, 1378–1389. doi: 10.1049/iet-epa.2018.5059

Du, S., Zhang, J., and Hu, G. (2020). A robust data-driven AVO inversion with logarithm absolute error loss function. *Acta Geophy.* 68, 445–458. doi: 10.1007/s11600-020-00416-1

Fang, J., Wang, F., Shen, Y., Li, H., and Blum, R. S. (2016). Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach. *IEEE Trans. Signal Process.* 64, 4649–4662. doi: 10.1109/TSP.2016.2572041

Ferguson, B. G. (1992). A ground-based narrow-band passive acoustic technique for estimating the altitude and speed of a propeller-driven aircraft. *J. Acoust. Soc Am.* 92, 1403–1407. doi: 10.1121/1.403934

Ferrari, A., Napoli, A., Fischer, J. K., Costa, N., D'Amico, A., Pedro, J., et al. (2020). Assessment on the achievable throughput of multi-band ITU-T g. 652. d fiber transmission systems. *J. Light. Technol.* 38, 4279–4291. doi: 10.1109/JLT.2020.2989620

Filippini, F., Colone, F., and De Maio, A. (2019). Threshold region performance of multicarrier maximum likelihood direction of arrival estimator. *IEEE Trans. Aerosp. Electron. Syst.* 55, 3517–3530. doi: 10.1109/TAES.2019.2909335

Gogineni, V. C., Talebi, S. P., Werner, S., and Mandic, D. P. (2020). Fractional-order correntropy filters for tracking dynamic systems in α-stable environments. *IEEE Trans. Circuits Syst. II: Express Briefs* 67, 3557–3561. doi: 10.1109/TCSII.2020.2995357

Gusland, D., Christiansen, J. M., Torvik, B., Fioranelli, F., Gurbuz, S. Z., and Ritchie, M. (2021). "Open radar initiative: Large scale dataset for benchmarking of micro-Doppler recognition algorithms," In *2021 IEEE Radar Conference (RadarConf21).* (Atlanta, GA, USA:IEEE) 1–6.

Haiying, W., and Zufeng, F. (2014). "A privacy preserving protocol for generating the common tangent of two circles," In *2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*, (Hangzhou, China:IEEE) 1. 227–231.

Hanif, A., Muaz, M., Hasan, A., and Adeel, M. (2022). Micro-Doppler based target recognition with radars: A review. *IEEE Sens. J* 22, 2948–2961. doi: 10.1109/JSEN.2022.3141213

Hara, T., Ajima, T., Tanabe, Y., Watanabe, M., Hoshino, K., and Oyama, K. (2018). Analysis of vibration and noise in permanent magnet synchronous motors with distributed winding for the PWM method. *IEEE Trans. Ind. Appl.* 54, 6042–6049. doi: 10.1109/TIA.2018.2847620

Jobst, W., Whited, L., and Smith, D. W. (2020). Acoustic clutter removal. *IEEE J. Ocean. Eng.* 46, 1000–1007. doi: 10.1109/JOE.2020.3014007

Le Besnerais, J., Lanfranchi, V., Hecquet, M., and Brochet, P. (2009). Characterization and reduction of audible magnetic noise due to PWM supply in induction machines. *IEEE Trans. Ind. Electron.* 57, 1288–1295. doi: 10.1109/TIE.2009.2029529

Lexa, M., Coraluppi, S., Carthel, C., and Willett, P. (2020). "Distributed MHT and ML-PMHT approaches to multi-sensor passive sonar tracking," In 2020 IEEE Aerospace Conference. (USA:IEEE) 1–12.

Li, H., Chen, S.-R., Qin, Y.-L., Wang, H.-Q., and Li, X. (2011). Detection and parameter estimation method for polyphase-coded pulse compression waveforms. *Syst. Eng. Electron.* 33, 310–314. doi: 10.3969/j.issn.1001-506X.2011.02.16

Li, X., Lu, B., Ali, W., Su, J., and Jin, H. (2021a). Passive sonar multiple-target tracking with nonlinear Doppler and bearing measurements using multiple sensors. *Int. J. Aerosp. Eng.* 2021, 1–11. doi: 10.1155/2021/4163766

Lindberg, A., Lövgren, B., Asp, J., Antoni, J., and Gällström, A. (2022). Short-time least squares spectral analysis of pass-by noise in water from a rigid inflatable boat. *J. Acoust. Soc Am.* 151, 1932–1943. doi: 10.1121/10.0009847

Liu, N., Gao, J., Jiang, X., Zhang, Z., and Wang, Q. (2016). Seismic time–frequency analysis *via* STFT-based concentration of frequency and time. *IEEE Geosci. Remote Sens. Lett.* 14, 127–131. doi: 10.1109/LGRS.2016.2630734

Li, Y., Wang, Z., Zhao, T., and Song, W. (2021b). An improved multi-ridge extraction method based on differential synchro-squeezing wavelet transform. *IEEE Access* 9, 96763–96774. doi: 10.1109/ACCESS.2021.3095054

Lo, K. W. (2017). Flight parameter estimation using instantaneous frequency and direction of arrival measurements from a single acoustic sensor node. *J. Acoust. Soc Am.* 141, 1332–1348. doi: 10.1121/1.4976091

Lo, W., Chan, C. C., Zhu, Z.-Q., Xu, L., Howe, D., and Chau, K. (2000). Acoustic noise radiated by PWM-controlled induction machine drives. *IEEE Trans. Ind. Electron.* 47, 880–889. doi: 10.1109/41.857968

Nielsen, M. C., Johansen, T. A., and Blanke, M. (2018). "Cooperative rendezvous and docking for underwater robots using model predictive control and dual decomposition," in *2018 European Control Conference (ECC).* (Limassol, Cyprus: IEEE) 14–19.

Qiao, X., Amin, M. G., Shan, T., Zeng, Z., and Tao, R. (2021). Human activity classification based on micro-Doppler signatures separation. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1016/j.neucom.2020.04.118

Qiao, C., Shi, Y., Diao, Y.-X., Calhoun, V. D., and Wang, Y.-P. (2020). Log-sum enhanced sparse deep neural network. *Neurocomputing* 407, 206–220. doi: 10.1016/j.neucom.2020.04.118

Railey, K., DiBiaso, D., and Schmidt, H. (2020). An acoustic remote sensing method for high-precision propeller rotation and speed estimation of unmanned underwater vehicles. *J. Acoust. Soc Am.* 148, 3942–3950. doi: 10.1121/10.0002954

Saffari, A., Zahiri, S.-H., and Khishe, M. (2022). Automatic recognition of sonar targets using feature selection in micro-Doppler signature. *Def. Technol* 2022, 1–14. doi: 10.1016/j.dt.2022.05.007

Stankovic, L., and Katkovnik, V. (1999). The wigner distribution of noisy signals with adaptive time-frequency varying window. *IEEE Trans. Signal Process.* 47, 1099–1108. doi: 10.1109/78.752607

Sun, W., Wang, H., Gu, Q., Rong, S., et al. (2021b). Exact and robust time-frequency estimation *via* accumulation of phase-difference power on multiple log-sum. *J. Latex Class Files* 14, 1–11. doi: 10.13140/RG.2.2.21036.59523

Sun, W., Wang, H., Gu, Q., Rong, S., and Fan, L. (2021a). Exact frequency estimation in the iid noise *via* KL divergence of accumulated power. *IEEE Commun. Lett.* 25, 2574–2578. doi: 10.1109/LCOMM.2021.3077315

Weber, I., and Oehrn, C. R. (2022). A waveform-independent measure of recurrent neural activity. *Front. Neuroinf.* 16. doi: 10.3389/fninf.2022.800116

Wong, K. T., and Chu, H. (2002). Beam patterns of an underwater acoustic vector hydrophone located away from any reflecting boundary. *IEEE J. Ocean. Eng.* 27, 628–637. doi: 10.1109/JOE.2002.1040945

Yang, Y., and Fang, S. (2021). Improved velocity estimation method for doppler sonar based on accuracy evaluation and selection. *J. Mar. Sci. Eng.* 9, 576. doi: 10.3390/jmse9060576

Yu, G., Yu, M., and Xu, C. (2017). Synchroextracting transform. 64, 8042–8054. doi: 10.1109/TIE.2017.2696503

Zhang, X., Li, H., and Himed, B. (2018). Maximum likelihood delay and Doppler estimation for passive sensing. *IEEE Sens. J.* 19, 180–188. doi: 10.1109/JSEN.2018.2875664

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). 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., 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 Nav* 16, 531–541. doi: 10.1049/rsn2.12200

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020b). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Nav.* 14, 1055–1060. doi: 10.1049/iet-rsn.2019.0477

Zhang, J., Zhang, T., Shin, H.-S., Wang, J., and Zhang, C. (2020a). Geomagnetic gradient-assisted evolutionary algorithm for long-range underwater navigation. *IEEE Trans. Instrum. Meas.* 70, 1–12. doi: 10.1109/TIM.2020.3034966

Zhou, P., Peng, Z., Chen, S., Yang, Y., and Zhang, W. (2018). Non-stationary signal analysis based on general parameterized time–frequency transform and its application in the feature extraction of a rotary machine. *Front. Mech. Eng.* 13, 292–300. doi: 10.1007/s11465-017-0443-0

Zhu, Y., Liang, S., Wu, X., and Yang, H. (2021). A random finite set based joint probabilistic data association filter with non-homogeneous Markov chain. *Front. Inf. Technol. Electron. Eng.* 22, 1114–1126. doi: 10.1631/FITEE.2000209

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fmars.2022.1027830

Check for updates

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Hamada Esmaiel,
Aswan University, Egypt
Pan Huang,
Weifang University, China

*CORRESPONDENCE
Jianfeng Chen
chenjf@nwpu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 25 August 2022
ACCEPTED 19 October 2022
PUBLISHED 10 November 2022

CITATION
Li X, Chen J, Bai J, Ayub MS, Zhang D, Wang M and Yan Q (2022) Deep learning-based DOA estimation using CRNN for underwater acoustic arrays. *Front. Mar. Sci.* 9:1027830. doi: 10.3389/fmars.2022.1027830

Deep learning-based DOA estimation using CRNN for underwater acoustic arrays

Xiaoqiang Li[1], Jianfeng Chen[1]*, Jisheng Bai[1], Muhammad Saad Ayub[1], Dongzhe Zhang[1], Mou Wang[1] and Qingli Yan[2]

[1]Joint Laboratory of Environmental Sound Sensing, School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China, [2]School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an, China

In the marine environment, estimating the direction of arrival (DOA) is challenging because of the multipath signals and low signal-to-noise ratio (SNR). In this paper, we propose a convolutional recurrent neural network (CRNN)-based method for underwater DOA estimation using an acoustic array. The proposed CRNN takes the phase component of the short-time Fourier transform of the array signals as the input feature. The convolutional part of the CRNN extracts high-level features, while the recurrent component captures the temporal dependencies of the features. Moreover, we introduce a residual connection to further improve the performance of DOA estimation. We train the CRNN with multipath signals generated by the BELLHOP model and a uniform line array. Experimental results show that the proposed CRNN yields high-accuracy DOA estimation at different SNR levels, significantly outperforming existing methods. The proposed CRNN also exhibits a relatively short processing time for DOA estimation, extending its applicability.

KEYWORDS

DOA estimation, array signal processing, underwater acoustic, convolutional recurrent neural network, deep learning

1 Introduction

Direction of arrival (DOA) estimation of an acoustic signal is of considerable interest in several applications, including environmental monitoring, defense, and information acquisition (Singer et al., 2009; Zhang et al., 2022; Kandimalla et al., 2022). Especially in the underwater environment, DOA estimation plays an important role in source tracking, coastal surveillance, and navigation. Underwater DOA estimation methods are based on the capability of sonar arrays to receive acoustic signals transmitted from the source. The use of sonar arrays has significant advantages, including a wide detection range, reduced power consumption, and increased safety. Such advantages are preferable in underwater target detection, recognition, and tracking (Han et al., 2019; Zhang and Yang, 2021).

Several effective methods for estimating DOAs in different complex scenarios have been developed, such as conventional beamforming (CBF), estimation of parameters *via* the rotational invariance technique (ESPRIT), minimum variance distortionless response (MVDR), multiple signal classification (MUSIC), and their variants (Li et al., 2019). However, the underwater environment has additional complexity in terms of changing time-space-frequency channels (Jia et al., 2012), which introduce signal attenuation during transmission, multipath effects, Doppler effects, and varying propagation delays (Yang, 2012). The multipath effects occur because underwater the signal propagates through multiple paths with an abundance of delay spreads that cause a lot of interference to the original signal. The resultant signal obtained at the sonar array is significantly faded and has the ambiguity of direction. Similarly, the complex time-variant nature of the ocean and the waves generated on the ocean surface result in a large Doppler spread that makes the estimation of DOAs a difficult task. Due to these challenging characteristics, the signals received by sonar arrays consist of anomalies that make it difficult for traditional algorithms to accurately determine the DOA of the source. Although several methodologies have been proposed for accurate DOA estimation in the underwater environment, their performance degrades in the presence of the above challenging characteristics. To improve the capability of sonar array systems in the challenging underwater environment, novel DOA estimation methodologies are required.

In recent years, the availability of big data has enabled machine learning and deep learning to be employed in research domains such as image processing, speech processing, and acoustic signal processing (Wang et al., 2019; Bai et al., 2019; Bai et al., 2021; Chen et al., 2022; Bai et al., 2022). Various methodologies based on deep learning have been highly effective in solving classification, localization, association, and functional approximation tasks (Ayub et al., 2021; Desai and Mehendale, 2022; Ayub et al., 2022). Deep learning has recently been applied to the estimation of various parameters from acoustic signals in an underwater environment using sonar array systems (Niu et al., 2017; Ferguson et al., 2017; Houégnigan et al., 2017; Wang and Peng, 2018; Bianco et al., 2019; Shen et al., 2020; Ozanich et al., 2020).These studies show that deep learning performs exceptionally well in comparison to traditional methods. Specifically, in the domain of DOA estimation in an underwater environment using sonar arrays, several methods based on convolutional neural networks (CNNs) have been developed (Liu et al., 2021; Cao et al., 2021). These methods take the DOA estimation task as a classification problem and employ a CNN to compute the DOA of the source. CNNs are good at modeling time and frequency invariances and have the capability to exploit temporal contexts. Nonetheless, they struggle to exploit longer temporal context information. To overcome this shortcoming, recurrent neural networks (RNNs) combine information from previous temporal windows, enabling theoretically unlimited contextual information to be incorporated (Sun et al., 2021). To combine the advantages of CNNs and RNNs, the two architectures can be employed together in the form of a single network with convolutional layers followed by recurrent layers. This structure is typically referred to as a convolutional recurrent neural network (CRNN).

To efficiently solve the underwater DOA estimation problem, this paper proposes a deep DOA estimation algorithm for underwater signals based on CRNNs. The model takes the phase components of the short-time Fourier transform (STFT) of the received signal at each sensor of the array. The CNN is used to efficiently extract high-level information, while the RNN is used to ensure that temporal context information is efficiently modeled to enable DOA estimation. To adopt challenging conditions like multipath propagation of high-frequency sound signals in a shallow water environment, the proposed method is validated on a synthetic dataset generated by BELLHOP Jing et al. (2018); Han et al. (2021); Li et al. (2022). Similarly, to validate the performance of the proposed methodology in actual complex underwater environmental effects we test it on real data obtained through experiments in the sea. The results show that our model obtains accurate DOA estimates for multipath signals using a sonar array, and is suitable for use in underdetermined scenarios that are overlooked by traditional methods.

The main contributions of the paper are listed as follows:

1. We propose to use deep learning-based method for underwater DOA estimation. The proposed method uses residual CNN to extract high-level information and RNN to model the temporal contexts of the received signals.
2. For data simulation, we use BELLHOP to simulate the array signals which are multipath signals and with low SNRs in the underwater environment.
3. We further conduct the proposed method on real sea data to validate the performance of DOA estimation.
4. The proposed method achieves a notable improvement in the performance of DOA estimation on simulated and real data as compared to traditional and deep learning-based methods

The remainder of this paper is organized as follows. Section 2 presents a review of existing traditional and learning-based methods for the problem of DOA estimation in underwater scenarios. In section 3, the signal model is formulated and the array formation, feature generation process, and proposed DOA estimation model are described. Section 4 presents experimental results and analysis to verify the effectiveness of the proposed methodology. Finally, this paper is concluded in section 5.

2 Related work

2.1 Traditional DOA estimation methods

Since the breakthrough of sonar array systems, determining the DOA has been the most important task in underwater applications. The earliest method for estimating the DOA is beamforming (Singer et al., 2009). This method employs the array to define the directivity and searches through spatial regions to determine the direction. Although the limited processing and analysis power at the time meant this method had low accuracy and performed poorly in the presence of noise, it laid the foundations for the development of novel estimation methods. During the 1960s, a new methodology of maximum entropy spectral estimation (Ables, 1974; Berger et al., 1996) was presented, which is the basis of most modern estimation methods. This class of methodologies possesses high resolution but has significant computational requirements. Later, Capon presented the MVDR methodology based on the maximum SNR criterion (Capon, 1969). MVDR estimates the spatial wave number spectrum as a means of improving the resolution and collectively increasing the noise suppression. Nonetheless, the computational complexity of this methodology is still high.

A methodology based on the time difference of arrival was proposed by Knapp and Carter (1976). This focused on minimizing the calculation complexity while increasing the resolution, but the performance deteriorates significantly in the presence of noise and reverberation (Zhang et al., 2020; Zhang et al., 2021). A breakthrough was achieved in the 1970s when Schimt proposed a methodology based on the spatial subspace (Schmidt, 1986). This MUSIC method laid the foundation for a new direction of research in the field of DOA estimation. The concept of estimation gave birth to the expansion of subspace class estimation methodologies. In subsequent years, many enhanced variants of MUSIC were proposed, including the weighted MUSIC method (Stoica and Nehorai, 1990; Xu and Buckley, 1992), root MUSIC algorithm (Barabell, 1983; Rao and Hari, 1989; Ren and Willis, 1997), and several others. In the 1980s, the ESPRIT framework was presented (Roy and Kailath, 1989). This methodology employed the phenomenon of rotation-invariance among the subspaces to compute the DOA. The methodology was further enhanced to produce the MI-ESPRIT (Swindlehurst et al., 1992) and weighted ESPRIT (Eriksson and Stoica, 1994) methods.

In the 1990s, Ottersten and Viberg developed weighted subspace fitting (Viberg et al., 1991), which consists of a combined structure for minimizing the error in the estimation of the covariance matrix. This methodology can distinguish the sources accurately and has enhanced resolution. Nonetheless, the methodology is computationally expensive in terms of computing the set parameters and is prone to fail in the presence of small errors. In 2006, Candes et al. presented a new concept based on sparse signal acquisition and recovery, known as compressed sensing (Donoho, 2006; Candès et al., 2006). Their methodology is based on the sparsity of signals and can be employed without fulfilling the Nyquist sampling theorem.

2.2 Deep learning-based DOA estimation methods

Recently, there have been continuous improvements in deep learning theory and methodologies for DOA estimation (Hu et al., 2020). The use of deep learning for DOA estimation can be segregated into two broad domains. The first domain is based on supervised learning and employs the learned projective relationship among the input measurements to give the DOA output. A single-layer network model that learned the DOA using the input features for the first time was presented by Xiao et al. (2015). Similarly, a CNN was employed by Chakrabarty and Habets (2017) to learn the DOAs from the input features. This methodology enhanced the accuracy of estimation in noisy and reverberant environments. Xiang et al. (2020) presented a methodology that employs phase enhancement to increase the accuracy of DOA estimation. They later proposed an LSTM-based DOA estimation method for moving targets, which achieved high robustness to array imperfections (Xiang et al., 2021).

The second domain is based on unsupervised learning. For instance, Yuan et al. (2021) proposed an unsupervised learning strategy for DOA estimation using a novel loss function. Although methods based on deep learning provide significant improvements in the estimation results, they cannot be easily generalized to the underwater environment, which is complex and exhibits temporal and spatial variations. Several methods have been proposed to target these changing conditions. Liu et al. (2021) proposed a DOA estimation method based on CNNs using sonar arrays. Their methodology consists of a two-channel CNN that estimates the DOA using the real and imaginary covariance matrices. This approach outperforms MUSIC in terms of accuracy and estimation time. Similarly, a deep transfer learning methodology in which a CNN-based network adapts to new environments has been proposed (Cao et al., 2021). This methodology uses a single vector sensor as opposed to a sonar array. However, the above approaches cannot efficiently exploit the temporal context information, which is essential in the underwater environment. To solve this problem, we propose a DOA estimation method based on a CRNN that can capture the temporal context information in addition to the time-frequency invariance. The proposed methodology achieves accurate DOA estimation with relatively low time and space complexity.

3 Methodology

In this section, we introduce the proposed method for underwater DOA estimation using an acoustic array. We first formulate the problem of underwater DOA estimation using a uniform sonar linear array. Secondly, we show the scheme overview of the proposed method. Then we introduce the feature extraction part of our method. Finally, we describe the proposed CRNN for underwater DOA estimation.

3.1 Problem formulation

This paper presents a methodology based on an underwater sonar array. The methodology considers a uniform sonar linear array for the reception of acoustic signals. The assumption of narrowband conditions requires the time during which the signal passes through the complete array to be less than the coherence time of the generated signal. It is also assumed that the generating source and the receiving array lie on the same plane. Similarly, the far-field condition is imposed and it is assumed that the signal reaches the array as a plane wave. Consider a total of N narrowband signals, denoted by $s_n(t)$, which are received by the sonar array consisting of M sensor elements with an inter-element distance of d . The wave path difference among the elements of the receiving array is denoted by D_m and can be expressed as:

$$D_m = (m-1)dsin\theta_n \tag{1}$$

The time difference of arrival among the elements of the array can be computed as:

$$\tau_m = \frac{D_m}{v} \tag{2}$$

In the above equation, v denotes the speed of signal propagation in the underwater environment. The phase-shift of the signals approaching the array can then be computed as:

$$\beta = e^{-j*2\pi f\frac{(m-1)dsin\theta_n}{v}} \tag{3}$$

This can be further elaborated by setting $f=v/\lambda$:

$$\beta = e^{-j*2\pi\frac{(m-1)dsin\theta_n}{\lambda}} \tag{4}$$

The signal received by the array can then be formulated as:

$$x_m(k) = \sum_{n=1}^{N} S_n(k)e^{-j*2\pi\frac{(m-1)dsin\theta_n}{\lambda}} + n_m(k) \tag{5}$$

In the above equation, $n_m(k)$ denotes the interference at the m th sensor element. The objective is to estimate the DOA of the source using the acquired signal. The methodology divides the set of incident angles [−70°,70°] into 141 distinct classes and computes the probability among the classes to give the model output.

3.2 Methodology overview

The overview of the proposed deep learning-based underwater DOA estimation method using CRNN is shown in Figure 1. Initially, we use the BELLHOP model to simulate the multipath array signals for the underwater marine environment. Then, we extract the acoustic features in the feature extraction stage. Next, we propose to use CRNN for modeling the local and temporal acoustic characteristics, and finally, the trained model is used to estimate the direction of the target source.

3.3 BELLHOP for underwater data simulation

BELLHOP is a well-known model for ocean environment simulations, allowing acoustic ray tracing to be performed by configuring the ocean environmental files and predicting the acoustic pressure fields in the ocean (Porter and Bucker, 1987). As BELLHOP provides detailed modeling of the underwater

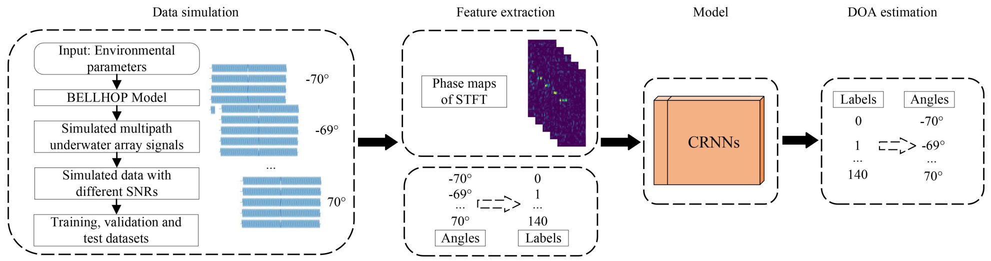


FIGURE 1
Overview of the proposed method.

environment, we used it in this study to generate underwater acoustic signals.

In the BELLHOP tool, the number of multipaths, incident angles, transmission losses, amplitudes, and delays are obtained by specifying the parameters of the channel geometry, velocity profile, submarine topography, and interface reflection loss (Porter, 2011).

3.4. Feature extraction

Feature extraction aims to extract an acoustic representation that enables the acoustic model to learn the mapping from array signals to a set of DOA values *via* training. In this paper, we use the phase map of STFT as the input feature instead of applying explicit feature extraction to calculate the input acoustic features of the network.

We first transform the array signals into STFTs with an N -point Fourier transform. The STFT feature of the array signals is computed as:

$$\boldsymbol{X} = [X_1(t,f), X_2(t,f), \ldots, X_m(t,f), \ldots, X_M(t,f)]^T \qquad (6)$$

where $\boldsymbol{X} \in \mathbb{C}^{M\times T\times F}$ is the STFT feature, T is the number of frames, F is the number of frequency bins, $X_m(t,f) = A_m(t,f)e^{j\phi_m}(t, f)$ is the complex component of $\boldsymbol{X}$ at the m -th element for the t -th frame and f -th frequency bin, and $A_m(t,f)$, $\phi_m(t,f)$ are the corresponding magnitude and phase, respectively. We directly use the phase part of the STFT as the input feature of our method, formulated as $\boldsymbol{Y} \in \mathbb{R}^{M\times T\times F}$, where $F=N/2+1$ can be up to the Nyquist frequency. The phase of the STFT is denoted as the STFT phase in this paper.

3.5 CRNN for underwater DOA estimation

CNNs have achieved significant success in computer vision tasks due to their feature extraction ability. CNNs have recently been used for audio pattern recognition tasks, such as speech recognition, environmental sound recognition, and DOA estimation. Conventional CNNs usually consist of convolution layers, downsampling layers, and fully connected layers. The CNNs adopted in the proposed CRNN are structured as follows.

We assume that the input to the proposed network is $\boldsymbol{Y}$, which is fed to the convolutional layers. We employ three convolutional layers in the CRNN, each having the same kernel size of 3×3 . After each convolutional layer, we apply batch normalization to accelerate and stabilize the training. Next, we adopt the ReLU (Nair and Hinton, 2010) nonlinear activation function. The operations on the input feature Y are expressed as:

$$\boldsymbol{O} = \sigma(\boldsymbol{W} \otimes \boldsymbol{Y} + b) \qquad (7)$$

where $\boldsymbol{O}$ is the output feature, $\boldsymbol{W}$ is the kernel, b is the bias $\otimes$ represents the convolution, and $\sigma(\cdot)$ denotes the ReLU. We apply max-pooling in the downsampling layer to reduce the dimensionality of the feature. The first to third convolutional layers have channels of 32, 64, and 64. The kernels and filters of the convolutional layers are used to learn the local patterns of the input features. High-level features can be extracted by the stacked convolutional layers.

To further improve the performance and simplify the training process, we introduce a residual connection into the proposed CRNN. We assume that the input of the residual connection is $\boldsymbol{O}$ and express the operations of the residual connection as:

$$O' = \sigma(\boldsymbol{W} \otimes \boldsymbol{O} + b) + \boldsymbol{O} \qquad (8)$$

where $\boldsymbol{O}'$ is the output of the residual connection.

CNNs sometimes struggle to capture the temporal dependency of the input features. Therefore, an RNN is used to model temporal sequences by storing historical information in the hidden states. As we are using the phase map of the STFT as the input feature, we adopt an RNN to model the STFT phase over the time dimension. The temporal information is useful for robust DOA estimation.

After the convolutional layers, we average the frequency dimension of the CNN output, denoted as $\boldsymbol{O}_c \in \mathbb{R}^{T'} \times C$, where T' is the time dimension and C is the number of channels of the final convolutional layer. The temporal features from $\boldsymbol{O}_c$ are iteratively modeled by multiple RNN units, which output a hidden feature $\boldsymbol{o}_t$ for each frame. The operations can be formulated as:

$$\begin{aligned} \boldsymbol{o}_t^1 &= F(\boldsymbol{o}_t^0, \boldsymbol{o}_{t-1}^1) \\ &\ldots \\ \boldsymbol{o}_t^l &= F(\boldsymbol{o}_t^{l-1}, \boldsymbol{o}_{t-1}^l), \\ \boldsymbol{o}_t &= \mathcal{G}(o_t^l), \end{aligned} \qquad (9)$$

where $\mathcal{F}$ and G represent the mapping functions of the RNN and l is the number of recurrent layers. In this paper, we use different RNN units, i.e., gated recurrent unit (GRU), bi-directional GRU (biGRU), long short-term memory (LSTM), and bi-directional LSTM (biLSTM).

The output of the final RNN layer is passed into a fully connected layer to predict the direction probabilities. Three different architectures used in this paper are shown in Figure 2.

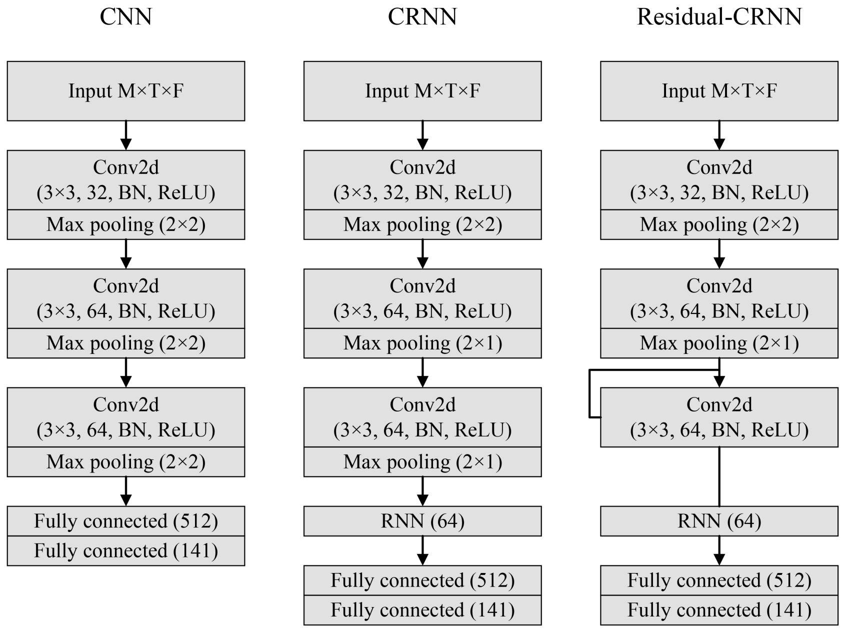


FIGURE 2
Three different architectures of the proposed networks.

4 Experiments

4.1 Dataset

The steps for generating the underwater acoustic signals using BELLHOP are elaborated here. We used a 1000 Hz sound source at a source depth of 75 m, with the receiver placed at a distance of 1.5 km and a depth of 75 m. The DOA varied from −70° to 70° . All other parameters are presented in Table 1. By setting the environmental parameters, the multipath signals (eigenrays) of the receiver were simulated. The received multipath signals consisted of impulse responses with different amplitudes and delays. The multipath signals are illustrated in Figure 3.

TABLE 1 Environmental parameters of BELLHOP.

Parameter	Value
Sound source frequency (Hz)	1000
DOA (°)	-70:1:70
Source depth (m)	75
Source range (km)	1.5
Receiving depth (m)	75
Water depth (m)	100
Sound speed (m/s)	1500
Density (g/cm^3)	1.3
Attenuation (dB/λ)	0.3
Sampling frequency (Hz)	5000
Duration of signal time series (s)	0.5

The dataset used to evaluate our proposed method was generated using the following procedures. We simulated the signal of the first element by applying the BELLHOP toolbox in MATLAB. Array signals were generated by delaying the signal of the first element in the uniform line array (ULA) with an inter-element distance of d = 0.75 m. The DOA angles ranged from −70° to 70° at intervals of 1° . We generated 500 samples for each direction, and so the total number of underwater ULA signals was 70,500. The samples were divided into training, validation, and test sets at a ratio of 7:2:1 . The training and validation sets were fed into the deep learning-based method, and the testing set was used to compare the performance of all methods.

4.2 Feature and training setups

We used the phase map of the STFT as the input acoustic features for the proposed CRNN. The duration of the underwater ULA signals was 0.5 s with a sampling rate of 5000 Hz. We calculated the STFT using a frame length of 26 ms and a hop length of 13 ms in a Hanning window. For each channel of the array signals, the size of the STFT-phase was 65×24 .

For training, we used the Adam optimizer with an initial learning rate of 0.001, which was reduced by a factor of 0.95

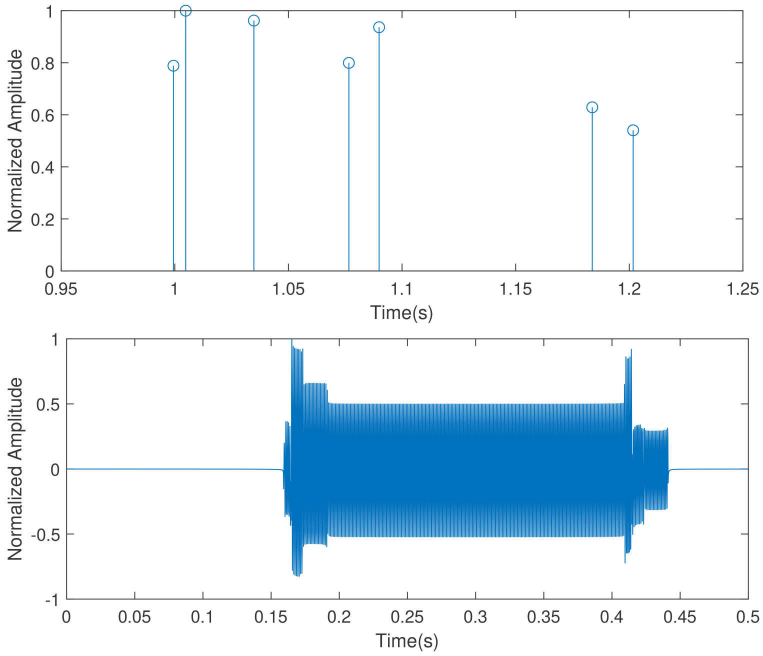


FIGURE 3
Multipath signals using BELLHOP. The picture above is underwater impulse responses with amplitudes and delays, and the picture below is received underwater multipath signals.

every two epochs. The batch size was set to 32. We trained the model for 50 epochs in all of the experiments.

4.3 Performance metrics

We evaluated the performance of our proposed method for the DOA estimation of underwater array signals *via* the classification accuracy (ACC) and root mean square error (RMSE), which is formulated as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} ||\theta_n - \hat{\theta}_n||_2^2} \tag{10}$$

where N is the number of test samples, θ_n is the true angle of the n -th sample, and $\hat{\theta}_n$ is the estimated angle of the n -th sample. The RMSE is calculated as the averaged mean value of 500 samples for all angles from -70° to 70° at intervals of 1°

4.4 DOA estimation of different methods

In this section, we present results from various DOA estimation methods to evaluate the effectiveness of our proposed algorithm.

Table 2 compares the performance of different methods for DOA estimation. CBF and MUSIC are conventional DOA estimation algorithms. The grid interval of these methods was

TABLE 2 Performance comparison of different methods.

SNR (dB)	**-10**		**-5**		**0**	
	ACC (%)	**RMSE (°)**	**ACC (%)**	**RMSE (°)**	**ACC (%)**	**RMSE (°)**
CBF	85.4	0.390	98.1	0.140	98.7	0.116
MUSIC	85.2	0.395	98.0	0.140	99.9	0.036
Two-channel CNN	83.0	0.434	98.9	0.110	99.9	0.029
DTL CNN	85.0	0.638	93.6	0.259	98.0	0.144
Dense U-net	92.3	0.351	99.2	0.090	99.9	0.038
CRNN (proposed)	90.3	0.313	99.3	0.080	99.9	0.024

set to 1° for comparison. The two-channel CNN (Liu et al., 2021) for underwater DOA estimation uses real and imaginary covariance matrices as the two-channel inputs to the network. The deep transfer learning (DTL) CNN was proposed for the DOA estimation using a single vector sensor Cao et al. (2021). The authors also used KRAKEN to simulate the underwater array signals, which is similar to the research target of this paper. This method contains 8 convolutional layers and 4 fully connected layers. We used the phase map of the STFT as input instead of the real and imaginary parts of the cross-spectrum due to the difference between the two types of acoustic arrays. The Dense U-net was proposed for high-resolution DOA estimation using a DenseBlock-based U-net structure with the bearing-time record Sun et al. (2022). In our experiment, we only used the contracting path because we do not focus on reconstructing the input feature. And we also use the phase map of the STFT as the input feature to make a fair comparison.

In Table 2, the results from the different methods are compared in terms of ACC and RMSE under different SNRs with 10 array elements. For all methods, ACC increases and RMSE decreases with increasing SNR. Generally, deep learning-based methods achieve better DOA estimation performance than conventional methods. That is, the deep learning-based methods achieve higher ACCs and lower RMSEs than the conventional methods. Compared with different deep learning-based methods, the proposed CRNN achieves high ACCs of 99.9% (0 dB), 99.3 % (-5 dB), and 90.3 % (-10 dB), and low RMSEs of 0.024° (0 dB), 0.080° (-5 dB), and 0.313° (-10 dB), outperforming the other deep learning-based methods. Dense U-net outperforms two-channel CNN when SNR is -10 and -5 dB but achieves worse results when SNR is 0 dB. Moreover, the DTL CNN fails to achieve good performance for DOA estimation The parameters of CNN are too large, so we assume that the network is overfitting. The above observations indicate that the proposed CRNN outperforms conventional and deep learning-based methods, and achieves stable DOA estimation performance under different SNRs.

4.5 Comparison of networks and features

We conducted ablation experiments to compare the DOA estimation performance of different features and networks. In Table 3, we first analyze the proposed CNN using the covariance matrix as the input feature. The proposed CNN does not achieve better results when the STFT-phase is used as the input feature instead of the covariance matrix. We then introduced various RNNs (GRU, biGRU, LSTM, and biLSTM) into our CNN architecture. The RMSE can be further reduced by introducing the GRU into our CNN model, indicating that RNNs can exploit temporal information in the STFT phase. From the experimental results with the various RNNs, we can see that bi-directional architectures do not produce lower RMSEs. Moreover, the proposed CRNN adopts a residual connection to improve the performance of DOA estimation. It can be seen that the proposed residual-CNN-GRU achieves an ACC of 99.9% and an RMSE of 0.024° , outperforming the other networks.

4.6 Comparison of array elements

For underwater DOA estimation in realistic conditions, the SNR is usually low. More array elements will enable more useful information to be obtained from the target signals. Therefore, we further explored the relationships between the number of array elements and the performance of deep learning-based DOA estimation at a low SNR of -10 dB. Table 4 presents the ACC and RMSE results of the proposed CRNN with 10, 16, and 20 array elements. The ACC increases and the RMSE decrease with the increasing number of array elements. This is because the STFT-

TABLE 3 Performance comparison of different networks and features.

Feature	Network	Metric	
		ACC (%)	RMSE (°)
Covariance matrix	CNN	99.9	0.034
STFT-phase	CNN	99.5	0.072
	CNN-GRU	99.9	0.030
	CNN-biGRU	99.6	0.067
	CNN-LSTM	99.9	0.034
	CNN-biLSTM	99.4	0.080
	Residual-CNN-GRU	99.9	0.024

TABLE 4 Performance comparison of different array elements under SNR of -10 dB.

M	Metrics	
	ACC (%)	RMSE (°)
10	90.3	0.313
16	98.4	0.212
20	99.7	0.052

phase channels increase with an increase in the number of array elements, and so more useful information is available for modeling between the directions and features.

4.7 Comparison of the estimation time

The underwater acoustic environment is complex and dynamic. Therefore, the processing time is important for practical DOA estimation methods. We compared the processing time of conventional methods against that of the proposed CRNN using only the CPU. As deep learning-based methods usually transform underwater signals into acoustic features, we included the processing time for feature extraction to ensure a fair comparison. Figure 4 shows the mean DOA estimation times of different methods with 10 array elements. The experiments were repeated 7,050 times for each method. The proposed CRNN achieves the best DOA estimation time of 3.21 ms. MUSIC has a processing time of 3.29 ms, which is five times faster than that of CBF. This is because the matrix operations are optimized in MUSIC, leading to lower processing times. The proposed CRNN simultaneously achieves high accuracy and fast processing for DOA estimation, making it suitable for real-time applications.

4.8 Analysis of angles within 1°

In realistic conditions, there will be non-integer source directions. Conventional methods can calculate non-integer directions by changing the search grid, but this increases the computation time. However, the proposed deep learning-based DOA estimation method has been trained with defined integer angles. Therefore, we performed an experiment with three different non-integer angles to investigate whether these directions could be correctly estimated by the proposed method. The DOA estimation results of three different non-integer angles are shown in Figure 5. The proposed CRNN tends to classify 0.3° as 0° and 40.7° as 41° . That is, the non-integer angles are mostly classified as the nearest integer value. This indicates that the proposed CRNN can classify non-integer angles to the closest integer angle with the least error. Thus, the deep learning-based method can achieve stable and accurate DOA estimation performance in the case of non-integer directions.

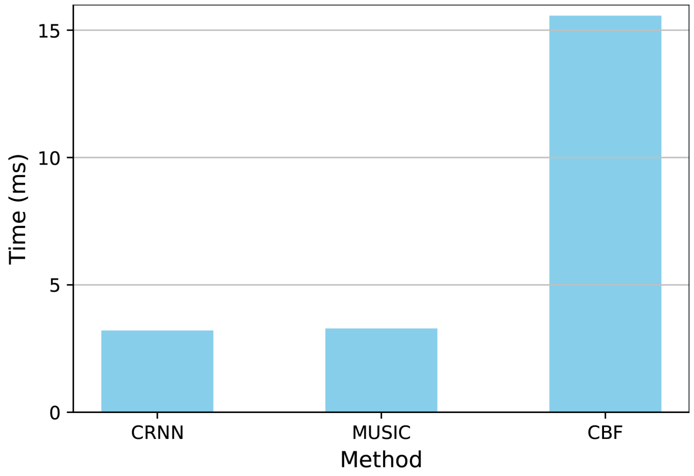


FIGURE 4
Estimation times of different methods.

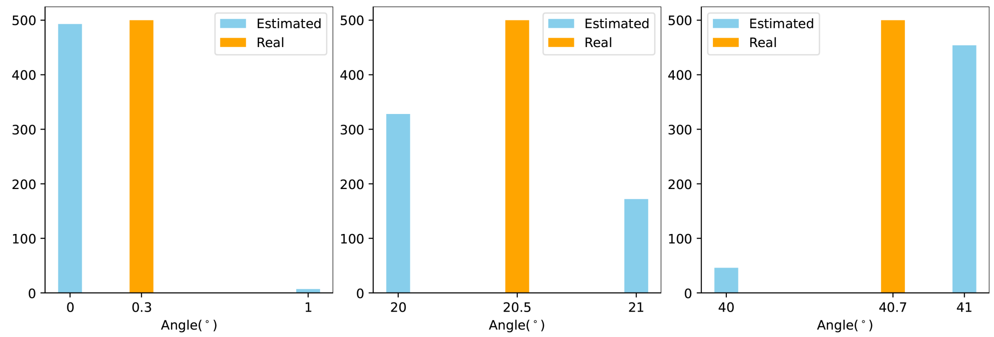


FIGURE 5
DOA estimation results in three different non-integer angles (0.3°, 20.5°, and 40.7°).

4.9 Results in real marine environment

In this section, we conduct experiments using the proposed method and the comparative deep learning-based methods on the actual data in the marine environment. The real marine underwater data was collected in January 2016 with a vertical ULA in the shallow water sea area around China Jiang et al. (2022). The target source is a real ship sailing on the sea. The ULA consists of 16 elements, the depth of the first element is about 205m, and the depth of the 16th element is about 303m. The GPS information of the ULA and the target ship is provided, so we can calculate the distance between the ULA and the target ship, and then calculate the direction of the target ship. The angle of the ship is ranged from 20° to 70°, and we divided the angles into 51 classes. We collected a total of 500 recordings each for a duration of 0.5s at a sampling rate of 17,067 Hz. We divided the dataset into training, validation, and test sets at a ratio of 7:2:1.

Table 5 compares the performance of different deep learning-based methods on the test set for DOA estimation in the real sea environment. The Dense U-net achieves an RMSE of 7.709°, which is higher than the other methods. While the DTL CNN performs better than the two-channel CNN, showing better DOA estimation performance than that on simulated data. And the proposed CRNN achieves an RMSE of 3.594°, outperforming the other methods on the real data.

Similarly, we calculate the RMSEs in different angle intervals for the deep learning-based methods, and the results are shown in Figure 6. It can be seen that the proposed CRNN achieves less than 2° RMSEs when the angle is in 20° - 30° and 40° - 50 and . The two-channel CNN achieves less than 2° RMSEs when the angle is in 20° - 30° and 30° - 40. Similarly, the DTL CNN has less than 2° RMSEs in 60° - 70°, however, the performance of this method degrades in other intervals. The Dense U-net achieves an RMSE of about 2° in 40° - 50 but fails to achieve robust performance in other intervals.

From the above observations, it is seen that the DOA estimation results in a real marine environment are not accurate as in simulation. Nevertheless, the proposed method can achieve robust performance for DOA estimation in different underwater environments, and deep learning-based methods can be applied to more complex underwater environments.

5 Conclusion

This paper has proposed a CRNN-based method for underwater DOA estimation employing an acoustic ULA. We used the phase component of the STFT as the input feature of the CRNN. The CRNN structure uses CNN layers to extract local invariant features and RNN layers to model the temporal dependencies of the input features. The method was validated on a dataset consisting of multipath signals, which was simulated using the BELLHOP model and a ULA. We compared the proposed CRNN with traditional and deep learning-based methods for DOA estimation. The simulations and

TABLE 5 Performance comparison of deep learning-based methods on real data.

Method	RMSE(°)
Dense U-net	7.709
Two-channel CNN	5.395
DTL CNN	4.943
CRNN (proposed)	3.594

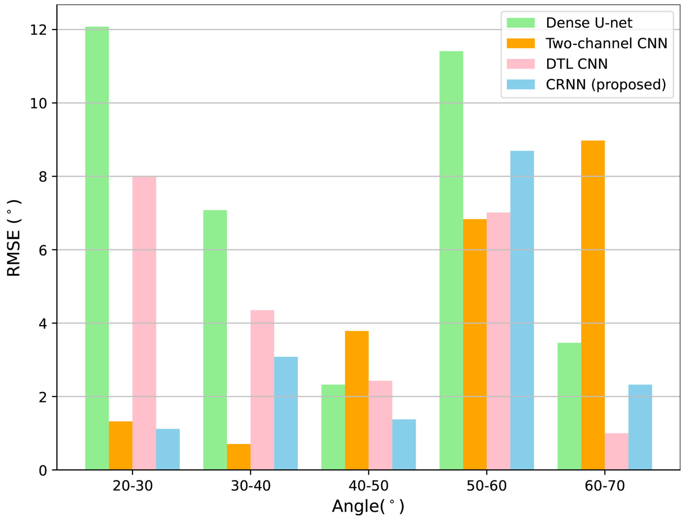


FIGURE 6
DOA estimation results in different intervals of angles.

experimental analysis at various SNRs indicate that the proposed method achieves high accuracy and low RMSEs compared with previous methods. We also experimented with different network architectures and found that the residual-CNN-GRU achieves the best DOA estimation performance. In a comparison of different array elements under a low SNR of -10 dB, it was observed that the DOA estimation could be improved by increasing the number of array elements. The proposed CRNN has a lower estimation time than other DOA methods. Similarly, experiments are also validated on real data captured from the sea. The observations and experimental results show that the proposed method is sufficiently robust and accurate for underwater DOA estimation in different underwater environments, and can be applied to various underwater monitoring tasks.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XL wrote the data processing code, conducted the experiments, and wrote the first draft of the manuscript. JC designed the experiments and revised the draft of the manuscript. JB conducted the experiments and analysis and created the figures. AM, DZ, MSA, and QY conducted the experiments. All authors contributed to manuscript revision, read and approved the submitted version, and were involved in the conception and design of the study.

Funding

This work is supported by the National Natural Science Foundation of China (Grant No. 62071383) and the Key Research and Development Plan of Shaanxi Province (Grant No. 2021NY-036).

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

Ables, J. (1974). Maximum entropy spectral analysis. *Astron. Astrophys. Supp. Ser.* 15, 383 t1.

Ayub, M. S., Jianfeng, C., and Zaman, A. (2021). "Multiple source data association for distributed acoustic sensor network in open environment," in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (NEW YORK, USA: IEEE) 01–06.

Ayub, M. S., Jianfeng, C., and Zaman, A. (2022). Multiple acoustic source localization using deep data association. *Appl. Acoust.* 192, 108731. doi: 10.1016/j.apacoust.2022.108731

Bai, J., Chen, C., and Chen, J. (2019). "A multi-feature fusion based method for urban sound tagging," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 1313–1317 (NEW YORK, USA: IEEE).

Bai, J., Chen, J., and Wang, M. (2022). Multimodal urban sound tagging with spatiotemporal context. *IEEE Trans. Cogn. Dev. Syst* 1-1. doi: 10.1109/TCDS.2022.3160168

Bai, J., Wang, M., and Chen, J. (2021). "Dual-path transformer for machine condition monitoring," in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1144–1148 (NEW YORK, USA: IEEE).

Barabell, A. (1983). "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 8. 336–339 (NEW YORK, USA: IEEE).

Berger, A., Della Pietra, S. A., and Della Pietra, V. J. (1996). A maximum entropy approach to natural language processing. *Comput. Linguist.* 22, 39–71.

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., et al. (2019). Machine learning in acoustics: Theory and applications. *J. Acoust. Soc. Am.* 146, 3590–3628. doi: 10.1121/1.5133944

Candès, E. J., Baraniuk, R. G., Candes, E., Nowak, R., and Vetterli, M.. (2006). "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, Vol. 3. 1433–1452 (Citeseer).

Cao, H., Wang, W., Su, L., Ni, H., Gerstoft, P., Ren, Q., et al. (2021). Deep transfer learning for underwater direction of arrival using one vector sensor. *J. Acoust. Soc Am.* 149, 1699–1711. doi: 10.1121/10.0003645

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* 57, 1408–1418. doi: 10.1109/PROC.1969.7278

Chakrabarty, S., and Habets, E. A. (2017). "Broadband DOA estimation using convolutional neural networks trained with noise signals," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 136–140 (NEW YORK, USA: IEEE).

Chen, J., Wang, M., Zhang, X.-L., Huang, Z., and Rahardja, S. (2022). "End-to-end multi-modal speech recognition with air and bone conducted speech," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing*. 6052–6056 (NEW YORK, USA: IEEE).

Desai, D., and Mehendale, N. (2022). A review on sound source localization systems. *Arch. Comput. Methods Eng.*, 1–12. doi: 10.1007/s11831-022-09747-2

Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* 52, 1289–1306. doi: 10.1109/TIT.2006.871582

Eriksson, A., and Stoica, P. (1994). Optimally weighted ESPRIT for direction estimation. *Signal Process* 38, 223–229. doi: 10.1016/0165-1684(94)90141-4

Ferguson, E. L., Ramakrishnan, R., Williams, S. B., and Jin, C. T. (2017). "Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2657–2661 (NEW YORK, USA: IEEE).

Han, X., Liu, M., Zhang, S., and Zhang, Q. (2019). A multi-node cooperative bearing-only target passive tracking algorithm *via* UWSNs. *IEEE Sens. J.* 19, 10609–10623. doi: 10.1109/JSEN.2019.2931885

Han, X., Liu, M., Zhang, S., Zheng, R., and Lan, J. (2021). A passive doa estimation algorithm of underwater multipath signals *via* spatial time-frequency distributions. *IEEE Trans. Veh. Technol.* 70, 3439–3455. doi: 10.1109/TVT.2021.3064279

Houégnigan, L., Safari, P., Nadeu, C., André, M., and van der Schaar, M. (2017). "Machine and deep learning approaches to localization and range estimation of underwater acoustic sources," in *2017 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)*. 1–6 (NEW YORK, USA: IEEE).

Hu, B., Liu, M., Yi, F., Song, H., Jiang, F., Gong, F., et al. (2020). DOA robust estimation of echo signals based on deep learning networks with multiple type illuminators of opportunity. *IEEE Access* 8, 14809–14819. doi: 10.1109/ACCESS.2020.2966653

Jia, F., Cheng, E., and Yuan, F. (2012). "The study on time-variant characteristics of under water acoustic channels," in *2012 International Conference on Systems and Informatics*. 1650–1654 (NEW YORK, USA: IEEE).

Jiang, J., Wu, Z., Huang, M., and Xiao, ,. Z. (2022). Detection of underwater acoustic target using beamforming and neural network in shallow water. *Appl. Acoust.* 189, 108626. doi: 10.1016/j.apacoust.2021.108626

Jing, H., Wang, H., Liu, Z., and Shen, X. (2018). Doa estimation for underwater target by active detection on virtual time reversal using a uniform linear array. *Sensors* 18, 2458. doi: 10.3390/s18082458

Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Front. Mar. Sci* 8, 2049. doi: 10.3389/fmars.2021.823173

Knapp, C., and Carter, G. (1976). The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Process* 24, 320–327. doi: 10.1109/TASSP.1976.1162830

Li, J., Tong, F., Zhou, Y., Yang, Y., and Hu, Z. (2022). Small size array underwater acoustic doa estimation based on direction-dependent transmission response. *IEEE Trans. Veh. Technol.*, 1–12. doi: 10.1109/TVT.2022.3197922

Liu, Y., Chen, H., and Wang, B. (2021). DOA estimation based on CNN for underwater acoustic array. *Appl. Acoust.* 172, 107594. doi: 10.1016/j.apacoust.2020.107594

Li, J., Wang, J., Wang, X., Qiao, G., Luo, H., and Gulliver, T. A. (2019). Optimal beamforming design for underwater acoustic communication with multiple unsteady sub-Gaussian interferers. *IEEE Trans. Veh. Technol.* 68, 12381–12386. doi: 10.1109/TVT.2019.2945007

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *ICML*.

Niu, H., Reeves, E., and Gerstoft, P. (2017). Source localization in an ocean waveguide using supervised machine learning. *J. Acoust. Soc Am.* 142, 1176–1188. doi: 10.1121/1.5000165

Ozanich, E., Gerstoft, P., and Niu, H. (2020). A feedforward neural network for direction-of-arrival estimation. *J. Acoust. Soc Am.* 147, 2035–2048. doi: 10.1121/10.0000944

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.

Porter, M. B., and Bucker, H. P. (1987). Gaussian Beam tracing for computing ocean acoustic fields. *J. Acoust. Soc Am.* 82, 1349–1359. doi: 10.1121/1.395269

Rao, B. D., and Hari, K. S. (1989). Performance analysis of root-MUSIC. *IEEE Trans. Acoust. Speech Signal Process* 37, 1939–1949. doi: 10.1109/29.45540

Ren, Q., and Willis, A. (1997). Fast root MUSIC algorithm. *Electron. Lett.* 33, 450–451. doi: 10.1049/el:19970272

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

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

Shen, Y., Pan, X., Zheng, Z., and Gerstoft, P. (2020). Matched-field geoacoustic inversion based on radial basis function neural network. *J. Acoust. Soc Am.* 148, 3279–3290. doi: 10.1121/10.0002656

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

Stoica, P., and Nehorai, A. (1990). MUSIC, maximum likelihood, and Cramer-rao bound: further results and comparisons. *IEEE Trans. Acoust. Speech Signal Process* 38, 2140–2150. doi: 10.1109/29.61541

Sun, D., Jia, Z., Teng, T., and Ma, C. (2022). Robust high-resolution direction-of-arrival estimation method using denseblock-based u-net. *J. Acoustic. Soc. America* 151, 3426–3436. doi: 10.1121/10.0011470

Sun, X., Jia, X., Zheng, Y., and Wang, Z. (2021). A data-driven method for estimating the target position of low-frequency sound sources in shallow seas. *Front. Inf. Technol. Electron. Eng.* 22, 1020–1030. doi: 10.1631/FITEE.2000181

Swindlehurst, A. L., Ottersten, B., Roy, R., and Kailath, T. (1992). Multiple invariance ESPRIT. *IEEE Trans. Signal Process* 40, 867–881. doi: 10.1109/78.127959

Viberg, M., Ottersten, B., and Kailath, T. (1991). Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Trans. Signal Process* 39, 2436–2449. doi: 10.1109/78.97999

Wang, Y., and Peng, H. (2018). Underwater acoustic source localization using generalized regression neural network. *J. Acoust. Soc Am.* 143, 2321–2331. doi: 10.1121/1.5032311

Wang, M., Zhao, M., Chen, J., and Rahardja, S. (2019). Nonlinear unmixing of hyperspectral data *via* deep autoencoder networks. *IEEE Geosci. Remote Sens. Lett.* 16, 1467–1471. doi: 10.1109/LGRS.2019.2900733

Xiang, H., Chen, B., Yang, T., and Liu, D. (2020). Phase enhancement model based on supervised convolutional neural network for coherent DOA estimation. *Appl. Intell.* 50, 2411–2422. doi: 10.1007/s10489-020-01678-4

Xiang, H., Chen, B., Yang, M., Xu, S., and Li, Z. (2021). Improved direction-of-arrival estimation method based on LSTM neural networks with robustness to array imperfections. *Appl. Intell.* 51, 4420–4433. doi: 10.1007/s10489-020-02124-1

Xiao, X., Zhao, S., Zhong, X., Jones, D. L., Chng, E. S., and Li, H. (2015). "A learning-based approach to direction of arrival estimation in noisy and reverberant environments," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2814–2818 (NEW YORK, USA: IEEE).

Xu, X.-L., and Buckley, K. M. (1992). Bias analysis of the MUSIC location estimator. *IEEE Trans. Signal Process* 40, 2559–2569. doi: 10.1109/78.157296

Yang, T. (2012). Properties of underwater acoustic communication channels in shallow water. *J. Acoust. Soc Am.* 131, 129–145. doi: 10.1121/1.3664053

Yuan, Y., Wu, S., Wu, M., and Yuan, N. (2021). Unsupervised learning strategy for direction-of-arrival estimation network. *IEEE Signal Process Lett.* 28, 1450–1454. doi: 10.1109/LSP.2021.3096117

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). 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. (2021). An improved imaging algorithm for multi-receiver SAS system with wide-bandwidth signal. *Remote Sens.* 13, 5008. doi: 10.3390/rs13245008

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 Nav.* 16, 531–541. doi: 10.1049/rsn2.12200

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Nav.* 14, 1055–1060. doi: 10.1049/iet-rsn.2019.0477

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fmars.2022.1066693

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Haocai Huang,
Zhejiang University, China
Enjin Zhao,
China University of Geosciences
Wuhan, China

*CORRESPONDENCE
Xiao-Hua Zhu
xhzhu@sio.org.cn

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 11 October 2022
ACCEPTED 21 November 2022
PUBLISHED 01 December 2022

CITATION
Xiao C, Zhu X-H, Zhu Z, Zhang C,
Zheng H, Zhang Z, Zhong J and Wei L
(2022) Effect of dam discharges and
tides to transport of Yangtze River
using coastal acoustic
tomography systems.
Front. Mar. Sci. 9:1066693.
doi: 10.3389/fmars.2022.1066693

Effect of dam discharges and tides to transport of Yangtze River using coastal acoustic tomography systems

Cong Xiao[1], Xiao-Hua Zhu[1,2,3]*, Zenan Zhu[1], Chuanzheng Zhang[1], Hua Zheng[1,3], Zhensheng Zhang[1], Jiwen Zhong[4] and Lixin Wei[4]

[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, [4]Lower Changjiang River Bureau of Hydrological and Water Resources Survey, Hydrology Bureau of Changjiang Water Resources Commission, Nanjing, China

The streamflow characteristics within the Yangtze River Basin have experienced substantial fluctuations in recent years because of the combined effects of environmental factors and intensive human activities. In this study, at the Datong station, two coastal acoustic tomography (CAT) systems were used to track the Yangtze River discharge from July 2018 to January 2021. The stage–discharge relationship presented large uncertainties because of the Three Gorges Dam (TGD) operations, whereas the CAT method performed effectively in discharge monitoring even during extreme flood events. The distribution of downstream discharge was concentrated because of the regulation by the TGD. Analysis of the potential drivers in the downstream river hydrology reveals that the effect of rainfall events (leading to a maximum of ~40% changes) was heavily influenced by the regulation by the TGD (at least 50% contribution). Additionally, the river–tide process is also sensitive to the discharge regulated by the TGD. The discharge induced by tidal waves was negligible (a maximum of 1.11% change). This work demonstrates that an acoustic method can effectively monitor the massive flood discharge in unsteady flow conditions in large rivers, thereby facilitating the management of large-scale dam- and tide-influenced river systems.

KEYWORDS

coastal acoustic tomography, discharge of Yangtze River, Three Gorges Dam, precipitation, tidal wave

Introduction

River discharge is a key hydrological issue for the river and water resources management. The continuous and real-time monitoring of water discharge and its variation at different time scales are of prime importance (Mei et al., 2019), especially for large river basins. Thus, establishing technology and methods for river discharge monitoring is pertinent. Usually, river discharge is estimated based on stage–discharge relationships using the rating curve (RC) method (Gore and Banning, 2017). Usually a one-to-one relationship between the river stage and water discharge is not established under unsteady flow conditions (Baldassarre and Montanari, 2009), thereby posing a challenge in measuring the cross-sectional velocity and discharge in unsteady flows or during extreme flood events. Recently, a variety of hydroacoustic instruments, for example, acoustic velocity meters (Ruhl and Derose, 2004) and horizontal acoustic Doppler profilers (Hoitink et al., 2009), have been employed to measure river discharge. However, these shipboard measurements lead to short duration estimates and are labor-intensive, and observations during extreme hydrological events are difficult and dangerous, especially for large river channels.

The Yangtze River (YR), China's largest river, has attracted extensive attention from various research fields (e.g., ecology, hydrology, and climatology) in recent decades, particularly since the Three Gorges Dam (TGD) commenced operations in 2003 (Dai et al., 2010; Guo et al., 2012; Sun et al., 2012; Lai et al., 2014; Lyu et al., 2018; Cai et al., 2019; Gao et al., 2021). Effective monitoring and control of floods in the Yangtze River Basin (YRB) ensure China's national water security (Xia and Chen, 2021). Historically, frequent and massive floods have significantly halted social advancement of the YRB (Zhang et al., 2021). Floods from the upstream Yichang station (~1,800 km upstream from the estuary) typically take at least 15–20 days to reach the estuary, and coupled with the incoming discharge from the middle–lower reaches, the floods remain for a long duration, which makes the middle–lower regions of the YRB vulnerable to water disasters (Jia et al., 2022). Therefore, an accurate estimation of water discharge in the middle–lower regions of the YRB, especially the precise determination of the instantaneous flood peak, is essential for the management and control of floods in the YRB (Luo et al., 2015).

Coastal acoustic tomography (CAT) is a novel flow measurement technique (Kaneko et al., 2020) that does not disturb shipping traffic and fishing activity, and it is extensively utilized to gauge water currents and discharge in estuaries and tidal rivers (Zhu et al., 2012; Kawanisi et al., 2017). Compared with traditional gauging stations or acoustic Doppler current profiler (ADCP) transect water discharge measurements, the depth- and section-averaged velocities along the sound ray path can be easily gauged using the acoustic tomography method at a high temporal resolution (Kawanisi et al., 2018). Unlike conventional methods, CAT can immediately and in real-time observe the cross-sectional average velocity based on its sound reciprocal travel time without using complicated post-processing procedures. A previous study demonstrated the feasibility of using CAT to estimate discharge at the Datong (DT) station of the YR for 2.5 years (Zhu et al., 2021) and they successfully reported two massive floods with a peak discharge of > 80,000 m^3/s; however, it did not discuss the roles of the TGD in discharge control and the tidal effect on the discharge. This inspired us to use the CAT dataset to study the temporal runoff characteristics at the DT hydrometric station. Consequently, the primary objective of this study was to estimate the continuous discharge in large river using CAT and investigate the hydrological runoff characteristics in a critical location in the large river basin along the YR. A comprehensive understanding of the influence of the dam and the barrier effects of tidal waves on YR discharge would contribute to better water resource use and flood prevention in the YRB.

Data and methods

Field observations

In this study, the CAT observation region was located at the DT station (Figure 1). The locations of the observation area and the two CAT systems are shown in Figure 1B. the horizontal distance between the two CAT stations is 3,015 m and the angle between the *NS* line and the north riverbank is ~31°

The DT station is the first hydrometric station along the main channel of the YR, situated ~1,100 km downstream of the TGD and ~630 km upstream from the estuary (Cai et al., 2019; Mei et al., 2021). Additionally, the DT station is the most widely acknowledged location of the topmost boundary of the tide, and the discharge is not altered by the tide (Shi et al., 2018). In the middle–lower reaches of the YR, the water levels frequently exceeded the warning levels owing to the continuous heavy rainfall during the rainy season. Thus, monitoring water discharge using CAT at the DT station is crucial for flood management in the middle–lower reaches of the YR.

The YR is ~6,300 km long and the area of YRB is 1.8 million km^2. The total annual water flow through the YR is 961.6 billion m^3, accounting for 36% of the total runoff in Chinese rivers (Yu et al., 2020). The world's largest hydroelectric project, the TGD, is situated in the middle reaches of the YR, it measures 2,309.5 meters in length and 185 m in height. There is an artificial lake (Three Gorges Reservoir) formed upon the accomplishment of the TGD, covering 1,084 km^2 with a total reservoir capacity of 39.3 billion m^3 (Wang et al., 2020).

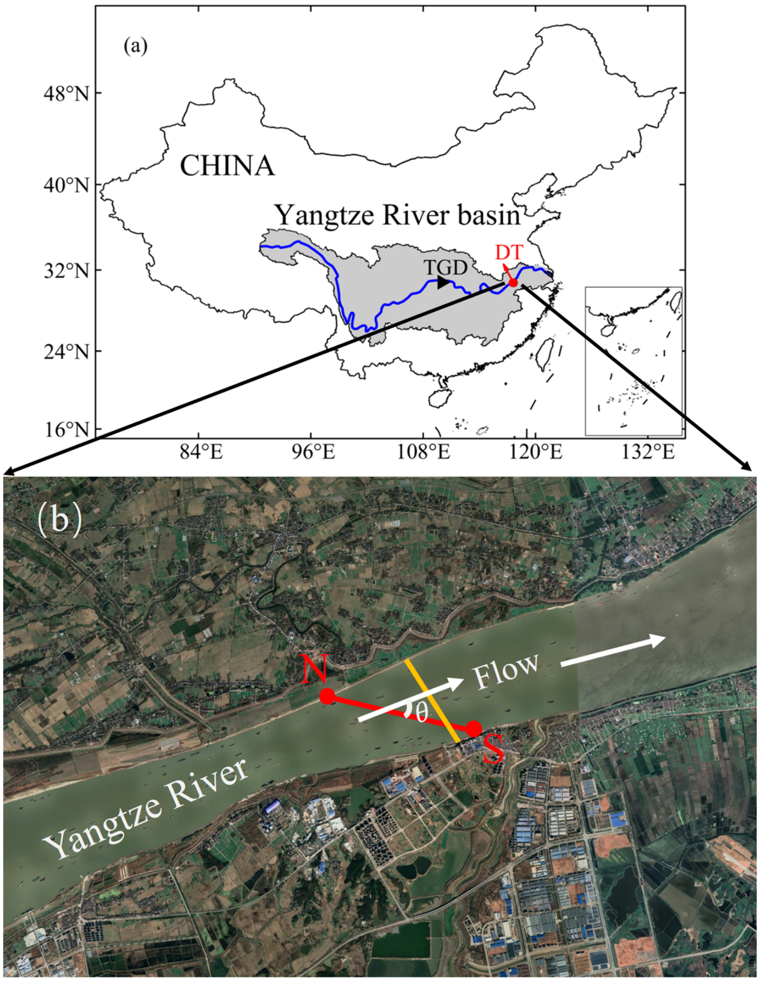


FIGURE 1
(A) Map of the Yangtze River Basin and the main channel of Yangtze River (blue line); **(B)** Map of the observation site and positions of two CAT stations (*N* and *S*), *θ* represents the angle between the main flow direction and ray path of two CAT stations projected to the horizontal plane, the station *S* is near the Datong (DT) hydrometric station, the yellow line denotes the transect of the moving-boat ADCP. The black triangle denotes the position of the Three Gorges Dam (TGD), and the red dot denotes the location of the DT station.

Precipitation in the YRB showed distinct seasonal patterns from 2018 to 2021 (Figure 2). Precipitation data were obtained from the fifth generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) hourly data (https://cds.climate.copernicus.eu) for global climate and weather. Precipitation in the YRB is typically high from March to May, followed by the plum-rain season over the middle–lower river areas from June to early July (Dai, 2021) (Figure 2B). During summer, the entire river catchment may be covered by the rain belt from June to August (Figure 2C). The runoff of the YRB exhibits a high consistency with rainfall and progressively declines from southeast to northwest (Gao et al., 2021), with low rainfall from November to February (Figure 2A) and high rainfall from June to August (Dai, 2021) (Figure 2C).

Overview and measurements using CAT

Discharge measurements were conducted using CAT for ~2.5 years (July 2018–January 2021) at the DT hydrometric station (near station *S*). As shown in Figure 1B, two CAT systems were set on both north (*N*) and south (*S*) riverbanks, with a distance of 3,015 m between the two stations. The monitoring systems were placed on a floating barge, and the CAT transducers were suspended 2 m under

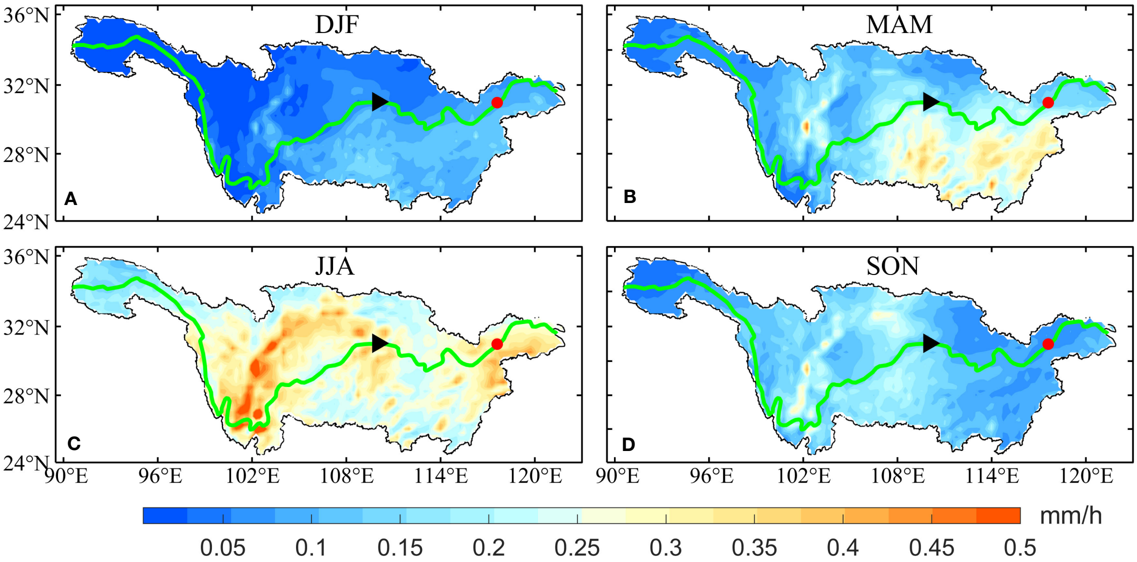


FIGURE 2
Seasonal climatological distributions of precipitation **(A)** December–January–February (DJF), **(B)** March–April–May (MAM), **(C)** June–July–August (JJA), and **(D)** September–October–November (SON) over the Yangtze River Basin from 2018 to 2021. The black triangle shows the position of the Three Gorges Dam (TGD), and the red dot shows the location of the Datong (DT) station.

water. The CAT systems were supplied with a clock from GPS receivers to maintain and guarantee simultaneous operations of the systems. During the observation period, the acoustic pulse was modulated by an 11-order M sequence, and acoustic pulses were simultaneously transmitted from both transducers every 5 min with a central frequency of 9 kHz.

The travel-time method was used in CAT, assuming the same reciprocal travel path between two stations; therefore, the travel time of the acoustic signals from site N to site S (t_{NS}) and site S to site N (t_{SN}) can be represented as

$$t_{NS} = \frac{L}{c_{CAT} + u_{CAT}} \tag{1}$$

$$t_{SN} = \frac{L}{c_{CAT} - u_{CAT}} \tag{2}$$

where L represents the oblique distance between the two stations, c_{CAT} and u_{CAT} represent the section-averaged sound speed and velocity, respectively.

The discharge obtained by CAT (Q_{CAT}) can be written as:

$$Q_{CAT} = A \cdot u_{CAT} \cdot \cos\theta \tag{3}$$

where θ is the flow angle, and A is the cross-sectional area. A was estimated using the mean water depth and mean distance, since two stations were fixed, and the error of A was mostly related to the water depth error. As Zhu et al. (2021) showed that an error of 0.19 m in mean water depth would induce an ~1% (~329 m^2) error in the estimation of mean area.

ADCP measurements and establishing the rating curve

During the CAT observation period, many moving-boat ADCP (600 kHz Rio Grande Workhorse, California, USA) observations were conducted to collect reference discharge data at the DT station. For each campaign, the ADCP was arranged under water (0.5 m), a differential GPS receiver was fixed above the ADCP to record the real-time position. During each campaign, ADCP sampling was conducted each 2 s using the ADCP bottom tracking mode. The bin number was set at 70 and each bin size was 0.5 m and the unmeasurable surface depth was 0.25 m. The accuracy of velocity was ± 0.25% of the measured velocity with a resolution of 0.1 cm/s.

Owing to the limitations of the moving-boat ADCP method, there are a blank area and the underwater depth of the ADCP in the surface layer and the near-bed flow velocity could not be measured by ADCP. Hence, water discharge in the surface and bottom layers and areas adjoining riverbanks were not measured. Generally, the discharge in these unmeasured areas can be ignored in small river systems; however, in large river systems such as the YR, estimation of the water discharge in these unmeasured areas is necessary. Hence, in this study, the velocities in the surface and near-bed layers were estimated using linear extrapolation of data of the nearest bin, that is, the surface velocity was estimated from data of the first bin and the bottom velocity was estimated from data of the deepest bin. The boat is difficult to start from the river bank, there will be a certain distance from the bank. The velocity in areas close to river banks was derived through extrapolation, assuming that the near-bank water column progressively became shallow with a steady water slope until it extended to the river bank (Gordon, 1989). Finally, the discharge of each ADCP transect was estimated *via* integration.

Missing ADCP data in the surface layer accounted for the largest fraction, with a range of 543 to 3,863 m^3/s. Missing data in the bottom layer varied from 650 to 3,389 m^3/s, whereas missing data in the left and right areas were negligible with their averages accounting for 0.2 and 0.3% of the total discharge, respectively. In total, 140 discharge estimates (Q_{ADCP}) were collected through the moving-boat ADCP during the CAT observation period. The Q_{ADCP} ranged from 11,969 to 83,055 m^3/s.

A HOBO water level logger with a pressure sensor (U20L-01) was attached at station *S* near the riverbed to measure the water depth (*H*) every 10 min with a precision of ± 0.1 cm and a typical error of ± 0.1% of the measured range of 1.0 cm. The collected water depth was further utilized to estimate the cross-sectional area and build an RC.

An RC relationship (Kawanisi et al., 2016) was established between the moving-boat ADCP discharges and water depths at station *S* as follows:

$$Q_{RC} = c_1(H + c_2)^2 \tag{4}$$

where c_1 and c_2 are the calibration constants and they are determined using the least-squares method.

Extraction of baseflow and stormflow

Understanding the response of runoff to precipitation events and anthropogenic activities is essential for hydrological applications and water resource management (Wang et al., 2022). Baseflow is an essential component of runoff in hydrological basins (Dai et al., 2010), whereas stormflow is usually at a higher volume over a short period following major rainfall events. Estimating baseflow and stormflow can provide important references for the sustainability of water resources in the YRB. Hence, the HydRun toolbox (Tang and Carey, 2017), which can automatically distinguish baseflow and stormflow based on the recursive digital filter technique, was utilized in this work. HydRun can also match stormflow with their accompanying rainfall events using a flexible time step. Further details are available in Tang and Carey (2017).

Grouped frequency distribution

The grouped frequency distribution is usually recommended when a large number of continuous variables must be evaluated (Mei et al., 2015), and it first groups parameters into different intervals based on their values, and then, gives each class a frequency. Thus, the relative frequency (*RF*) is determined by

$$RF = \frac{N_i}{N} \tag{5}$$

where N_i represents the number that falls in a specific class and *N* is the amount of data.

Grouped frequency analysis was applied at the DT station (Q_{CAT}) to determine the occurrence probabilities of water discharge and TGD outflow at different intervals.

Results

Rating curve and index velocity methods

Figure 3A shows the relationship between water depth (*H*) at station *S* and moving-boat ADCP discharge (Q_{ADCP}). The power function in Eq. 4 is fitted to the relation to establish the RC, the RC (red line) and the 95% confidence intervals (gray lines) are presented in Figure 3A. In addition, a linear relationship (R^2 = 0.997) was established between the moving-boat ADCP-determined cross-sectional area and the water depth (*H*) at station *S* (Figure 3B). This empirical relationship was used to estimate the time series of the cross-sectional area.

Kawanisi et al. (2013) stated that the incorrect resolve of angle (θ) between the transmission line and the main flow direction might cause a considerable discharge estimation error and revealed that variations of ±1° could induce relative discharge errors of −5.4 and 5.5%. Hence, this study utilized an index velocity for u_{CAT} to avoid estimating flow direction. The index velocity ratings are shown in Figure 4. The index velocity curve for CAT was resolved using linear regression, the measured section-averaged velocity $u_{ADCP} = Q_{ADCP}/A_{ADCP}$, with Q_{ADCP} representing the moving-boat ADCP discharge and A_{ADCP} representing the sectional area along the ADCP transect. Finally, the streamflow of CAT can be computed by:

$$Q_{CAT} = (0.9922u_{CAT} + 0.009) \times (1973H + 19947) \tag{6}$$

Time series of discharge and TGD outflow

CAT- and RC-derived discharges are shown in Figure 5A. Discharges of the YR ranged from 11,059 to 82,206 m^3/s from July 2018 to January 2021 and displayed an evident seasonal variation. The discharge exhibits high consistency with the rainfall pattern, with fluctuations of high discharge and precipitation in summer and low discharge and precipitation in winter (Figure 5A). CAT shows a good consistency with the

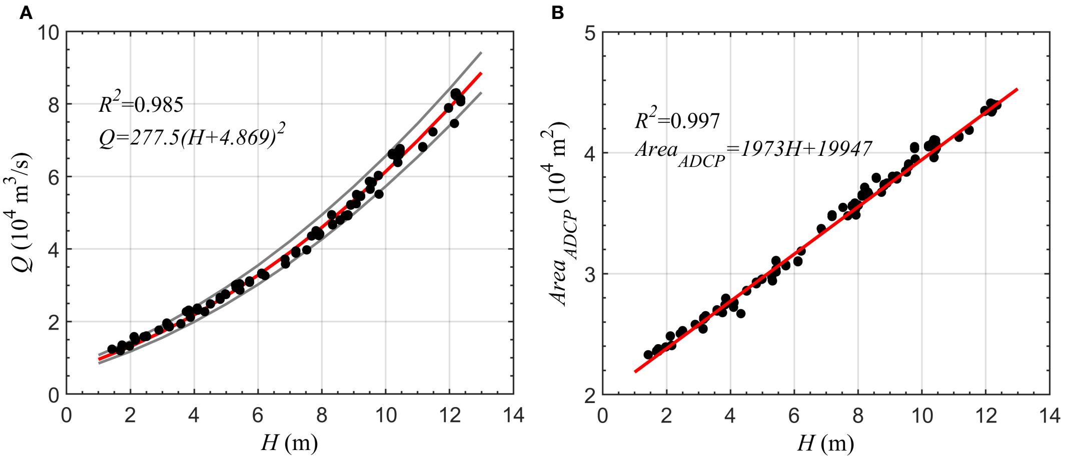


FIGURE 3
(A) RC relationship (red line) build using moving-boat ADCP results, gray lines represent the 95% confidence level; **(B)** Linear empirical relationship between the water depth **(H)** and moving-boat ADCP measured area ($Area_{ADCP}$).

moving-boat ADCP results. The comparison between Q_{ADCP} and Q_{CAT} showed a good correlation with R^2 = 0.999 and RMSE = 891.9 m^3/s (Figure 6). Compared with previous studies (Zhu et al., 2021), the results of this study are higher by approximately 1%, which is mainly due to the index method and the consideration of the flow angle.

Figure 5B shows the outflow, water storage, and water level in the TGD. TGD data were collected from the National Rain and Water Discharge website (http://xxfb.mwr.cn/index.html); however, only data with 6-h intervals during the wet season (June to September) were collected. As shown in Figure 5B, during the observation period, the TGD outflow varied from 7,900 to 49,400 m^3/s. Since 2010, the water level of the Three Gorges Reservoir has been planned to control downstream floods and minimize downstream sedimentation. Generally, the release of water from the TGD is scheduled to maintain a flood-restricted level of 145 m during the flood season (Figure 5B).

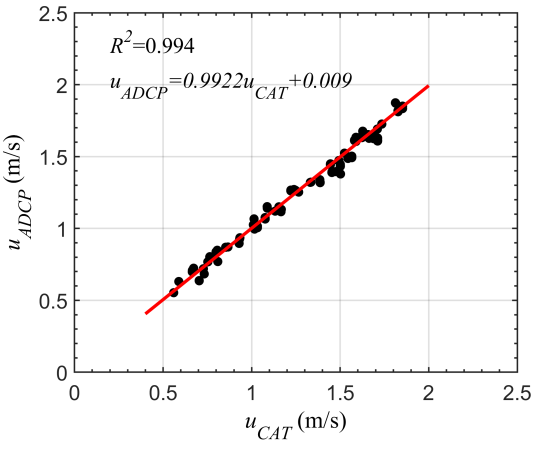


FIGURE 4
Index velocity relationship of u_{CAT} and u_{ADCP}.

Significantly, a detail example of Q_{ADCP}, Q_{CAT}, and Q_{RC} presents in Appendix, during the period of water release from the TGD (July–September) and during the period of higher precipitation, the discrepancy between Q_{RC} and Q_{ADCP} were obvious, whereas Q_{CAT} continued to match well with the results of Q_{ADCP} (Figure 5). Specifically, during the period of the TGD outflow, the uncertainty of Q_{RC} compared to Q_{ADCP} was within ± 20%, whereas the uncertainty was within ± 7% of Q_{CAT} compared to Q_{ADCP}. This revealed that the dam outflow or huge rainfall events induced dramatic fluctuations in the water level at the DT station, indicating that the uncertainties and errors obtained by RC were too large to be ignored, especially under unsteady flow conditions, and demonstrated the advantages of using CAT in large river channels.

Baseflow and stormflow

Baseflow and stormflow were estimated from Q_{RC} and Q_{CAT}, respectively. Although both revealed similar fluctuations, the stormflow from *the* Q_{RC} was smoother, and it was difficult to capture the short- and high-frequency variations induced by rainfall events. Baseflow and stormflow obtained from Q_{CAT} at the DT station during the observation period are shown in Figure 7. The baseflow varied from 11,059 to 62,231 m^3/s with a mean of 25,485 m^3/s. The stormflow, which mostly resulted from precipitation, varied from 0 to 34,822 m^3/s with a mean of

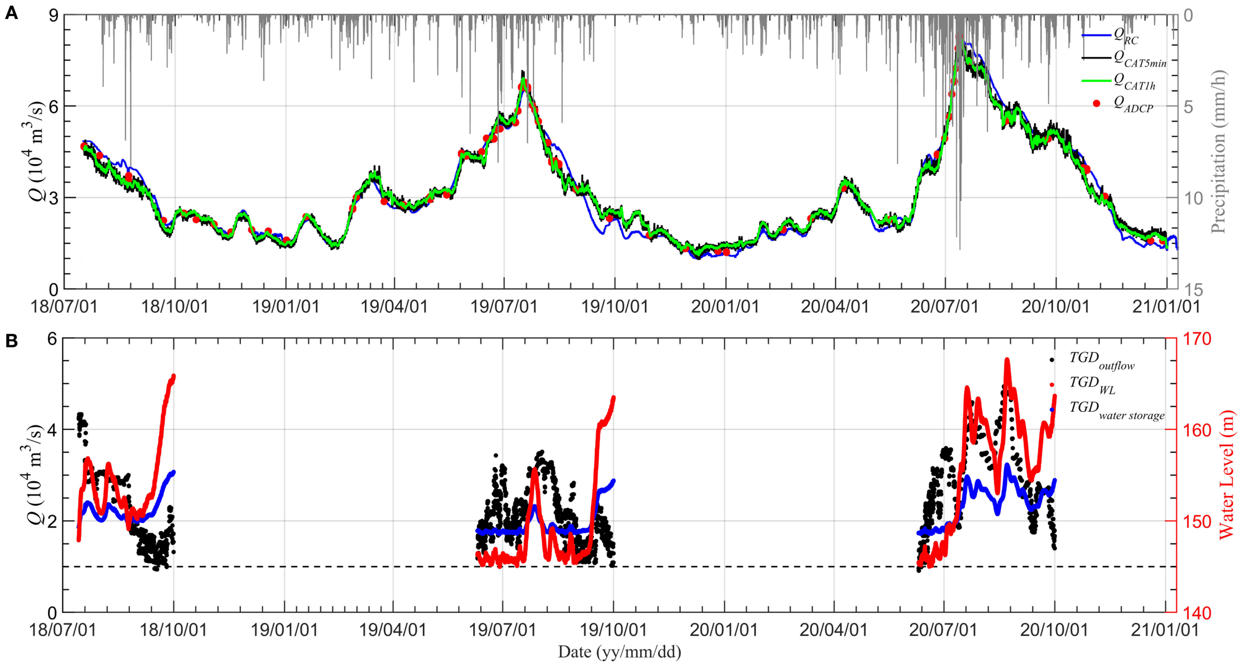


FIGURE 5
(A) Discharge time series during the observation period (July 2018 to January 2021) for different methods, RC (blue) and CAT. The black line represents the raw 5-min interval of CAT data and the green line represents the 1-h moving average. Red dots represent results obtained using the moving-boat ADCP. The gray bar shows hourly precipitation data near the Datong station obtained from ERA5 data; **(B)** TGD outflow, water storage, and water level time series during the observation period.

5,918.1 m^3/s. The mean proportions of baseflow and stormflow were 81.1 and 18.8%, respectively. The ratios of baseflow and stormflow in the flood and dry seasons were 80.9 and 81.7% and 19.1 and 18.3%, respectively. These values are similar to previous studies; for example, Dai et al. (2010) investigated the proportion of baseflow at several stations along the YR and revealed that its mean proportion at the DT station was ~82% and in the flood and dry seasons, it was ~81 and ~83%, respectively.

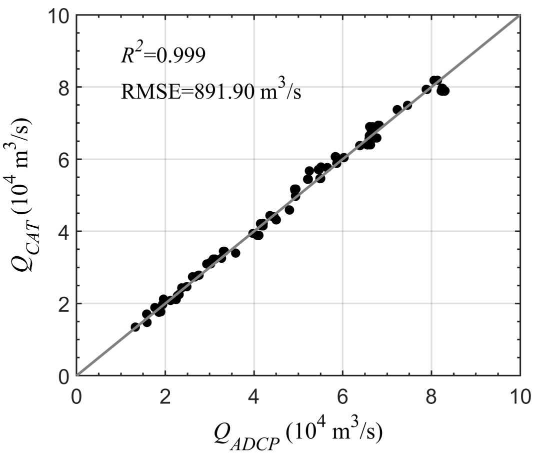


FIGURE 6
Comparison of discharge obtained by CAT and moving-boat ADCP.

Discussions

Effects of the TGD on runoff at the Datong station

Grouped frequency analysis was applied to the TGD outflow, and the class interval for discharge was set to 10,000 m^3/s in this study. The contribution of each month during the TGD outflow period is shown in Figure 8. At the DT station, during the TGD dam outflow period, the water discharge is in the range of 30,000 to 60,000 m^3/s, accounting for 70% of total intervals (Figure 8A). The occurrence probabilities of large flood events exhibited declining trends. The proportion of extreme discharge (> 60,000 m^3/s) reduced from 6.9 to 0.7% (Figure 8A). In addition, the grouped distribution of the TGD outflow and the contributions of the TGD outflow to the grouped discharge at the DT station (Q_{DT}) are also presented in Figures 8B, C. The TGD outflow occurred mainly in the range of 20,000 to 30,000

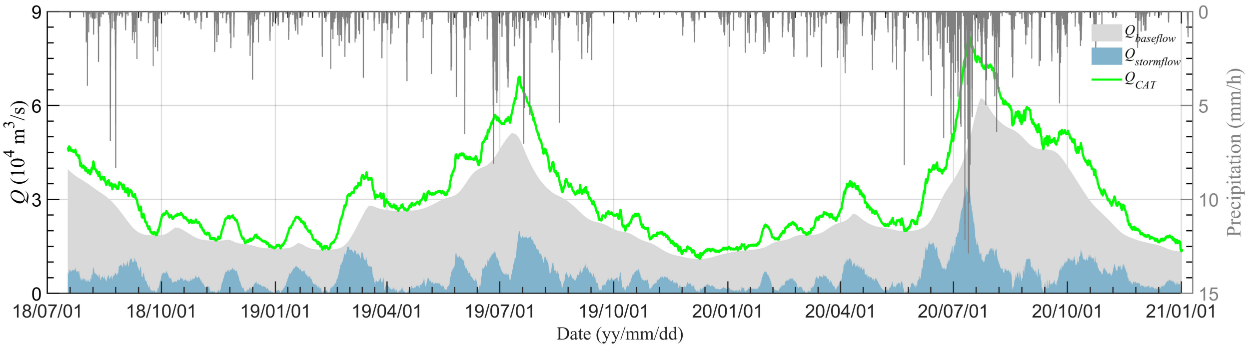


FIGURE 7
Time series of baseflow and stormflow obtained from Q_{CAT} during the observation period (July 2018 to January 2021).

m^3/s with a proportion of 43.5–77.6% over the 2.5-year period of observation (Figure 8B). For the class Q_{DT} > 30,000 m^3/s, Q_{DT} increased from 30,000 to 80,000 m^3/s and the proportion of TGD outflow gradually decreased from 54.2 to 43.5% in June, from 77.6 to 24.7% in July, and from 73.6 to 51.5% in August (Figures 8B, C). Taken together, the TGD organizes downstream runoff by boosting the proportion of normal streamflow and lowering the proportion of exceptional streamflow. Consequently, the management of the TGD forces the distribution of downstream runoff to become more centralized.

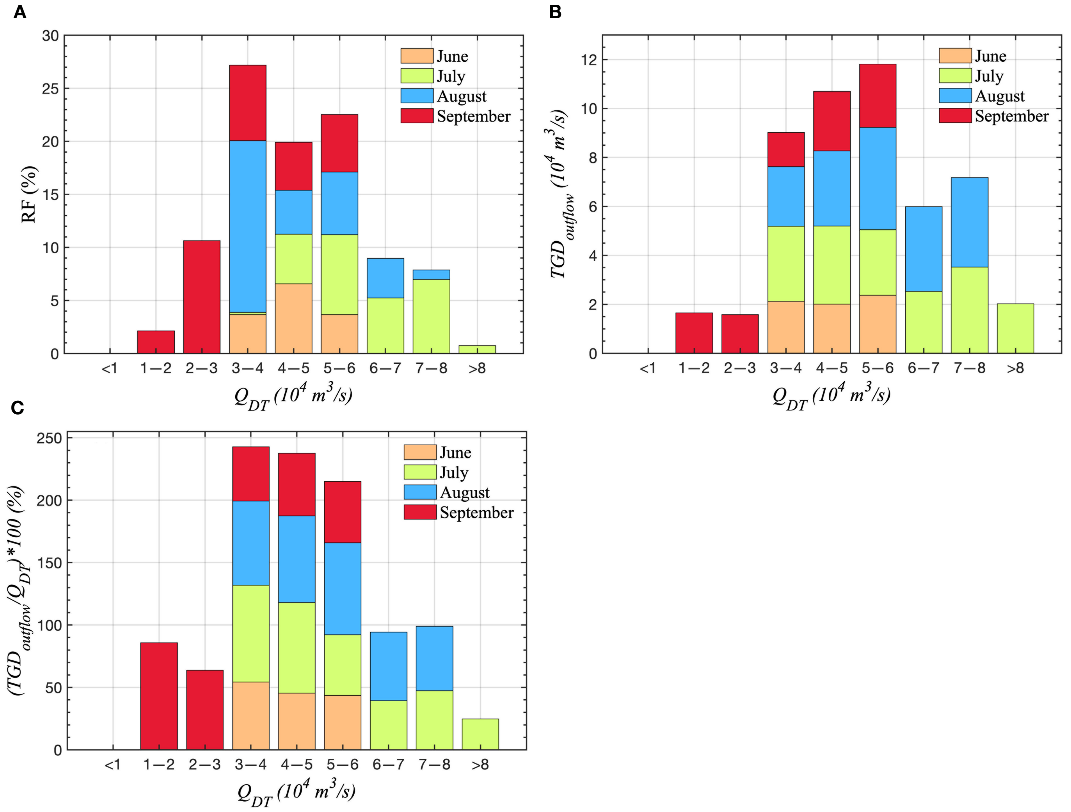


FIGURE 8
Grouped distributions of TGD outflow varied with the DT discharge (Q_{DT}) during the wet season (June–September). **(A)** Relative frequency (*RF*) of TGD outflow; **(B)** Grouped TGD outflow in each Q_{DT} class; **(C)** The proportions of TGD outflow in each Q_{DT} class.

The TGD has been in operation since 2003. Previous studies have stated that the TGD has significantly flattened river discharge hydrographs, such as by increasing the discharge of 500–2,000 m^3/s in the dry season and decreasing the annual flood peak flow by up to 20,000 m^3/s (Guo et al., 2015). Especially, the regulation and management of the TGD have considerably eased river flow processes in the middle–lower reaches of the YR (Cai et al., 2019). Additionally, Mei et al. (2015) reported that from low- to high-runoff situations, obvious declining trends have led to low- and normal-runoff situations from 2003. Instead, the high-runoff situation presented uniform states without noticeable gradient patterns. Consequently, the role of the TGD on downstream hydrological processes is substantially more significant in the dry season than those in the flood season (Chen et al., 2016).

Effects of rainfall on runoff at the Datong station

Seasonal areal precipitation characteristics during the observation period are shown in Figure 2. Along the YR, during June–July–August (JJA), two maximum precipitation areas occurred upstream of the TGD and near the DT, respectively; therefore, the strong flood processes at the DT station during summer can be partially clarified by areal precipitation, but additional factors might affect changes in discharge downstream of the YR (such as the TGD effect). As shown in Figure 7, compared with the baseflow, the stormflow, which mainly results from precipitation, accounted for 0 to 47.7% with a mean of 18.8%. Mei et al. (2015) argued that the hydrological processes of the YR were largely structured by the TGD and that rainfall events had little force on the discharge.

To enrich the cognition of the role of rainfall on the Q_{DT} of the YR, we followed Liang, 2014; Liang, 2015) to evaluate the cause–effect relationship between precipitation and discharge at the DT station (Q_{DT}). The information flow (*IF*) in the natural unit of information (nat) per unit time (nat/h) during the CAT observation period was estimated. Causality was computed as the time rate of *IF* from one time series to another. When only two time series, that is, X_1 (Q_{DT}) and X_2 (T_{rain}) are considered, a model to fulfill the *IF* evaluation was given by Liang (2014), and the rate of *IF* from T_{rain} to Q_{DT} can be written as

$$IF_{rain \rightarrow Q} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2} \tag{7}$$

where C_{ij} (i, j = 1 (i.e., Q), 2 (i.e., rain)) denotes the covariance between T_{rain} and Q_{DT}, $C_{i,dj}$ represents the covariance between X_i and $X_j^{'}$, and $X_j^{'}$ is the difference approximation of dX_j/dt using the Euler forward scheme.

$IF_{rain \rightarrow Q}$ represents the rate of *IF* from T_{rain} to Q_{DT}, which can be zero or nonzero. If $IF_{rain \rightarrow Q}$ is zero, T_{rain} does not cause Q_{DT}; if the value is nonzero, it is causative. When there is a causal relationship, two cases can be discerned as stated by the sign of the *IF*: the positive value suggests that T_{rain} works to cause Q_{DT} more uncertain, whereas the negative value denotes that T_{rain} would stabilize Q_{DT}. Additionally, $IF_{rain \rightarrow Q}$ can be normalized to assess the importance of the influence of T_{rain} on Q_{DT}relative to other processes (Liang, 2015).

One-way causality analysis was used to assess the sensitivity of discharge at the DT station to the regions and changes of precipitation in the YRB. Positive *IF* ranging from 0 to 0.05 nat/h (Figure 9A) occurred in the upper and lower areas of the YRB, with the highest values around the DT station. The computed relative *IF* rates were up to ~40% near the DT station (Figure 9B). These demonstrate that during summer the flood processes at the DT station are partially caused by areal precipitation in the downstream reaches of the YR. A positive *IF* in the upstream can be interpreted as runoff caused by precipitation gradually entering the baseflow and eventually affecting the streamflow at the DT station. On the contrary, the non-positive *IF* exhibited in the middle areas of the YRB revealed the role of the TGD, which means that the runoff induced by the rainfall events in the middle areas would be controlled by the TGD, and the TGD would stabilize the downstream discharge. Mei et al. (2015) investigated and compared the characteristics of monthly areal precipitation during the pre-TGD and post-TGD periods and found that precipitation showed a minor decreasing trend since 2003 and stated that this decreasing trend was not adequately strong to support the vast changes in runoff along the YR, proving that TGD regulation significantly disturbed the hydrology of the YR.

In summary, we investigated the contributions of the TGD (human activities) and rainfall (natural factor) on the discharge and revealed the relationships between the river flow, rainfall, and dam outflow (Figures 8, 9). The YR is the China's largest river and covers a vast river basin, natural factors (e.g., Typhoon, rainfall) are easy to induce flood events and enhance flood risk to human activities. Generally, variations of discharge are consistent with those of precipitation in natural hydrological conditions; nevertheless, the low and high flows are strongly altered by the TGD. Therefore, in this study, a quantitative understanding of the contributions of rainfall and the TGD are helpful for the development of effective flood control strategies.

Relationships between runoff and tide at the Datong station

The DT station is a widely recognized site for the tidal limit position; however, the geographical and temporal distributions of runoff in the YRB has changed significantly over the past decades owning to many large-scale water conservation projects along its course. Some studies suggested that the location of the

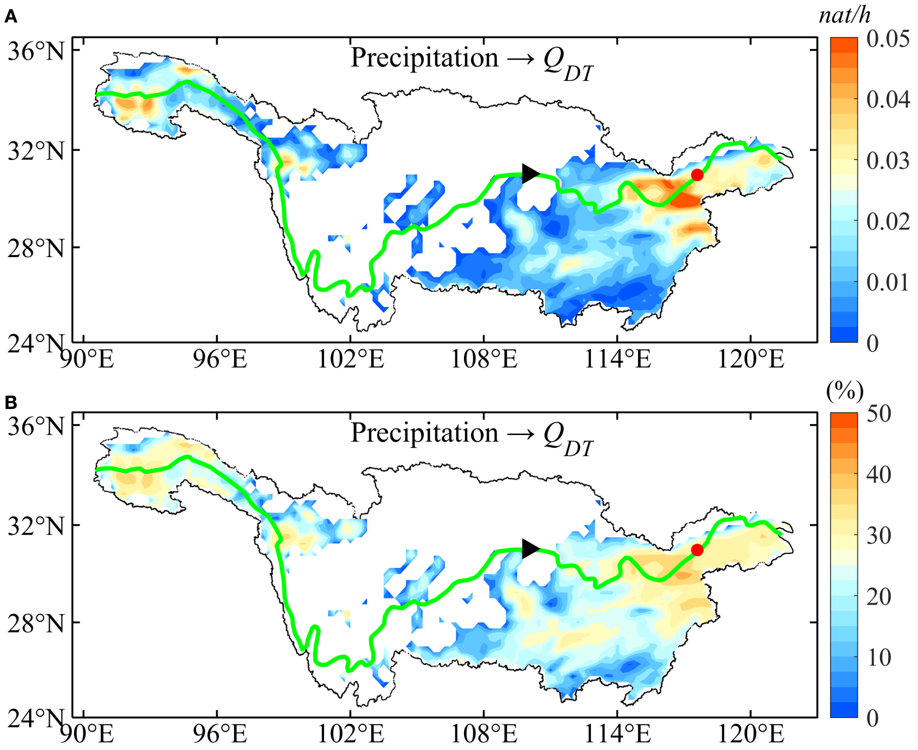


FIGURE 9
Distributions of results of the one-way causality analysis in the Yangtze River Basin. **(A)** Rate of information flow (*IF*) from precipitation to discharge at the DT station (Q_{DT}); **(B)** Relative *IF* from precipitation to discharge at the DT station (Q_{DT}). Since the precipitation data is of hourly interval, the hourly Q_{DT} was picked up to match the time of precipitation data. Only positive values are presented in the figure, while the non-positive values are masked by white.

recent tidal limit in the YR fluctuates from Jiujiang (~220 km upstream from the DT station) to DT. During the dry season, the upper limit of the tide is near Jiujiang, whereas during the flood season, the lowest bound of the tidal zone is near DT (Shi et al., 2018).

The continuous wavelet transform (CWT) method was used to investigate the characteristics of tidal species and the temporal evolution of tidal frequency spectra at the DT station (Sassi and Hoitink, 2013). CWT analysis can determine different tidal species, for example, the semi-diurnal (D_2) tidal species, but specific tidal components (e.g., M_2) cannot be resolved (Guo et al., 2015). More information on the CWT method and settings used to determine the specific tidal species are described in Guo et al. (2015). Here, the CWT method was applied to data on water depth at station S, velocity obtained by CAT (u_{CAT}), and discharge at the DT station (Q_{CAT}) to decompose the time series data into the D2 tidal species because the semi-diurnal tide is evident along the YR.

At the DT station, the tidal signal was very weak, but the seasonal signals, which were driven by fluctuating river discharges, remained noteworthy (Figure 10). The temporal fluctuations in the D_2 amplitudes are shown in Figure 10A. Evidently, D_2 showed a clear seasonal fluctuation of high amplitudes in winter and low amplitudes in summer. A high river discharge improves tidal amplitudes by boosting the nonlinear effect and promoting the transfer of astronomical tidal energy to shallow-water tides, which is the primary cause of lower amplitudes during the wet season (Godin, 1985). In addition, the attenuation of tidal amplitudes is related to frequency; generally, lower-frequency tidal components damp more slowly than higher-frequency tidal components (Godin, 1985). The amplitudes of water level in the D_2 bands present an evident significant seasonal signal, with values fluctuating from 0 to 0.08 m, respectively. In the flood and dry seasons, D_2 contributed 0.30% and 1.16%, respectively, to the water level at the DT station (Figure 10A). The tidal velocity in the D_2 band also presented a significant seasonal signal with a maximum value of 0.086 m/s, and the mean proportions of the velocity during the flood and dry seasons were 2.90% and 6.27%, respectively (Figure 10B). Both were directly driven by fluctuating river discharges.

Significant fortnightly tidal signals were detected at the DT station simultaneously (Figure 10). A similar phenomenon was exhibited by Guo et al. (2015), who stated that this clear fortnightly signal was mostly generated inside the estuary owing to tidal interactions rather than river flow or oceanic tides. As oceanic tides and river flow have limited energy at the fortnightly frequency, tidal interactions are central sources of tidal semi-monthly or monthly changes. For example, the interaction between M_2 and S_2 causes fortnightly (spring and neap) variations. In such a long river tidal system, this result strongly suggests the spatial evolution of tidal dynamics caused by mixed tidal and fluvial processes.

The CWT analysis of the discharge showed fluctuations in the D_2 bands (Figures 10C, D) that were different from the amplitude and velocity (Figures 10A, B), which mostly resulted from the tidal damping processes. The high (low) discharges in summer (winter) lead to a stronger (weaker) tidal damping process and generate a low (high) tidal discharge in summer (winter). The discharge in D_2 bands varied from 19.4 to 502.9 m^3/s, respectively, accounting for 0.11–1.11% of the total discharge. The velocity resolution (u_r) of CAT is as follows:

$$u_r = \frac{C^2}{2L \cdot 2f} \tag{8}$$

where f means the frequency and L and C are the distances between two stations and sound speed, respectively. The velocity resolution considered in this study is 0.02 m/s, the precision of the water level is 0.5 cm, assuming the distance (~3,015 m) between two stations caused by the tidal wave was negligible, the theoretical minimum discharge estimated using the velocity-area method is ~0.3 m^3/s, and thus, the CWT results are within the observation resolution and are reliable.

Results of CWT demonstrate that the effects of tidal waves on river transport were insignificant at the DT station, probably because the tidal wave propagated to the DT station; however, owing to the river discharge–controlled systems, the direction and velocity of flow did not change significantly, and the changes in water level and cross-sectional area caused by the tidal wave were negligible. In fact, tides assert a barrier impact on the discharge of the YR, which can be attributed mainly to the Stokes flux. Nevertheless, in the middle–lower reaches of the YR (i.e., at the DT station), river flow significantly controls river and tidal dynamics (Guo et al., 2015).

Taken together, the discharge regulated by the TGD takes a significant influence on the tidal propagation process along the YR; that is, a decrease in discharge can cause the tidal wave to move upstream and lead to an increase in tidal amplitude, even if the TGD is situated ~1,800 km upstream from the mouth of the YR. Hence, more attention should be paid to this man-made effect on tidal processes and the river–tide interactions since it would be easily overlooked as the TGD is sited far away from the river mouth. Additionally, understanding tidal wave effects in the middle–lower reaches of the YR is of significant social value for tidal prediction and water management in channels.

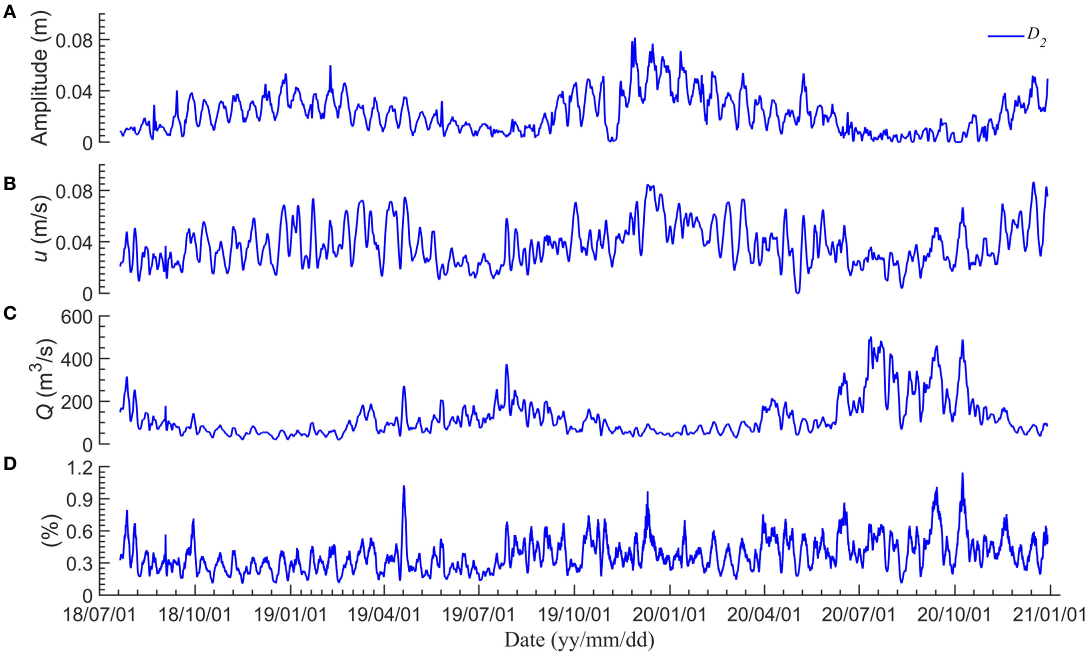


FIGURE 10
Time series of the results of the continuous wavelet transform (CWT) results at the Datong station. **(A–C)** Amplitude, velocity, and discharge in the semi-diurnal (D_2) band; **(D)** Proportion of the semi-diurnal discharge in the total discharge.

Review of the present study and scope of the future studies

One objective of this study was to highlight the performance of CAT in continuously monitoring large river discharges in YR. The main drawback of traditional instruments (e.g., ADCP) is the insufficient number of velocity sample points in the cross-section for estimating cross-sectional average velocity. Unlike the traditional hydroacoustic systems, the CAT enables continuous and real-time observations of the depth- and range-averaged velocity with a transect, without disturbing shipping traffic in YR. The method of CAT in real-time could help in the increased ability to monitor the flood events and provide reference data to flood risk management.

Although some of our findings are preliminary, our results are encouraging. The CAT is capable of capturing discharge in large river basin; therefore, future studies should consider longer periods to determine the variations of floods in the YR and the role of the TGD. In addition, a preliminary result of discharge induced by tidal wave was revealed here, however, the YR has a vast area affected by tides, monitoring the tidal discharge in different sections of the YR is useful to study the river-tidal interactions in the YR.

Conclusions

Since the construction of the TGD, several studies on the hydrological processes of the YRB. We investigated and analyzed streamflow behavior in the YR considering the effects of a large dam (TGD) and other natural activities using the CAT system at the DT hydrometric station. We installed two CAT systems at the DT station to continuously monitor YR discharge from July 2018 to January 2021. The RC method for the YR leads to large uncertainties in the wet season, especially during the TGD outflow period. In contrast, the CAT method can directly measure the sectional-average velocity, allowing for the precise acquisition of instantaneous flood discharge; for example, two huge flood events, with peak discharges of 69,164 and 82,206 m^3/s were demonstrated by CAT. The comparison between the discharge estimated by ADCP and CAT showed a good correlation with an RMSE of 891.9 m^3/s.

Changes in discharge of the YR at the DT station induced by the TGD and rainfall are discussed in detail. Results demonstrated that the streamflow of the YR is primarily influenced by the TGD. Moreover, results of causality analysis proved that the discharge at the DT station was partially influenced by rainfall in the middle–lower areas of the YRB, with a maximum contribution of ~40% changes in runoff. As a result of the regulation by the TGD, the distribution of the downstream discharge at DT became more centralized in the wet season. Additionally, discharge controlled by the TGD can affect the tidal propagation process along the YR, and the decrease in discharge during the dry season can cause an increase in downstream tidal amplitude. Results of CWT revealed that although the tidal wave affected the water level at the DT station, the tidal wave-induced discharge was negligible, with a maximum of 1.11% change.

This study offers an acoustic method (CAT) for accurate, continuous, and real-time monitoring of the discharge of large rivers, even during massive flood events. Hence, the application of CAT over a longer time scale (e.g., decades) and monitoring tidal discharge in tidal rivers is necessary to determine various patterns of flood events and hydrological processes. Results of this study suggest a strong need for strategies to balance the role of the TGD in flood control and management of the YR.

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 and X-HZ contributed to conception and design of the study. ZNZ, CZ and ZSZ organized the observational data. CX performed the statistical analysis. ZNZ, CZ, HZ, ZSZ, JZ and LW obtain the field data. CX, X-HZ, ZNZ and CZ wrote the first draft of the article. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (grants 41920104006, 41906024, 52071293), the Scientific Research Fund of the Second Institute of Oceanography, MNR (JZ2001), the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (SOEDZZ2106 and SOEDZZ2207), and the Innovation Group Project of the Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai (No. 311020004).

Acknowledgments

The authors are grateful to the National Rain and Water Discharge (http://xxfb.mwr.cn/index.html) for sharing their data. We acknowledge the students in the laboratory for their assistance in data collection. We are thankful to Editage (www.editage.com) for providing assistance with English language editing.

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.2022.1066693/full#supplementary-material

References

Baldassarre, G., and Montanari, A. (2009). Uncertainty in river discharge observations: a quantitative analysis. *Hydrol. Earth Syst. Sci.* 13, 913–921. doi: 10.5194/hess-13-913-2009

Cai, H., Zhang, X., Zhang, M., Guo, L., Liu, F., and Yang, Q. (2019). Impacts of three gorges dam's operation on spatial-temporal patterns of tide-river dynamics in the Yangtze river estuary, China. *Ocean Sci.* 15, 583–599. doi: 10.5194/os-15-583-2019

Chen, J., Finlayson, B. L., Wei, T., Sun, Q., Webber, M., Li, M., et al. (2016). Changes in monthly flows in the Yangtze river, China - with special reference to the three gorges dam. *J. Hydrol.* 536, 293–301. doi: 10.1016/j.jhydrol.2016.03.008

Dai, Z. (2021). *Changjiang riverine and estuarine hydro-morphodynamic processes* (Gateway East: Singapore Springer). doi: 10.1007/978-981-16-3771-1

Dai, Z. J., Chu, A., Du, J. Z., Stive, M., and Hong, Y. (2010). Assessment of extreme drought and human interference on baseflow of the Yangtze river. *Hydrol. Process.* 24, 749–757. doi: 10.1002/hyp.7505

Gao, Y., Chen, L., Zhang, W., Li, X., and Xu, Q. (2021). Spatiotemporal variations in characteristic discharge in the Yangtze river downstream of the three gorges dam. *Sci. Total Environ.* 785, 147343. doi: 10.1016/j.scitotenv.2021.147343

Godin, G. (1985). Modification of river tides by the discharge. *J. Waterw. Port, Coastal, Ocean Eng.* 111, 257–274. doi: 10.1061/(ASCE)0733-950X(1985)111:2(257)

Gordon, R. L. (1989). Acoustic measurement of river discharge. *J. Hydraul. Eng.* 115, 925–936. doi: 10.1061/(ASCE)0733-9429(1989)115:7(925

Gore, J. A., and Banning, J. (2017). "Discharge measurements and streamflow analysis," in *Methods in stream ecology*, Volume 1. (New York, United States: Academic Press), 49–70.

Guo, H., Hu, Q., Zhang, Q., and Feng, S. (2012). Effects of the three gorges dam on Yangtze river flow and river interaction with poyang lake, China: 2003-2008. *J. Hydrol.* 416–417, 19–27. doi: 10.1016/j.jhydrol.2011.11.027

Guo, L., van der Wegen, M., Jay, D. A., Matte, P., Wang, Z. B., Roelvink, D., et al. (2015). River-tide dynamics: Exploration of nonstationary and nonlinear tidal behavior in the Yangtze river estuary. *J. Geophys. Res. Ocean.* 120, 3499–3521. doi: 10.1002/2014JC010491

Hoitink, A. J. F., Buschman, F. A., and Vermeulen, B. (2009). Continuous measurements of discharge from a horizontal acoustic Doppler current profiler in a tidal river. *Water Resour. Res.* 45, 1–13. doi: 10.1029/2009WR007791

Jia, H., Chen, F., Pan, D., Du, E., Wang, L., Wang, N., et al. (2022). Flood risk management in the Yangtze river basin —comparison of 1998 and 2020 events. *Int. J. Disaster Risk Reduct.* 68, 102724. doi: 10.1016/j.ijdrr.2021.102724

Kaneko, A., Zhu, X.-H., and Lin, J. (2020). *Coastal acoustic tomography* (Amsterdam, Netherlands: Elsevier). doi: 10.1016/B978-0-12-818507-0.00003-2

Kawanisi, K., Al Sawaf, M. B., and Danial, M. M. (2018). Automated real-time streamflow acquisition in a mountainous river using acoustic tomography. *J. Hydrol. Eng.* 23, 04017059. doi: 10.1061/(asce)he.1943-5584.0001604

Kawanisi, K., Bahrainimotlagh, M., Al Sawaf, M. B., and Razaz, M. (2016). High-frequency streamflow acquisition and bed level/flow angle estimates in a mountainous river using shallow-water acoustic tomography. *Hydrol. Process.* 30, 2247–2254. doi: 10.1002/hyp.10796

Kawanisi, K., Razaz, M., Yano, J., and Ishikawa, K. (2013). Continuous monitoring of a dam flush in a shallow river using two crossing ultrasonic transmission lines. *Meas. Sci. Technol.* 24, 1–10. doi: 10.1088/0957-0233/24/5/055303

Kawanisi, K., Zhu, X. H., Fan, X., and Nistor, I. (2017). Monitoring tidal bores using acoustic tomography system. *J. Coast. Res.* 33, 96–104. doi: 10.2112/JCOASTRES-D-15-00172.1

Lai, X., Jiang, J., Yang, G., and Lu, X. X. (2014). Should the three gorges dam be blamed for the extremely low water levels in the middle-lower Yangtze river? *Hydrol. Process.* 28, 150–160. doi: 10.1002/hyp.10077

Liang, X. S. (2014). Unraveling the cause-effect relation between time series. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 90, 1–11. doi: 10.1103/PhysRevE.90.052150

Liang, X. S. (2015). Normalizing the causality between time series. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 92, 1–6. doi: 10.1103/PhysRevE.92.022126

Luo, P., He, B., Takara, K., Xiong, Y. E., Nover, D., Duan, W., et al. (2015). Historical assessment of Chinese and Japanese flood management policies and implications for managing future floods. *Environ. Sci. Policy* 48, 265–277. doi: 10.1016/j.envsci.2014.12.015

Lyu, Y., Zheng, S., Tan, G., and Shu, C. (2018). Effects of three gorges dam operation on spatial distribution and evolution of channel thalweg in the yichang-chenglingji reach of the middle Yangtze river, China. *J. Hydrol.* 565, 429–442. doi: 10.1016/j.jhydrol.2018.08.042

Mei, X., Dai, Z., Darby, S. E., Zhang, M., Cai, H., Wang, J., et al. (2021). Landward shifts of the maximum accretion zone in the tidal reach of the changjiang estuary following construction of the three gorges dam. *J. Hydrol.* 592, 125789. doi: 10.1016/j.jhydrol.2020.125789

Mei, X., Dai, Z., Gelder, P. H. A. J. M., and Gao, J. (2015). Linking three gorges dam and downstream hydrological regimes along the Yangtze river, China. *Earth Sp. Sci.* 2, 94–106. doi: 10.1002/2014EA000052

Mei, X., Zhang, M., Dai, Z., Wei, W., and Li, W. (2019). Large Addition of freshwater to the tidal reaches of the Yangtze (Changjiang) river. *Estuaries Coasts* 42, 629–640. doi: 10.1007/s12237-019-00518-0

Ruhl, C. A., and Derose, J. B. (2004). Investigation of hydroacoustic flow-monitoring alternatives at the Sacramento river at Freeport, California: results of the 2002-2004 pilot study. *US Geol. Surv.* 5172, 1–18. doi: 10.3133/sir20045172

Sassi, M. G., and Hoitink, A. J. F. (2013). River flow controls on tides and tide-mean water level profiles in a tidal freshwater river. *J. Geophys. Res. Ocean.* 118, 4139–4151. doi: 10.1002/jgrc.20297

Shi, S., Cheng, H., Xuan, X., Hu, F., Yuan, X., Jiang, Y., et al. (2018). Fluctuations in the tidal limit of the Yangtze river estuary in the last decade. *Sci. China Earth Sci.* 61, 1136–1147. doi: 10.1007/s11430-017-9200-4

Sun, Z., Huang, Q., Opp, C., Hennig, T., and Marold, U. (2012). Impacts and implications of major changes caused by the three gorges dam in the middle reaches of the Yangtze river, China. *Water Resour. Manage.* 26, 3367–3378. doi: 10.1007/s11269-012-0076-3

Tang, W., and Carey, S. K. (2017). HydRun: A MATLAB toolbox for rainfall–runoff analysis. *Hydrol. Process.* 31, 2670–2682. doi: 10.1002/hyp.11185

Wang, S., Cao, Z., Luo, P., and Zhu, W. (2022). Spatiotemporal variations and climatological trends in precipitation indices in shaanxi province, China. *Atmosphere (Basel).* 13, 1–22. doi: 10.3390/atmos13050744

Wang, H., Sun, F., and Liu, W. (2020). Characteristics of streamflow in the main stream of changjiang river and the impact of the three gorges dam. *Catena* 189, 104498. doi: 10.1016/j.catena.2020.104498

Xia, J., and Chen, J. (2021). A new era of flood control strategies from the perspective of managing the 2020 Yangtze river flood. *Sci. China Earth Sci.* 64, 1–9. doi: 10.1007/s11430-020-9699-8

Yu, X., Zhang, W., and Hoitink, A. J. F. (2020). Impact of river discharge seasonality change on tidal duration asymmetry in the Yangtze river estuary. *Sci. Rep.* 10, 1–17. doi: 10.1038/s41598-020-62432-x

Zhang, Y., Fang, G., Tang, Z., Wen, X., Zhang, H., Ding, Z., et al. (2021). Changes in flood regime of the upper Yangtze river. *Front. Earth Sci.* 9. doi: 10.3389/feart.2021.650882

Zhu, X. H., Zhang, C., Wu, Q., Kaneko, A., Fan, X., and Li, B. (2012). Measuring discharge in a river with tidal bores by use of the coastal acoustic tomography system. *Estuar. Coast. Shelf Sci.* 104–105, 54–65. doi: 10.1016/j.ecss.2012.03.022

Zhu, Z.-N., Zhu, X.-H., Zhang, C., Chen, M., Zheng, H., Zhang, Z., et al. (2021). Monitoring of Yangtze river discharge at datong hydrometric station using acoustic tomography technology. *Front. Earth Sci.* 9. doi: 10.3389/feart.2021.723123

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fmars.2022.1086799

Observation of a mesoscale warm eddy impacts acoustic propagation in the slope of the South China Sea

Wen Chen, Yongchui Zhang*, Yuyao Liu, Yanqun Wu, Yun Zhang and Kaijun Ren*

College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China

Check for updates

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Changming Dong,
Nanjing University of Information Science and Technology, China
Xiaohui Liu,
Ministry of Natural Resources, China

*CORRESPONDENCE
Yongchui Zhang
zyc@nudt.edu.cn
Kaijun Ren
renkaijun@nudt.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 01 November 2022
ACCEPTED 01 December 2022
PUBLISHED 15 December 2022

CITATION
Chen W, Zhang Y, Liu Y, Wu Y, Zhang Y and Ren K (2022) Observation of a mesoscale warm eddy impacts acoustic propagation in the slope of the South China Sea. *Front. Mar. Sci.* 9:1086799. doi: 10.3389/fmars.2022.1086799

Acoustic rays are modified while propagating through oceanic eddies. However, due to the lack of field synchronous observation, the impact of mesoscale eddy on the acoustic propagation is less clarified. To address the issue, an eddy-acoustic synchronous observation (EASO) field experiment for a mesoscale warm eddy was carried out in the slope of the South China Sea (SCS) in October, 2021. During the field experiment, a total of 105 conductivity-temperature-depth (CTD) stations, as well as a zonal acoustic survey line through the center of the warm eddy, were obtained. The vertical structures of temperature and salinity indicate that the warm eddy is surface-intensified with temperature and salinity cores confined within depths from 70 m to 200 m and 10 m to 70 m, respectively. The acoustic observation shows two obvious convergency zones (CZs) at about 39 km and 92 km in the eastern half acoustic line, and one convergency zones (CZ) at about 25 km in the western half acoustic line. By comparing with the none eddy circumstance, the respective impacts of the topography and warm eddy are quantitatively analyzed with a ray-tracing model. The results indicate that the topography shortens the horizontal span of the CZ by 11.4 km, while the warm eddy lengthens it by 1.7 km. Additionally, the warm eddy shallows the depth and broadens the width of the CZ by 32 m and 1.4 km, respectively. The anisotropy of 3D sound fields jointly influenced by the warm eddy and the local topography show that the distance differences of the first CZs in different horizontal directions can be as long as 31 km.

KEYWORDS

mesoscale eddy, acoustic propagation, field experiment, topography, anisotropy of sound fields

1 Introduction

In the global oceans, mesoscale eddies are ubiquitous (Chelton et al., 2011). They are usually accompanied by temperature and salinity anomalies, and thus distort the sound speed profile (SSP) (Jian et al., 2009; Chen et al., 2022). Due to the significant abnormal sound fields, the impacts of mesoscale eddies on the acoustic properties gained considerable research attention.

The earliest studies about the eddy-induced anomalies in sound fields were conducted by Vastano and Owens (1973) and Weinberg and Zabalgogeazcoa (1977), based on the hydrological observed data measured from a cold Gulf Stream ring detected in 1967. In the relatively uniform acoustic environment of the Sargasso Sea, Vastano and Owens (1973) first observed the significant acoustic field perturbation and dispersion phenomenon of ray path caused by the cold ring through the ray-tracing model. Gemmill conducted research on the effects of the eddy on the convergence mode of sound propagation in the following year, and discovered that the cold eddy refracts sound rays into the deep sound channel and destroys the cyclic distribution of convergence zones. Following that, the effects of eddy on ray travel time, which reflects the sound arrival structure, were investigated using a range-dependent model presented by Weinberg and Zabalgogeazcoa (1977). However, due to the scarcity of *in-situ* eddy observations, the corresponding sensitivity investigations of sound propagation on eddy property variation could not be conducted. To solve this problem, Henrick et al. (1977) developed a parametric eddy model that was qualitatively validated by specific Gulf Stream ring observation. The model was used to investigate the effects of geometric size, peak rotation speed, and eddy intensity on sound propagation. Due to the favorable performance of the model in describing eddy sound speed structure, it was employed in the research of Baer (1981), who revealed the eddy-caused significant changes in vertical arrival structure. However, due to the limited subsurface observations of eddies and the immature acoustic propagation models, researches of the effects of eddy on sound fields during this period are at the stage of theoretical predictions.

Benefiting from the advancement in observation schemes and facilities in recent years, three-dimensional (3D) structures of the mesoscale eddies and the underwater acoustic propagation were studied. For instance, Nan et al. (2017) detected an extra-large subsurface anticyclonic eddy with a horizontal scale of 470 km, which showed a lens-shaped vertical structure with shoaling of the seasonal and deepening of the main thermoclines. Zhang et al. (2019) conducted a high-resolution field observation of a cyclonic eddy in the Kuroshio Extension and obtained the anatomy of a cyclonic eddy. The observed eddy showed vertical thermal monopole and haline dipole structure, respectively. A series of acoustic experiments, such as SLICE89, ATOC, PhiSea09, PhiSea10, and OBSAPS have been carried out since 1989 to study the impacts of environmental variability on acoustic propagation (Worcester and Spindel, 2005; Worcester et al., 2013; Colosi et al., 2019), especially the PhiSea10 (Ramp et al., 2017), wherein a strong fluctuation in the travel time caused by the intense eddy activity was observed. Those observed eddy-induced abnormal sound fields emphasized its significance in underwater communication, positioning, and so on (Zhang et al., 2020; Zhang et al., 2021; Wu et al., 2022), and thus further motivated the development of acoustic experiments specialized for eddies.

In recent years, the effects of eddies on the sound field were widely studied based on field experiments and composited eddies research. For example, Chen et al. (2019) studied the effects of the eddy on the surface duct energy leakage phenomenon, which were verified with acoustic data measured in the South China Sea (SCS). Liu et al. (2021a) analyzed a cyclonic eddy that was found in the Pacific Northwest to investigate the effect of eddy on the coupling coefficient of various orders of normal modes. Gao et al. (2022) investigated the effects of the eddy on horizontal and vertical spatial coherence in deep water with the Gaussian eddy model presented by Calado et al. (2006). However, due to the movements of eddies, their corresponding acoustics observation data need to be measured through specialized design experiments. Thus, based on the composite research, which combines Argo floats and satellite altimetry, the 3D structures of mesoscale eddies and their impacts on the acoustic characteristics are derived. Chen et al. (2022) established a region-dependent parametric model for eddy-induced sound speed anomaly structure based on abundant Argo profiles. The parametric model can fast reconstruct the underwater sound speed field only using the satellite altimetry data.

However, due to the movement of mesoscale eddies and the complicated characteristics of sound field, the synchronous survey between the eddy and acoustics are less performed. In this study, a field experiment for a mesoscale warm eddy was conducted in the slope of the SCS in October, 2021. Through a series of fine design, both the oceanography and acoustics were simultaneously observed, which can greatly improve the understanding of the propagation characteristics of sound field in a specific marine environment. To the best of our knowledge, no experiments dedicated to the 3D eddy investigation and synchronized acoustic propagation were conducted in the SCS thus far, especially in the rough topography, which can provide more comprehensive insights in revealing the effects of the eddy on sound fields.

The rest of the paper is organized as follows. The datasets involved in this work and methodology are introduced in section 2. In section 3, the temperature and salinity structures of the observed warm eddy are presented. Simultaneously, the synchronized acoustic observations are illustrated and compared with the simulated results in the none eddy circumstance. In section 4, the differences between the results with and without eddy circumstances are interpreted using the ray-tracing model (Bellhop) and the results about the anisotropy of sound fields are disscused. The conclusions are presented in section 5.

2 Data and methods

2.1 Field experiment

An eddy-acoustic synchronous observation (EASO) field experiment was conducted in the SCS slope over the area 111˚–116˚E, 15˚–20˚N, (Figure 1) from October 1 to 30, 2021. The experiment consists of the synchronized hydrological survey and acoustic propagation measurement. The hydrological survey section contains a total of 105 conductivity-temperature-depth (CTD) measurement stations, which measured temperature (T) and salinity (S) profiles at different locations of the warm eddy over one month. The four groups of relatively complete CTD observation lines (namely lines A, B, C, and D, respectively) are shown in Figure 1 (yellow dot) (the locations of other CTD stations are not shown). Lines A, B, C and D appropriately represent complete sections in the northern, eastern, western and southern sides of the warm eddy respectively, and the observation times are October 3, 20, 21 and 29, 2021, respectively.

2.2 Other data

2.2.1 EN.4.2.2

To obtain the eddy-induced T and S perturbations structures (hereafter referred to as eddy structure), the associated climatological profiles need to be removed. The EN.4.2.2 dataset (available at https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-2.html) provides monthly mean gridded T/S profiles with a spatial resolution of 1°, which was distributed by the Met Office Hadley Center. The monthly mean gridded T/S profiles are derived from various *in situ* observation data like Argo floats and drifting buoys using objective analysis (Good et al., 2013).

In the field experiment, most of the observation depths of CTD are below 1500 m which is still far from the sea bottom. To calculate the sound field, the T/S profiles need to be extended to the whole depth. In this study, the EN.4.2.2 dataset is used to fill in the missing T/S data. According to the time of the EASO field experiment, the monthly T/S profiles in October 2021 are employed.

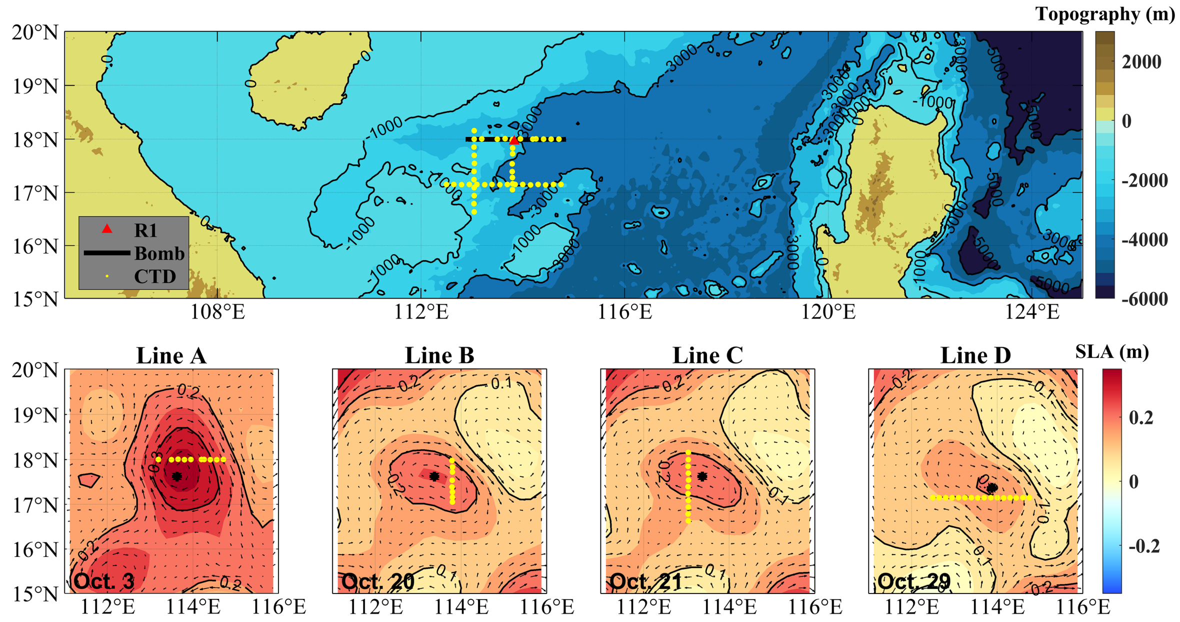


FIGURE 1
Map showing the environment configuration and survey stations in the EASO field experiment. The topographies (unit: m) are provided by the ETOPO2 dataset (Doi: 10.7289/V5J1012Q). The black line indicates the trajectory of explosive sound sources from the west to east. The red triangle and yellow dots indicate the vertical linear array (VLA) (marked as R1) and conductivity-temperature-depth (CTD) observation stations. The positions of the CTD observation stations are shown in the lower panel, where the shaded areas indicate the daily surface level anomalies (SLAs), and vector arrows indicate daily geostrophic currents anomalies (GCAs). According to the observation dates of CTD stations, they are named lines A, B, C and D, respectively.

2.2.2 Sea surface anomaly

Daily sea surface level anomaly (SLA) and surface geographic current (GCA) are the critical parameters for identifying and tracking the mesoscale eddies. In the northern (southern) hemisphere, warm eddies are characterized as positive SLAs and counterclockwise (clockwise) GCAs. In the EASO field experiment, they are used to determine the CTD measurement stations located in the observed warm eddy. Additionally, the daily SLA and GCA can also reflect the variation of the position and intensity of the observed warm eddy with time.

Daily SLA and GCA with a spatial resolution of 0.25° is provided by the Copernicus Marine Environment Monitoring Service (CMEMS) with the product identifier: SEALEVEL_GLO_PHY_L4_NRT_OBSERVATIONS_008_046 (available at https://resources.marine.copernicus.eu/product-detail/SEALEVEL_GLO_PHY_L4_NRT_OBSERVATIONS_008_046/INFORMATION). The SLA was calculated by removing a twenty-year (from 1992 to 2022) mean, and gridded data was estimated by optimal interpolation.

2.2.3 ETOPO2 Dataset

As an essential bottom boundary condition for the sound propagation models, the topography profoundly affects the distribution pattern of sound energy. Especially in 3D sound propagation models, the topography is one of the most crucial factors leading to the horizontal refraction of sound rays (also called the 3D effect of sound field) compared to inhomogeneous water media (Chiu et al., 2011; Dossot et al., 2019; Liu et al., 2021b).

The ETOPO dataset merges various observation data from different measurement equipment, such as satellites, shipboard echo-sounding measurements, and so on. It provides global topography and bathymetry with different spatial resolutions, such as 0.5', 1' and 2'. In this work, the 2' topography data (ETOPO2v2c, available at https://ngdc.noaa.gov/mgg/global/relief/ETOPO2/ETOPO2v2-2006/) is used to calculate the 3D sound fields in section 4.3 (different resolutions do not change the main conclusions).

2.3 Methodology

2.3.1 Ray-tracing model

To reveal the relative roles of the mesoscale eddy and topography, an acoustic propagation model is employed. Numerous sound propagation models, such as the ray-tracing model (Porter, 2019), normal mode model (Westwood et al., 1996), parabolic equation model (Collins and Werby, 1989) as well as some hybrid models are developed. Among them, the ray-tracing model is a high frequency approximation of the wave equation, which is suitable for predicting deep-sea high-frequency sound fields where the sound speeds are range-dependent. Furthermore, the ray-tracing model has a clear physical meaning, which interprets the sound field as the superposition of a series of sound rays with different propagation paths. These rays launch from the sound source to receiver points and obey the generalized Snell's law (Porter, 2019). Considering the above factors, the ray-tracing model is employed in this study. Readers interested in detailed derivation of the ray-tracing model can refer Jensen et al. (2011).

The sound pressure calculation formula in Cartesian coordinate system is directly written as

$$p(x) = e^{i\omega\tau(x)}\sum_{j=1}^{\infty}\frac{A_j}{(i\omega)^j} \qquad (1)$$

where, $\tau(s) = \tau(0) + \int_0^s cs'ds'$ represents the travel time of sound ray, s' denotes the ray trajectory, and its arclength is denoted by s. τ is derived from "Eikonal Equation" in ray coordinates. $\boldsymbol{x} = (x, y, z)$ denotes the position of the receiving point, j is the number of the sound ray, and A_j is the amplitude jth of the sound ray. c denotes the sound speed of position $\boldsymbol{x}$, and ω denotes the angle frequency.

Bellhop and Bellhop3D, developed by Porter, 2011; Porter, 2016) are highly efficient ray-tracing programs for predicting two-dimensional (2D) and 3D acoustic pressure fields. They are used to simulating the sound field presented in section 3 and 4, and to better depict the acoustic field in shadow and caustic regions, Gaussian beam are selected in this work.

When simulating the sound fields using Bellhop and Bellhop3D, the whole-depth sound speed profiles (SSPs) of the warm eddy circumstance are necessary. Using the T, S, and depth data measured by CTD, SSPs from 0 m to 1500 m can be obtained through Mackenzie formula (Mackenzie, 1981), which is given by

$$\begin{aligned} c &= 1448.96 + 4.591T - 5.304 \times 10^{-2}\mathrm{T}^2 + 2.374 \times 10^{-4}\mathrm{T}^3 \\ &+ 1.34(\mathrm{S}-35) + 1.63 \times 10^{-2}D + 1.675 \times 10^{-7}\mathrm{D}^2 - 1.025 \\ &\times 10^{-2}(\mathrm{T} \times (\mathrm{S}-35)) - 7.139 \times 10^{-13}TD^3 \end{aligned} \qquad (2)$$

where, D indicates the depth, S and T indicate salinity and temperature, respectively.

Although the measured depth range of the CTD can cover the sound fixing and ranging (SOFAR) axis, it is still far from the local seafloor. Thus, the missing depth from 1500 m to the seafloor are filled using the climatological SSPs estimated from T and S profiles provided by the EN.4.2.2 dataset.

2.3.2 Acoustic reciprocity theorem

Figure 2 shows the experimental configuration (upper panel) and simulation conditions (lower panel), which are consistent in terms of the point-to-point acoustic response. The sound ray from $\mathrm{S_H}$ to RD1 (blue line) is employed to explain the acoustic

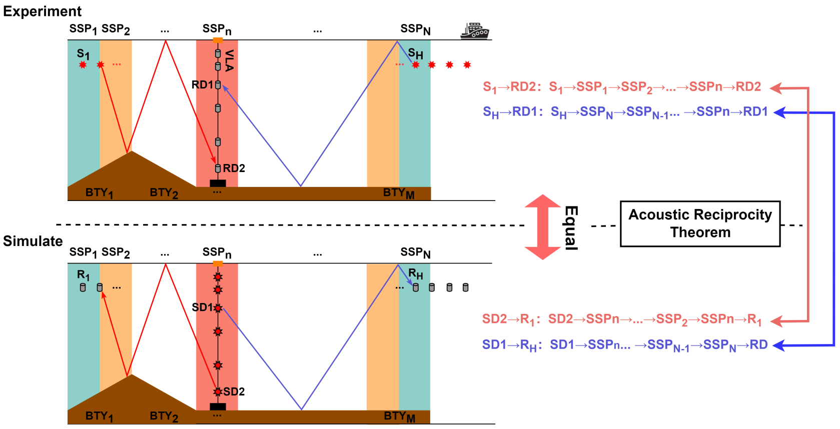


FIGURE 2
The diagram of acoustic reciprocity theorem of the field experiment and simulation. The upper panel shows the position of the VLA (marked with grey cylinders) and the trajectory of the explosive sound sources (marked with red eight-pointed stars) with a fixed explosive depth of 200 m. The colorful shaded areas represent range-dependent sound speeds in the presence of the warm eddy, and the brown shaded area represents the sea bottom. The red (blue) solid line indicates the sound ray launched from S_1 (S_H) and received by the RD2 (RD1) receiver. The lower panel shows the simulation configuration when analyzing the sound field using 2D and 3D ray-tracing models.

reciprocity theorem (Jensen et al., 2011). In the field experiment, it is launched from the sound source S_H, which eventually undergoes the range-dependent SSPs (from No. 1 to No. n), and arrival at the receiver RD1. In the simulation, the sound ray from SD1 to R_H (blue line) travels the total opposite path. Apparently, the system responses of the above two paths are equal. Specifically, the simulation can be consistent with the experiment by swapping the positions of the sound source and receivers.

2.3.3 Processing methods of the received acoustic signals

The VLA is a type of equipment commonly used in acoustic survey experiments (Ge and Kirsteins, 2017; Song and Wang, 2022). It mainly consists of a series of hydrophones, underwater signal recorders (USR), and floats, which are combined *via* a Kevlar rope. The deployed depth of the hydrophone can be designed according to the experimental purpose and environment.

To obtain the transmission losses (TL) at different depths, the received signals measured by hydrophones deployed at different depths need to be processed. The specific process has been optimized into the following six steps.

Step1: Truncate the signal emitted from the sound source from the discrete time signal recorded by USR and noted as $s_n = s(t_n)$, where $s = (n - 1)\Delta t$, $(n = 1,\ldots,N)$ denotes the discrete time, and n is the number of sample point. f_s denotes the sampling rate of USR and $\Delta t = 1/f_s$ denotes the discrete time interval. The start time of the truncated signal can be roughly estimated from the recorded time of the bomb explosion, the horizontal distance between the sound source and the receiver, and the sound speed of the local sea water. Simultaneously, to obtain the complete signal of the explosion, the length of the window for truncating should be determined by amplitudes of recorded signals.

Step 2: Calculate the frequency spectrum of the discrete time signal s_n using Discrete Fourier Transform (DFT) method, which is written as

$$S_k = \sum_{n=0}^{N=1} s_n \exp\left[-i\frac{2\pi}{N}kn\right] \tag{3}$$

where, k denotes the point number of the DFT.

Step 3: Calculate the total energy of the truncated signal in the 1/3 Octave frequency band of the center frequency f_0. The total energy is given by

$$E(f_0) \approx \frac{2}{Nf_s}\sum_{k=n_1}^{n_2}|S_k|^2 \tag{4}$$

where, $n_1 = f_l/d_f + 1$, $n_2 = f_H/d_f + 1$, $d_f = f_s/N$. $f_L = 2^{-1/6}f_0$ and $f_H = 2^{1/6}f_0$ denote the lower and upper band of center frequency f_0.

Step 4: Compute the mean energy of the truncated signal in the narrow band from f_L to f_H. The formula is written as

$$\hat{E}(f_0) = \frac{1}{f_H - f_L}E(f_0) \approx \frac{2}{(n_2 - n_1)f_S^2}\sum_{k=n_1}^{n_2}|S_k|^2 \tag{5}$$

Step 5: Compute the source level (SL). The SL (unit: dB) at distance R and center frequency f_0 is given by

$$SL(f_0, R) = 10\lg\left[\hat{E}(f_0)\right] - M_V - m + 20\lg(R) \tag{6}$$

where, M_V denotes the sensitivity of hydrophone in dB, m denotes the magnification of the receiving system in dB, and R is the distance from source to receiver in meter.

Step 6: Calculate the TL at f_0 center frequency. TL indicates the degree of attenuation of sound energy from source to the receiver, thus, the difference between SL and $\hat{E}(f_0)$ is the TL. The formula is given by

$$TL(f_0, r) = SL(f_0) - \left(10\lg\left[\hat{E}(f_0)\right] - M_V - m\right) \qquad (7)$$

3 Analysis of observational data

3.1 Temperature and salinity structures

The surveyed eddy is accompanied with positive SLA and counterclockwise geostrophic currents as shown in Figure 1. With the evolution of SLAs in the four days, it was observed that: (1) SLAs associating with the eddy gradually decrease with time; (2) The shape of the eddy appears as an ellipse pattern changing temporally. Specifically, the long axis of the ellipse is east-west oriented at October 3, 2022, while in the other days, it becomes northwest-southeast oriented.

Based on the *in-situ* data measured in lines A, B, C, and D through the northern, eastern, western and southern sides of the warm eddy in Figure 1 (yellow dots), the corresponding T, temperature anomaly (TA), S, salinity anomaly (SA), density (Den), and density anomaly (DA) vertical sections of the eddy are exhibited in Figure 3, respectively. Following the naming rules of CTD lines in Figure 1, the four vertical sections here are also named sections A, B, C, and D, respectively. Among them, T and S are directly measured by CTD observation stations, Den is derived from T, S and pressure, and the TA, SA and DA are computed by removing the associated climatological profile provided by the EN.4.2.2 dataset.

Through the distributions of SAs and TAs in the four vertical sections, it was observed that the eddy-induced TA and SA occur at different depths. The TA is confined to the depths from 70 m to 200 m with a maximum TA of about 6°C, while the SA is confined to the depths from 10 m to 70 m with a maximum SA of about 1 PSU. Simultaneously, isohalines and isotherms of the warm eddy exhibit lenticular structures with an upper convexity and a lower concavity, respectively. Based on the T and S characteristics accompanied with the SLAs, it was concluded that it is a surface-intensified warm eddy (Chaigneau et al., 2011).

There are different behaviors among the four vertical sections, indicating significant differences in terms of SA and

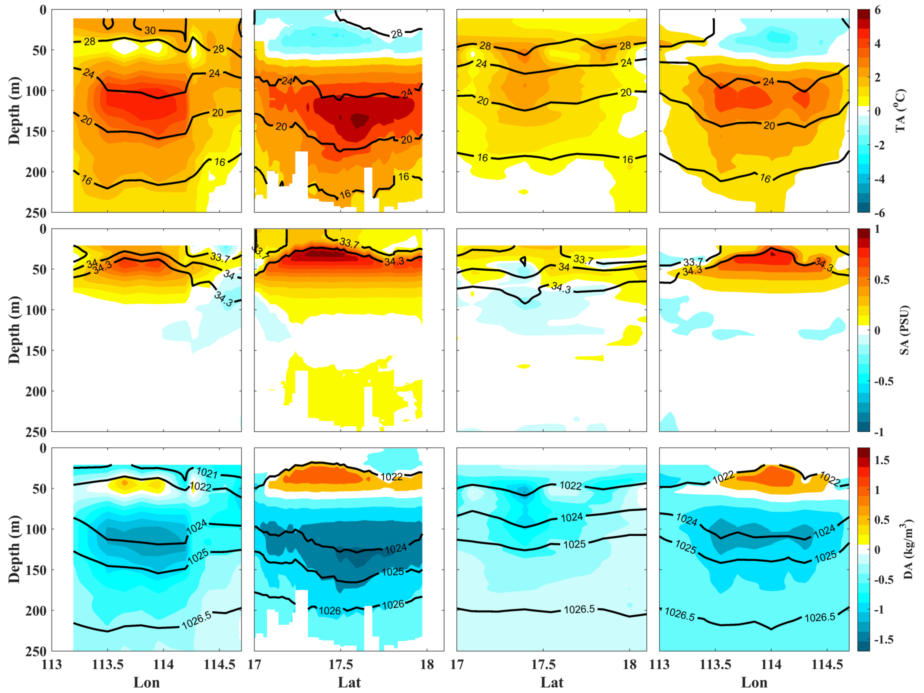


FIGURE 3
The four investigated thermohaline vertical sections of the warm eddy measured by A, B, C and D from left to right, respectively. The T (unit: °C) structures of the investigated vertical sections are shown in upper panels, where the shaded areas show the magnitude of TA (unit: °C), and the solid black lines indicate isotherms. The S (unit: PSU) structures of the investigated vertical sections are shown in middle panels; the shaded areas show the magnitude of SA (unit: PSU), and the solid black lines indicate iso-salinity lines. The Den (unit: kg/m^3) structures of the investigated vertical sections are shown in bottom panels; the shaded areas show the magnitude of DA (unit: kg/m^3), and the solid black lines indicate isopycnal lines. The x-axis indicates latitudinal and longitudinal ranges. The y-axis indicates the water depth from 0 m to 250 m.

TA structures at different times and positions of the warm eddy. Specifically, the dipole structures of TA are demonstrated in vertical sections B and D, where the cold and warm cores occur at depths of about 40 m and 130 m, respectively. Similar to the TA, DAs in vertical sections, B and D also show dipole structures. On the contrary, the monopole structures of TA are demonstrated in vertical sections A and C with warm core depths of 120 m and 50 m, respectively.

The measurement of vertical sections B and C were conducted in two consecutive days (Oct 20 and 21, respectively), of which the difference in measurement times is negligible relative to the slow propagation of the mesoscale eddy. Therefore, it is reasonable to use vertical sections B and C to represent the structures at different positions of the eddy at the same moment. Further, the difference between vertical sections B and C can prove that the western and eastern sides of the eddy are still significantly different in T and S structures even simultaneously. For instance, on the western side of the eddy (section C), the TA is small, and even a slight negative SA (-0.18 PSU) appears in the center of the TA. While on the eastern side of the eddy (section B), the most significant positive TA (6°C) and SA (1 PSU) are observed.

By comparing the SLAs in Figure 1 and subsurface anomalies (Figure 3) associated with the four vertical sections, it shows that the subsurface features could reflect the surface characteristics. For instance, the upraising (negative) and depressing (positive) of isopycnal lines (SLA) basically follow the first-order baroclinic mode relationship.

As mentioned in Section 2.1, a total of 105 CTD station observations were performed during the entire EASO field experiment. Among them 55 CTD (not shown), from October 20 to October 30, 2021, are selected to composite the 3D structure of the observed eddy. As shown in Figure 1, SLAs and GCAs demonstrate that the surface shape and position of the eddy had moderate changes during that period. Thus, the corresponding gridded observation stations are selected to composite the 3D structure, which are sufficient to represent the impact of the eddy on the sound propagation. Using all the 105 stations will not change the main conclusions of the study (figures not shown).

After interpolating the T and S observations of 55 CTD stations onto the 3D grid points, the 3D T and S structures of the warm eddy are obtained, as shown in Figures 4A, B. The 3D TA and SA structures in Figures 4C, D are obtained by removing the associated climatological profile provided by the EN4.2.2 dataset from the 3D T and S of the eddy, respectively. From 110 m to 170 m depths, the shapes of T and TA show good consistency with the surface shape of the eddy defined by the contour of SLA. As shown in Figure 1, from October 20 to October 30, 2021, the shape of the eddy at the surface is ellipse-like, with the long axis of the ellipse-oriented northwest-southeast. Simultaneously, the shapes of T and TA also exhibit elliptical patterns, with the long axis of the ellipse oriented in the same direction as that of the surface. Furthermore, the composited 3D eddy maintains the features in core depths of TA and SA, which are confined in 70 m to 200 m and 10 m to 70 m, respectively. The composited 3D

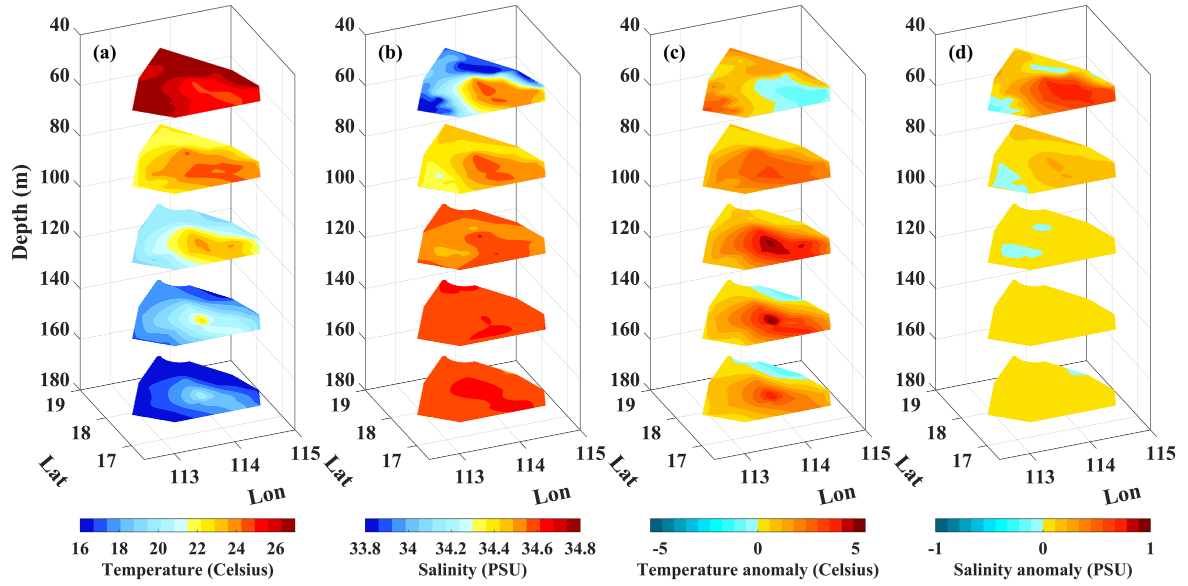


FIGURE 4
3D structures of the warm eddy. The **(A)** T (unit: °C), **(B)** S (unit: PSU), **(C)** TA (unit: °C), and **(D)** SA (unit: PSU) slices at depths from 50 m to 170 m with an interval of 30 m are shown from left to right panels.

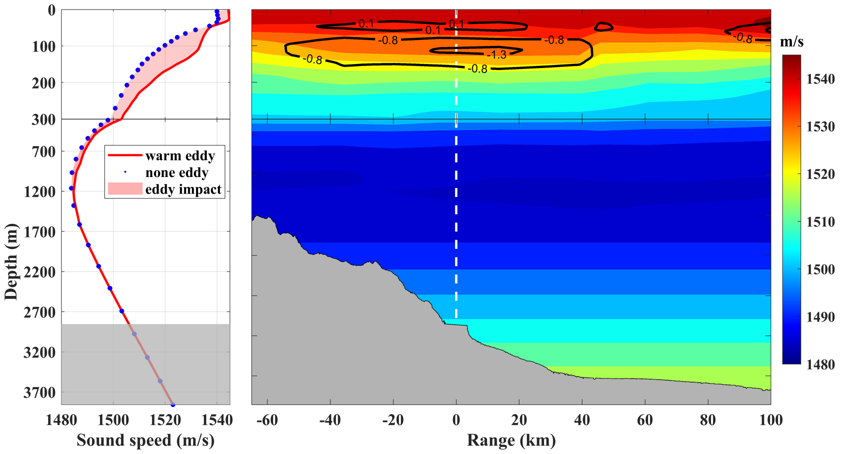


FIGURE 5
Sound speed (unit: m/s) of the section A. The SSPs in warm eddy (red solid line) and none eddy (blue dotted line) circumstances are shown in the left panel, and the red shadow area indicates the impacts of the warm eddy on sound speeds. The 2D sound speed distribution is shown in the right panel, an the black solid lines represent the anomalies of density induced by the warm eddy. The white dotted line indicates the position of VLA (R1), where SSP corresponds to the red one in the left panel.

eddy here will support the study of its 3D sound field in section 4.3.

3.2 Acoustic characteristics

Figure 5 (right panel) illustrates the whole-depth vertical section of SSPs associated with line A, as well as the acoustic survey line, where the R1(VLA) is arranged at 0 km, and the x-axis is eastward. The black solid lines indicate the eddy-induced density anomalies. Simultaneously, bathymetries measured by the shipboard echo sounders are shown in Figure 5, agreeing well with the bathymetries provided by the ETOPO2 dataset (Figure 1, upper panel). Furthermore, to exhibit impacts of the warm eddy, SSPs in the warm eddy (red line) and none eddy (blue dots) circumstances are also presented in Figure 5 (left panel). The impacts of the warm eddy on sound speeds are indicated by the red shallow area. The SSP in the eddy center is marked with the white dotted line in Figure 5 (right panel); while the SSP representing the none eddy case selected from the monthly mean SSP is located at the same position. Obviously, both the two SSPs in Figure 5 are incomplete sound channel environments due to the local bathymetries.

Some differences between the warm eddy SSP and none eddy SSP were observed, illustrated by the following aspects: (1) The warm eddy leads to a higher T in the upper layer. Thus, the value at the surface of the warm eddy SSP (1544 m/s) is slightly faster than that of none eddy SSP (1540 m/s). (2) The warm eddy SSP shows two rapid vertical changes from 29 m to 45 m and from 115 m to 300 m depths. (3) The impact of the surface-intensified eddy is mainly concentrated at the depths from 0 m to 300 m, and the most significant difference between the warm eddy SSP and none eddy SSP is about 15 m/s occurring at a depth of 135 m. In depths deeper than 135 m, the difference in sound speed caused by the warm eddy gradually decreases and disappears at a depth of 1200 m, which is close to the SOFAR axis. In terms of the vertical section of sound speed, iso-speed lines show a downward concave pattern within the warm eddy and gradually lift as moving to both sides of the eddy center.

Through the processing procedure mentioned in Section 2.3.3, the TL curve at the receiving depth (RD) of 185 m and 300 Hz center frequency for the EASO field experiment is presented in the bottom panel of Figure 6, marked with black cross symbols. For the R1-T2 survey line, two obvious CZs with about 70 dB and 78 dB TL at about 39 km and 92 km, respectively, were observed. As for the T1-R1 survey line, only a CZ with about 72 dB TL at about -25 km influenced by the rough topographies was observed.

4 Discussion

4.1 Comparisons of sound fields between warm eddy and none eddy circumstances

To analyze the acoustic experiment, the simulations based on the acoustic reciprocity theorem with Bellhop (Porter, 2011) are performed. As shown in Figure 6 (top panel), the simulated 2D TL with the warm eddy circumstance is computed within the narrow band determined by the 1/3 octave of the center frequency of 300 Hz, where the simulated sound source is set

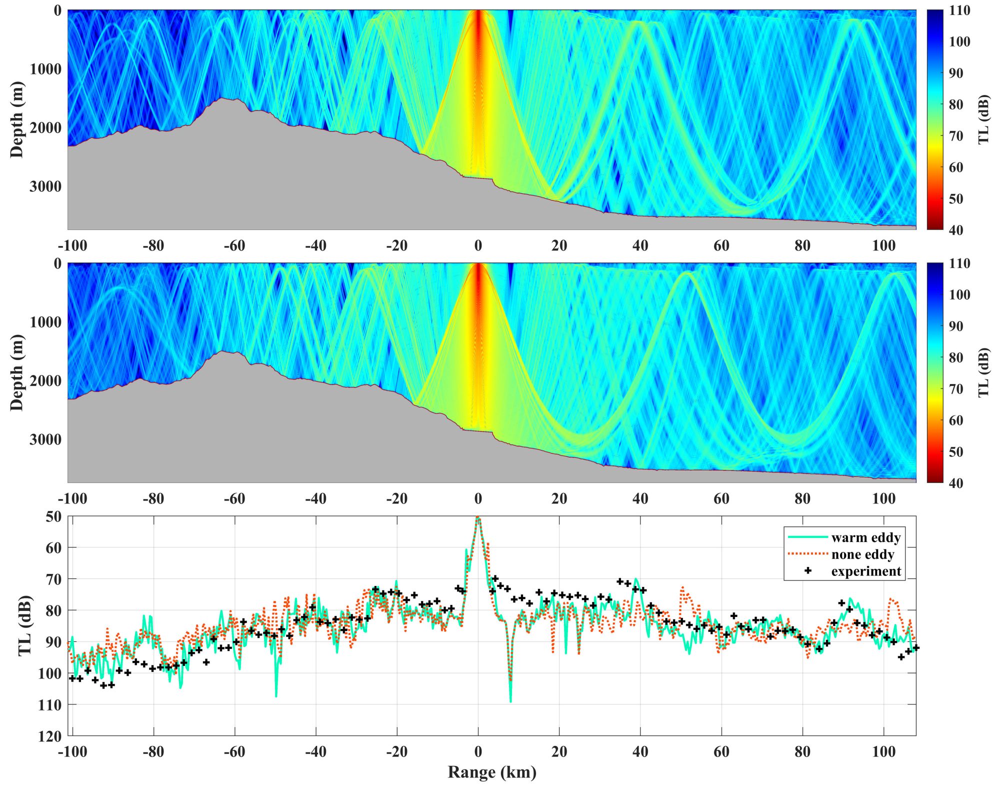


FIGURE 6
The comparison of sound field between the warm eddy and none eddy circumstances. The 2D sound TL in the r-z plane with warm eddy (none eddy) circumstance is shown in the upper (middle) panel, where the source is deployed at 185 m depth and 0 km distance, and the grey shaded area represent the sea bottom. The TL (unit: dB) curves are arranged in the bottom panel with the fixed RD of 200 m. The green solid line, red dotted line, and black cross symbols indicate the TLs for the simulated warm eddy, simulated none eddy, and in the experiment, respectively.

at a depth of 185 m. The coherent Gaussian Beams are employed in Bellhop. The acoustic absorption coefficient of the water body dominated by frequency can be estimated with the following formula (Jensen et al., 2011), written as

$$\alpha = 3.3 \times 10^{-3} + \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 3.0 \times 10^{-4}f^2 \quad (8)$$

where, f denotes frequency in kHz, and the unit of α is dB/km. Due to the incomplete sound channel, sound rays frequently interact with the sea floor. Therefore, the acoustic absorption coefficient of the local sea floor is a crucial parameter. However, there are no sediment observations in the EASO field experiment. Thus, it is represented by a fixed value of 0.2 dB/wavelength. As shown in Figure 6 (bottom panel), there is a good consistency between the simulated (Figure 6, green line) and experiment TL curves in the presence of the warm eddy.

From the simulated 2D TL (Figure 6, top panel) with the warm eddy circumstance, the possible causes for the difference between TL curves of T1-R1 and R1-T2 can be determined. The topographies in R1-T2 are flatter and deeper, in favor of the formation of CZs. Contrarily, topographies in T1-R1 are rougher and shallower, leading to the high-frequency interaction between sound rays and sea floors, and thus destroy the structure of the CZ.

To demonstrate the effect of warm eddy on the sound field, the simulated TL results (Figure 6, middle and bottom panels) with the none eddy circumstance are calculated as a comparison, where the simulation conditions are the same as that in the warm eddy circumstance except for the range-independent SSP (the blue dotted line in Figure 5) deriving from the EN.4.2.2 dataset through Formula (2). From the 2D sound field of the none eddy case in R1-T2, it was observed that two CZs at about 50.4 km and 101.7 km are present. As for the T1-R1 line, only one CZ with about 72 dB TL at about -25 km is present.

By comparing the TL curves with and without eddy (red dotted line) in R1-T2, it was observed that the first and second CZs in the warm eddy case exhibit noticeable forward shifts with about 11.4 km and 9.7 km, respectively. However, in the T1-R1 survey line, it is hard to find obvious difference in CZs. Additionally, the differences in sound fields between the two circumstances are also manifested in the depth and width variations of the CZ. As shown in Figure 6 (upper panel), the second CZ (150 m) shows a downward lifting of about 32 m

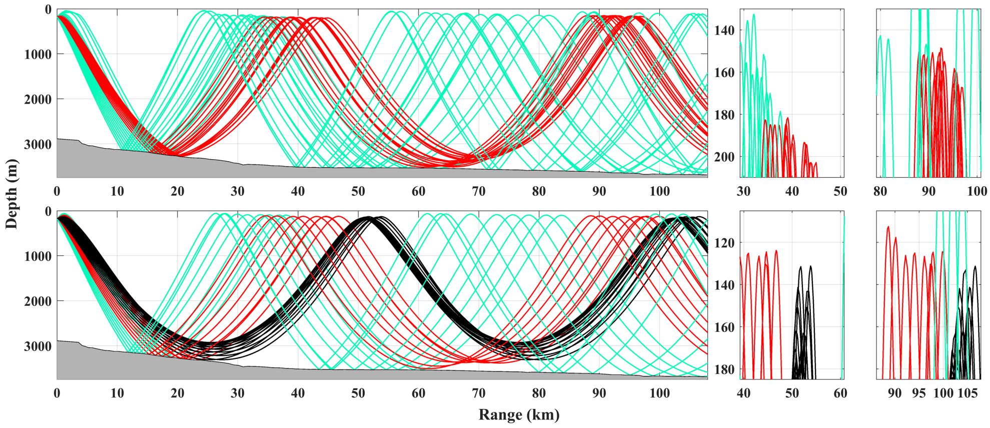


FIGURE 7
The comparison of simulated sound ray traces with 300 Hz frequency between the warm eddy (upper panel) and none eddy (lower panel) circumstances. The two right columns show the enlarged and detailed characteristics.

compared to that of the first (182 m), while such a phenomenon cannot be found in the none eddy case. In the warm eddy case, the widths of the first and second CZs are 10.4 km and 9 km, respectively, showing a change of 1.4 km.

4.2 Roles of topography and eddy

In this section, the respective roles of topography and warm eddy in changing sound fields will be discussed. Figure 6 shows that both the warm eddy and topography are responsible for the differences in sound fields between warm eddy and none eddy circumstances. In the T1-R1 survey line, the rough and shallow topographies result in the frequent interaction between sound rays and the bottom, and thus obscure the effects of the warm eddy on sound rays, which render the differences in TL curves imperceptible. Contrarily, in the R1-T2 survey line, the sound rays only hit the relatively flat and deep bottom one time before the first CZ in the warm eddy circumstance. Between the first and second CZs, the sound rays only experienced the refraction of the warm eddy-induced water body. Therefore, the fields before and after the first CZ in the R1-T2 survey line can illustrate the impacts of the warm eddy and topography on sound fields, respectively.

The sound ray traces simulated by Bellhop are presented to further interpret the differences between the warm eddy (upper panel) and none eddy (lower panel) circumstances in the R1-T2 line, as shown in Figure 7. Here, different coloured lines are used to identify purely refracted paths that do not hit surface and bottom boundaries (black), one bottom-reflected paths (red), more than one bottom-reflected paths (green), and both surface and bottom paths (blue), respectively.

Considering the relative roles of topography and warm eddy on the horizontal span of CZs, as shown in Figure 7 (upper panel), in the warm eddy circumstance, the sound rays launched from source and hit the bottom one time and then generate the first CZ in 39 km. Thereafter, the bottom-reflected sound ray only experienced refraction and generate the second CZ in 92 km. In the none eddy circumstance, both the first (50.4 km) and second CZs (101.7 km) are mainly generated by the purely refracted rays (black lines), representing the fields avoiding the influence of the topography and warm eddy. Thus, the differences in the horizontal spans of the two CZs between the warm and none eddy circumstances represent the influences dominated by the topography and warm eddy, respectively. The comparison results indicate that the topography shortens the horizontal span of the CZ by 11.4 km, while the warm eddy lengthens it by 1.7 km.

Considering the impact of warm eddy on the vertical shift of CZ, as only the warm eddy impacts the sound rays between the first and second CZs in the warm eddy circumstance, the difference between the first and second CZ can directly reflect the effect of the warm eddy. As shown in the detailed views of Figure 7 (upper panel), the first and second CZs are at depths of 182 m and 150 m, showing an upward vertical shift of about 32 m resulting from the warm eddy. As the first CZ is close to the center of the warm eddy, it can be concluded that the depth of the CZ decreases as it moves away from the center of the warm eddy. Specifically, the warm eddy decreases the depth of the CZ. Considering the impact of warm eddy on the width of CZ, from the detailed views of Figure 7 (upper panel), it was observed that the widths of the first and second CZs are 10.4 km and 9 km, respectively, showing a decrease of about 1.4 km. The comparison results indicate that the width of the CZ decreases

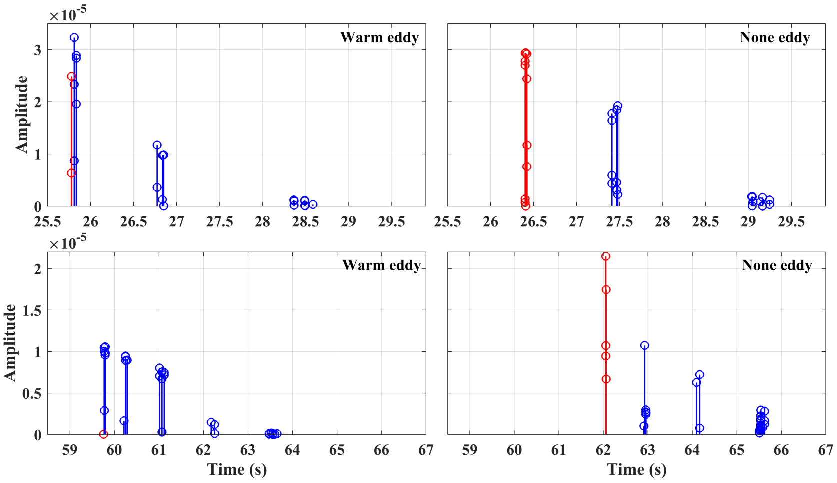


FIGURE 8
Arrivals from the sound source at 185 m depth and 0 km distance to the receiver at 200 m depth and 39 km (upper panel) and 92 km (lower panel) distances with warm eddy (left panel) and none eddy (right panel) circumstances.

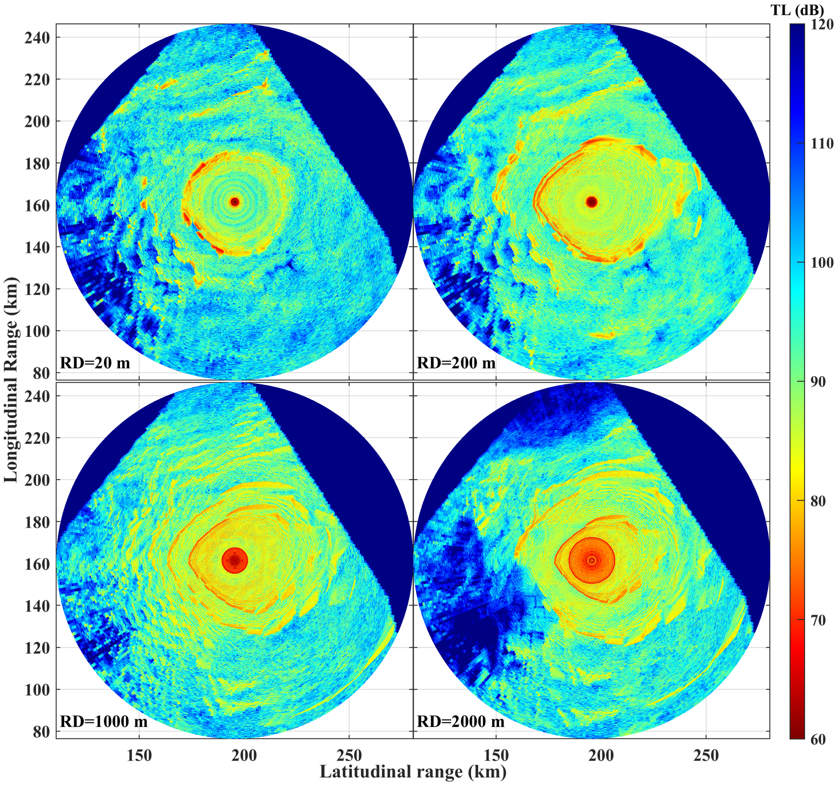


FIGURE 9
The 3D sound field at cartesian coordinate system in the presence of warm eddy and complex topography. The sound source is deployed at 195.5 km, 161.4 km, and 0.185 km with 300 Hz frequency. Top views of the TLs at receiver depths of 20 m, 200 m, 1000 m and 2000 m are shown, respectively.

as moving away from the center of the warm eddy. Specifically, the warm eddy increases the width of the CZ.

To provide more insights into the sound field impacted by the warm eddy, the arrival structures and travel times of rays at the first (39 km) and second CZs (92 km) with warm eddy (left panels) and none eddy (right panels) circumstances are presented in Figure 8. It was observed that travel times in the warm eddy case are shorter than those in the none eddy case due to the higher sound speed induced by the warm eddy. Additionally, the arrival structure of the second CZ in the warm eddy case is more complex than that in the none eddy case.

4.3 3D Effects of the warm eddy on sound fields

In the incomplete sound channel environment, both SSPs and topographies should be taken into consideration for predicting the sound field. To study the anisotropy of the 3D sound field jointly dominated by the eddy and topographies, the 3D sound field calculated with the Bellhop3D (Porter, 2011) in the receiving depths of 20 m, 200 m, 1000 m, and 2000 m are illustrated (Figure 9). In Figure 9, the Cartesian coordinate system is transformed from the earth coordinate system originating at 16.5° N, 112°E, and the eastward x-axis and northward y-axis denote latitudinal and longitudinal ranges, respectively.

The 3D sound speed of the warm eddy is derived from the 3D T and S from Figures 4A, B through Formula (2), and the 3D topography is provided by the ETOPO2 dataset. The sound source is deployed at 195.5 km, 161.4 km, and 0.185 km with 300 Hz frequency. Additionally, the azimuth angle in horizontal plane is set from 0° (eastward) to 360° with an interval of 1°, and the pitch angle in vertical plane is set from -180° (downward) to 0° with an interval of 1°. The acoustic survey line T1–T2 corresponds to the line parallel to the x-axis and passing through the source.

From the simulation results, it was observed that all the first CZs are the reflection type, which implies that the topography is the primary factor determining the sound field, while the warm eddy-induced 3D sound speed field is the second factor. By comparing the sound fields in the mixed layer (20 m) and the thermocline (200 m), it is found that there are significant differences in the distribution patterns of sound fields. In the mixed layer, the distances of the first CZs are relatively uniform. But in the thermocline, the distances of the first CZs exhibit significant non-homogeneous. For the first reflection CZ, its horizontal distance to the source increases with the depth of the seafloor. For the TLs at a depth of 200 m, the CZs positions vary with the horizontal propagation directions, which are referred to as the anisotropy of the sound field. Specifically, the distances of the first CZs on the eastern side of the sound source are longer than that on the western side. The nearest (20 km) and farthest (51 km) CZs occur in the 213° and 15° directions of the sound source, respectively, which correspond to the rough and flat topographies in the study area. Furthermore, in the 3D sound fields, the most significant shadow zones induced by the rough and shallow topography, with most TLs higher than 110 dB, are observed in the southwest and northwest corners at a depth of 2000 m.

5 Conclusions

An eddy-acoustic synchronous observation field experiment was performed in the slope of the South China Sea in October, 2021, to study the effects of the warm eddy on sound fields. A total of 105 CTD profiles and an acoustic survey line with about 210 km length are measured in the experiment. In the hydrological part of the investigation, four complete vertical sections precisely crossing the northern, southern, eastern and western sides of the center of the observed warm eddy were present. From the four vertical sections of the eddy, the lenticular structures of isohalines and isotherms were observed. The subsurface and surface characteristics suggested that the observed warm eddy is surface-intensified and ellipse-like in shape. Subsequently, 55 CTD profiles were chosen to composite the 3D structures of the observed eddy, which showed consistent structures with surface and the largest TA and SA distributing in 70 m to 200 m and 10 m to 70 m, respectively.

In the acoustic part of the experiment, the acoustic observation showed the presence of two obvious CZs at about 39 km and 92 km in R1-T2 line with flatter and deeper bottom, and one CZ at about 25 km in T1-R2 with rough and shallower bottom, indicating that the topography and warm eddy are jointly responsible for the differences. To quantitatively analyze the acoustic behaviors of the topography and warm eddy, a ray-tracing acoustic model is adopted to compare the sound fields with and without eddy circumstances, as well as the different topographies. The comparison results showed that the topography shortens the horizontal span of the CZ by 11.4 km, while the warm eddy lengthens it by 1.7 km. Additionally, the warm eddy decreases the depth of the CZ by 32 m, and increases its width 1.4 km. Finally, based on the composited 3D eddy, anisotropy of 3D sound fields is found through Bellhop3D. The results show that the distances differences of the first CZs in different horizontal directions can be up to 31 km.

The effects of topographies on the sound field, such as the submarine mountain, slop, and ridge, were widely studied in the traditional acoustical works, suggesting that topography has the most significant effect on the sound field (Chiu et al., 2011; Dossot et al., 2019; Liu et al., 2021b). However, the effects of inhomogeneous water bodies were not discussed thoroughly in these studies due to the complicated oceanic processes. In this study, benefiting from the well-designed field experiment, quantitative effects of the topography and warm eddy are obtained. In future works, the topography-induced horizontal refraction effects in the sound fields will be emphasized on different oceanic systems, such as fronts, internal waves, sub-mesoscale processes, and so on.

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

WC performed the data analysis and wrote the initial draft. YCZ proposed the main ideas and dramatically modified the manuscript. YL and KR all participated in the discussion and contributed to the improvement of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the Science and Technology Innovation Program of Hunan Province (2022RC3070) and the National Natural Science Foundation of China under Grant No. 62101578.

Acknowledgments

The authors thank all of the experiment participants for their great efforts to make this experiment successful in 2021. The authors thank Wei Guo and Guojun Xu for their reasonable advices on acoustic calculation.

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

Baer, R. N. (1981). Propagation through a three-dimensional eddy including effects on an array. *J. Acoust. Soc Am.* 69 (1), 70–75. doi: 10.1121/1.385253

Calado, L., Gangopadhyay, A., and da Silveira, I. C. A. (2006). A parametric model for the Brazil current meanders and eddies off southeastern Brazil. *Geophys. Res. Lett.* 33 (12), L12602. doi: 10.1029/2006GL026092

Chaigneau, A., Le Texier, M., Eldin, G., Grados, C., and Pizarro, O. (2011). Vertical structure of mesoscale eddies in the eastern south pacific ocean: A composite analysis from altimetry and argo profiling floats. *J. Geophys. Res.: Oceans.* 116 (C11), C11025. doi: 10.1029/2011JC007134

Chelton, D. B., Schlax, M. G., and Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.* 91 (2), 167–216. doi: 10.1016/j.pocean.2011.01.002

Chen, C., Jin, T., and Zhou, Z. (2019). Effect of eddy on acoustic propagation from the surface duct perspective. *Appl. Acoust.* 150, 190–197. doi: 10.1016/j.apacoust.2019.02.019

Chen, W., Zhang, Y., Liu, Y., Ma, L., Wang, H., Ren, K., et al. (2022). Parametric model for eddies-induced sound speed anomaly in five active mesoscale eddy regions. *J. Geophys. Res.: Oceans.* 127 (8), e2022JC018408. doi: 10.1029/2022JC018408

Chiu, L. Y. S., Lin, Y.-T., Chen, C.-F., Duda, T. F., and Calder, B. (2011). Focused sound from three-dimensional sound propagation effects over a submarine canyon. *J. Acoust. Soc Am.* 129 (6), EL260–EL266. doi: 10.1121/1.3579151

Collins, M. D., and Werby, M. F. (1989). A parabolic equation model for scattering in the ocean. *J. Acoust. Soc Am.* 85 (5), 1895–1902. doi: 10.1121/1.397896

Colosi, J. A., Cornuelle, B. D., Dzieciuch, M. A., Worcester, P. F., and Chandrayadula, T. K. (2019). Observations of phase and intensity fluctuations for low-frequency, long-range transmissions in the Philippine Sea and comparisons to path-integral theory. *J. Acoust. Soc Am.* 146 (1), 567–585. doi: 10.1121/1.5118252

Dossot, G. A., Smith, K. B., Badiey, M., Miller, J. H., and Potty, G. R. (2019). Underwater acoustic energy fluctuations during strong internal wave activity using a three-dimensional parabolic equation model. *J. Acoust. Soc Am.* 146 (3), 1875–1887. doi: 10.1121/1.5125260

Gao, F., Xu, F.-H., and Li, Z.-L. (2022). The effects of mesoscale eddies on spatial coherence of middle range sound field in deep water. *Chinese. Phys. B* 31 (11). doi: 10.1088/1674-1056/ac6014

Ge, H., and Kirsteins, I. P. (2017). Characterization and exploitation of lucky scintillations in HLA and VLA data from the 2006 shallow water experiment. *J. Acoust. Soc Am.* 141 (5), 3918–3918. doi: 10.1121/1.4988853

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res.: Oceans.* 118 (12), 6704–6716. doi: 10.1002/2013JC009067

Henrick, R. F., Siegmann, W. L., and Jacobson, M. J. (1977). General analysis of ocean eddy effects for sound transmission applications. *J. Acoust. Soc Am.* 61 (S1), S11–S11. doi: 10.1121/1.2015408

Jensen, F. B., Kuperman, W. A., Porter, M. B., Schmidt, H., and Tolstoy, A. (2011). *Computational ocean acoustics* (Springer).

Jian, Y. J., Zhang, J., Liu, Q. S., and Wang, Y. F. (2009). Effect of mesoscale eddies on underwater sound propagation. *Appl. Acoust.* 70 (3), 432–440. doi: 10.1016/j.apacoust.2008.05.007

Liu, J., Piao, S., Gong, L., Zhang, M., Guo, Y., and Zhang, S. (2021a). The effect of mesoscale eddy on the characteristic of sound propagation. *J. Mar. Sci. Eng.* 9 (8). doi: 10.3390/jmse9080787

Liu, W., Zhang, L., Wang, W., Wang, Y., Ma, S., Cheng, X., et al. (2021b). A three-dimensional finite difference model for ocean acoustic propagation and benchmarking for topographic effects. *J. Acoust. Soc Am.* 150 (2), 1140–1156. doi: 10.1121/10.0005853

Mackenzie, K. V. (1981). Nine-term equation for sound speed in the oceans. *J. Acoust. Soc Am.* 70 (3), 807–812. doi: 10.1121/1.386920

Nan, F., Yu, F., Wei, C., Ren, Q., and Fan, C. (2017). Observations of an extra-Large subsurface anticyclonic eddy in the northwestern pacific subtropical gyre. *J. Mar. Sci. Res. Dev.* 7, 234. doi: 10.4172/2155-9910.1000234

Porter, M. B. (2011) *The BELLHOP manual and user's guide PRELIMINARY DRAFT*. Available at: http://oalib.hlsresearch.com/Rays/HLS-2010-1.pdf.

Porter, M. B. (2016) *Bellhop3D user guide*. Available at: https://usermanual.wiki/Document/Bellhop3D20User20Guide202016725.1524880335/view.

Porter, M. B. (2019). Beam tracing for two- and three-dimensional problems in ocean acoustics. *J. Acoust. Soc Am.* 146 (3).2016 doi: 10.1121/1.5125262

Ramp, S. R., Colosi, J. A., Worcester, P. F., Bahr, F. L., Heaney, K. D., Mercer, J. A., et al. (2017). Eddy properties in the subtropical countercurrent, Western Philippine Sea. *Deep. Res. Part I Oceanogr. Res. Pap.* 125, 11–25. doi: 10.1016/j.dsr.2017.03.010

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

Vastano, A. C., and Owens, G. E. (1973). On the acoustic characteristics of a gulf stream cyclonic ring. *J. Phys. Oceanogr.* 3 (4), 470–478. doi: 10.1175/1520-0485(1973)003<0470:OTACOA>2.0.CO;2

Weinberg, N. L., and Zabalgogeazcoa, X. (1977). Coherent ray propagation through a gulf stream ring. *J. Acoust. Soc Am.* 62 (4), 888–894. doi: 10.1121/1.381609

Westwood, E. K., Tindle, C. T., and Chapman, N. R. (1996). A normal mode model for acousto-elastic ocean environments. *J. Acoust. Soc Am.* 100 (6), 3631–3645. doi: 10.1121/1.417226

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

Worcester, P. F., and Spindel, R. C. (2005). North pacific acoustic laboratory. *J. Acoust. Soc Am.* 117 (3), 1499–1510. doi: 10.1121/1.1854780

Wu, Y., Zhang, W., Hu, Z., Zhang, W., Zhang, B., Wang, J., et al. (2022). Directional response of a horizontal linear array to an acoustic source at close range in deep water. *Acoust. Aust.* 50 (1), 91–103. doi: 10.1007/s40857-021-00250-5

Zhang, Y., Chen, X., and Dong, C. (2019). Anatomy of a cyclonic eddy in the kuroshio extension based on high-resolution observations. *Atmosphere* 10 (9). doi: 10.3390/atmos10090553

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

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Navig.* 14 (7), 1055–1060. doi: 10.1049/iet-rsn.2019.0477

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fmars.2022.1049761

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Jiahua Zhu,
National University of Defense Technology, China
Weitao Sun,
The 54th Research Institute of China Electronics Technology Group Corporation, China

*CORRESPONDENCE
Pan Huang
wfxyhp@wfu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 21 September 2022
ACCEPTED 23 November 2022
PUBLISHED 16 December 2022

CITATION
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

Synthetic aperture imagery for high-resolution imaging sonar

Pan Huang[1]* and Peixuan Yang[2]

[1]School of Mathematics and Information Science, Weifang University, Weifang, China, [2]Research & Development Department, Acoustic Signal and Electronics Science and Technology Corporation, Lanzhou, China

Synthetic aperture sonar (SAS) can provide high-resolution underwater images. Traditional fast imaging algorithms designed for multi-receiver synthetic aperture sonar (MSAS) are complex because the point target reference spectrum (PTRS) deduction and imaging algorithm development are complicated. This paper proposes an imaging algorithm for the MSAS system to solve this issue. The proposed method first approximates the two-round slant range based on the phase center approximation method. The PTRS, including the quasi-monostatic and bistatic deformation terms, can be easily deduced. After compensating for the bistatic deformation term based on the interpolation and complex multiplication with the preprocessing step, the MSAS imagery can be simplified to the focus of the traditional monostatic SAS. Therefore, the conventional imaging algorithms designed for traditional monostatic SAS can be used directly. The proposed method providing high-resolution imaging results is more efficient than the traditional methods.

KEYWORDS

synthetic aperture sonar, high-resolution, imaging algorithm, phase center approximation, imaging performance

1 Introduction

The complex ocean environment (Hunter, 2006; Davis et al., 2009; Xia et al., 2016; Zhang et al., 2020) makes its observation challenging. Therefore, underwater imaging is essential for scientists and engineers working in ocean engineering. Currently, the high-resolution imaging equipment named synthetic aperture sonar (SAS) (Gough and Hawkins, 1996; Gough and Hawkins, 1997; Gough and Hawkins, 1998; Zhang and Ying, 2022; Dillon and Steele, 2022; Yang and Liu, 2022) can generate high-resolution images. Compared with traditional imaging sonar, such as side scan and multibeam sonar (Cutrona, 1975), SAS can provide high resolution in the along-track dimension independent of the frequency and detection range. Therefore, the SAS image can be used for seabed mapping (Barclay et al., 2005; Tan et al., 2019) and target detection (Dobeck, 1999; Rankin et al., 1999; Cong et al., 2000; Gough et al., 2006), such as searching for shipwrecks, airplane wrecks (LeHardy and Larsen, 2015), and pipelines (Sæbø et al., 2010).

Synthetic aperture image formation is a crucial technique for the SAS system. The monostatic SAS transmitter is still used as the receiver. The point target reference spectrum (PTRS) can be easily deduced using the stationary phase method. Based on the PTRS of the monostatic SAS, fast imaging algorithms such as the range-Doppler (RD) algorithm (Zhong and Liu, 2006; Neo et al., 2008; Wu et al., 2014; Zhang et al., 2017b), chirp scaling (CS) algorithm (Zhong and Liu, 2009; Zhong and Liu, 2010; Zhang et al., 2013), and range migration algorithm (RMA) (Liu et al., 2009; Shin and Lim, 2009; Zhang et al., 2021c) can be easily designed. For pulsed sonar, a longer detection range in the cross-track dimension requires a larger pulse repetition interval (PRI). However, higher resolution in the along-track dimension requires a lower PRI. Therefore, there is a significant contradiction between the cross-track dimension mapping swath and the along-track dimension resolution.

A multi-receiver synthetic aperture sonar (MSAS) (Rolt and Schmidt, 1992; Gough and Miller, 2003; Zhang et al., 2014a; Zhang et al., 2021c), which includes a single transmitter and many receivers in the along-track dimension, is proposed to solve this problem. MSAS can provide a wider mapping swath at a given resolution than the monostatic SAS system. Unfortunately, the use of the MSAS system involves two major challenges. Firstly, the two-round slant range of MSAS is characterized by double square root terms, which makes it difficult to deduce the PTRS of MSAS based on the stationary phase method (Wang et al., 2008; Wang et al., 2011). Secondly, all of the imaging algorithms have been designed for the monostatic SAS system and cannot be directly used for MSAS systems. To deduce the PTRS based on the stationary phase method, several approximations, including the Taylor approximation of the two-round slant range and Loffeld's bistatic formula (LBF) (Loffeld et al., 2004; Zhang et al., 2021b), are used.

In addition, the imaging algorithms used for monostatic SAS are further extended to the MSAS system. In general, these extensions can be grouped into two classes. The first involves developing a new algorithm according to the RD, CS, and RMA. This operation often employs the series approximation of PTRS, which is exceedingly tedious. The other involves directly using the monostatic SAS system imaging algorithms after converting the MSAS data into a monostatic SAS-equivalent signal. The second approach is much easier to achieve when comparing the advantages and disadvantages of both methods. The PTRS of MSAS must be decomposed into a quasi-monostatic term and a bistatic deformation term in order to directly use the imaging algorithms designed for monostatic SAS. After compensating for the bistatic deformation term, traditional imaging algorithms can be used directly.

Based on the assumption that the transmitter and the receiver contribute equally to the Doppler phase in the along-track dimension, Loffeld et al. (2004) developed a PTRS (often called LBF) that can be decomposed into a quasi-monostatic term and a bistatic deformation term. The bistatic deformation term of LBF depends significantly on the range, instantaneous frequency, and Doppler frequency. The sub-block processing method (Zhang et al., 2019) in the cross-track dimension is usually exploited to cancel this term; therefore, the sub-block width determines the imaging performance. In other words, a wider sub-block width reduces performance, while a narrower width increases performance. Unfortunately, the use of a narrow sub-block width is time-consuming.

In general, there is a significant contradiction between imaging performance and efficiency when the LBF is used. The phase center approximation (PCA) method (Gough et al., 2000; Gough and Hayes, 2005) supposes an element located at the midpoint between the transmitter and the receiver. This element is equivalent to a monostatic transducer used for signal transmission and reception. However, this operation results in an approximation error, called the PCA error (Zhang et al., 2021a). Traditional methods do not compensate for this approximation error, and image distortion is introduced at a close range. Some methods neglect the stop-and-hop error (Zhang et al., 2017a; Zhang et al., 2018), seriously affecting long-range SAS imaging results.

This paper presents an imaging method based on the PCA method. Firstly, the two-round slant range was approximated by the second-order Taylor approximation and inverse Taylor approximation. Then, the slant range equivalent to the traditional monostatic SAS was obtained. Because the PTRS of the monostatic SAS can be easily obtained, the stationary phase method was exploited to deduce the PTRS of MSAS, including the quasi-monostatic and bistatic deformation terms. After compensating for the bistatic deformation term, the MSAS data were converted into monostatic SAS-equivalent data, considered the input to the RD algorithm in this paper. Lastly, simulations were performed to validate the proposed method. The remainder of this paper is organized as follows. *Section 2* discusses the imaging geometry. *Section* 3 describes the proposed method. In *Section 4*, the simulations are described in detail. Finally, the last section summarizes the conclusions of this study.

2 MSAS imaging geometry

Figure 1 shows the MSAS imaging geometry. The sonar system is towed in the along-track direction, and the horizontal axis is in the cross-track direction. The fast time is denoted by τ and the slow time by t. An ideal target was assumed in this imaging scenario. The coordinate in the cross-track dimension is r, while that in the along-track dimension is 0. The transmitter and receiver arrays were deployed in the along-track dimension, where the black rectangle is the transmitter and the remaining

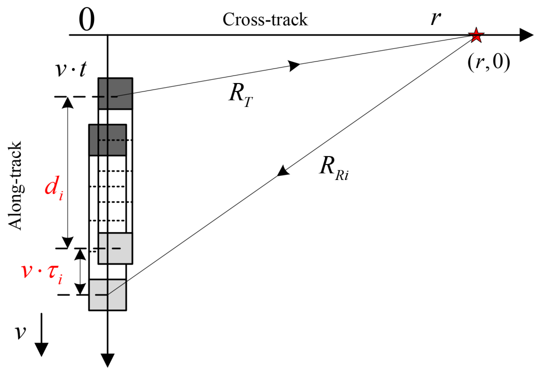


FIGURE 1
Multiple-receiver synthetic aperture sonar (MSAS) imaging geometry.

rectangles are the receivers, which were uniformly spaced at equal distances. The distance between the transmitter and the ith receiver is d_i, where i ($i\in[1,M]$) is the receiver index. M is the total number of receivers.

Based on Figure 1, the two-round slant range between the transmitter to the target and back to the ith receiver is

$$R_i(t;r) = R_T(t) + R_{Ri}(t) = \sqrt{r^2 + (vt)^2} + \sqrt{r^2 + (vt + d_i + 2v\cdot r/c)^2}. \quad (1)$$

In Equation 1, $R_T(t) = \sqrt{r^2 + (vt)^2}$ is the distance between the transmitter and the target. The distance between the target and the ith receiver is denoted by $R_{Ri}(t) = \sqrt{r^2 + (vt + d_i + 2v\cdot r/c)^2}$, where v is the velocity of the sonar platform and c is the sound velocity in water. Traditional methods neglect the stop-and-hop approximation (Zhang et al., 2017a) on the SAS system because traditional SAS systems work with low resolution and narrow mapping swath. A lot of high-resolution SAS systems with wide mapping swaths are currently being developed. If the stop-and-hop error is neglected, distortions such as higher side lobes and along-track shifting are introduced into the SAS image. Compared with the traditional signal model, Equation 1 considers the influence of the stop-and-hop approximation (Zhang et al., 2017a) on the SAS system.

The SAS system often uses a chirp signal. The received signal of the ith receiver is given by

$$ss_i(\tau,t) = p\left(\tau - \frac{R_i(t;r)}{c}\right)\omega_a(t)\cdot\exp\left\{-j2\pi f_c\frac{R_i(t;r)}{c}\right\} \quad (2)$$

In Equation 2, the transmitted chirp is denoted by $p(\tau)$. The beam pattern of the SAS system is denoted by $\omega_a(t)$. Because ω_a (t) does not affect the SAS focusing, it is suppressed in the following discussions. The carrier frequency is f_c . Because the two-round slant range in Equation 1 has two square root terms, it is challenging to deduce the PTRS of MSAS using the stationary phase method.

3 MSAS imaging algorithm

3.1 MSAS PTRS

The PTRS (Wu et al., 2011) should be obtained first to develop the MSAS imaging algorithm. Based on Equation 1, the second-order Taylor expansion is used, and the result is

$$\begin{aligned} R_i(t;r) &\approx \sqrt{r^2 + (vt)^2} + \sqrt{r^2 + \left(vt + d_i + v\tfrac{2r}{c}\right)^2} \\ &= 2r + \tfrac{v^2}{r}\left[t + \left(\tfrac{r}{c} + \tfrac{d_i}{2v}\right)\right]^2 + \tfrac{1}{4r}\left(v\tfrac{2r}{c} + d_i\right)^2. \end{aligned} \quad (3)$$

With the inverse Taylor expansion, Equation 3 can be further reformulated as

$$\begin{aligned} R_i(t;r) &\approx 2r + \tfrac{v^2}{r}\left[t + \left(\tfrac{r}{c} + \tfrac{d_i}{2v}\right)\right]^2 + \tfrac{1}{4r}\left(v\tfrac{2r}{c} + d_i\right)^2 \\ &\approx 2\sqrt{r^2 + (vt + v\tfrac{r}{c} + \tfrac{d_i}{2})^2} + \tfrac{(v\frac{2r}{c} + d_i)^2}{4r} \end{aligned} \quad (4)$$

The square root term in Equation 4 is similar to the slant range of the monostatic SAS system. The second term in Equation 4 denotes the error introduced by PCA and the stop-and-hop approximation. With the second term in Equation 4 compensated by the preprocessing, the following work focuses on the MSAS data using the traditional imaging algorithms of a monostatic SAS system.

The current work deduces the PTRS. Starting with Fourier transformation (FT) in the cross-track dimension for the echo signal (Equation 2),

$$Ss_i(f_\tau, t) = P(f_\tau)\exp\left\{-j2\pi(f_c + f_\tau)\frac{R_i(t;r)}{c}\right\} \tag{5}$$

In Equation 5, $P(f_\tau)$ is the spectrum of the transmitted chirp signal. The instantaneous frequency is denoted by f_τ. Performing an FT in the along-track dimension for Equation 5,

$$SS_i(f_\tau, f_t) = P(f_\tau) \cdot \int_{-T_s/2}^{T_s/2} \exp\left\{-j2\pi(f_c + f_\tau)\frac{R_i(t;r)}{c} - j2\pi f_t t\right\} dt \tag{6}$$

In Equation 6, the Doppler frequency is denoted by f_t, and the synthetic aperture time is represented by T_s. The exponential phase in Equation 6 is further denoted by

$$\varphi_i(f_\tau, f_t; r) = -2\pi(f_c + f_\tau)\frac{R_i(t;r)}{c} - 2\pi f_t t. \tag{7}$$

Performing the stationary phase method (Wang et al., 2009) on Equation 7 yields

$$\frac{\partial R_i(t_i;r)}{\partial t} + \frac{cf_t}{f_c + f_\tau} = 0. \tag{8}$$

In Equation 8, the stationary point is denoted by t_i. Based on Equation 4, Equation 8 is then reformulated as

$$\frac{2v\left(vt_i + v\frac{r}{c} + \frac{d_i}{2}\right)}{\sqrt{r^2 + \left(vt_i + v\frac{r}{c} + \frac{d_i}{2}\right)}} + \frac{cf_t}{f_c + f_\tau} = 0. \tag{9}$$

The solution to Equation 8 is expressed as

$$t_i = -\frac{crf_t}{v(f_c + f_\tau)} \cdot \frac{1}{\sqrt{4v^2 - \left[\frac{cf_t}{(f_c + f_\tau)}\right]^2}} - \frac{r}{c} - \frac{d_i}{2v} \tag{10}$$

Substituting Equation 10 into Equation 7, we obtain the PTRS phase given by

$$\begin{aligned}\varphi_i(f_\tau, f_t; r) &= -2\pi(f_c + f_\tau)\tfrac{R_i(t_i;r)}{c} - 2\pi f_t t_i \\ &= -2\pi(f_c + f_\tau)\frac{2\sqrt{r^2 + (vt_i + v\frac{r}{c} + \frac{d_i}{2})^2 + \frac{(v\frac{2r}{c} + d_i)^2}{4r}}}{c} - 2\pi f_t t_i \\ &= -4\pi\tfrac{f_c}{c}r\sqrt{\left(1 + \tfrac{f_\tau}{f_c}\right)^2 - \tfrac{c^2 f_t^2}{4v^2 f_c^2}} + \pi f_t\tfrac{d_i}{v} + 2\pi f_t\tfrac{r}{c} - \pi(f_c + f_\tau)\frac{\left(2\frac{v}{c}r + d_i\right)^2}{2rc}\end{aligned} \tag{11}$$

Inspection of Equation 11 reveals that the first term is similar to the PTRS of traditional monostatic SAS. The spatial sampling of each receiver causes the second term. The third term denotes the azimuthal shifting caused by the stop-and-hop approximation. The last term is the Doppler error and micro-range migration caused by the PCA and stop-and-hop errors. Based on the idea of the monostatic conversion-based method, the phase error related to the displaced distance, d_i, is the bistatic deformation term, which is

$$\Psi_i(f_\tau, f_t; r; d_i) = -\pi f_c\frac{\left(2\frac{v}{c}r + d_i\right)^2}{2rc} - \pi f_\tau\frac{\left(2\frac{v}{c}r + d_i\right)^2}{2rc} + \pi f_t\frac{d_i}{v} \tag{12}$$

The remaining phase denotes the quasi-monostatic term given by

$$\Phi(f_\tau, f_t; r) = -4\pi\frac{f_c}{c}r\sqrt{\left(1 + \frac{f_\tau}{f_c}\right)^2 - \frac{c^2 f_t^2}{4v^2 f_c^2}} + 2\pi f_t\frac{r}{c} \tag{13}$$

Subsequently, Equation 11 can be further reformulated as

$$\varphi_i(f_\tau, f_t; r) = \Phi(f_\tau, f_t; r) + \Psi_i(f_\tau, f_t; r; d_i) \tag{14}$$

It is apparent from Equation 14 that the MSAS imagery can be converted into monostatic SAS focusing after compensating for the bistatic deformation term, $\Psi_i(f_\tau, f_t; r; d_i)$.

3.2 MSAS imagery

The MSAS imagery includes two primary operations. The first operation compensates for the bistatic deformation term, $\Psi_i(f_\tau, f_t; r; d_i)$, and is called preprocessing. The second operation conducts the MSAS data focusing based on the quasi-monostatic term.

The compensation of the bistatic deformation term is now discussed. The first step is the compression in the cross-track dimension. The matched filtering function is written as

$$H_{\text{cross}} = P^*(f_\tau) \tag{15}$$

In Equation 15, the superscript asterisk denotes the complex conjugate operation.

The second step is the compensation of the Doppler phase, the first term in Equation 12, in the two-dimensional (2D) time domain. The filtering function is expressed as

$$\Gamma_i(\tau, d_i) = \exp\left\{\pi f_c\frac{\left(2\frac{v}{c}r + d_i\right)^2}{2rc}\right\} \tag{16}$$

The second step corrects the micro-range migration, the second term in Equation 12, which is caused by the PCA and stop-and-hop errors. From Equation 12, the micro-range migration in the 2D time domain corresponding to the second term in the 2D frequency domain is given by

$$\Delta_i(\tau, d_i) = \frac{\left(2\frac{v}{c}r + d_i\right)^2}{4rc} \tag{17}$$

The term shown in Equation 17 is range-variant. Interpolation is exploited to correct the micro-range migration.

The third step of the preprocessing operation compensates for the spatial sampling error, the last term in Equation 12. This can be achieved by a complex multiplication in the cross-track time, along-track Doppler domain. The filtering function is expressed as

$$\Lambda_i(\tau, d_i) = \exp\left\{-\pi f_t \frac{d_i}{v}\right\} \tag{18}$$

The last step is the recovery of the chirp characteristic. Filtering is conducted in the cross-track frequency domain, and the filtering function directly uses the spectrum of the transmitted signal.

A block diagram of the preprocessing operation is depicted in Figure 2 based on the steps discussed above.

After conducting the preprocessing operations, the remaining work involves MSAS focusing using the RD algorithm based on the quasi-monostatic term in Equation 13. The use of the RD algorithm is enabled by approximating the first term in Equation 13 by the second-order Taylor expansion. After the series approximation, Equation 13 is reformulated as

$$\Phi(f_\tau, f_t; r) \approx -\frac{4\pi r\beta}{\lambda} - \frac{4\pi r}{c\beta} f_\tau + 2\pi \frac{r\lambda}{c^2}\left(\frac{1}{\beta^3} - \frac{1}{\beta}\right) f_\tau^2 + 2\pi f_t \frac{r}{c} \tag{19}$$

Where

$$\beta = \sqrt{1 - \frac{c^2 f_t^2}{4v^2 f_c^2}} \tag{20}$$

The first term in Equation 19 denotes the along-track modulation, the second term is the range migration, and the third term denotes the cross-track modulation. The last term is the along-track shifting.

The first step in the RD algorithm requires performing the cross-track compression in the 2D frequency domain based on the third term in Equation 19. Because the third term in Equation 19 is slightly range-variant, this term can be compensated using the sub-block processing method (Zhang et al., 2019). For the nth sub-block, the filtering function is expressed as

$$H_n(f_\tau, f_t; r_n) = \exp\left\{j\pi \frac{f_\tau^2}{\gamma_e(f_t; r_n)}\right\} \tag{21}$$

Where

$$\gamma_e(f_t; r_n) = \frac{1}{\gamma} - 2\lambda r_n \frac{1 - \beta^2}{c^2\beta^3} \tag{22}$$

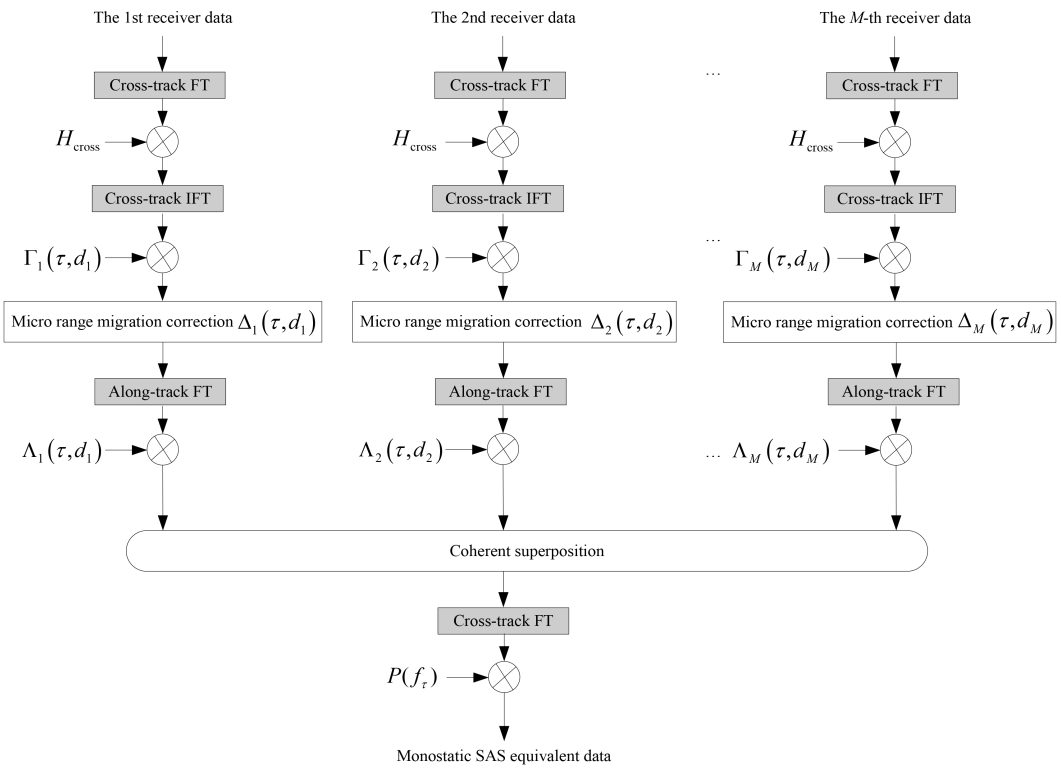


FIGURE 2
Block diagram of multiple-receiver synthetic aperture sonar (MSAS) preprocessing.

In Equation 21, the subscript n ($n\in[1,N]$) is the sub-block index and N is the total number of sub-blocks. In Equation 22, γ is the chirp rate of the transmitted signal and r_n is the center range of the n^{th} sub-block. In practice, Equation 21 is also called the secondary range compression (SRC), which is range-variant. To some degree, the phase error within each sub-block width should not exceed $\pi/16$, and this can be expressed as $|\gamma_e(f_t;r_n)-\gamma_e(f_t;r_n\pm 0.5\Delta r)|<\pi/16$, where Δr is the sub-block width. The sub-block width can be easily determined with this method. In general, the SRC depends on the range, to some extent. The data can now be coarsely segmented into several sub-blocks in the cross-track dimension.

The second term in Equation 19 denotes the range migration, which can be corrected by interpolation in the range-Doppler domain. Based on the second term in Equation 19, the range migration in the cross-track time, along-track Doppler domain is denoted by

$$\Delta R(f_t;r)=\frac{2r}{\sqrt{1-\left(\frac{\lambda f_t}{2v}\right)^2}}-2r \tag{23}$$

After this step, the coupling between the cross-track and along-track dimensions can be decoupled. Subsequently, the along-track compression can be performed based on the first term in Equation 19. The matched filtering function in the along-track dimension is denoted by

$$H_{\text{along}}(f_t;r)=\exp\left\{j4\pi\frac{\beta}{\lambda}r\right\} \tag{24}$$

The last step involves correcting the along-track shifting based on the last term in Equation 19. The filtering function is given by

$$H_{\text{shifting}}(f_t;r)=\exp\left\{-j2\pi f_t\frac{r}{c}\right\} \tag{25}$$

Correction of the along-track shifting can also be carried out with the along-track compression. A block diagram based on the RD algorithm discussed above is shown in Figure 3.

4 Simulation results

The simulations presented in this section validate the proposed method. The MSAS parameters are listed in Table 1, which indicates that there are 32 receivers.

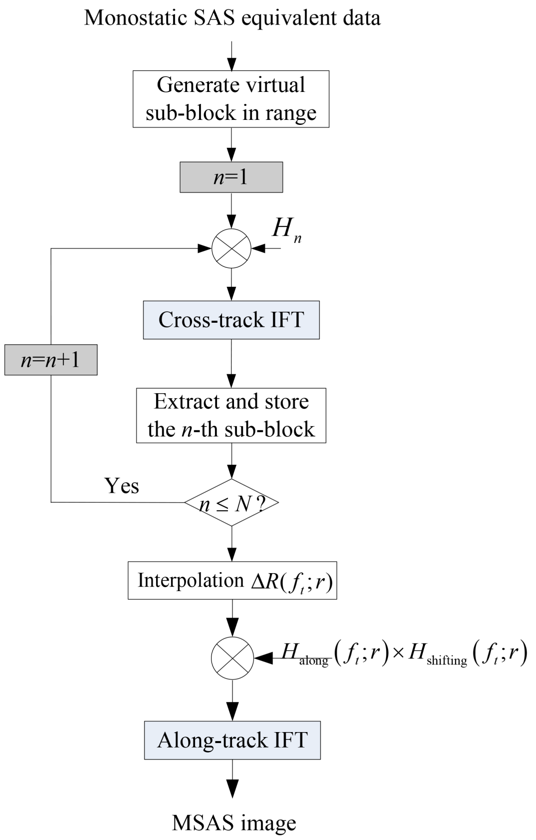


FIGURE 3
Block diagram of the multiple-receiver synthetic aperture sonar (MSAS) range-Doppler (RD) algorithm.

TABLE 1 Synthetic aperture sonar (SAS) parameters.

Parameter	Value	Unit
Platform velocity	2	m/s
Pulse repetition interval	0.32	s
Signal bandwidth	20	kHz
Carrier frequency	150	kHz
Receiver array length	1.28	m
Receiver width	0.04	m
Transmitter width	0.08	m

There are six ideal point targets in the imaging scenario depicted in Figure 4. These targets are numbered T1, T2, …, and T6 to simplify the description.

The back-projection (BP) algorithm (Zhang et al., 2014b; Wang et al., 2015) can provide high-performance results based on the interpolation theory and is therefore used as the benchmark. After focusing the simulated MSAS data, the imaging results from the proposed method are shown in Figures 5A, B shows the results from the BP algorithm. Comparing the imaging results of the proposed method to those of the BP algorithm (Zhang et al., 2014b), it can be seen that the performance of the proposed method is very similar to that of the BP algorithm, indicating that the proposed method can accurately reconstruct the targets.

The along-track slices are shown in Figure 6. It is readily apparent that the along-track slices of the present method agree well with the slices from the BP algorithm. These results further confirm our conclusion that the present method can accurately recover the targets.

Figure 7 shows the along-track resolution and peak side lobe ratio (PSLR) of the focused targets. These results show that the proposed method can obtain high-resolution imaging results consistent with those of the BP algorithm.

The PSLR, integrated side lobe ratio (ISLR), and the along-track resolution (AR) were calculated to further evaluate the imaging performance of the proposed method. The quality parameters of the proposed method are listed in Table 2, indicating that the difference in the PSLRs between the proposed method and the BP algorithm was about 0.19 dB. The difference in the ISLRs between the proposed method and the BP algorithm was about 0.08 dB. The degradation in the quality parameters was very modest, and the performance loss can be neglected. Both methods obtained essentially the same along-track resolution. Therefore, it can be concluded that the proposed method can focus the targets well. Furthermore, the proposed method can obtain high-resolution imaging results similar to those of the BP algorithm.

5 Processing results for real data

In this section, real data were used to verify the proposed method. The real data included 800 sampling points in the range dimension. Because the BP algorithm is considered precise, the BP results were again used as the benchmark. Figure 8A depicts the results of the proposed method, while Figure 8B shows the results of the BP algorithm. It can be observed that the proposed method can obtain nearly the same results as the BP algorithm.

Table 3 lists the processing times of the proposed method and the BP algorithm, which showed that the BP algorithm is extremely time-consuming. However, the proposed method can significantly improve the imaging efficiency by a factor of 62 in terms of the processing speed. Consequently, it is concluded that the proposed method can accurately focus the targets without loss of imaging performance. In addition, the imaging efficiency can be improved using the proposed method.

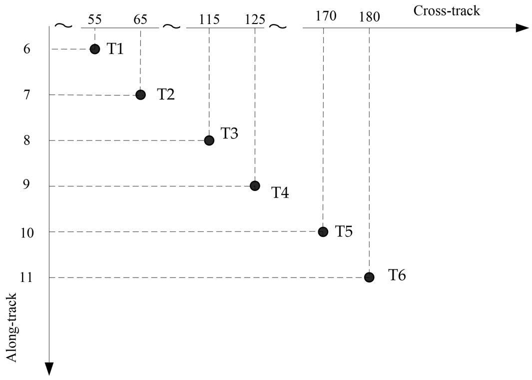


FIGURE 4
Point target distribution in the imaging scenario.

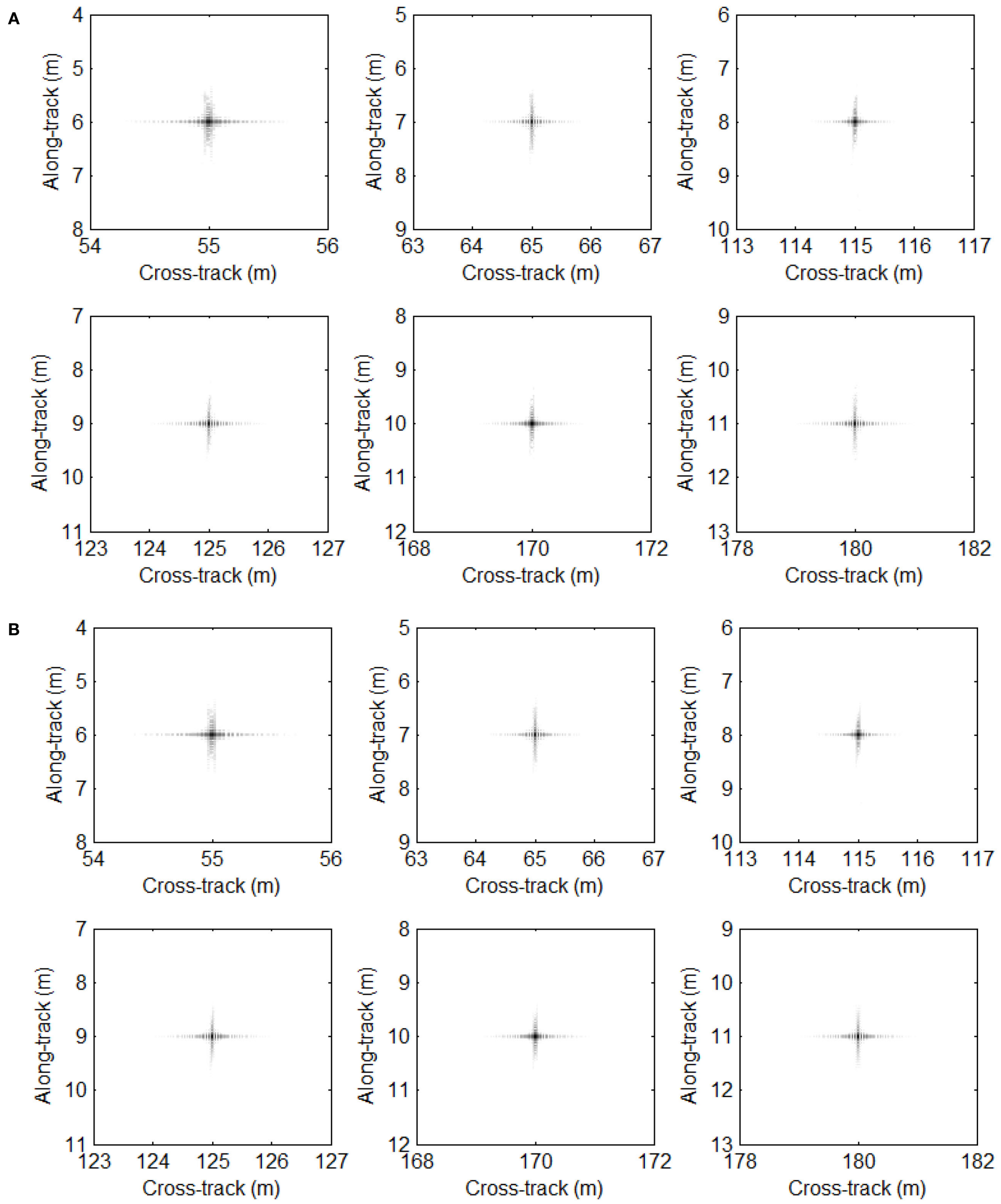


FIGURE 5
Imaging results using simulated data. **(A)** Proposed method. **(B)** Back-projection (BP) algorithm.

6 Conclusions

MSAS imagery is challenging because the two-round slant range is characterized by double square root terms. The method proposed in this paper transforms the two-round slant range into a single square root term based on the second-order Taylor expansion and inverse Taylor expansion to solve this problem. Subsequently, the PTRS can be decomposed into quasi-monostatic and bistatic deformation terms. The bistatic deformation term can be

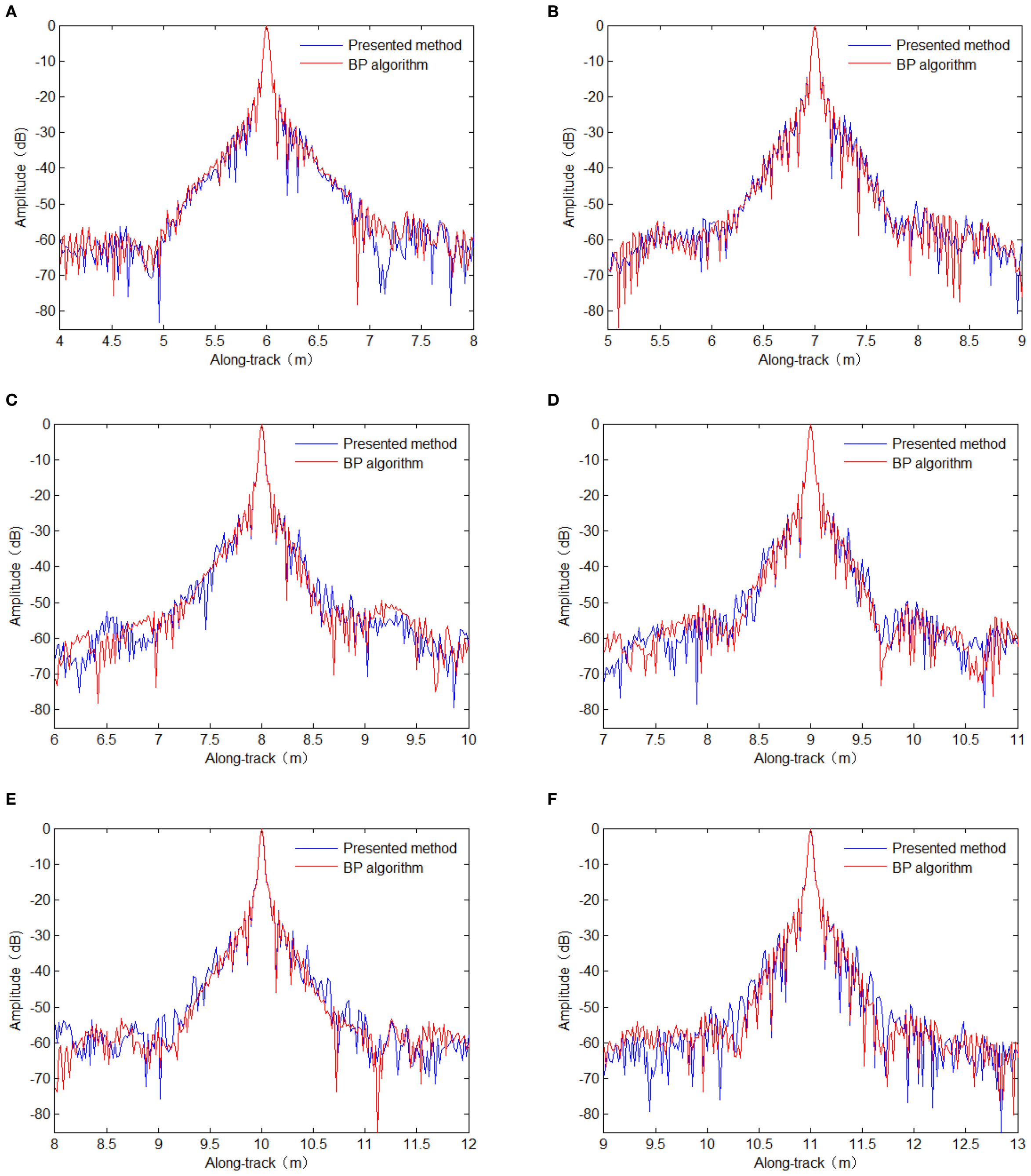


FIGURE 6
Along-track slices of the focused targets. **(A)** T1. **(B)** T2. **(C)** T3. **(D)** T4. **(E)** T5. **(F)** T6. *BP*, back-projection.

canceled with the preprocessing step based on complex multiplication, interpolation, coherent superposition, and dechirping operations. Consequently, the MSAS data can be considered a monostatic SAS-equivalent signal, which is used as the input to an RD algorithm. Considering the cross-track variance of the SRC, this paper exploits the sub-block processing method in the cross-track dimension. The proposed method can generate high-resolution results very

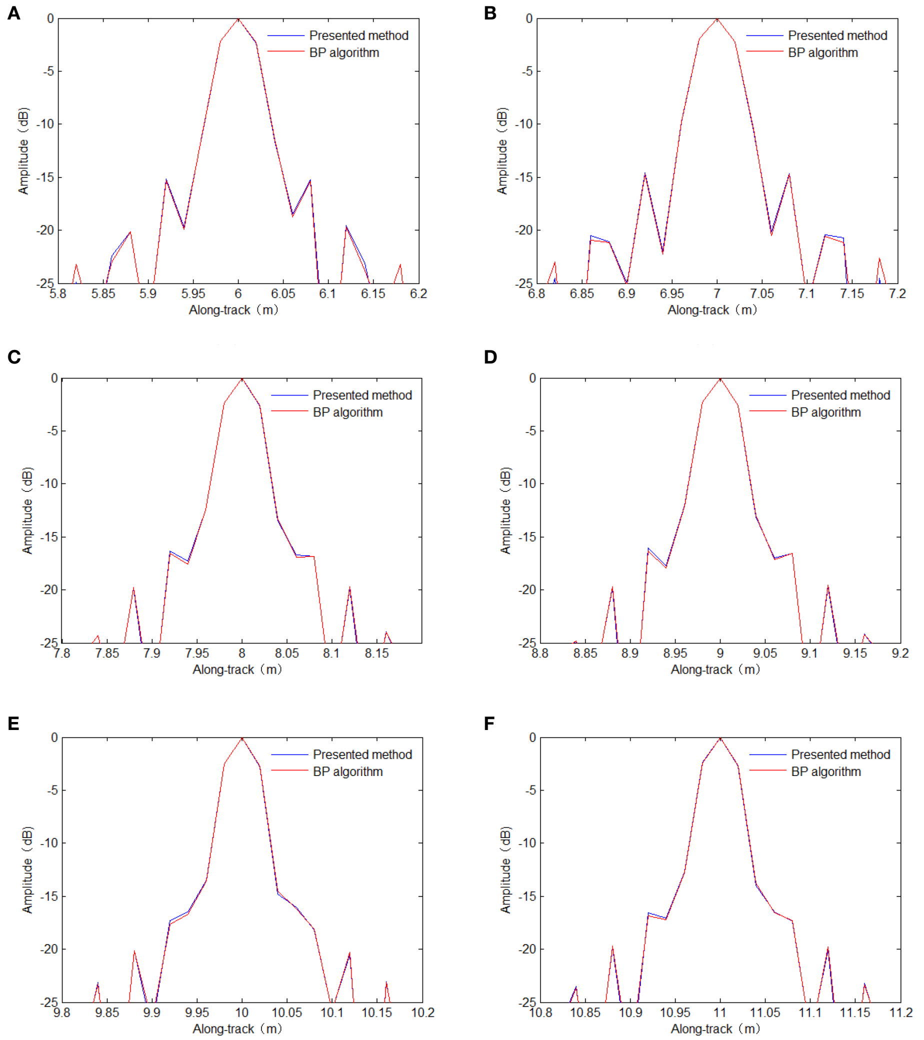


FIGURE 7
Along-track resolution. **(A)** T1. **(B)** T2. **(C)** T3. **(D)** T4. **(E)** T5. **(F)** T6. *BP*, back-projection.

TABLE 2 Quality parameters of the focused targets.

Method	PSLR (dB)	ISLR (dB)	AR (m)	Target
Proposed method	−14.44	−8.61	0.04	T1
BP	−14.57	−8.69	0.04	
Proposed method	−14.34	−8.09	0.04	T2
BP	−14.47	−8.11	0.04	
Proposed method	−14.64	−9.51	0.04	T3
BP	−14.77	−9.56	0.04	
Proposed method	−14.63	−9.36	0.04	T4
BP	−14.74	−9.41	0.04	
Proposed method	−14.65	−9.95	0.04	T5
BP	−14.84	−10.00	0.04	
Proposed method	−14.63	−9.65	0.04	T6
BP	−14.76	−9.69	0.04	

PSLR, peak side lobe ratio; ISLR, integrated side lobe ratio; AR, along-track resolution; BP, back-projection.

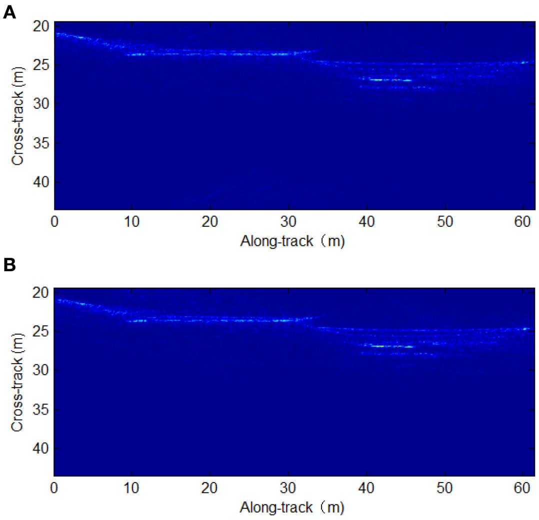


FIGURE 8
Processing results for real data. **(A)** Proposed method. **(B)** Back-projection (BP) algorithm.

TABLE 3 Processing times of the imaging algorithms.

	Proposed method	BP algorithm
Processing time (s)	28	1,761

BP, back-projection

similar to those of the BP algorithm benchmark in a significantly shorter processing time. Simulations using a simulated imaging scenario and real sonar data further validated the proposed method.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

PH and PY designed the study, wrote the manuscript, managed communication between the authors, and analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Shandong Province (grant no. ZR2019MA022) and the Enterprise Marine Project (grant no. 20210701).

Acknowledgments

The authors thank the reviewers and editors for their valuable comments that helped improve this paper.

Conflict of interest

PY is employed by the Acoustic Signal and Electronics Science and Technology Corporation.

The remaining author declares 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, P. J., Hayes, M. P., and Gough, P. T. (2005). "ML estimation of seafloor topography using multi-frequency synthetic aperture sonar," in *Europe Oceans*. (Brest, France: IEEE), vol. 1, 579–584. doi: 10.1109/OCEANSE.2005.1511779

Cong, W., Huo, G., and Zheng, Y. (2000). "Anti-reverberation target detection and identification in synthetic aperture sonar," in *Proceedings of the 2000 international symposium on underwater technology (Cat. No.00EX418)*. (Tokyo, Japan: IEEE), 54–61. doi: 10.1109/UT.2000.852515

Cutrona, L. J. (1975). Comparison of sonar system performance achievable using synthetic aperture techniques with the performance achievable by more conventional means. *J. Acoust. Soc Am.* 58, 336–348. doi: 10.1121/1.380678

Davis, B. J., Gough, P. T., and Hunt, B. R. (2009). Modeling surface multipath effects in synthetic aperture sonar. *IEEE J. Ocean. Eng.* 34, 239–249. doi: 10.1109/JOE.2009.2017796

Dillon, J., and Steele, S. (2022). "Square SAS: multi-aspect imaging with a towed synthetic aperture sonar," in *Oceans*. (Chennai, India: IEEE), 2022 1–6. doi: 10.1109/OCEANSChennai45887.2022.9775390

Dobeck, G. J. (1999). "Fusing sonar images for mine detection and classification," in *Detection and remediation technologies for mines and minelike targets IV*. (Orlando, FL: SPIE), 602–614. doi: 10.1117/12.357082

Gough, P. T., and Hawkins, D. W. (1996). "On the use of mapping operators in SAR and SAS," in *Conference record of the thirtieth asilomar conference on signals, systems and computers*. (Pacific Grove, CA: IEEE) vol. 81, 86–89. doi: 10.1109/ACSSC.1996.600834

Gough, P. T., and Hawkins, D. W. (1997). Imaging algorithms for a strip-map synthetic aperture sonar: minimizing the effects of aperture errors and aperture undersampling. *IEEE J. Ocean. Eng.* 22, 27–39. doi: 10.1109/48.557537

Gough, P. T., and Hawkins, D. W. (1998). "A short history of synthetic aperture sonar. IGARSS '98. sensing and managing the environment," in *1998 IEEE international geoscience and remote sensing. symposium proceedings. (Cat. No.98CH36174)*. (Seattle, WA, USA: IEEE), vol. 2, 618–620. doi: 10.1109/IGARSS.1998.699529

Gough, P. T., and Hayes, M. P. (2005). "Fast Fourier techniques for SAS imagery," in *MTS/IEEE oceans conference*. (Brest, France: IEEE), 563–568. doi: 10.1109/OCEANSE.2005.1511776

Gough, P. T., Hayes, M. P., and Wilkinson, D. R. (2000). "An efficient image reconstruction algorithm for a multiple hydrophone array synthetic aperture sonar," in *Proceedings of the 5th European conference on underwater acoustics (ECUA2000)* (Lyon: Luxembourg: Office for Official Publications of the European Communities), 395–400.

Gough, P. T., and Miller, M. A. (2003). "The SAR map-drift algorithm extended for a multi-hydrophone SAS," in *Oceans 2003: Celebrating the Past. Teaming toward the future (IEEE cat. No.03CH37492)*. (San Diego, CA, USA: IEEE), vol. 5, 2427–2432. doi: 10.1109/OCEANS.2003.178292

Gough, P. T., Noonchester, M. A., Hunter, A. J., and Hayes, M. P. (2006). "Towards multi-frequency imaging and analysis of sub-surface targets using SAS," in *Oceans 2006*. (Boston, MA, USA: IEEE). doi: 10.1109/OCEANS.2006.306797

Hunter, A. (2006). *Underwater acoustic modeling for synthetic aperture sonar* (Christchurch, New Zealand: University of Canterbury).

LeHardy, P. K., and Larsen, J. (2015). "Deepwater synthetic aperture sonar and the search for MH370," in *MTS/IEEE oceans conference*. (Washington, DC, USA: IEEE), 1–4. doi: 10.23919/OCEANS.2015.7401838

Liu, B., Wang, T., Wu, Q., and Bao, Z. (2009). Bistatic SAR data focusing using an omega-K algorithm based on method of series reversion. *IEEE Trans. Geosci. Remote Sens.* 47, 2899–2912. doi: 10.1109/TGRS.2009.2017522

Loffeld, O., Nies, H., Peters, V., and Knedlik, S. (2004). Models and useful relations for bistatic SAR processing. *IEEE Trans. Geosci. Remote Sens.* 42, 2031–2038. doi: 10.1109/TGRS.2004.835295

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, 14–21. doi: 10.1109/TGRS.2007.909090

Rankin, W., Cooper, M. T., Wood, J. L., and Lathrop, J. D. (1999). "Demonstration of advanced underwater sensors for military and civilian applications," in *Detection and remediation technologies for mines and minelike targets IV* (Orlando, FL: SPIE), 647–660. doi: 10.1117/12.357047

Rolt, K. D., and Schmidt, H. (1992). Azimuthal ambiguities in synthetic aperture sonar and synthetic aperture radar imagery. *IEEE J. Ocean. Eng.* 17, 73–79. doi: 10.1109/48.126956

Sæbø, T. O., Callow, H. J., and Hagen, P. E. (2010). "Pipeline inspection with synthetic aperture sonar," in *The 33th Scandinavian symposium on physical acoustics* (Geilo, Norway), 1–6. Available at: https://www.yumpu.com/en/document/read/22516981/pipeline-inspection-with-synthetic-aperture-sonar.

Shin, H.-S., and Lim, J.-T. (2009). Omega-k algorithm for airborne spatial invariant bistatic spotlight SAR imaging. *IEEE Trans. Geosci. Remote Sens.* 47, 238–250. doi: 10.1109/TGRS.2008.2002954

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

Wang, R., Deng, Y., Loffeld, O., Nies, H., Walterscheid, I., Espeter, T., et al. (2011). Processing the azimuth-variant bistatic SAR data by using monostatic imaging algorithms based on two-dimensional principle of stationary phase. *IEEE Trans. Geosci. Remote Sens.* 49, 3504–3520. doi: 10.1109/TGRS.2011.2129573

Wang, R., Loffeld, O., Nies, H., Knedlik, S., and Ender, J. H. (2008). Chirp-scaling algorithm for bistatic SAR data in the constant-offset configuration. *IEEE Trans. Geosci. Remote Sens.* 47, 952–964. doi: 10.1109/TGRS.2008.2006275

Wang, R., Loffeld, O., Nies, H., Medrano-Ortiz, A., and Knedlik, S. (2009). Bistatic point target reference spectrum in the presence of trajectory deviations. *IET Radar Sonar Navig.* 3, 177–185. doi: 10.1049/iet-rsn:20080053

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 (IIICEC)*. (Xi'an, China: Atlantis Press), 1610–1615. doi: 10.2991/iiicec-15.2015.355

Wu, J. J., Yang, J. Y., Huang, Y. L., Liu, Z., and Yang, H. G. (2011). A new look at the point target reference spectrum for bistatic SAR. *Prog. Electromagn. Res.* 119, 363–379. doi: 10.2528/PIER11050704

Wu, Q., Zhang, Y., Amin, M., and Himed, B. (2014). Focusing of tandem bistatic SAR data using range Doppler algorithm. *IEEE Radar Conference Cincinnati*. (Cincinnati, OH, USA: IEEE) 927–931. doi: 10.1109/RADAR.2014.6875724

Xia, X., Zhang, X., and Chen, X. (2016). "Parameter estimation for Gaussian mixture processes based on expectation-maximization method," in *2016 4th international conference on machinery, materials and information technology applications (ICMMITA 2016)*. (Xi'an, China: Atlantis Press), 519–523. doi: 10.2991/icmmita-16.2016.96

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), 109–113. doi: 10.1109/ICCSN55126.2022.9817582

Zhang, X., Chen, X., and Qu, W. (2017a). "Influence of the stop-and-hop assumption on synthetic aperture sonar imagery," in *2017 IEEE 17th international conference on communication technology (ICCT 2017)*. (Chengdu, China: IEEE), 1601–1607. doi: 10.1109/ICCT.2017.8359901

Zhang, X., Dai, X., and Fang, B. (2019). A range-Doppler imaging method for the multireceiver synthetic aperture sonar. *Geomat. Inf. Sci. Wuhan Univ.* 44, 1667–1673. doi: 10.13203/j.whugis20180076

Zhang, X., Dai, X., and Yang, B. (2018). Fast imaging algorithm for the multiple receiver synthetic aperture sonars. *IET Radar Sonar Navig.* 12, 1276–1284. doi: 10.1049/iet-rsn.2018.5040

Zhang, X., Huang, H., Ying, W., Wang, H., and Xiao, J. (2017b). An indirect range-Doppler algorithm for multireceiver synthetic aperture sonar based on Lagrange inversion theorem. *IEEE Trans. Geosci. Remote Sens.* 55, 3572–3587. doi: 10.1109/TGRS.2017.2676339

Zhang, X., Liu, Y., and Deng, X. (2021a). "Influence of phase centre approximation error on SAS imagery," in *2021 IEEE the 6th international conference on computer and communication systems (ICCCS 2021)* (Chengdu, China: IEEE), 352–356. doi: 10.1109/ICCCS52626.2021.9449222

Zhang, X., Tang, J., and Ouyang, J. (2014a). "Imaging processor for multi-receiver SAS in the presence of partially failed receivers," in *2014 international conference on vehicle & mechanical engineering and information technology (VMEIT 2014)*. (Beijing, China: TTP), 2225–2228. doi: 10.4028/www.scientific.net/AMM.543-547.2225

Zhang, X., Tang, J., Wang, F., Bai, S., and Liu, D. (2014b). Accurate back projection imaging algorithm for multi-receiver SAS in engineering application. *J. Nav. Univ. Eng.* 26, 20–24. doi: 10.7495/j.issn.1009-3486.2014.02.005

Zhang, X., Tang, J., Zhang, S., and Zhong, H. (2013). Chirp-scaling imaging algorithm for multi-receiver synthetic aperture sonar. *Syst. Eng. Electron.* 35, 1415–1420. doi: 10.3969/j.issn.1001-506X.2013.07.11

Zhang, X., Wu, H., Sun, H., and allying, W. (2021b). 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 Ying, W. (2022). Influence of the element beam pattern on synthetic aperture sonar imaging. *Geomat. Inf. Sci. Wuhan Univ.* 47, 133–140. doi: 10.13203/j.whugis20190148

Zhang, X., Ying, W., Liu, Y., and Deng, X. (2021c). "Processing multireceiver SAS data based on the PTRS linearization," in *IEEE International geoscience and remote sensing symposium (IGARSS 2021)*. (Brussels, Belgium: IEEE), 5167–5170. doi: 10.1109/IGARSS47720.2021.9553688

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of impulsive noise with class b model. *IET Radar Sonar Navig.* 14, 1055–1060. doi: 10.1049/iet-rsn.2019.0477

Zhong, H., and Liu, X. (2006). "A fourth-order imaging algorithm for spaceborne bistatic SAR," in *2006 IEEE international geoscience and remote sensing symposium (IGARSS)*. (Denver, CO, USA: IEEE), 1196–1199. doi: 10.1109/IGARSS.2006.309

Zhong, H., and Liu, X. (2009). An extended nonlinear chirp-scaling algorithm for focusing large-baseline azimuth-invariant bistatic SAR data. *IEEE Geosci. Remote Sens. Lett.* 6, 548–552. doi: 10.1109/LGRS.2009.2021676

Zhong, H., and Liu, X. (2010). An effective focusing approach for azimuth invariant bistatic SAR processing. *Signal Process* 90, 395–404. doi: 10.1016/j.sigpro.2009.07.005

Frontiers in Marine Science

TYPE Methods
PUBLISHED 16 January 2023
DOI 10.3389/fmars.2022.1097003

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Yan Yong,
Bengbu Officer College, China
Dong Huixu,
Air Force Aviation University, China
Pan Huang,
Weifang University, China

*CORRESPONDENCE
Dawei Li
lee_dw01@163.com

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 13 November 2022
ACCEPTED 28 December 2022
PUBLISHED 16 January 2023

CITATION
Li DW, Wu MH, Yu L, Han JH and Zhang H
(2023) Single-channel blind source
separation of underwater acoustic signals
using improved NMF and FastICA.
Front. Mar. Sci. 9:1097003.
doi: 10.3389/fmars.2022.1097003

Single-channel blind source separation of underwater acoustic signals using improved NMF and FastICA

Dawei Li*, Minghui Wu, Liang Yu, Jianhui Han and Hao Zhang

Aviation Operations Service Academy, Aviation University, Yantai, Shandong, China

When automatic monitoring buoys receive mixed acoustic signals from multiple underwater acoustic targets, the statistical blind source separation (BSS) task is used to separate the signals and identify vessel features, which is overly complex and needs improvement, especially noting that noise cancellation and stealth technologies are advancing rapidly. To fill this gap in capability, an improved non-negative matrix factorization (NMF) based BSS algorithm is built on a FastICA machine learning backbone. With this tool, the spatial and spectral correlation of underwater acoustic signals is introduced into the NMF algorithm improved by to resolve the non-convex and feature correlation problems commonly encountered by contemporary NMF algorithms. Moreover, the improved modulation feature adaptability of the NMF increases the local expressivity and independence of the decomposed base matrix, which is proven to meet the requirements of FastICA and used to improve the BSS effect of the FastICA. Simulated and empirical results show that compared with state-of-the-art FastICA and NMF based BSS algorithms, our novel approach obtains better signal-to-noise reduction and separation accuracy while maintaining superior target signal recognition features.

KEYWORDS

single-channel multi-target underwater acoustic signal, multi-target signal separation, improved NMF, FastICA, local expression and independent characteristics

Introduction

As a floating automatic monitoring platform for monitoring the comprehensive marine environment, such as marine meteorological, hydrological and ecological parameters, ocean observation sonar buoy plays an important role in the monitoring and research of marine environment, and is the basis for caring about the ocean, understanding the ocean and managing the ocean. Maintaining and improving this capability is crucial to the protection of maritime and littoral waterways. With the development of marine resources and the strengthening of maritime trade, the identification and judgment of marine and underwater targets, based on underwater acoustic signals received by buoys, has gradually become an important direction to analyze the impact of the development of modern ship technology and the prosperity of maritime trade on marine ecology.

Underwater vessels (i.e., submarines) and surface vessels create and transmit acoustic signals with their screws (sometimes referred to as "propellers" by folks outside the field), which are transmitted over long distances through the water medium. So, when multi-target signals are received by single buoys, separating and processing them requires statistical blind source separation (BSS) techniques (Dianmant et al., 2019; Zhang et al., 2020) prior to any recognition, analysis, and localization efforts (Zhang and Yang, 2021). The number of signals, source and its proportion of each underwater acoustic target signal in the mixed signal received by a single passive sonar buoy is unknown, Hence, a single-channel statistical BSS framework is used (Wildeboer et al., 2020).

Although nonlinear signals are most often encountered in real maritime scenarios, it is incredibly difficult to deconstruct and process them, even with machine learning tools. Therefore, linear models are widely applied based on mathematical approximations (Huang et al., 2019). Notably, the powerful non-negative matrix factorization (NMF) method, which was the hotspot in the linear model, ingests a data structure that fits the linear mixed model reasonably well (Rathnayake et al., 2020). Thus, dependence on prior knowledge is reduced and the dimension of high-dimensional massive data is reduced (Zhang et al., 2022).. However, owing to the non-convexity of the NMF objective function, the globally optimal solution is difficult to guarantee. Therefore, various constraints must be imposed according to the signal characteristics to narrow the solution space. The remainder of the introduction leverages the extant literature to explain why NMF-type models are the only viable solutions to mixed-signal BSS tasks. Subsequently, Section 2 provides additional information from the literature to provide the current NMF machine learning advancements that illuminate the present capability gap while also explaining algorithmic methods.

A neighborhood spatial information constraint was added into NMF to improve its convergence and classification in Ref (Lu et al., 2013)., by analyzing the manifold structural features of the signal image; however, the sparse constraint of the method's norm was vulnerable to noise and overly sensitive to the initial value. Hence, bias was unavoidable. Li (Li and Wang, 2019) applied a quadratic sparse constraint to the NMF based on manifold attributes to make better use of the sparse features of the NMF's coefficient matrix. Wang (Wang et al., 2020) and Lu (Lu et al., 2020) used similar adjacent pixels to extract spatial structure information by clustering time-frequency images. The spatial similarity, strengthened by cluster labels, was then used to optimize the constraint of the NMF's objective function, and the BSS efficacy was improved (Zhang et al., 2021). Similarly, the adaptive spectral local neighborhood information of pixels was extracted by Chen (Chen and Lv, 2021) and added to the NMF in the form of adaptive weights, and the blind unmixing performance was improved.

Advancements outside the maritime domain have contributed to the current operational capability. Wang et al. (2019) highlighted the local features and achieved the successful separation of multiple coupled fault signals by combining the variational mode decomposition optimized by energy convergence with local NMF whose optimal base matrix dimensionality was calculated by using the adjacent feature dominance method A deep-transduction NMF was later developed by Li (Yurong et al., 2020) to separate the mixed speech signals from multiple speakers, and Sriharsha (Sriharsha and Abhijit, 2018). integrated the homotopy optimisation with perturbation and ensemble and denoising autoencoder into an NMF for single-channel audio source separation. The objective speech quality was significantly improved (Zhang et al., 2022), and the techniques were useful for underwater target signal separation. Huang (Huang et al., 2019) fused the L_{21} and $L_{1/2}$ norms and added the composite to the NMF model, improving the model's anti-noise performance against band- and pixel-level noise.

From the examples annotated above, it is clear that NMF methods are essential to BSS applications in the fields of hyperspectral unmixing, composite fault signal separation, and multi-speaker voice separation (Huang et al., 2019), and it is clear that the NMF model is essential to the BSS of underwater acoustic multi-target signals. Hence, this article reports on the joint BSS algorithm based on our improved NMF and FastICA backbone. The main improvement of the proposed algorithm is to improve the NMF based on the features of underwater acoustic target signal, and then, the NMF base matrix is used as signal input of FastICA, realizing BSS. Simulated and empirical results show that, compared with state-of-the-art NMF and FastICA BSS models, our new method obtains better signal noise reduction and separation accuracy while maintaining superior target recognition signal characteristics. Hence, the verification of underwater targets and their statuses is improved for maritime targets detection and location.

Current application of NMF in multi-target underwater acoustic signals

NMF achieves part-to-whole representations using non-negative constraints on decomposed objects (Weiderer et al., 2020). For a non-negative data matrix, $V = [v_1, v_2, \cdots, v_n] \in R_+^{m\times n}$, where v_i is a m-dimensional vector, the NMF algorithm decomposes it into two non-negative matrices, $W = [w_1, w_2, \cdots, w_r] \in R_+^{m\times r}$ and $H = [h_1, h_2, \cdots, h_n] \in R_+^{r\times n}$, *via* cyclic iterations, according to Eq. (1):

$$V_{m\times n} \approx W_{m\times r} H_{r\times n} \tag{1}$$

Where W and H are the base and coefficient matrices, respectively, r is the dimensionality of the base vector (usually $r \ll n$). Obviously, the NMF method can obtain data dimension reduction at the same time. To measure the reconstruction effects of Eq. (1), many optimization improvements to the objective function have been proposed (Yurong et al., 2020), such as the Kullback–Leibler (KL) divergence:

$$KL(V \mid WH) = \sum_{i,j} V_{i,j} log \frac{V_{i,j}}{(WH)_{ij}} - V_{i,j} + (WH)_{ij} \tag{2}$$

The iterative solution of Eq. (2) can be found by using the multiplication criterion (Hien and Gillis, 2021).

When performing underwater acoustic multi-target BSS with the classical NMF method, the low-frequency analysis and range (LOFAR) spectrum, $V \in R_+^{F\times N}$, of the mixed signal is calculated first, where N indicate the frame numbers of received mixed signal and F is the FFT dimensions of signal frames. After estimating the optimal number r of base vectors, the base matrix, $W = \{\sum_i W_i \in R_+^{F\times r}\}$, can be calculated using Eq. (2), where W_i is the base matrix of the target i, and r reflects the description performance of the base vector to the original signal. And then, if the basis matrix W_i of each source signal is separated by using some BSS algorithm, such as FastICA, from the above W, the source signal spectrum will be constructed using the W_i and its corresponding coefficient matrix. In this way, the final separation of the mixed signals received by signal buoy is achieved. This process is shown in Figure 1, in which, the Feature Dictionary Set W is the whole base matrix calculated from the received mixed signals and W_I and e is the base matrix of target $I$and target e

Obviously, the classical NMF algorithm is suitable for underwater acoustic multi-target signal decomposition; however, the algorithm itself is non-convex, often causing it to fall into local optimum (Li and Yang, 2016). Moreover, the characteristics of the underwater acoustic target signals also affect the decomposition effect. First, the base matrix obtained by NMF comprises an unordered array of base vectors, which makes it difficult to reconstruct the source signal directly. Second, the multi-source components cause a lot of interference, which quickly leads to data redundancy. Therefore, the performance of the NMF BSS algorithm needs to be improved by deeply mining the relevant features contained in the underwater acoustic target signals and establishing suitable constraints based on the actual signal features. To answer this need, our joint BSS algorithm for underwater acoustic multi-target signals based on an improved NMF with a FastICA backbone is proposed.

Improving NMF using signal features

Optimize the NMF base vector using signal features

Improving the adaptability of objective function to signal feature will significantly improve NMF decomposition performance (Li and Yang, 2016). Thus, when extracting the base vector of the received underwater acoustic multi-target signals, if the feature correlation between source signals is low, the local independence of the base vector improves, as does the efficacy of the NMF algorithm (Li et al., 2016; Zhang et al., 2022). However, underwater acoustic target signals show significant feature diversity within the same class and feature correlations between different classes. Notably, the line spectrum feature of underwater acoustic signals is an important characteristic for target recognition, leading to the limited use of that the traditional KL divergence, F-norm, Lp-norm, and other target functions that is easily influenced by high-energy line spectrum components within the LOFAR spectrum (Sadeghi et al., 2021). Therefore, the proposed algorithm improves the NMF objective function with dual constraints, that is determinant constraints and У β-divergence, to reduce the method's dependence on the source signal structure. Hence, the uniqueness and independence of the base matrix are reinforced.

The formula of the β-divergence is

$$d_\beta(y,x) = \begin{cases} \frac{y^\beta}{\beta(\beta-1)} + \frac{x^\beta}{\beta} - \frac{yx^{\beta-1}}{(\beta-1)}, & \beta \in R/\{0,1\} \\ y\ln\frac{y}{x} - y + x, & \beta = 1 \\ \frac{y}{x} - \ln\frac{y}{x} - 1, & \beta = 0 \end{cases} \tag{3}$$

Where $R/\{0,1\}$ is the set of real numbers, excluding zero and one. It can be seen from Eq. (3) that the β-divergence satisfies

$$d_\beta(\lambda y, \lambda x) = \lambda^\beta d(y,x) \tag{4}$$

When the value of β is zero, the value of Eq. (4) is scale shift invariant. That is, it is unrelated to scale factor λ, indicating that when NMF is performed, the line spectrum components in the LOFAR image of mixed signal will have equal weight across the continuous spectrum components, so as to avoid the influence of high-energy line

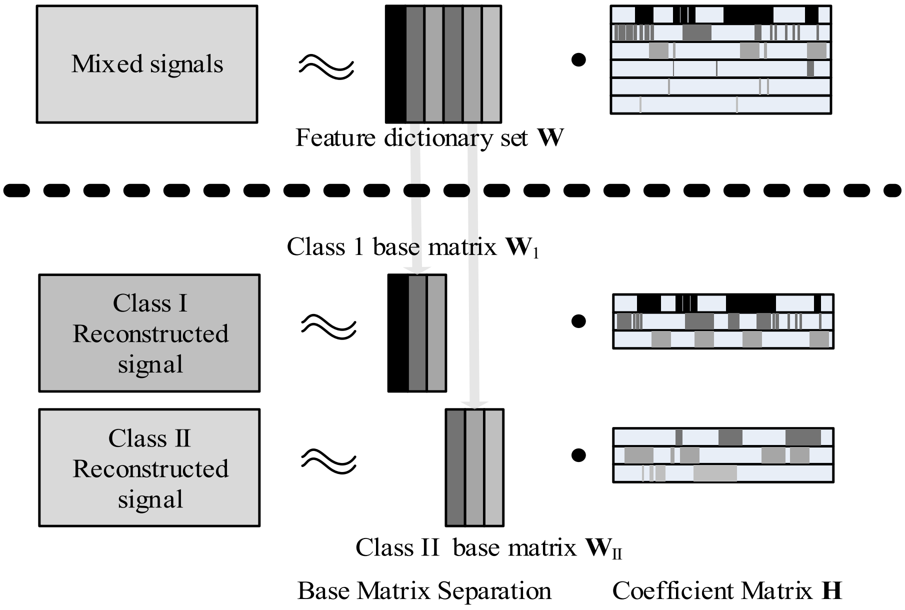


FIGURE 1
The NMF process of Underwater acoustic multi-target signal.

spectrum on the results,. However, when $\beta \neq 0$, β-divergence will still be affected by the line spectrum components.

To improve the independence of the NMF base vectors and mitigate the impact of their feature redundancy on the signal reconstruction, determinant constraints are added. Let the volume of the space stretched by matrix $W = [w_1, w_2, \cdots, w_r] \in R_+^{m \times r}$be expressed as

$$vol(W) = \begin{cases} \sqrt{\det(WW^T)}, & m < r \\ |\det(W)|, & m = r \\ \sqrt{\det(W^T W)}, & m > r \end{cases} \tag{5}$$

When the volume, vol(W) , in Eq. (5) is smallest, each vector $w_i \in W$ can be uniquely determined (Sadeghi et al., 2021). Thus, the improved NMF objective function based on β-divergence and determinant constraints can be expressed as

$$\min(J) = d_{\beta=0}(V, WH) + \alpha \cdot vol(W) \tag{6}$$

Optimize the NMF coefficient matrix by spatial similarity

The radiated noise of underwater acoustic targets is the main source of passive buoy detection (Liu et al., 2021). The shape, displacement, structure, and other features of acoustic targets comprise the main factors of signal characteristics, that is, when the acoustic targets remains a relatively stable sailing state, the spectrum distribution of its radiated signal will show good short-term stability, as shown in Figure 2. Owing to the short interval between two adjacent frames, even if environmental noise exists, It still shows good short-term similarity by using of a fitted spectrum distribution and an interframe alias is useful (Li et al., 2016),.

Figure 2 shows the spectrum of multi-target mixed signals observed by a single marine environment monitoring buoy in a certain time period. Although the frequency spectrums of different targets under different working conditions are different, and the frequency spectrums of mixed signals formed by these signals are also diverse, it is usually difficult for each target to have a large change in a short period of time within the monitoring range of the marine environment monitoring buoy. Therefore, the mixed signals will also maintain a certain characteristic stability in a short time. Of course, the spectrum of the mixed signal is different for different periods of time or for different targets, the spectrum is relatively stable only in a short period of time.

When calculating the LOFAR spectrum of the received signal, fast Fourier transform (FFT) resolution inefficiencies will lead to a certain degree of spectral smoothness between adjacent frequencies. Therefore, in addition to the short-time similarity between the signal frames, as shown in Figure 2, the eight neighbors of the current pixel are taken as local neighborhood candidate regions. Then, adaptive local neighborhood weighting is used to analyze the similarity contributions of the neighboring candidate pixels to the current pixel to make full use of the spatial similarity between them for the benefit of the NMF model.

Within the eight-neighborhood range of the current pixel, the weight contribution of pixel $\boldsymbol{y}_j$ to $\boldsymbol{y}_i$ is expressed as

$$c_{ij} = \Gamma^{-1} + \rho \tag{7}$$

where Γ describes the neighborhood position of pixel $\boldsymbol{y}_j$, and P reflects the similarity between two pixels. The calculation formula is

$$\begin{cases} \Gamma = |x_i - x_j| + |y_i - y_j| \\ \rho = \langle H_i, H_j \rangle \end{cases} \tag{8}$$

where $\langle\cdot\rangle$ represents the inner product of the vector, and $\boldsymbol{s}_i$ is the coefficient vector corresponding to the pixel. Thus, when the spatial similarity of two pixels is high, their weight value, c_{ij}, is also large, and vice versa. Therefore, the weight contribution of the eight pixels in the neighborhood of the current pixel, $\boldsymbol{y}_i$, can be expressed as

$$C_i = \sum_{j \in N(i)} c_{ij} \tag{9}$$

Thus, an adaptive calculation of the weight contribution of the pixels in the neighborhood can be realized. The improved NMF model that incorporates the adaptive weighted spatial similarity is

$$\min(J) = d_{\beta=0}(V, WH) + \frac{\alpha}{2} \cdot \left(vol(W) + \left\|\tilde{W} - W\right\|^2\right)$$
$$+ \lambda \sum_{i=1}^{N} \sum_{j \in N(i)} c_{ij} \cdot \left\|H_i - H_j\right\|_2^2 + \beta \|H\|_{1/2} \tag{10}$$

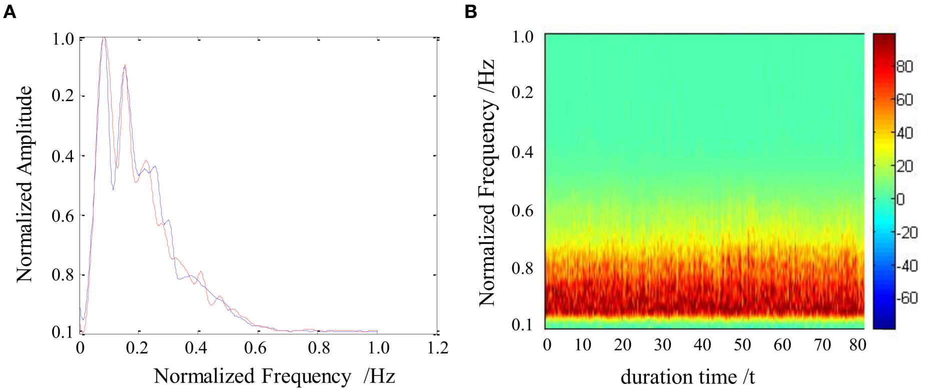


FIGURE 2
Spectrogram of radiated noise from a surface target: **(A)** trend diagram of two adjacent signals in a frame; **(B)** spectrum trend of the surface target.

where W ≥ 0, H ≥ 0, $1_r^{\mathrm{T}}H = 1_n^{\mathrm{T}}$, α, β, and λ are regularization coefficients. $\|\cdot\|_{1/2}$ is the $L_{1/2}$-norm, which is used to add a sparse constraint to the coefficient matrix to prevent overfitting and noise residue. $\| \tilde{W} - W \|^2$ describes the smoothness of a single source signal, which is used to increase the separability between multi-targets, and $\tilde{W}$ is the moving average matrix of the calculated values of the previous iteration of W (Liu et al., 2021). The optimal solution of Eq. (10) can be achieved by the derivative of a variable or the least-squares method under the Karush–Kuhn–Tucker condition (Chen and Lv, 2021), that is:

$$W \leftarrow W \frac{\left[V(WH)^{-2} + \alpha\tilde{W}\right]H^{\mathrm{T}}}{(WH)^{-1}H^{\mathrm{T}} + \alpha W}$$

,

$$H \leftarrow H \frac{W^{\mathrm{T}}\left[V(WH)^{-2} + \lambda H_C\right]}{W^{\mathrm{T}}(WH)^{-1} + \lambda CH + \frac{\beta}{2}H^{-\frac{1}{2}}} \tag{11}$$

The dimensionality of the base matrix is very important for NMF decomposition as it directly affects target feature extraction, especially for acoustic target signals. The noise interference of the receiving signal is large, and the signal features overlap to some extent. Hence, when the dimensionality is too large, the NMF decomposition base matrix introduces too much noise. However, if the dimensionality is too small, the decomposition fidelity of the base matrix will be reduced, leading to the non-uniqueness of the local expression characteristics and insufficient separation accuracy. Therefore, the nearest-neighbor eigenvalue dominance method (Wang et al., 2019) is used to estimate the dimensionality of the base matrix.

Joint blind separation using improved NMF and FastICA

The base matrix combines the base vector of all source signals in the mixed signal received by a single buoy. Because the base vectors are not ordered, it is difficult to determine which base vectors are from the same source signal. Therefore, the FastICA algorithm is used here to separate the base matrix into base vector groups based on their independent sources.

According to the central limit theorem (Krishna et al., 2020), the mixed signal obtained from independent source signals has a high Gaussian distribution. From information theory, the stronger the Gaussian property of the equivariance random variable, the greater its entropy. The FastICA algorithm was explicitly built to maximize the non-Gaussian property of mixed observation signals, and negative entropy is used as the parameter. When negative entropy reaches its maximum, the non-Gaussian property of each source signal also reaches its maximum, indicating that each independent component is well-separated (Krishna et al., 2020).

Let signal $\mathbf{Y}=\mathbf{B}^{\mathrm{T}\cdot}\ \boldsymbol{v}$ be the matrix of the source signal separated from the whitened observation signal, $\boldsymbol{v}$. Its negative entropy can then be expressed by its differential entropy. However, for that the probability density of random variables is difficult to estimate, the calculation is simplified with a common approximation formula:

$$\begin{aligned} J(Y) &= (E(g(Y)) - E(g(Y_{Gauss})))^2 \\ &= (E(g(B^T \cdot v)) - E(g(Y_{Gauss}))))^2 \end{aligned} \tag{12}$$

where $\boldsymbol{B}$ is the separation matrix, $b_i \in B$ is a column vector in the matrix, and $\|b_i\|=1$. $g(\cdot)$ is an arbitrary non-quadratic function and its reasonable selection leads to a good approximation of the source signal (Abdalla and Alrufaiaat, 2021). When the y^3 form is selected, it has already been proven to be optimal for separation performance, accuracy, and convergence when used by the FastICA model (Xiumin et al., 2020). To solve Eq. (12), the Lagrangian function is constructed as

$$\begin{aligned} L(B,\beta) &= J(B) - \beta(\|B\|^2 - 1) \\ &= (E(g(B^T \cdot v)) - E(g(Y_{Gauss})))^2 + \beta\|B\|^2 \end{aligned} \tag{13}$$

where $\beta = E(B^T v g'(B^T \cdot v))$, and $g'(\cdot)$ is the derivative of $g(\cdot)$. It can be seen that the maximization of Eq. (12) can be converted to the derivative of Eq. (13):

$$E(v \cdot g(B^T \cdot v)) - \beta B = 0 \tag{14}$$

As $\boldsymbol{v}$ is the whitened data, according to the third-order Newton iteration method, the iteration formula of the FastICA algorithm is

$$\begin{cases} B^{\star} = E(vg(B^T v)) - E(g'(B^T v))B \\ B = E(vg(B^T v)) + E(vg(B^{\star T} v))) - E(g'(B^T v))B - \beta B^{\star} \end{cases} \tag{15}$$

where $B^{\star}$ is the intermediate value of the iterative calculation of new B. FastICA is the main model used for BSS tasks. However, the number of observed signals must be greater than or equal to the number of source signals. Although the multivariate LOFAR time-frequency spectrum can be obtained from the short-time Fourier transform, it cannot be directly applied to FastICA for BSS because all column vectors in the spectrum come from the same channel.

However, after NMF decomposition of the LOFAR spectrum, the characteristics of each independent source signal from the mixed signal can be reflected by NMF base vectors. Thus, the base matrix can be viewed as a combination of the base vectors of each independent source signal. Nevertheless, the order of the base vectors is messy, making it impossible to determine ownership.

Therefore, after improving the NMF method, we designate the base and coefficient matrices of the received observation signal, V, as W, and H, respectively, where $W=\{w_1,w_2,\cdots,w_r\}$, and r is the dimensionality of the base matrix. If the ownership of the base vectors in $W=\{w_1,w_2,\cdots,w_r\}$ is obtained, then W can be expressed as $W=\{W_1,W_2,\cdots,W_i\}$, such that

$$\begin{aligned} V_j &= \{W_1, W_2, \cdots, W_i\}\left\{h_{j1}^{\mathrm{T}}, h_{j2}^{\mathrm{T}}, \cdots, h_{ji}^{\mathrm{T}}\right\}^{\mathrm{T}} = W_1h_1 + W_2h_2 + \cdots + W_ih_i \\ &= Y_1b_1 + Y_2b_2 + \cdots + Y_ib_i = bY = h_jW^{\mathrm{T}} \end{aligned} \tag{16}$$

where Y is the source signal matrix, i is the number of source signals, and W_i is the base matrix of the source signal, which comprises the base vector belonging to the same independent source. It is, therefore, obvious that $W_i$has disturbed the order of the original w_i in W. Thus, h_{ji} becomes the new column vector of the coefficient matrix, H, corresponding to the new order of w_i (i.e., corresponding to the order of W_i). $h_i = h_{ji}^{\mathrm{T}}$ and its dimensionality is thus consistent with that of W_i, and b_i is a vector separated from h_i. Since there is a part value h_i' in h_i, making $W_i\mathrm{h}_i' = Y_i$, the remaining part b_i of h_i after removing h_i' can be regarded as the contribution of

Y_i when V_j is generated, Thus, we have the parameter of its corresponding mixed matrix.

Alongside the derivation process of FastICA (Abdalla and Alrufaiaat, 2021), it can be further deduced from Eq. (16) that

$$Y = b^{-1} h_j W^{\mathrm{T}} = BW^{\mathrm{T}} \tag{17}$$

Accordingly, the source signal, *Y*, can be calculated from the base matrix, *W* (obtained from the improved NMF), and the unmixing matrix, *B*.

The input matrix, *v*, of the proposed algorithm from Eq. (14) can be obtained by recombining *W* and one column, V_i, of the LOFAR image of the observed mixed signal. Then, the unmixing matrix, *B*, in Eq. (17) can be obtained *via* iterative calculations, as shown in Eq. (14). Finally, the BSS of the acoustic multi-target signals received by the buoy can be achieved.

From the above analysis, the proposed BSS algorithm that joint improved NMF and FastICA is shown in Figure 3. Note that after signal separation, $Y=Bv^{\mathrm{T}}$, because the dimensionality of the base matrix obtained by the improved NMF is greater than the number of source signals, there will be different sequences of the same source signal in the separated results. Thus, it will be necessary to further select and combine them using correlation analysis or other algorithms.

Experimental verification and analysis

To verify the efficacy of the proposed BSS algorithm used for underwater acoustic multi-target signal received by a single buoy, the normalized cross-correlation coefficient (NCC) and logarithmic spectral distance (LSD) are used as evaluation criteria. The larger the LSD value, the better the signal reconstruction performance (Li and Wang, 2019). The LSD formula is

$$f_{\mathrm{LSD}} = \sqrt{\frac{1}{N} \sum_{n=0}^{N/2-1} \left| L(S(l,n)) - L(S_d(l,n)) \right|^2} \tag{18}$$

where $L(S(l,n))$ is the logarithmic spectrum between two signals, and *N* is the number of data. The $\boldsymbol{S}$ and $\boldsymbol{S}_d$ indicate source signals and reconstructed signals respectively

The spectrum features of an acoustic target signal mainly include its line spectrum, continuous spectrum, and envelope modulation spectrum (Li et al., 2016). When the sailing speed of underwater acoustic target exceeds its critical speed, its propeller will periodically modulate it radiated signal, representing as envelope line spectrum. Therefore, the modulation line spectrum, which reflects the axial frequency and other information of targets, is an important feature for target recognition, that is why the modulated line spectrum is needed; and it can be obtained by extracting the envelope of the signal and performing FFT. Thus, the simulation signal model with periodic modulation is

$$s(t) = \sum_i \left(1 + A_m \cdot \cos\{-2\pi f_0 (t-\tau_i)\}\right) \cos\left(2\pi f_c (t - iT - \tau_i)\right) \tag{19}$$

Where A_m is the modulation amplitude, and its cosine signal is used to simulate the screw modulation, in which f_0 is the modulation frequency reflecting the axial frequency. Because the main test feature is the axial frequency information of the underwater acoustic target, other signals (e.g., hydrodynamic) are simplified to cosine or sine signals with frequency f_c. In our experiment, the mixed signals of three simulation signals were used. The sampling rate, f_s=1/T, was 6,000, and so there were 4,096 sampling point. the values of f_0 and f_C in each simulation signal are listed in Table 1.

The mixed matrix, $\omega = [0.756\ 0.871\ 0.559]$, was generated randomly. *n*(t) was the superimposed noise, and *i*=20. The time-domain waveforms of all signals and mixed signals are shown in Figure 4.

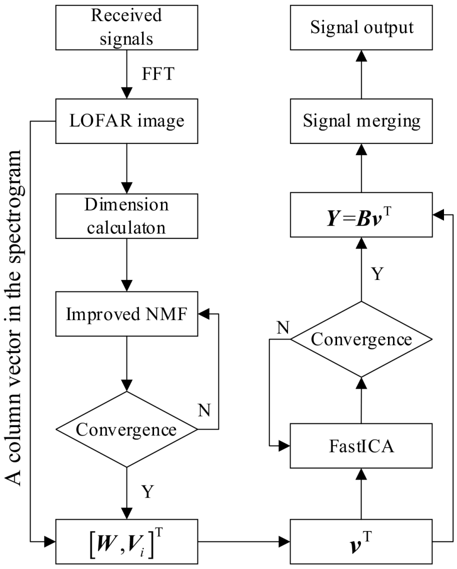


FIGURE 3
Application of the proposed improved non-negative matrix factorization (NMF) and FastICA for the blind source separation of underwater acoustic multi-target signals.

TABLE 1 Simulated signal parameters for the experiment.

f_{01}	f_{02}	f_{03}	f_{C1}	f_{C2}	f_{C3}
20	35	45	2,500	2,000	1,500

The demodulation line spectra of the mixed signal were calculated *via* FFT on the extracted envelope, $S(t)$, as shown in Figure 5. The spectrum information within 100 Hz is displayed for simplicity.

In Figure 5, the modulation line spectrum and its octave harmonics reflect the screw frequency information of different acoustic targets, but they are interlaced. The proposed BSS algorithm based on the improved NMF and FastICA was used as described in the previous sections. The signals were divided into frames, and FFT was performed first with all 4,096 data points and a 50% overlap between the frames. Thus, the LOFAR spectrum of the experimental mixed signals was generated. Next, the improved NMF was used to decompose the LOFAR spectrum using an 18-dimensional base matrix with 300 iterations. FastICA was then used for the BSS. After that the obtained decomposed signals were analyzed for similarity, and those with high similarity were merged into the same source signal.

The time domain waveform and envelope spectrum of the decomposed source signal are shown in Figures 6 and 7, respectively. From Figure 7, owing to the β -divergence, the proposed algorithm better maintains the spectrum components of the mixed signal, and their three modulation components (i.e., 20, 35, and 45 Hz) and harmonic components were effectively separated. This verifies the efficacy of our proposed algorithm.

The cross-correlation coefficient between the separated and source signals was calculated, as listed in Table 2. The cross-correlation between the separated signal and its source is large, but that of the separated signal and the other two source signals is small, indicating that the algorithm has good BSS accuracy.

To further understand the robustness of the proposed method, white noise was added at different signal-to-noise ratios to the simulated mixed source signals. The improved NMF with an Improved Crest Factor (ICF) ICF operator (NMF+ICF) (Li and Yang, 2016), improved FastICA (iFastICA) (Guotao et al., 2021), and the combined classical NMF and FastICA (NMF+FastICA) were compared. The NCC and LSD of each used were analyzed under different signal-to-noise ratios (Yurong et al., 2020), and the results are illustrated in Figures 8 and 9, where each value is the average of the three source signals from several experimental trials. The KL divergence is used as the objective function of classical NMF in this experiment.

$$R_{SNR} = 10\lg\left(\frac{E\left(x^{\mathrm{T}}x\right)}{E\left(n^{\mathrm{T}}n\right)}\right) \tag{20}$$

From Figures 8 and 9, iFastICA’s BSS results were not ideal for all SNRs because its original FastICA component requires mixed signals in multiple channels. However, recall that the LOFAR spectrum is essentially one channel signal. Therefore, although the FastICA algorithm was improved in iFastICA, its blind separation effect is still not ideal.

Compared with iFastICA, the signal separation accuracy of NMF +ICF was improved to some extent, but it is still not ideal. On the one hand, the algorithm improved the NMF and its adaptability to modulated pulse signals. However, because there was no reasonable sparse constraint on the NMF coefficient matrix, the decomposed base matrix suffered an insufficient independent local expression ability for the source signal. On the other hand, NMF+ICF used the correlation-based ICF factor to select the base vector with the highest ICF value as the final separated signal, which easily causes the loss of important signal components.

NMF+FastICA combined the advantages of classical NMF and FastICA algorithms and achieved the best separation results.

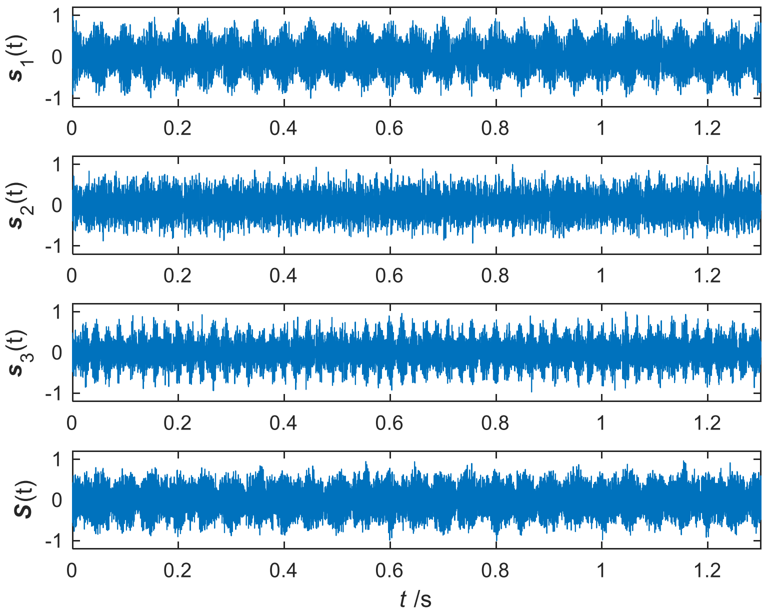


FIGURE 4
Time domain waveform of the simulated signal for the experiment.

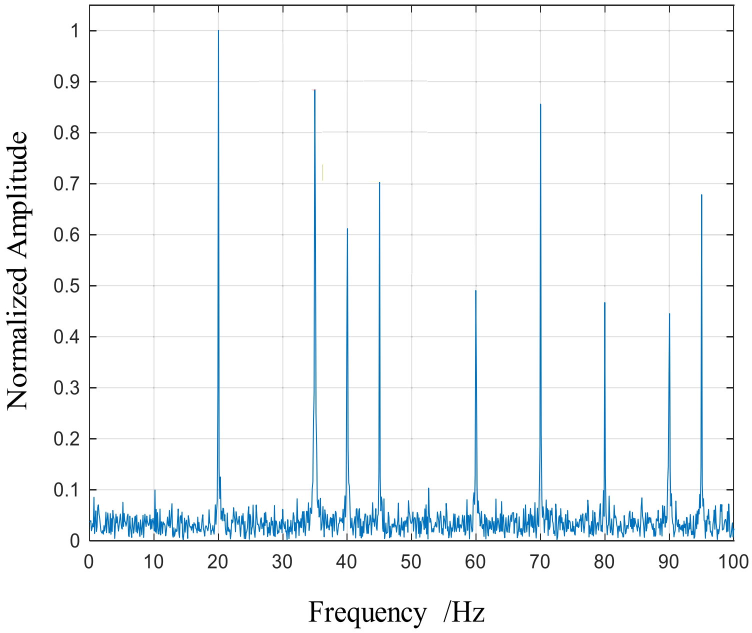


FIGURE 5
Envelope spectrum of the hybrid simulation signal.

However, as the classic NMF is not optimized for underwater acoustic target signals, the local and dependent characteristics of the basis matrix affected the final BSS effect. The proposed algorithm (iNMF +FastICA) achieved the best separation effect under various SNRs; because of that, it better retains the separable characteristics of the *via* the improved NMF. Thus, the separated signal more approximated the source signal.

Conclusion

Based on the improved NMF and FastICA algorithm, this study addressed an important BSS problem used for mixed signals from underwater acoustic multi-targets are received by automatic monitoring buoys Our method was successful because it addressed the non-convex and feature correlation problems encountered by the classical NMF algorithm with spatial and spectral correlation optimization. This improved the adaptability of the classic NMF to handle the modulation characteristics of underwater acoustic target signals while improving the local expressability and independence of the NMF base matrix. Then, the advantages of the improved NMF and FastICA algorithms were combined to achieve superior BSS of underwater acoustic multi-target signals.

The simulation signal experiment results showed that compared with state-of-the-art BSS algorithms based on NMF and FastICA, the novel proposal achieves better signal separation

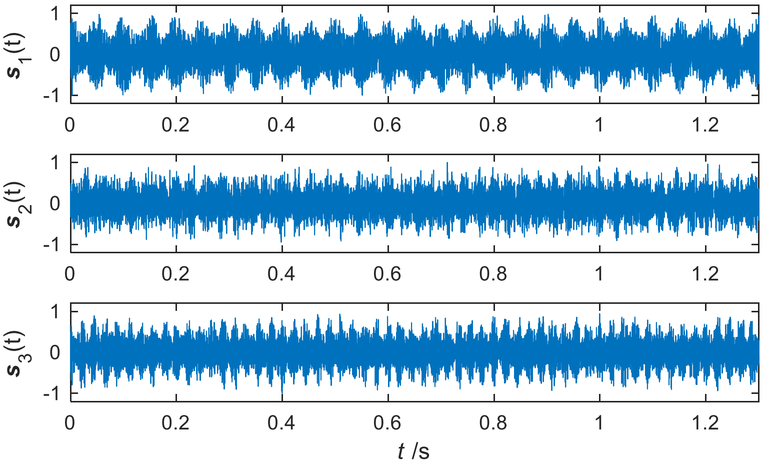


FIGURE 6
Time-domain waveform of separated signals.

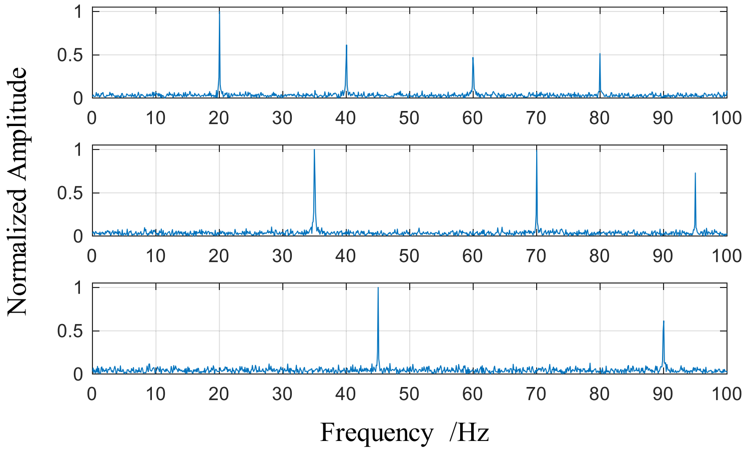


FIGURE 7
Envelope spectrum of the separated signals.

TABLE 2 Cross-correlation coefficient between separated and source signals.

Signal	Source 1	Source 2	Source 3	Mixed Signal
Separated signal 1	0.925	0.106	0.197	0.515
Separated signal 2	0.109	0.195	0.893	0.612
Separated signal 3	0.210	0.913	0.115	0.609

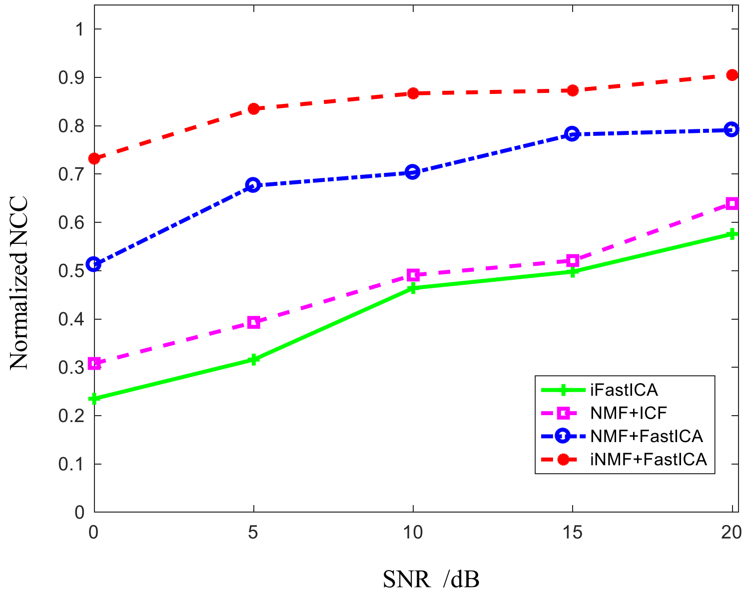


FIGURE 8
Normalized cross-correlation coefficient under different signal-to-noise ratios.

accuracy while maintaining the modulation characteristics of the original signal. Furthermore, it reduces SNR and enhances signal separation, which leads to better feature interpretation and target recognition.

Presently, when the proposed method is applied to underwater acoustic target signals with obvious lines or modulation spectra, the BSS is good. However, with the advancement of noise reduction technologies, the radiated signals of underwater acoustic targets have

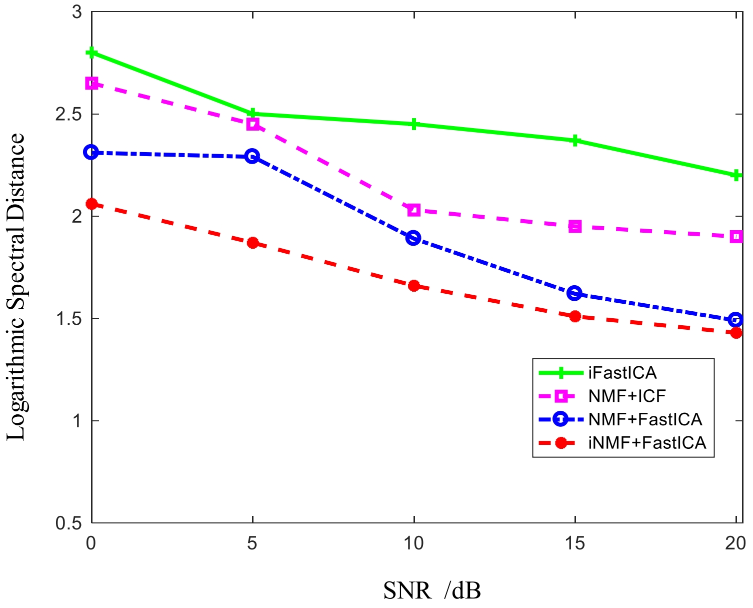


FIGURE 9
Logarithmic spectral distance of each algorithm under different signal-to-noise ratios.

become more difficult to detect. Therefore, Future work should further improve and optimize these methods by combining extant signal detection, feature extraction, noise reduction, and signal enhancement techniques to improve their applicability.

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

LD, the first author of the paper, led most of the main work, such as algorithm design, theoretical derivation, experimental verification, and finally completed the writing of the paper. WM, the second author of the paper, mainly modified the algorithm and made great contributions to the reasonable integration based on NMF and FastICA. As the third author of the paper, YL is mainly involved in the design and modification of the algorithm in the paper. He has deep attainments in the research of NMF algorithm feature line spectrum maintenance, and has great guidance for algorithm design. HJ is the fourth author of the paper. He designed verification experiments based on the algorithm of the paper, collected and simulated relevant experimental data. ZH, the fifth author of the paper, participated in the collation of experimental data, the revision and correction of the paper. All authors contributed to the article and approved the submitted version.

Funding

This paper is funded by the University Civil Affairs Fund (I32102003).

Acknowledgments

We would like to thank Editage (www.editage.cn) for English language editing.

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

Abdalla, K., and Alrufaiaat, S. (2021). A new robust decoding technique of four transmitters MIMO STBC system based on FastICA algorithm. *Int. J. Intelligent Eng. Systems.* 14 (1), 181–191. doi: 10.22266/ijies2021.0228.18

Chen, S., and Lv, J. (2021). A nonnegative matrix factorization method based on adaptive local neighborhood weighted constraint and its application in hyperspectral unmixing. *J. Signal Process.* 37 (5), 804–813. doi: 10.16798/j.issn.1003-0530.2021.05.014

Dianmant, R., Kipnis, D., Bigal, E., Scheinin, A., Dan, T., Pinchasi, A., et al. (2019). "An active acoustic track-before-detect approach for finding underwater mobile targets," in *IEEE Journal of Selected Topics in Signal Processing*, (United States: IEEE) Vol. 13. 104–119. doi: 10.1109/JSTSP.2019.2899237

Guotao, W., Qiuxi, J., and Fangzheng, L. (2021). An improved FastICA radar signal sorting method. *J. Detection Control* 43 (6), 43–49.

Hien, L., and Gillis, N. (2021). Algorithms for nonnegative matrix factorization with the kullback–leibler divergence. *J. Sci. Computing.* 87 (3), 417–439. doi: 10.1007/s10915-021-01504-0

Huang, R., Li, X., and Zhao, L. (2019). "Spectral–spatial robust nonnegative matrix factorization for hyperspectral unmixing," in *IEEE Transactions on Geoscience and Remote Sensing*, (United States: IEEE) Vol. 57. 8235–8254. doi: 10.1109/TGRS.2019.2919166

Krishna, A., Nimbal, A., Makam, A., and Rao, V. (2020). Implementation of fast independent component analysis on field-programmable gate array for resolving the slot collision issue in the space-based automatic identification system. *Int. J. Satellite Commun. Networking.* 38 (6), 480–498. doi: 10.1002/sat.1362

Liu, X., Wu, M., Zheng, X., and Li, D. W. (2021). Underwater acoustic target signal enhancement algorithm optimized by improved NMF. *Electron. Optics Control.* 28 (9), 6–5+53. doi: 10.3969/j.issn.1671-637X.2021.09.002

Li, D., and Wang, Z. (2019). Improved spatial information constrained nonnegative matrix factorization method for hyperspectral unmixing. *Laser Optoelectronics Progress.* 56 (11), 111006. doi: 10.3788/LOP56.111006

Li, D., and Yang, R. (2016). Detection of speech signal in strong ship-radiated noise based on spectrum entropy. *J. Vibroengineering.* 18 (1), 661–670. doi: 10.3969/j.issn.1001-2400.2016.05.024

Li, D., Yang, R., and Han, J. (2016). Study of speech enhancement in the background of ship-radiated noise. *J. Xidian University.* 43 (5), 133–138.

Lu, X., Dong, L., and Yuan, Y. (2020). "Subspace clustering constrained sparse NMF for hyperspectral unmixing," in *IEEE Transactions on Geoscience and Remote Sensing*, (United States: IEEE) Vol. 58. 3007–3019.

Lu, X., Wu, H., Yuan, Y., Yan, P., and Li, X. (2013). "Manifold regularized sparse NMF for hyperspectral unmixing," in *IEEE Transactions on Geoscience and Remote Sensing*, (United States: IEEE) Vol. 51. 2815–2826.

Rathnayake, B., Ekanayake, E., Weerakoon, K., Godaliyadda, G., and Herath, H. (2020). "Graph-based blind hyperspectral unmixing *via* nonnegative matrix factorization," in *IEEE Transactions on Geoscience and Remote Sensing*, (United States: IEEE) Vol. 58. doi: 10.1109/TGRS.2020.2976799

Sadeghi, M., Behnia, F., and Amiri, R. (2021). "Optimal geometry analysis for TDOA-based localization under communication constraints," in *IEEE Transactions on Aerospace and Electronic Systems*, (United States: IEEE) Vol. 57. 3096–3106. doi: 10.1109/TAES.2021.3069269

Sriharsha, K., and Abhijit, K. (2018). Homotopy optimization based NMF for audio source separation. *IET Signal Processing.* 12, 1099–1106. doi: 10.1109/ACCESS.2020.2985842

Wang, W., Qian, Y., and Liu, H. (2020). "Multiple clustering guided nonnegative matrix factorization for hyperspectral unmixing," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (United States: IEEE) Vol. 13. 5162–5179.

Wang, M., Wang, H., Dong, F., Ren, B., and Song, L. (2019). A method of compound fault signal separation based on EVMD-LNMF. *J. Vibration Shock.* 38 (16), 146–152. doi: 10.13465/j.cnki.jvs.2019.16.021

Weiderer, P., Tomé, A., and Lang, E. (2020). A NMF-based extraction of physically meaningful components from sensory data of metal casting processes. *J. Manufacturing Systems.* 54, 62–73. doi: 10.1016/j.jmsy.2019.09.013

Wildeboer, R., Schoot, B., Wijkstra, H., Mischi, M., and Salomon, G. (2020). "Blind source separation for clutter and noise suppression in ultrasound imaging: Review for different applications," in *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, (United States: IEEE) Vol. 67. 1497–1512.

Xiumin, C., Shanjun, L., and Xingjian, D. (2020). Simulation. analysis of nonlinear function performance of FastICA algorithm. *Comput. Appl. Software* 37 (6), 277–282+333. doi: CNKI:SUN:JYRJ.0.2020-06-049

Yurong, L., Jie, L., Lailin, L., Gong, C., and Wang, Y. (2020). Parallel algorithm of deep transductive non-negative matrix factorization for speech separation. *Comput. Sci.* 47 (8), 49–55. doi: 10.11896/jsjkx.190900202

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). "Multireceiver SAS imagery based on monostatic conversion," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (United States: IEEE) Vol. 14. 10835–10853.

Zhang, X., and Yang, P. (2021). An improved imaging algorithm for multi-receiver SAS system with wide-bandwidth signal. *Remote Sensing.* 13 (24), 5008. doi: 10.3390/rs13245008

Zhang, X., Yang, P., Feng, X., and Sun, H. (2022). Efficient imaging method for multireceiver SAS. *IET Radar Sonar Navigation.* 16 (9), 1470–1483.

Zhang, X., Yang, P., Huang, P., Huang, P., Sun, H., Ying, W., et al. (2022). Wide-bandwidth signal-based multireceiver SAS imagery using extended chirp scaling algorithm. *IET Radar Sonar Navigation* 16 (3), 531–541.

Zhang, X., Yang, P., and Sun, M. (2022).). experiment results of a novel sub-bottom profiler using synthetic aperture technique. *Curr. Science.* 12 (4), 461–464.

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Navigation.* 14 (7), 1055–1060. doi: 10.1049/iet-rsn.2019.0477

Frontiers in Marine Science

TYPE Methods
PUBLISHED 18 January 2023
DOI 10.3389/fmars.2022.1005959

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Arun Prakash,
Motilal Nehru National Institute of
Technology Allahabad, India
Kai Liu,
Beijing Institute of Technology, China

*CORRESPONDENCE
Yunlong Wei
weiyunlong@fzu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 28 July 2022
ACCEPTED 23 December 2022
PUBLISHED 18 January 2023

CITATION
Wei Y and Wang D (2023) Graph coloring-based multichannel MAC protocol in distributed underwater acoustic sensor networks.
Front. Mar. Sci. 9:1005959.
doi: 10.3389/fmars.2022.1005959

Graph coloring-based multichannel MAC protocol in distributed underwater acoustic sensor networks

Yunlong Wei[1]* and Deqing Wang[2]

[1]Department of Electronic Information Engineering, School of Physics and Information Engineering, Fuzhou University, Fuzhou, China, [2]Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University), Ministry of Education, Xiamen, China

In this paper, the multichannel medium access control (MAC) problem in distributed underwater acoustic sensor networks (UASNs) were investigated. Compared with single-channel MAC protocols in terrestrial radio networks, there exist multichannel hidden terminal problem and long-delay hidden terminal problem in multichannel MAC protocol due to long propagation delay in UASNs. In addition, energy constraint makes channel allocation a challenging problem in distributed UASNs. To solve these aforementioned problems, a new multichannel MAC protocol, called graph coloring-based multichannel MAC protocol (GCMAC) is present. The protocol GCMAC is a synchronized MAC protocol which splits the time into three phases, namely, channel negotiation phase, channel selecting phase and data transmission phase. Specially, the rule for selecting channel is carefully designed based on graph coloring theory to avoid collision and maximize the utilization rate of channels in channel selecting phase. Simulation results show that GCMAC can greatly improve the system throughput and energy efficiency by effectively solving the hidden terminal problems and channel allocation problem.

KEYWORDS

underwater acoustic sensor networks, multichannel, medium access control, graph coloring, system throughput, energy efficiency

1 Introduction

Underwater acoustic sensor networks (UASNs) have attracted extensive attention in the last couples of years (Cui et al., 2006; Jahanbakht et al., 2021). On the one hand, UASNs have been found in wide aquatic applications, such as oceanographic data collection (e.g., temperature, salinity, and zonality), field monitoring and disaster prevention. On the other hand, the adverse underwater environments posed severe challenges at almost every level of the protocol stacks, among which efficient medium access control (MAC) is one of the most fundamental issues (Ansari et al., 2015).

Compared with the terrestrial counterpart, there are some differences which will result in significant influence on network protocol design in underwater environment. Firstly, the propagation speed of acoustic signals in water is about $1.5{\times}10^3 m/s$, which is five orders of magnitude lower than radio propagation speed. This brings obvious propagation delay in underwater communications. Secondly, the available bandwidth is extremely limited and it is distance/frequency dependent. For example, there are only a few hundred kHz for short range, a few 10kHz for medium distance, and a few hundred Hz for long range (Huang et al., 2018). Lastly, severe multipath spread and high Doppler shift may cause low communication rate and unreliable physical layer. Due to these unique characteristics owned by underwater acoustic channels, the MAC protocols designed for terrestrial radio networks could not be adapted directly, and a significant amount of research efforts (Chen et al., 2014; Jiang, 2018) for MAC in UASNs have been inspired in the last ten years. Most of those protocols focus on single-channel network scenarios, whereas the MAC protocols with a single channel could not eliminate the interference between control packets and data packets due to long propagation delay. Consider that the multichannel MAC protocols have the capability of transmitting data in parallelly, leading to higher network throughput, lower channel access delay, smaller energy consumption, some research on multichannel MAC protocols for UASNs have been conducted. Therefore, we focus on the design of multichannel MAC protocols adapted to the unique characteristics of underwater acoustic channels. Specifically, the contribution of this paper mainly includes:

- A multichannel MAC, called Graph Coloring-based Multichannel MAC (GCMAC), is put forth to tackle the special hidden terminal problem resulted by the unique underwater acoustic channels. The hidden terminal problem consists of multichannel hidden terminal one and long delay hidden terminal one. In GCMAC, a synchronized scheduling mechanism is adopted and the data transmissions are originated by the receiving nodes.
- In order to increase the channel utilization, we identified the channel allocation problem based on the theory of graph coloring by taking physical transmission model into account. Furthermore, a channel-selection rule is designed to assist the data negotiation for GCMAC.

The rest of this paper is organized as follows. In Section 2, we review related work on multichannel MAC protocols for UASNs. Then in Section 3, we discuss multichannel problems needing to be solved in this paper. After that, we present a system model and describe the GCMAC in detail in Section 4 and 5. Simulation results are presented in Section 6. Finally, Section 7 gives a summary of this paper.

2 Related work

Multichannel MAC protocols for underwater acoustic networks have also aroused significant research interest recently. The authors in (Zhou et al., 2012) proposed a new MAC protocol called Cluster based Underwater MAC(CUMAC) for long-delay multichannel underwater sensor networks. They investigated triple hidden terminal problems, i.e., multichannel hidden terminal, long-delay hidden terminal and traditional multi-hop hidden terminal. In order to increase energy efficiency, the authors in (Chao et al., 2015) proposed a dynamic duty-cycled multiple-rendezvous multichannel medium access control (DMM–MAC) protocol that is suitable for transmitting bursty traffic in a duty-cycled UWSN. Consider that the contention-free approach can achieve high performance by avoiding the collisions at the MAC layer, The authors in (Hsu et al., 2009) proposed a TDMA (Time Division Multiple Access) -based MAC scheduling scheme called ST-MAC for underwater sensor networks. They constructed the Spatial-Temporal Conflict Graph (ST-CG) to describe the conflict delays among transmission links explicitly, thus the protocol ST-MAC is modeled as a new vertex coloring problem of ST-CG. Similarly, GC-MAC (Alfouzan et al., 2019) employs TDMA-like approach by assigning separate time-slots, colors, to every individual sensor node in every two-hop neighborhood. The protocol MC-UMMAC (Bouabdallah et al., 2017) is also an energy efficient MAC protocol that aims at achieving a collision free communication. It operates on a single slotted control channel to avoid the missing receiver problem and multiple data channels to improve the network throughput.

Among these works mentioned above, we could find that the channel allocation is considered a critical issue left to be solved in multichannel MAC for UASNs.

Channel or spectrum allocation has a significant impact on the performance of multichannel communication. The authors in (Wang and Liu, 2005) studied the dynamic channel allocation for open-spectrum wireless networks. Multi-frequency Media Access Control for Wireless Sensor Networks protocol (MMSN) is the first multi-frequency MAC protocol specially designed for wireless sensor networks. They also conducted a complete study on the tradeoffs among physical frequency requirements, potential conflict reduction and communication overhead during frequency assignment. While the authors in (Fitzpatrick et al., 2004) described two new online channel assignment algorithms for networks based on a regular hexagonal layout of cells. DRAND (Rhee et al., 2009) is a randomized time slot scheduling algorithm, which is a distributed implementation of RAND (Ramanathan and Engineer, 1997). The algorithm is suitable for a wireless network and could be used for frequency or code scheduling. The authors in (Saifullah et al., 2014) proposed a set of distributed algorithms for near optimal channel allocation in wireless sensor networks with theoretical bound. Inspired by these aforementioned works, we address the channel allocation problem based on graph coloring theory (Kuhn and Wattenhofer, 2006) in distributed topology to achieve high energy efficiency.

3 Problems in multichannel MAC protocol

3.1 Hidden terminal problem

There is multichannel hidden terminal problem and long delay hidden terminal problem in multichannel protocols (Zhou et al., 2012).

Figure 1 shows that when node a is to send data to node b, Request to send (RTS) and Clear to send (CTS) control packets may interact in the control channel. Data channel 1 is selected as the data transmission channel, and the data transmission is switched to data channel 1. In the channel negotiation process of node a and node b, node c is communicating with node d through data channel 2, and it does not know that nodes a and b have selected data channel 1 as the ensuing data transmission. Therefore, if node c selects data channel 1 in the next round of communication, then there may be data packet collision with node a. This is the multichannel hidden terminal problem.

In Figure 2, nodes a and b select data channel 1 when interacting control packet. At the same time, nodes c and d are also interacting control packet to select an idle data channel. Due to long transmission delay, the control packet of node b reaches node d after node d responds node c. Before that, node d mistakes data channel 1 as an idle one, and selects it as the channel for ensuing data transmission. It is clear that the data packets sent by nodes a and d collide at node b. This is the long delay hidden terminal problem.

3.2 Channel allocation problem

In (Ramanathan and Engineer, 1997), a unified framework in the multi-hop network environment was defined based on summarizing previous research. Since the assignable channels are discrete, they can be represented by different colors. Therefore, a channel assignment problem is converted into a graph coloring problem. Firstly, the following definition is made.

Definition 1. A graph G, denoted as $G = (V,E)$, with node set $V = \{v_1,v_2,\cdots\cdots,v_n\}$ representing the network node set; and directional edge set $E = \{e_1,e_2,\cdots\cdots,e_m\}$ representing the inter-node links. The nodes associated with the same edge are neighboring vertices in the network.

Definition 2 (*k*-coloring). The k vertex coloring of G represents a distribution of k kinds of colors, i.e.,1,2,⋯,k, to all

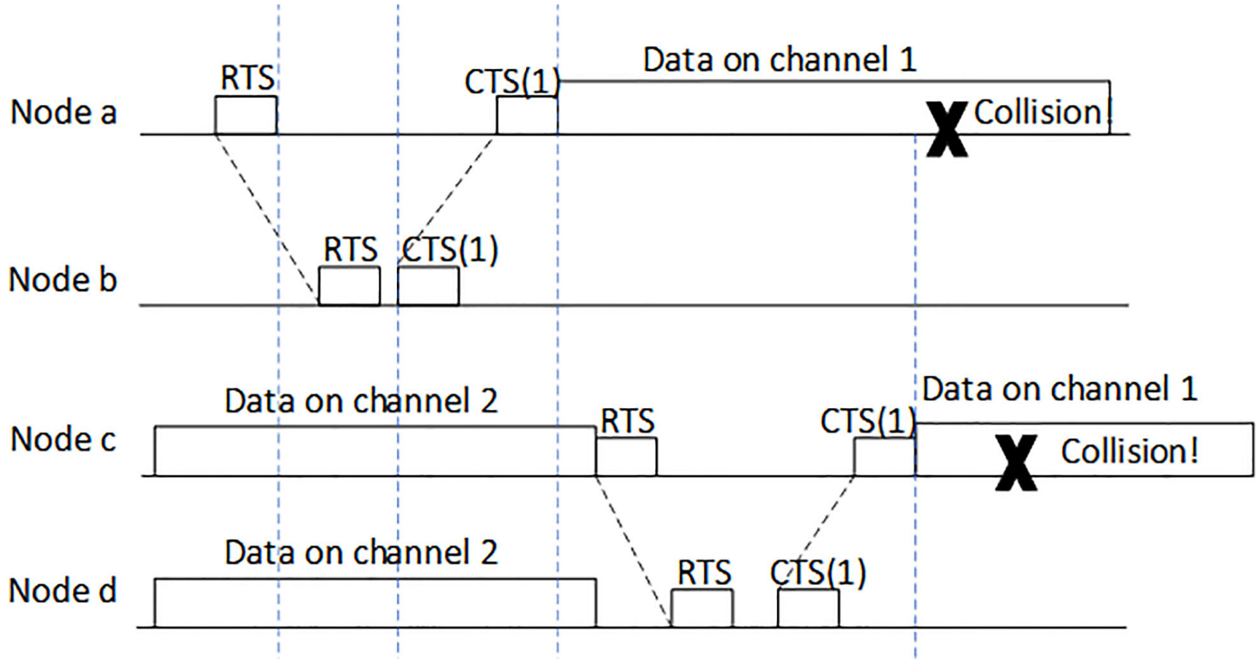


FIGURE 1
Multichannel hidden terminal problem.

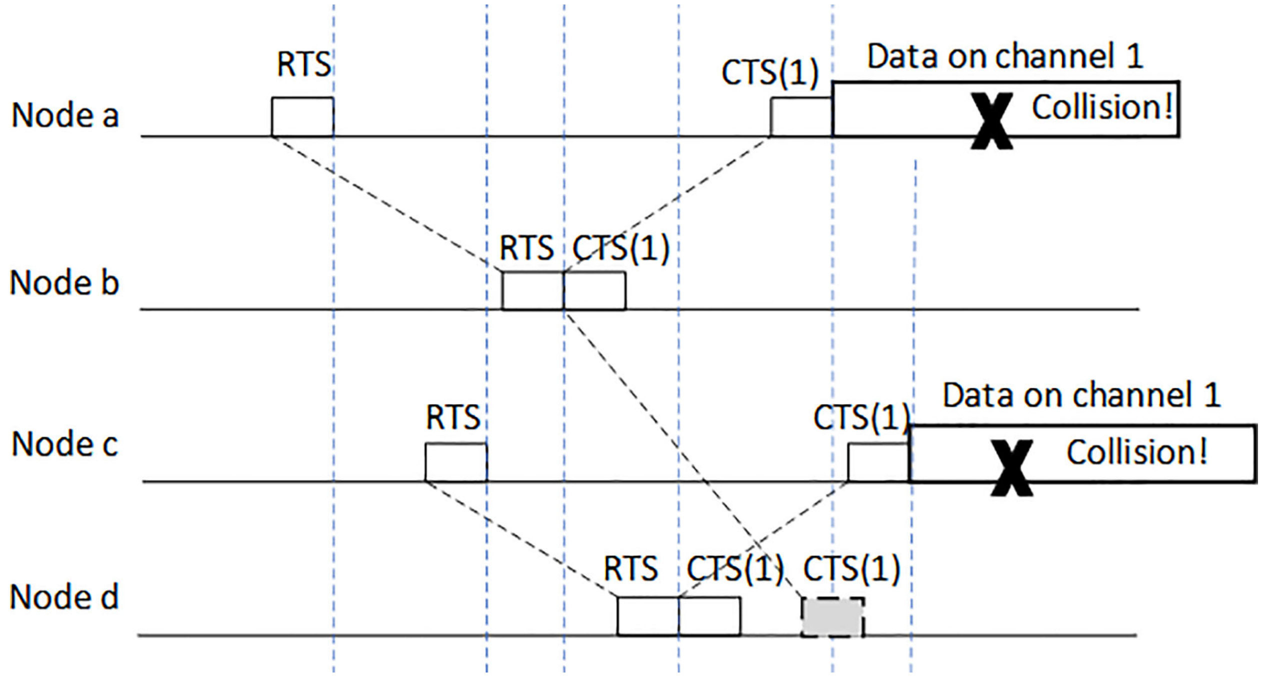


FIGURE 2
Long delay hidden terminal problem.

nodes of *G*. If any two neighboring nodes get different colors, the coloring is called normal. With respect to a network, the channel allocation scheme for neighboring nodes is available, since the neighboring nodes are assigned different channel number.

Definition 3 (Color number). The color number χ of G represents the minimum k such that G is k-color.

Definition 4 (Maximum degree). The maximum degree Δ of G v is any node of G, and the number of edges associated with it is called the degree of v, and denoted as $d(v)$. Then, $\Delta = \max\{d(v)|v \in V\}$.

Definition 5 (Hop distance). The hop distance $d(u,v)$ between two vertices u,v is the length of the reachable path linking the two vertices (using the reachable length of a single hop as the unit). The two nodes with a hop distance being 1 are neighboring nodes mutually. Neighboring nodes can directly communicate or become potential interfering nodes.

According to the above definitions, channel allocation problem is converted into a coloring problem of graph vertices, where the constraint condition is the premise that the channel allocation is available.

It can be known from the underwater acoustic signal attenuation model that the attenuation degrees of signals with different frequencies are different under the condition that communication distance is unchanged. We define the sound source emission spectrum level as *SL*,define the attenuation loss spectrum level as *TL*, define the environment noise spectrum level as *NL*, and define the signal bandwidth as *B*. Then the signal-to-noise ratio (SNR) (in dB) represented in spectrum level is

$$SNR(dB) = SL - TL - (NL + 10{*}\log 10(B)) \tag{1}$$

When receiving node receives signals, the other nodes using the same frequency band to communicate will interfere with the data being received. These nodes are called interference nodes. We define the distance between the sending and receiving node pair as communication distance, denoted as R_d, and define the distance between interference node and receiving node as interference distance, denoted as R_i. The distance between the sending and receiving node pair in communication in the form of signal to interference noise ratio (SINR) represented by energy is

$$SINR = 10\lg\frac{P_{rs}}{P_{ri} + \sigma_n^2} \tag{2}$$

where P_{rs} is the power of received signal; P_{ri} is the power of interference signal; and σ_n^2 is the power of environment noise. Figure 3 shows the SINR versus interference distance. In the figure, the transmission power is adjusted according to different communication distances, so that the interference-free SINR is 10dB. Meanwhile, to facilitate the comparison of $SINR_s$ with different communication distances, the interference distances R_i are normalized with respect to communication distances R_d.

When the interference node is far away from the receiving node, SINR will be larger. When R_i is two times of R_d, SINR is

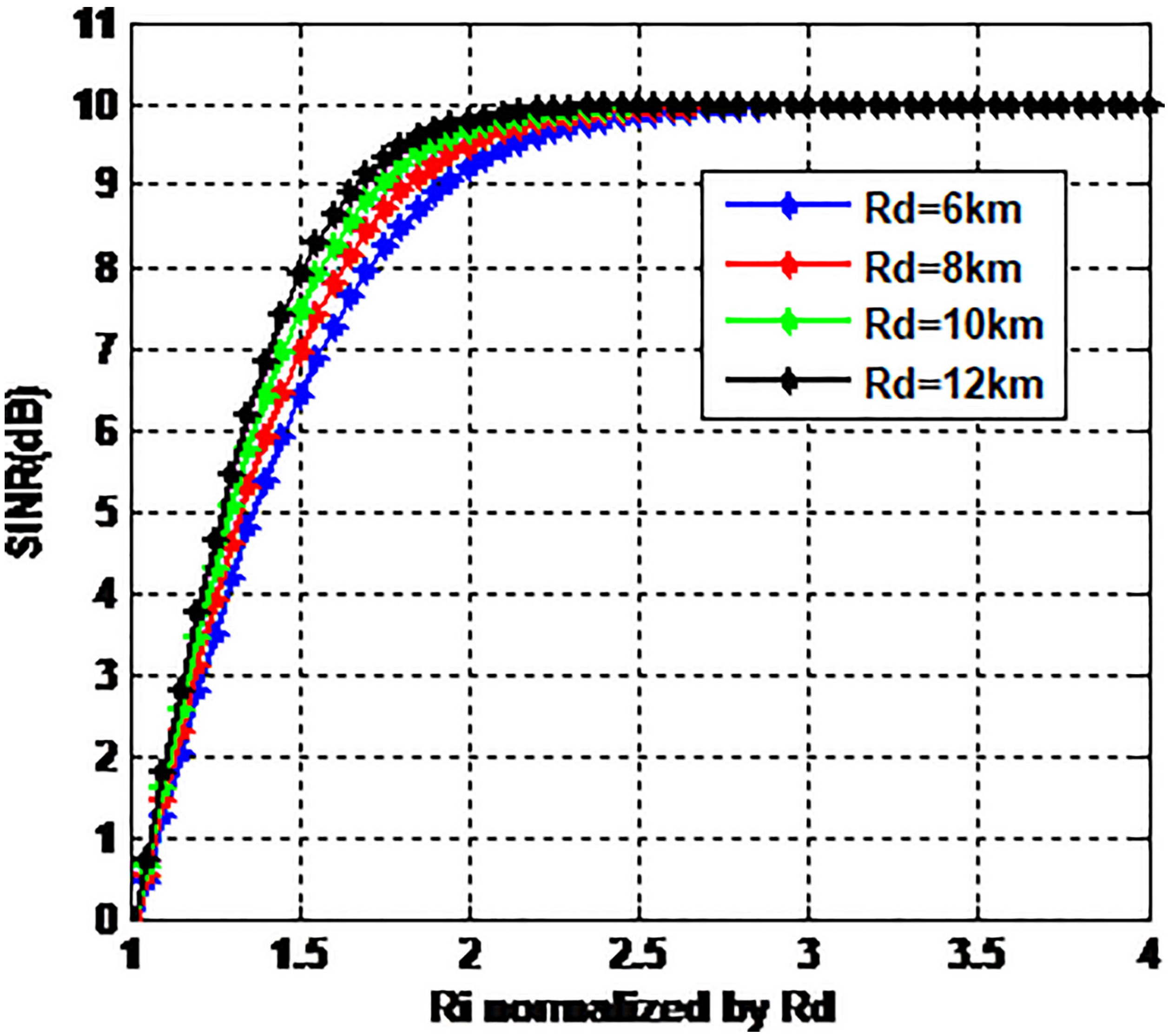


FIGURE 3
SINR versus interference distance.

close to that of interference-free cases. In such situations, the interference nodes have almost no effect on receiving nodes. So, the following interference constraint condition is defined.

Definition 6 (Interference constraint condition). Let f be the mapping for distributing the vertices $V(G)$ of graph G, with channel number set {0,1,2,⋯}. The mapping is denoted as $f:V(G)\rightarrow\{0,1,2,\cdots\}$. It satisfies: for any two nodes x,y, if $d(x,y) = 1$, then $|f(x) - f(y)| \geq 2$; if $d(x,y) = 2$, then $|f(x) - f(y)| \geq 1$.

The constraint condition of Definition 6 is the graph theory model in frequency allocation research——$L(2,1)$ label (Chang and Kuo, 1996). The $L(2,1)$ label problem is an NP-complete problem. In this paper, the research object is the MAC protocol in distributed networks. Therefore, distributed algorithms are adopted for channel allocation. Compared with concentrated algorithms, distributed algorithms can only obtain the information of neighboring nodes. Therefore, in limited communication phases and channels, it is more difficult to obtain coloring scheme satisfying the requirements. FirstFree is a basic distributed Δ+1-coloring algorithm, in which each node collects the existing coloring information of neighboring nodes and chooses an unused color for coloring. Since the maximum degree of a graph is Δ, it is guaranteed that we can use Δ+1 colors for available coloring. Its drawback is that if there are n nodes in the graph, then it is needed to execute the FirstFree algorithm n times, and no two or more nodes can be simultaneously colored in each round of coloring. It is clear that multiple rounds of communication reduces energy efficiency, prolongs control phase, and reduces the effectiveness of data transmission. Ideally, one round of communication is adopted to complete channel allocation. It can be mathematically proved that there exist possible coloring schemes when using only one round of communication. For example, it was proved in [31] that the upper bound of color number using one round of communication is $5\Delta^2\log n$. A distributed single-round coloring algorithm with an upper color number bound of $\Omega(\Delta^2/\log^2\Delta+\log m)$ is proposed, where m is the initial color number. In this case, the number of needed colors is much greater than $\Delta+1$, and the available narrow bandwidth of underwater acoustic environment makes the underwater acoustic nodes unable to provide the required number of channels. Therefore, completing the channel allocation for a limited number of channels in only one round of communication is the major difficulty and key to multichannel MAC protocol design for UASNs.

4 System model

Consider an underwater acoustic sensor network for environment surveillance. $N = I \times J$ sensor nodes (SNs) are evenly distributed in a sea area to monitor the physical quantities, such as temperature, salinity and water velocity. The raw sensed data of each SN are fused at the aggregation node away from the surveillance region. The chart of the physical quantities is drawn and shown in Figure 4A [33].

In this study, the SNs in the surveyed sea region communicate in the form of virtual clusters. Several SNs constitute a virtual cluster and select a cluster head (CH). The CH firstly collects sensor data in the cluster, and multi-hop communication is performed between clusters. The multichannel protocol established in this study completes the inter-cluster communication, as shown in Figure 4B, where arrows indicate the routing of data transmission.

The protocol is based on the following hypotheses:

(1) Each node is only equipped with a half-duplex transceiver, and sending and receiving cannot be simultaneously performed. In the same time, a node can only work in one channel. When a node is intercepting a certain channel, it cannot perform carrier sense for other channels. The transceiver can dynamically switch needed channels.

(2) According to different frequencies, n sub-channels are divided, with nodes transmitting data packets in sub-channels. No special control channel is set for control packets. Control packets are transmitted in the whole band.

(3) The whole network has a unified clock.

(4) Each CH node has at most N neighboring CHs within one hop as its neighboring nodes. A maximum transmission range is also set, so that the maximum

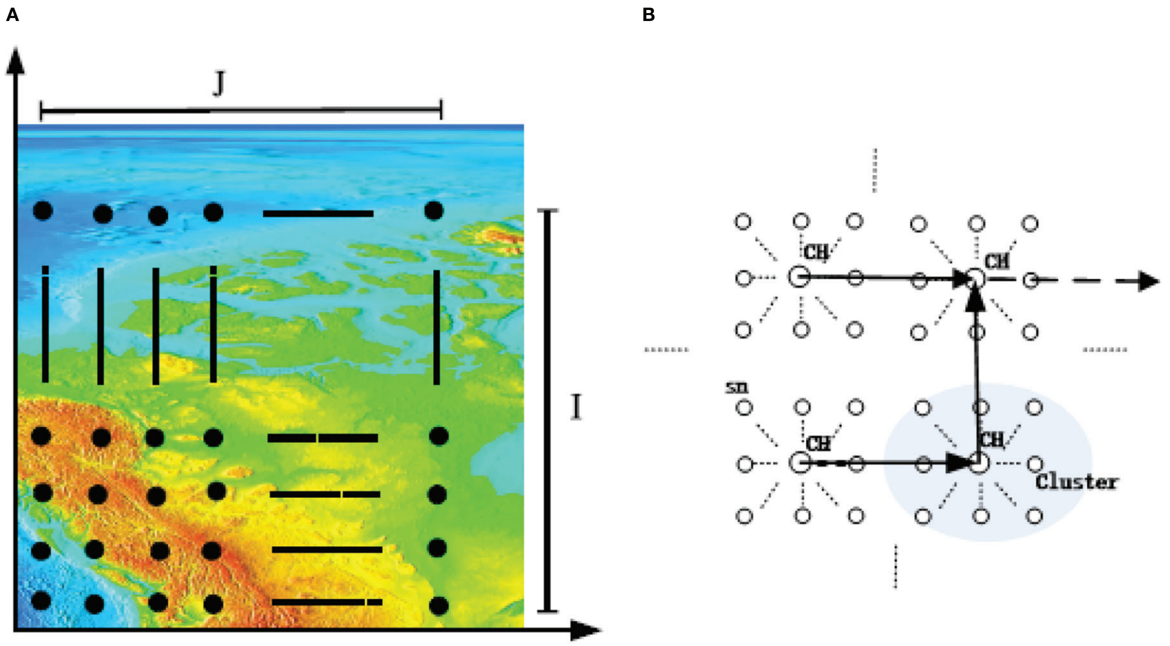


FIGURE 4
Network model. **(A)** Surveillance region where N SNs are distributed; **(B)** Clustered communication of regular network.

transmission delay between neighboring nodes is known, and denoted as τ_{max}.

(5) In one data transmission cycle, the underwater acoustic links are bilateral symmetry, and the link quality remains basically unchanged.

5 Protocol description

The MAC protocol designed in this study tries to solve the channel hidden terminal problem and channel allocation problem in multichannel protocols. To solve the channel hidden terminal problem, a synchronized scheduling mechanism is adopted. The receiving node is set as the viewpoint to intercept the working state information of the interference nodes among neighboring nodes, and choose channels without interference nodes to transmit data. The solution of the channel allocation problem is based on graph coloring theory, and a two-step strategy is adopted, namely, "initial allocation" and "dynamic adjustment".

5.1 Synchronization working process

As shown in Figure 5, the horizontal axis represents time, denoted as symbol t; while the vertical axis represents frequency, denoted as symbol f. The protocol divides the time axis into repeated frames, denoted as Frame 1, Frame 2, and so on. Each frame represents a beacon period, and each beacon period is composed of three stages, namely, transmission negotiation stage, channel selection stage and data transmission stage. The transmission negotiation stage is made up of transmission request and transmission response, with T_{TR} representing the duration. The propagation delay of transmission request and transmission response is set to the maximum propagation delay. The channel selection stage is made up of selection request and selection response, with T_{CS} representing the duration. The propagation delay of selection request is set to the twice of maximum propagation delay, while the propagation delay of selection response is set to the maximum propagation delay. The data transmission duration is denoted by T_D, with the propagation delay of data transmission set to the maximum propagation delay. The protocol divides the available frequency

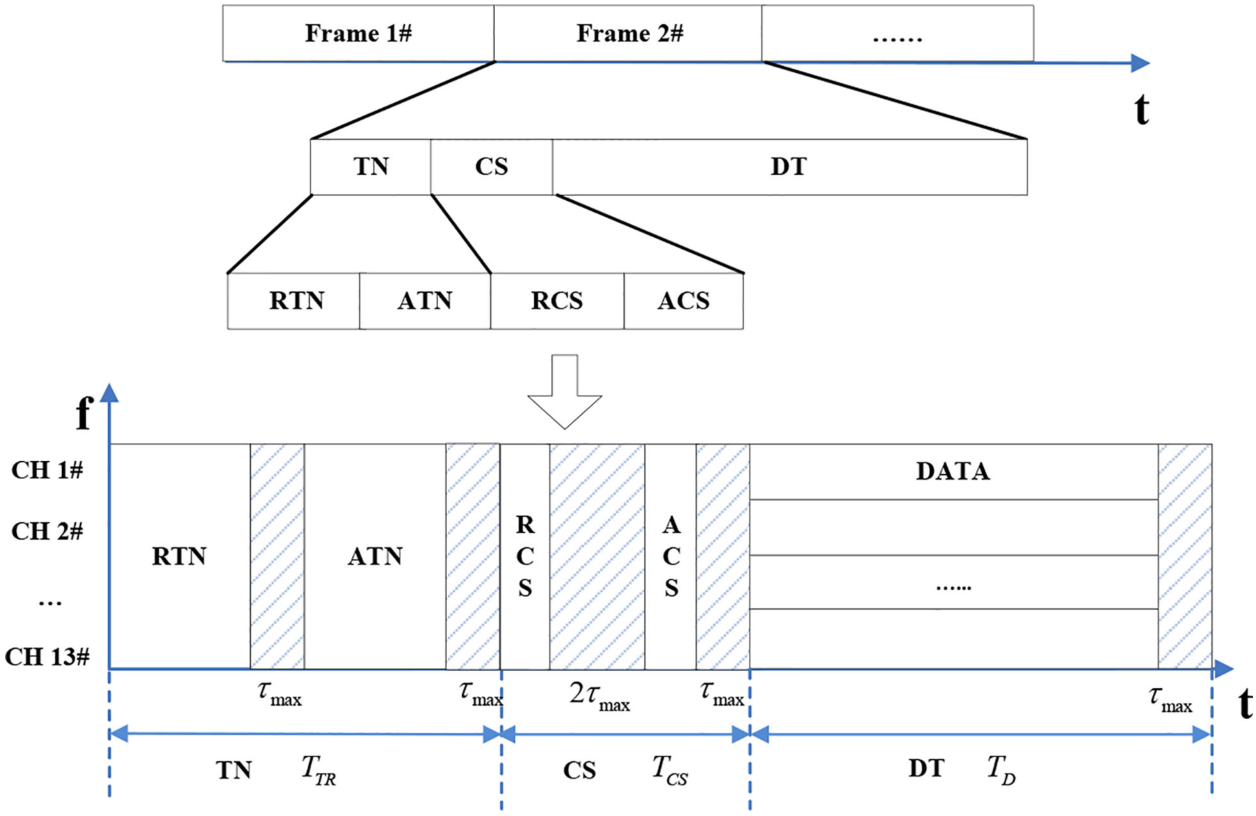


FIGURE 5
Synchronized work timing with frames being cycles.

resource into 13 channels, which are channel 1 to 13, shown in Figure 5. The frequency resource is used as distributable resource. However, the distributable resource is not limited to frequency resource, it also includes time, code channel, etc.

First, the nodes with sending requests send control packet RTN in the transmission negotiation stage. The ID information of the source node and sink node is loaded in the control packet RTN. ID is the identification of nodes, and each node has a unique ID. The sink node adopts a kind of appropriate scheduling strategy to respond to control packet ATN. The ID information of source node and sink node is loaded in the control packet ATN, and the relationship of transmission node pair is thereby established. Second, the sending node of the sending and receiving node pair with established transmission relationship sends channel selection request control packet RCS in the channel selection stage. The ID information and coordinate information of the source node and sink node are loaded in the control packet RCS. The receiving node of the node pair not only receives the RCS packet from the sending node, but also receives other RCS packets from neighboring nodes. At this moment, sink nodes allocate channels, and respond with control packet ACS. The Channel group information ready for transmission is loaded in the control packet ACS. Finally, the data are transmitted, and transmission nodes transmit data packets in the channel group determined by the negotiations of sending part and receiving part.

It can be concluded from the above timing design that the designed multichannel MAC protocol is a kind of protocol which first competes for transmission and then chooses channels, which is contrary to Multichannel MAC protocol (MMAC) (So and Vaidya, 2004). The reason is that in the underwater acoustic network environment with long propagation delay, performing competitive transmission first not only makes it easy to sense the working status of each node, but also provides the ensuing channel selection with accurate global information and ensures that the selected channels do not collide among neighboring nodes. Therefore, the negotiation mechanism in the transmission negotiation stage is essential.

5.2 Transmission negotiation

Nodes with sending tasks randomly select a moment to send control packet RTN in transmission request sub stage. The ID information of the source node and sink node is loaded in the control packet RTN. At this stage, sending/receiving collision or receiving/receiving collision may happen. Sending/receiving collision means that a node has successfully sent RTN packet and received RTN packet from another node. Receiving/receiving collision means that a node has successfully received multiple RTN packets from other nodes. As shown in the bottom left of Figure 6, node A has successfully received RTN packets sent from nodes B and C successively, and it has to select one node of the two to respond. This is called receiving/receiving collision. As shown in the top left of Figure 6, node A has successfully received RTN packet sent from node C, and also has sent RTN packet to node B. It has to decide whether to respond to node C. This is called sending/receiving collision.

The solution to sending/receiving collision is to give priority to respond the sending request of the neighboring node, as shown in the top right of Figure 6. Node A selects node C as the sink node for sending response packet ATN. The solution to receiving/receiving collision is to respond to one sending request of the neighboring nodes, as shown in the bottom right of Figure 6. Node A selects node B as the sink node for sending response packet ATN, and ignores node C.

5.3 Rulers for selecting channel

For convenience, the network graph is redefined as follows:

Definition 7. Network graph G = (V,E,a,b), where V is the node set; E is the one-hop link set of nodes; a is the minimum

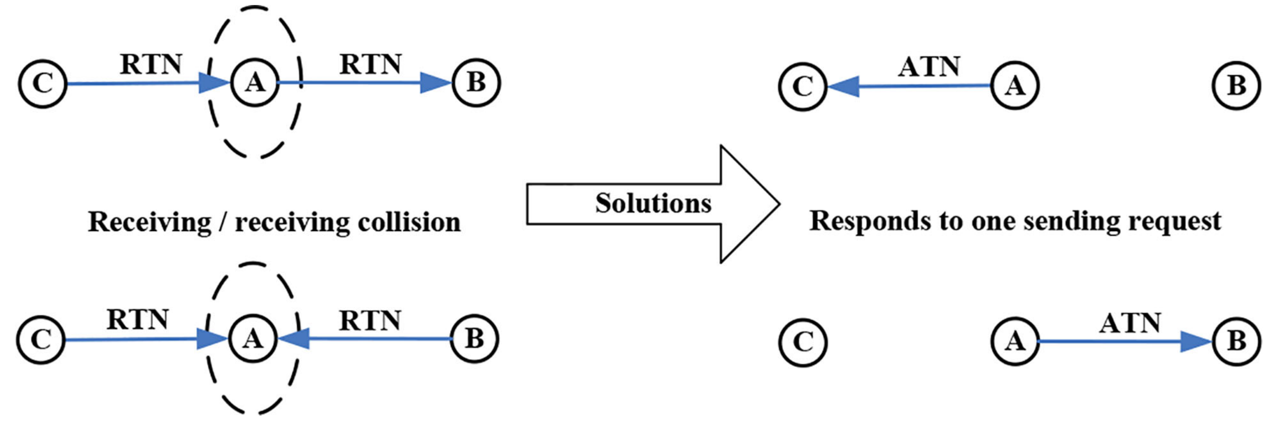


FIGURE 6
Transmission negotiation mechanism.

interval between the channels allocated for neighboring nodes, and b represents the minimum interval between the channels allocated for two nodes with hop distance being 2. According to the above definition, the channel allocation algorithm in this study is divided into two stages, namely, "initial allocation" stage and "dynamic adjustment" stage. First, when the clusters are formed, each CH sets the sink as the center, and defines a coordinate for itself according to its geographical location. Since a regular rectangle network is adopted in this study, each node can easily obtain its coordinate according to its location. The labeling result of nodes is shown in Figure 7.

Then, allocate a channel number $f(x,y)$ for the node with coordinate (x,y), which is expressed in (3):

$$\begin{cases} f(x,y) = x \cdot a + y \cdot (3a+b)(mod \quad N) \\ N = 5a + 3b \end{cases} \tag{3}$$

From the interference constraint condition described in Definition 6, the parameters in Definition 7 are set as $a = 2,b = 1$. Substitute the condition $a = 2,b = 1$, and we can obtain

$$f(x,y) = 2x + 7y(mod 13) \tag{4}$$

The above "initial allocation" regularly allocates a channel for each node, and the total number of channels used in the whole network is 13. Although the inter-node interference is suppressed, channel resource may be wasted. Therefore, in the second stage, a dynamic adjustment method is adopted. The dynamic adjustment strategy is to borrow unused channel numbers from idle neighboring nodes. The more the channels that can be borrowed, the higher the channel utilization rate will be. However, borrow may break the constraint condition in "initial allocation". Therefore, the design should consider the allocation scheme of "dynamic adjustment" and real-time channel situations.

The idea of algorithm design is that in a handshake information exchange process, the unused channels of idle neighboring nodes are first selected. Then, the channels possibly used by potential interference nodes are predicted, and the channels which may interfere with receiving are removed from the selected channels. The detailed allocation steps are as follows:

(1) Establish neighboring node list. Define the 8 nodes listed in Table 1 as neighboring nodes, with each neighboring node having its information, including its ID, coordinate, initially allocated channel number and current status, etc. The nodes can determine the status of their neighbors after RCS packet sending in channel negotiation stage. Assume the coordinate of current node is (i, j). The initial channel number calculated by (4) is iCH. The initial channel numbers of neighbors are calculated as $nbCH = \{nbCH_1, nbCH_1, \cdots, nbCH_8\}$.

(2) Establish usable channel vector. By default, a node can use channel iCH and $nbCH$. Therefore, the usable channel vector is $L = \{iCH\} \cup nbCH$.

(3) After the RCS packet in channel negotiation stage, the nodes fill the status information of neighboring nodes according to their received RCS packet information. When a node receives the RCS packet sent by its neighbors, it becomes sending node; when the receiving node information of the RCS packet is a neighboring node, the neighboring node is a receiving node. Otherwise, it is an idle node.

(4) Traverse the neighboring node list. If the neighboring node with ID being $nbID_i$ is either a sending node or a receiving node, then the usable channel vector is adjusted to $L = L - \{nbCHi\}$.

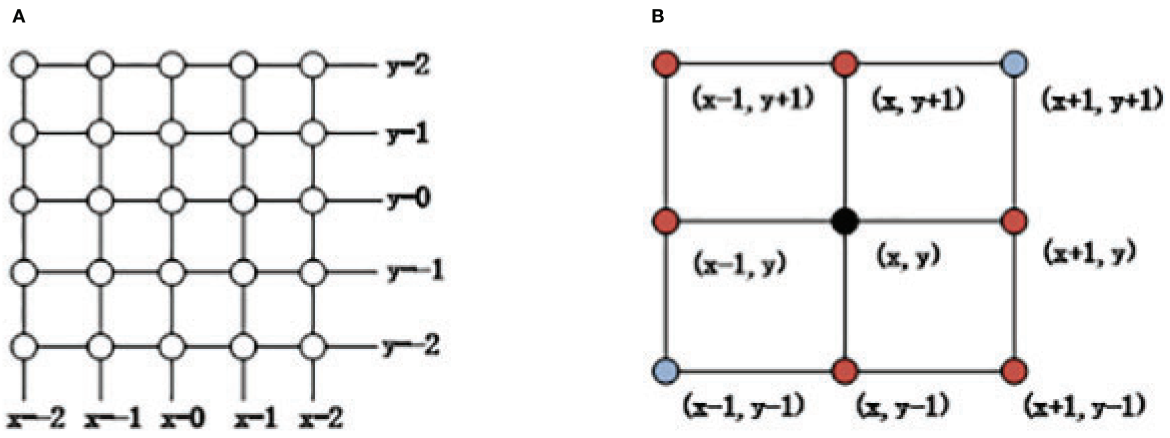


FIGURE 7
Labeling based on geographical locations of nodes. **(A)** Node coordinates **(B)** Coordinate value relations of neighboring nodes.

TABLE 1 Neighboring node list.

ID	Coordinate	Initial channel number $nbCH_i$	Status
$nbID_1$	(i+1,j)	$nbCH_1$	Idle/Sending/Receiving
$nbID_2$	(i-1,j)	$nbCH_2$	
$nbID_3$	(i,j+1)	$nbCH_3$	
$nbID_4$	(i,j-1)	$nbCH_4$	
$nbID_5$	(i+1,j-1)	$nbCH_5$	
$nbID_6$	(i-1,j+1)	$nbCH_6$	
$nbID_7$	(i-1,j-1)	$nbCH_7$	
$nbID_8$	(i+1,j+1)	$nbCH_8$	

(5) Define symbol xRCS as the RCS packet of sink node other than current node. xRCS packet includes the ID information of sending node and receiving node as well as their corresponding coordinate information. The potential interference node list is established according to this information, as shown in Table 2. Since the coordinates of receiving nodes are known, it is easy to solve the coordinates of their neighbors, and thereby solve the channel initially allocated to their neighbors. Moreover, the term “whether unilaterally intercept” in the table means whether receiving node xrID can intercept the RCS packet sent by the sending nodes corresponding to the current node. It can be determined by the location information of nodes.

(6) Establish the potential interference channel matrix *IL*. This matrix represents the channels that may be used by receiving node xrID. Assume the channel used by xrID is $xrCH_i$, and the channel used by its neighbor is $xrbCH_i$, then $IL_i = xrCH_i xrbCH_i$. Each potential interference node corresponds to a row of the interference matrix.

(7) Among the potential interference nodes, the ones that can “unilaterally intercept” have the highest priority to select channels. Others’ priority is discriminated by ID. The nodes with smaller IDs have higher priority. In other words, the current node must first consider whether the nodes with higher priority are using the same channel. By default, those nodes with higher priority may use all usable channels. Therefore, the rows with higher priority in the matrix *IL* are traversed to solve the usable channel $L = L - (L \cap IL_i)$.

6 Performance evaluation

6.1 Simulation settings

According to the network model described in Section 4, the regular network simulation scenario is shown in Figure 8. The whole network is composed of 24 CHs and 1 Fusion Center (FC). CHs collect the data of each cluster sensor and send the data to the FC according to the path pointed by the arrows. The distance between CHs is 6km, and this distance is set to the one-hop transmission range between nodes. The network model is performed by the professional software OPNET for network simulations.

The other parameters are shown in Table 3. When the sending power is 33W, the SNR of the receiving node 6km away is 10dB.

6.2 Simulation results

6.2.1 Effectiveness and reliability of multichannel protocol under unsaturated status

Define packet arrival rate as the number generated by each frame, and this rate indicates the network traffic volume, denoted by the symbol “Packet Arrival Rate per Frame”. Define channel occupation rate as the ratio of actual transmission rate to channel volume, and this rate indicates the channel multiplexing capability in multi-hop environment, denoted by the symbol “Average Channel Occupancy per Frame”, revealing the

TABLE 2 Potential interference node list.

Potential interference node ID	Receiving node ID	Receiving node coordinate	Whether unilaterally intercept
xsID	xrID	(x,y)	Yes/No
…	…	…	

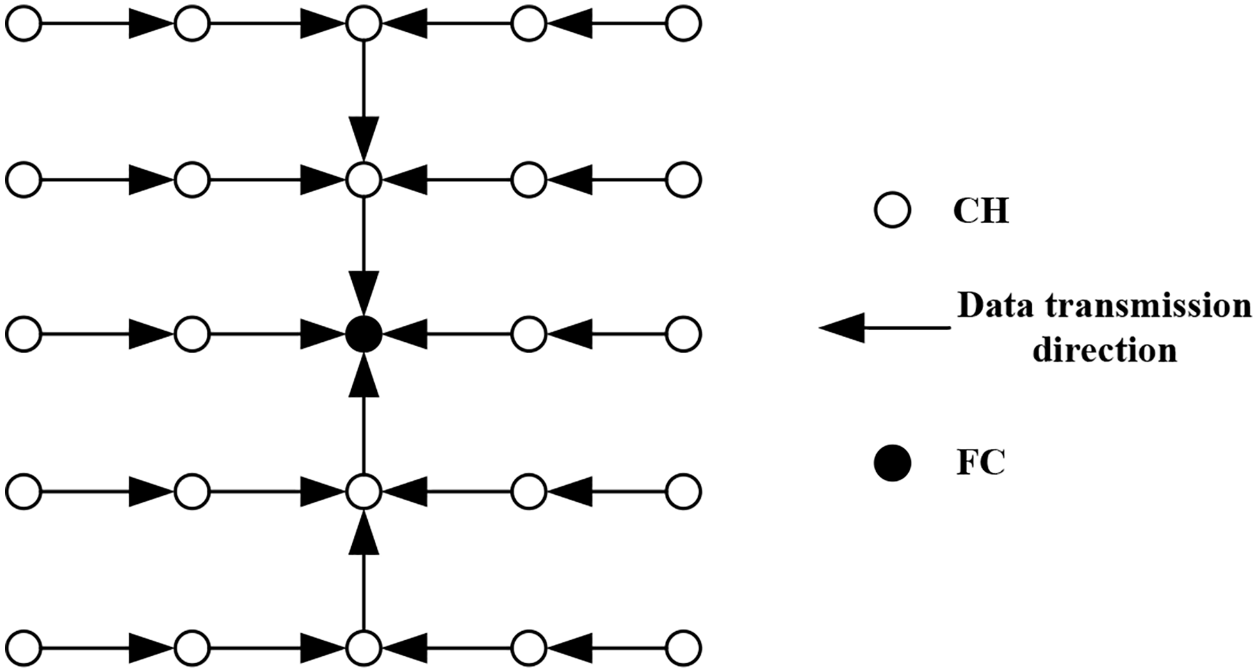


FIGURE 8
Network simulation scenario.

effectiveness of the protocol in the network. The symbol "Data Successful Receiving Rate" represents the rate of successful data packet receiving, which is defined as the ratio of the number of successfully received data packets to that of data packets sent in the whole network, revealing the reliability of the protocol in the network. In the network application scenario in this study, the packet arrival rate of each CH is different. The nodes away from FC have small data amount waiting for transmission and, therefore, the packet arrival rate of them is relatively low, since these nodes do not need to retransmit data sent by CHs in last hop or only retransmit a few. Conversely, the packet arrival rate is relatively high. When the packet arrival rate is relatively low, not every CH has data to send in the frames of each slot, which is called the unsaturated status of a network. In Figure 9, "MulCh" represents the multichannel MAC protocol established in this study, while "SingleCh" represents single-channel MAC protocol.

It is clear in Figure 9A that with the increase of packet arrival rate, the channel occupation rate of the multichannel MAC protocol also increases. When the packet arrival rate reaches 1, all network nodes are in saturated status, and the channel occupation rate also tends to be saturated at this moment. It can be known from the previous demonstration that in underwater acoustic environment, the neighboring nodes of the single-channel MAC protocol,if they are out of one hop and within two hops, may interfere with the receiving of data packets. In this situation, even the collision avoidance negotiation of control packets cannot avoid the data packet collision between the neighboring nodes of single-channel MAC protocol. Therefore, when traffic volume is relatively small, the channel space multiplexing rate is relatively low. There are fewer data packet collisions, and the possibility of successful data packet receiving increases, showing relatively high channel occupation rate. Otherwise, when the traffic volume is

TABLE 3 Network simulation parameters.

Variable	Parameter	Variable	parameter
Control packet RTN length (bits)	40	Control packet ATN length (bits)	56
Control packet RCS length (bits)	72	Control packet ACS length (bits)	56
Data packet DATA length (bits)	1072	Number of data channels	13
Control channel rate (bits/s)	1000	Data channel rate (bits/s)	200
Sending power (W)	33W	Receiving power (W)	1

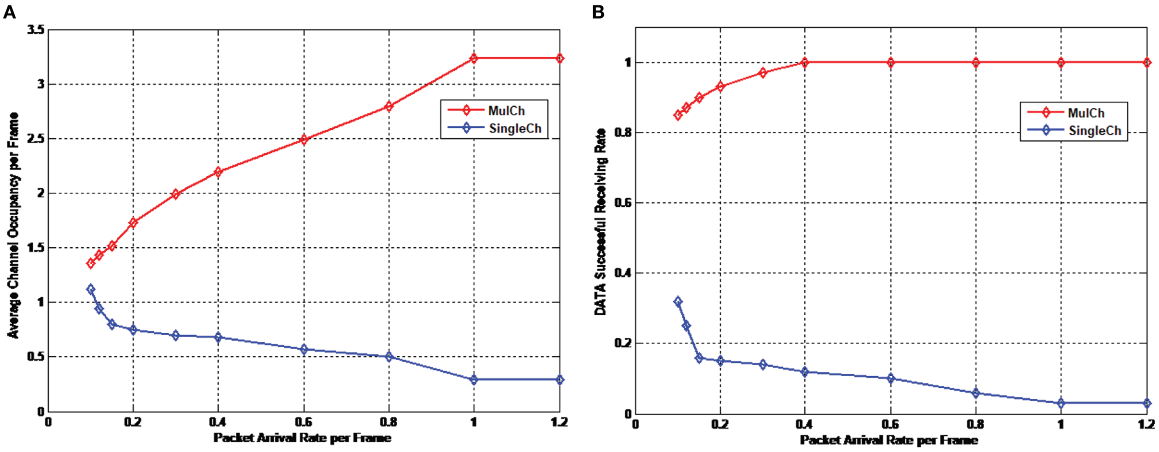


FIGURE 9
Effectiveness and reliability of multichannel protocol in unsaturated status. **(A)** Channel occupation rate **(B)** Successful data packet receiving rate.

relatively large, the control packet interaction within the traditional single-hop range is unable to avoid space multiplexing, so that the possibility of data packet collision increases, resulting in relatively low channel occupation rate. In comparison, the proposed multichannel MAC protocol uses the channel request packet within two-hop range to broadcast the channel utilization status of the current node, and adopts reasonable channel allocation algorithm to guarantee that no neighboring nodes within two-hop range can use the same data channel to send data packets, dramatically increasing the successful data packet receiving rate, and exhibiting relatively high channel occupation rate.

The successful data packet receiving rate in Figure 9B also verifies our analysis. In the figure, no matter what the packet arrival rate is, the successful data packet receiving rate of the multichannel protocol is over 80%. Moreover, with the increasing of packet arrival rate, all nodes in the network have data to transmit, instead of being limited to the CHs at FCs. Consequently, the transmission node pairs spread more sparsely, and the successful data packet receiving rate also increases. When the packet arrival rate reaches 0.4, it is possible to successfully receive all data packets sent.

6.2.2 Effectiveness of multichannel protocol in saturated status

In ideal channel allocation algorithms, a multichannel protocol does not change the channel occupation rate in the network due to the control packet collision in the transmission negotiation stage. This is because although the RTN packet and ATN packet collide in the transmission negotiation stage, the only result is that fewer nodes have channel selection request in channel selection stage. At this moment, directed by channel allocation algorithm, each transmission node pair obtains more channels for data transmission, and keeps the channel occupation rate unchanged. To verify the adaptability of the proposed channel allocation algorithm in the control of packet collision, a method is adopted in the simulation experiment in which the duration of transmission request sub-stage is reduced, so that the RTN packet collision is intensified. The variation of the average channel occupation rate under different RTN packet collisions is observed and shown in Figure 10. In the figure, the horizontal axis represents the number of time slots when sending RTN packets in the transmission request sub-stage, denoted by the symbol “RTN Slots”. The longer the time of control packet interaction is, the more the allocable slots will be; and the fewer the time slots are, the larger the possibility of RTN packet collision is. In Figure 10A, the symbol “RTN Successful Receiving Rate” represents the successful RTN packet receiving rate. In Figure 10B, the symbols “Mulch @Saturation” and “Single @Saturation” represent the performance of multichannel protocol and single channel protocol under saturated status, respectively. “Mulch@up” and “Single@down” represent the upper boundary of multichannel protocol performance and the lower boundary of single channel protocol performance, respectively.

It is clear in the figure that with the increase of the RTN slots, the successful RTN packet receiving rate increases, and so does the channel occupation rate of the multichannel protocol, approaching the channel occupation rate in ideal situation. The performance curve shows a relatively flat shape, which demonstrates that the channel allocation algorithm can relatively well adapt to the control packet collision, and does not dramatically reduce the channel occupation rate due to the control packet collision. The channel occupation rate of the single-channel protocol decreases with the increase of successful control packet receiving rate. This is because the more the transmission node pairs are, the higher the possibility that the data packets are interfered with is, and the lower the successful

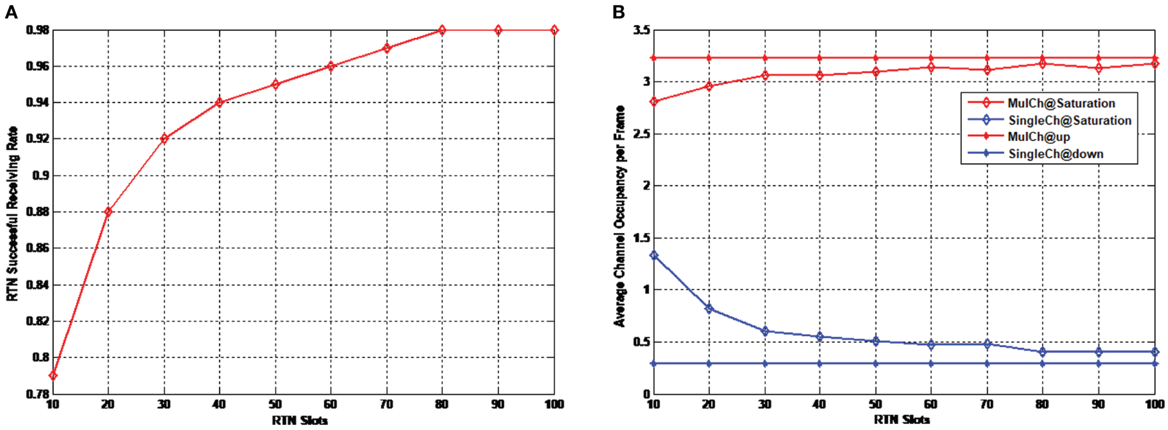


FIGURE 10
Effectiveness of multichannel protocol under saturated status. **(A)** Successful RTN packet receiving rate **(B)** Average channel occupation rate.

receiving rate will be. When there is no collision of control packets, the channel occupation rate approaches the lower bound of performance.

6.2.3 Energy efficiency

The energy consumption in the network includes the energy consumed by control packet sending and receiving as well as that consumed by data packet sending and receiving. We define average energy consumption per bit as the ratio of the total energy consumption to the number of successful received bits, with Joule being its unit, denoted as "Average Energy Consumption per Bit" in Figure 11. And also we define energy efficiency as the ratio of the energy consumption of successful data packet transmission to the total energy consumption, denoted as "Energy Efficiency". The higher the energy efficiency is, the smaller the extra cost brought by data packet collision or control packet interaction will be. "MulCh" and "SingleCh" represent the performance of multichannel protocol and single-channel protocol, respectively. "MulCh @up" and "SingleCh @up" represent the performance's upper bounds of multichannel protocol and single-channle protocol, respectively. Note that they are the performance in ideal status, which means there is no any cost spent.

It is clear in Figure 11A that since the data transmission rate of the multichannel protocol is much lower than that of the single-channel protocol (the ratio in this protocol is 1 to 13). Therefore, the energy consumption per bit of the single-channel protocol in ideal status is much lower. However with the increase of packet arrival rate, the energy consumption per bit of the multichannel protocol tends to be stable, and approaches to the energy consumption in ideal status. In comparison, the energy consumption per bit of the single-channel protocol increases with the increase of packet arrival rate. Moreover, when the packet arrival rate is larger than 0.6, the energy consumption of

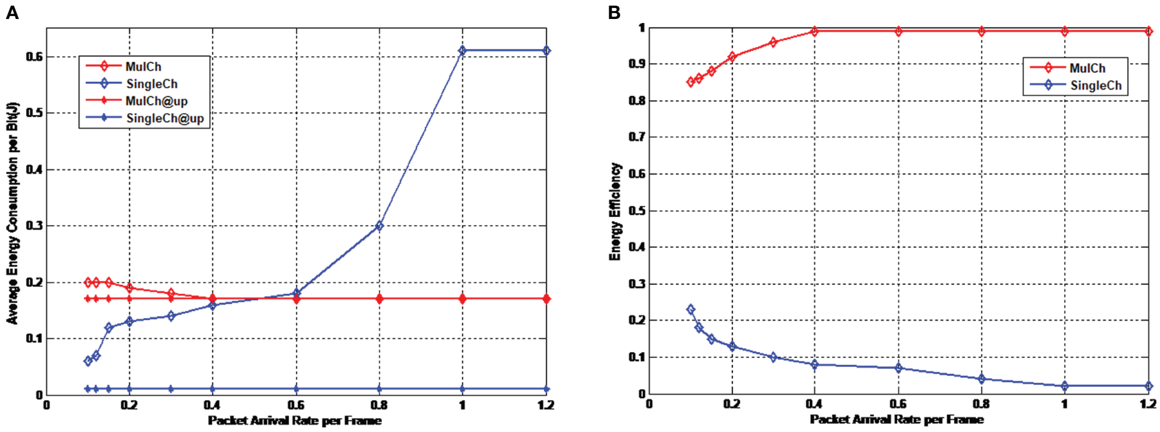


FIGURE 11
Energy efficiency of multichannel protocol. **(A)** Average energy consumption per bit **(B)** Energy efficiency.

single-channel protocol exceeds that of multichannel protocol. From the aspect of energy efficiency shown in Figure 11B, the energy efficiency of the multichannel protocol is significantly better than the single-channel protocol in which more energy is consumed by extra costs (including data packet collision and control packet interaction), demonstrating the high reliability of the multichannel protocol.

It can be concluded that compared with single-channel protocol, the multichannel protocol designed in this study has relatively high channel occupation rate and successful data packet receiving rate under different traffic volumes. Moreover, the proposed multichannel protocol is adapted to control packet collision, and the channel occupation rate approaches the ideal value, demonstrating that the designed channel allocation algorithm can not only avoid collisions, but also utilize all usable channels. Although the increase of channels reduces the data transmission rate and the theoretical energy consumption per bit is much higher than that of single-channel protocol, the increase of successful data packet receiving rate could compensate the energy consumption caused by the data transmission delay

7 Conclusion

In this study, aiming at the multichannel hidden terminal problem and long propagation delay hidden terminal problem in the multichannel MAC protocol design for long-delay UASNs, a multichannel MAC protocol GCMAC is proposed based on graph coloring theory, which is adapted to distributed underwater acoustic network environment. The GCMAC protocol adopts a synchronized work method. It performs negotiation and channel selection before data transmission, and then determines the transmission node pairs and channels to be used. It selects as may idle channels as possible for each transmission node pair based on coloring theory. Simulation results show that the proposed multichannel MAC protocol has relatively high channel occupation rate and successful data packet receiving rate, as well as much higher energy efficiency compared with single-channel MAC protocol.

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

YW and DW contributed to conception and design of the study. DW organized the database. YW performed the statistical analysis. DW wrote the first draft of the manuscript. YW wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 62271427).

Acknowledgments

The authors would like to thank the support from the teachers and students in the Underwater Acoustic Communication Lab of Xiamen University.

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

Alfouzan, F. A., Shahrabi, A., Ghoreyshi, S. M., and Boutaleb, T. (2019). A collision-free graph coloring MAC protocol for underwater sensor networks. *IEEE Access* 7, 39862–39878. doi: 10.1109/ACCESS.2019.2906555

Ansari, S., Poncela Gonzalez, J., Otero, P., and Ansari, A. (2015). Analysis of MAC strategies for underwater applications. *Wireless personal communications: An Int J.* doi: 10.1007/s11277-015-2743-1

Bouabdallah, F., Zidi, C., Boutaba, R., and Mehaoua, A. (2017). Collision avoidance energy efficient multi-channel MAC protocol for UnderWater acoustic sensor networks. *IEEE Trans. Mobile Comput.* 18 (10), 2298–2314. doi: 10.1109/TMC.2018.2871686

Chang, G. J., and Kuo, D. (1996). The L(2,1)-labeling problem on graphs. *SIAM J. Discrete Math.* 9 (2), 309–316. doi: 10.1137/S0895480193245339

Chao, C., Lu, M., and Lin, Y. (2015). Energy-efficient multichannel MAC protocol design for bursty data traffic in underwater sensor networks. *IEEE J. Ocean. Eng.* 40 (2), 269–276. doi: 10.1109/JOE.2014.2304311

Chen, K.C.K., Ma, M.M.M., Cheng, E.C.E., Yuan, F.Y.F., and Su, W.S.W. (2014). A survey on MAC protocols for underwater wireless sensor networks. *IEEE Communications surveys and tutorials* 16, 3, 1433–1447. doi: 10.1109/SURV.2014.013014.00032

Cui, J., Kong, J., Gerla, M., and Zhou, S. (2006). The challenges of building scalable mobile underwater wireless sensor networks for aquatic applications. *IEEE Netw.* 20 (3), 12–18. doi: 10.1109/MNET.2006.1637927

Fitzpatrick, S.S.U.C., Janssen, J.J.M.D., and Nowakowski, R.R.M.D. (2004). Distributive online channel assignment for hexagonal cellular networks with constraints. *Discrete Appl. Math.* 143 (1-3), 84–91. doi: 10.1016/j.dam.2003.06.005

Hsu, C., Lai, K., Chou, C., and Lin, K. C. (2009). *ST-MAC: Spatial-temporal MAC scheduling for underwater sensor networks* (Rio de Janeiro, Brazil: Institute of Electrical and Electronics Engineers Inc), 1827–1835.

Huang, J., Wang, H., He, C., Zhang, Q., and Jing, L. (2018). Underwater acoustic communication and the general performance evaluation criteria. *Front. Inf. Technol. electron. Eng.* 19 (8), 951–971. doi: 10.1631/FITEE.1700775

Jahanbakht, M., Xiang, W., Hanzo, L., and Rahimi Azghadi, M. (2021). Internet Of underwater things and big marine dataanalytics: A comprehensive survey. *IEEE Commun. Surv. Tut.* 23 (2), 904–956. doi: 10.1109/COMST.2021.3053118

Jiang, S. (2018). State-of-the-Art medium access control (MAC) protocols for underwater acoustic networks: A survey based on a MAC reference model. *IEEE Commun. Surveys Tutorials* 20 (1), 96–131. doi: 10.1109/COMST.2017.2768802

Kuhn, F., and Wattenhofer, R. (2006). *On the complexity of distributed graph coloring* (Denver, CO, United states: Association for Computing Machinery), 7–15.

Ramanathan, S., and Engineer, I.O.E.A (1997). “A unified framework and algorithm for (T/F/C)DMA channel assignment in wireless networks,” in *INFOCOM '97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution., Proceedings IEEE.* 900–907.

Rhee, I., Warrier, A., Min, J., and Xu, L. (2009). DRAND: Distributed randomized TDMA scheduling for wireless *ad hoc* networks. *IEEE Trans. Mobile Comput.* 8 (10), 1384–1396. doi: 10.1109/TMC.2009.59

Saifullah, A., Xu, Y., Lu, C., and Chen, Y. (2014). Distributed channel allocation protocols for wireless sensor networks. *IEEE Trans. Parallel Distributed Syst.* 25 (9), 2264–2274. doi: 10.1109/TPDS.2013.185

So, J., and Vaidya, N. (2004). *Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver* (Tokyo, Japan: Association for Computing Machinery). 222–233. doi: 10.1145/989459.989487

Wang, W., and Liu, X. (2005). *List-coloring based channel allocation for open-spectrum wireless networks* (Dallas, TX, United states: Institute of Electrical and Electronics Engineers Inc), 690–694. doi: 10.1109/VETECF.2005.1558001

Zhou, Z., Peng, Z., Cui, J., and Jiang, Z. (2012). Handling triple hidden terminal problems for multichannel MAC in long-delay underwater sensor networks. *IEEE Trans. Mobile Comput.* 11 (1), 139–154. doi: 10.1109/TMC.2011.28

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fmars.2022.1107184

Check for updates

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Changlin Chen,
Fudan University, China
Haocai Huang,
Zhejiang University, China

*CORRESPONDENCE
Xiao-Hua Zhu
✉ xhzhu@sio.org.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 24 November 2022
ACCEPTED 22 December 2022
PUBLISHED 18 January 2023

CITATION
Dai L, Xiao C, Zhu X-H, Zhu Z-N, Zhang C, Zheng H, Chen M and Li Q (2023) Tomographic reconstruction of 3D sound speed fields to reveal internal tides on the continental slope of the South China Sea.
Front. Mar. Sci. 9:1107184.
doi: 10.3389/fmars.2022.1107184

Tomographic reconstruction of 3D sound speed fields to reveal internal tides on the continental slope of the South China Sea

Longhao Dai[1,2†], Cong Xiao[2†], Xiao-Hua Zhu[1,2,3*], Ze-Nan Zhu[2], Chuanzheng Zhang[2], Hua Zheng[1,2], Minmo Chen[2] and Qiang Li[4]

[1]School of Oceanography, Shanghai Jiao Tong University, Shanghai, China, [2]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, [3]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), , Zhuhai, China, [4]Graduate School at Shenzhen, Tsinghua University, Shenzhen, China

Mirror-type coastal acoustic tomography (MCAT) is an advancement of coastal acoustic tomography (CAT) with a mirror function that enables the nearshore transmission of offshore observation data. Subsurface MCATs are used to monitor small-scale oceanographic phenomena over large areas on the continental slope of the South China Sea (300–1,000 m). This study investigated the performance of MCAT network observation for internal tide monitoring using 3D tomographic inversion. MCAT facilitates communication with each other and is networked. Results of 3D tomographic inversion were highly consistent with the Stanford unstructured nonhydrostatic terrain-following adaptive Navier–Stokes (SUNTANS) model data. The root-mean-square difference of the sound speed obtained by the inversion method was< 1.0 m/s. The high sound speeds induced by the internal tide were well captured by the results of 3D inversion. A high sound speed band of > 1,516 m/s was used as an index of internal tides, which propagated at a speed of 1.9 m/s. Sensitivity experiments demonstrated that a swing of the anchor tether of< 5 m had less impact on sound speed inversion, while distance drifts caused by human activities of > 50 m had a significant influence. The missing of one or several stations owing to natural factors did not generate significant variations in the results of network observation. This study demonstrated that the MCAT network observation strategy, equipped with numerous stations, is more effective in capturing the characteristics of internal tides.

KEYWORDS

mirror-type coastal acoustic tomography, 3D inversion, sound speed, internal tide, continental slope of the South China Sea

1 Introduction

Acoustic waves are known to be the best signals for long-distance transmission in the ocean, and ocean dynamic processes such as internal waves, tidal currents, and mesoscale eddies are affected by them. Ocean acoustic tomography (OAT) is an effective tool for measuring mesoscale phenomena in oceans. OAT has a special advantage in obtaining the sound speed and current velocity fields from acoustic travel time measured in mesoscale ranges (Wunsch, 1979). In 1981, the Woods Hole Oceanographic Institution led the first ocean acoustic tomography experiment in the Bermuda Sea (area = 300 × 300 km) south of the Gulf Stream and successfully obtained the three-dimensional (3D) structure of mesoscale eddies, which proved the feasibility of the acoustic tomography method for ocean observation (The Ocean Tomography Group, 1982). Later, a series of OAT experiments was conducted in the United States, Japan, and Europe (Cornuelle et al., 1989; Worcester et al., 1993; Send et al., 1995; Yuan et al., 1999) and mesoscale ocean phenomena such as the Kuroshio meander and deep-sea convection were successfully observed. OAT is a novel approach to simultaneously observe the 3D sound speed fields of ocean mesoscale phenomena. However, owing to the constraints of human and material resources, it is difficult to organize a large number of OAT in the observation domain of several hundred kilometers for small- and medium-scale dynamic processes.

The coastal acoustic tomography (CAT) system was designed by the Hiroshima University based on the advanced technologies of OAT for the observation of small-scale ocean dynamic phenomena in coastal seas. Many CAT observations have been successfully achieved near the shore (Zheng et al., 1997; Zheng et al., 1998; Yamoaka et al., 2002; Yamaguchi et al., 2005; Zhu et al., 2013). Compared with OAT, CAT observations typically cover only tens or several tens of square kilometers; however, it has a higher observation frequency and finer spatiotemporal resolution to meet the requirements of coastal sea observations. CAT has demonstrated its ability to capture small-scale tidal vortices. Yamoaka et al. (2002) successfully captured the processes of initiation, growth, translation, and decay of tidal vortex pairs with five CAT stations in the Nekoseto Strait, Japan. Zhu et al. (2013) conducted a CAT experiment with seven stations in Zhitou Bay, Zhoushan, and observed the dynamic processes of small-scale tidal current circulation over a sea area of 12 × 12 km. On the other hand, Park and Kaneko (2001) and Zhu et al. (2017) assimilated CAT data into the ocean circulation model and reconstructed the fine flow field in the inland sea and around the island, respectively. Zhang et al. (2015) and Chen et al. (2017) discussed upwelling after typhoon passage by inversion and assimilation of CAT data obtained in Hiroshima Bay, Japan, respectively. The CAT is an innovative observation method for small-scale flow fields.

Mirror coastal acoustic tomography (MCAT) is an advanced form of CAT that is enhanced with a mirror module for travel time summation. MCAT is designed to transfer offshore observation data nearshore in real time using travel time summation. Chen et al. (2018) introduced the working principle of MCAT in detail and successfully conducted MCAT experiments in the Nekoseto Strait, Japan, thereby verifying its feasibility. Subsequently, Syamsudin et al. (2019) operated MCATs to successfully capture the subsurface structure of large waves, which were generated over a sill near the southern inlet of the Lombok Strait, Indonesia and propagated northward, demonstrating the superiority of MCAT. However, to date, the network observation of MCATs has never been realized for propagating internal waves.

The unique dual-ridge topography of the Luzon Strait in the South China Sea (SCS) interacts with barotropic tides to generate energetic internal tides and waves. Internal tides and waves generated in the Luzon Strait spread to > 1,600 km southwestward along the deep-sea basin of the SCS and reach the area around the Xisha Islands (Buijsman et al., 2010; Zhao, 2015). Some of them propagate to the western continental shelf of the SCS at shallower water depths. The internal tides amplified and their crests were steepened to generate internal solitary waves (Li and Farmer, 2011). The internal solitary waves continued to propagate westward and eventually disappear nearshore owing to bottom friction. The vicinity of islands and reefs in the SCS is also a place where internal waves and other small- and medium-scale dynamic phenomena are relatively active. In CAT, travel time data are generally stored in individual subsurface stations, and data analyses are performed after system recovery. It is relatively easy to recover the systems during normal estuarine observation; however, it is difficult to observe ocean phenomena in the SCS. The moored MCAT in real time is suitable for SCS for effectively monitoring the small- and medium-scale dynamic processed in the latter.

This study is based on the virtual MCAT network observation with the aim of simulating the 3D sound speed fields of mesoscale and sub-mesoscale dynamic phenomena (e.g., internal tides and waves) in the SCS. The possibility of observing internal wave propagation using an MCAT network based on results of simulation is discussed, and a theoretical basis is provided for future MCAT network observation.

2 Forward problem

2.1 Numerical model for internal tides

Hourly oceanographic conditions, with temperatures, salinities, and currents in the SCS, are modeled using the parallel, unstructured grid, non-hydrostatic ocean model termed Stanford unstructured nonhydrostatic terrain-following adaptive Navier–Stokes simulator (SUNTANS) (Fringer et al., 2006), with a mean resolution of 1,190 m, in the Luzon Strait, high-resolution grids with a minimum resolution of 89 m that were settled. The detailed description of the model configuration and performance analysis were presented in Zhang et al. (2011). A total of 336 h of data were obtained (21:00 January 3–21:00 January 17, 2018).

Figure 1A shows the amplitude of the isotherm induced by internal tides, which was estimated from the amplitude of the 15° C isotherm (Zhao et al., 2010; Zhang et al., 2011). The internal tides generated in the Luzon Strait (the amplitude of the 15°C isotherm was > 70 m) eventually propagate into the SCS. In this study, an experimental site (encircled with a black rectangle in Figure 1) was selected near the western edge of the high-amplitude area (~50 m) and adjacent areas of the Dongsha Islands. The high variations in temperature and salinity fields were attributed to the temporal variability of internal tides. Figure 1B shows an example of the sound speed field at hour 243, demonstrating that the experimental area was located in the area with a depth range of 300–1,000 m and featured energetic internal tides with a wavelength of ~82.1 km and phase speed of 1.9 m/s.

2.2 Arrangement of virtual MCAT stations

Following the tomographic concept, numerous MCAT systems were designed for the experimental area. MCAT equipped a mirror function, which is similar to mirror transponder (MT) technology that has been used in seafloor geodetic centimeter-level positioning by shipboard and triangular MT arrays (Chen et al., 2018).

Virtual MCAT stations were determined following the principle that, for a fixed observation area, the greater the number of received acoustic rays, the stronger the level of

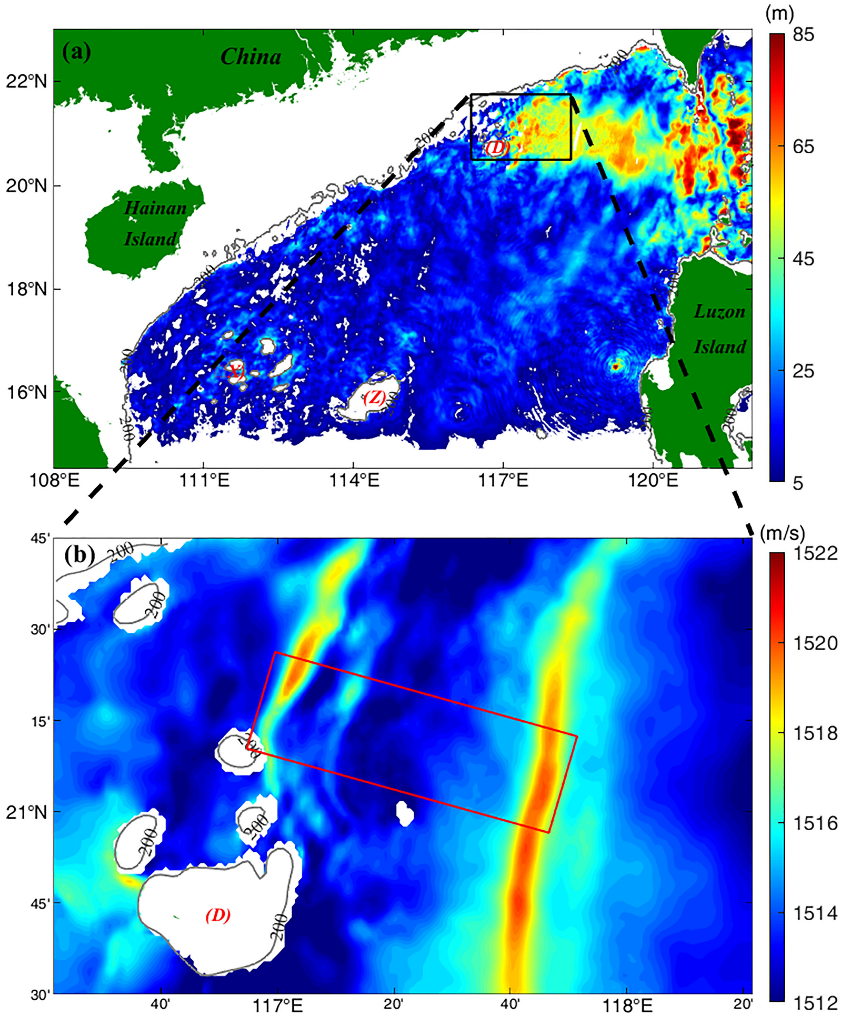


FIGURE 1
(A) Amplitude map of internal tides generated at hour 243 in the South China Sea during the application period of the SUNTANS model; **(B)** Magnified map of the area denoted by the black rectangle in **(A)** with color maps of sound speed at a depth of 200 m. The red rectangle represents the MCAT area. (D), (X), and (Z) indicate Dongsha, Xisha, and Zhongsha Islands, respectively.

received sound. Rays were simulated using Bellhop (Porter, 2011; Taniguchi and Huang, 2015). Eventually, 50 virtual MCAT stations were arranged in the selected rectangular observation area of 100 × 30 km (Figure 2). The distance between two adjacent stations varied from 5 to 12 km. The bathymetry in the middle of the observation area is complex to follow the characteristics of internal wave propagation and seafloor features, and hence, the layout of virtual MCAT stations was considered such that the mesh of the middle area was denser than that of the western and eastern areas (Figure 2). The considered depth ranges were 0–140, 140–280, 280–420, 420–560, and 560 m to the seafloor for the first, second, third,

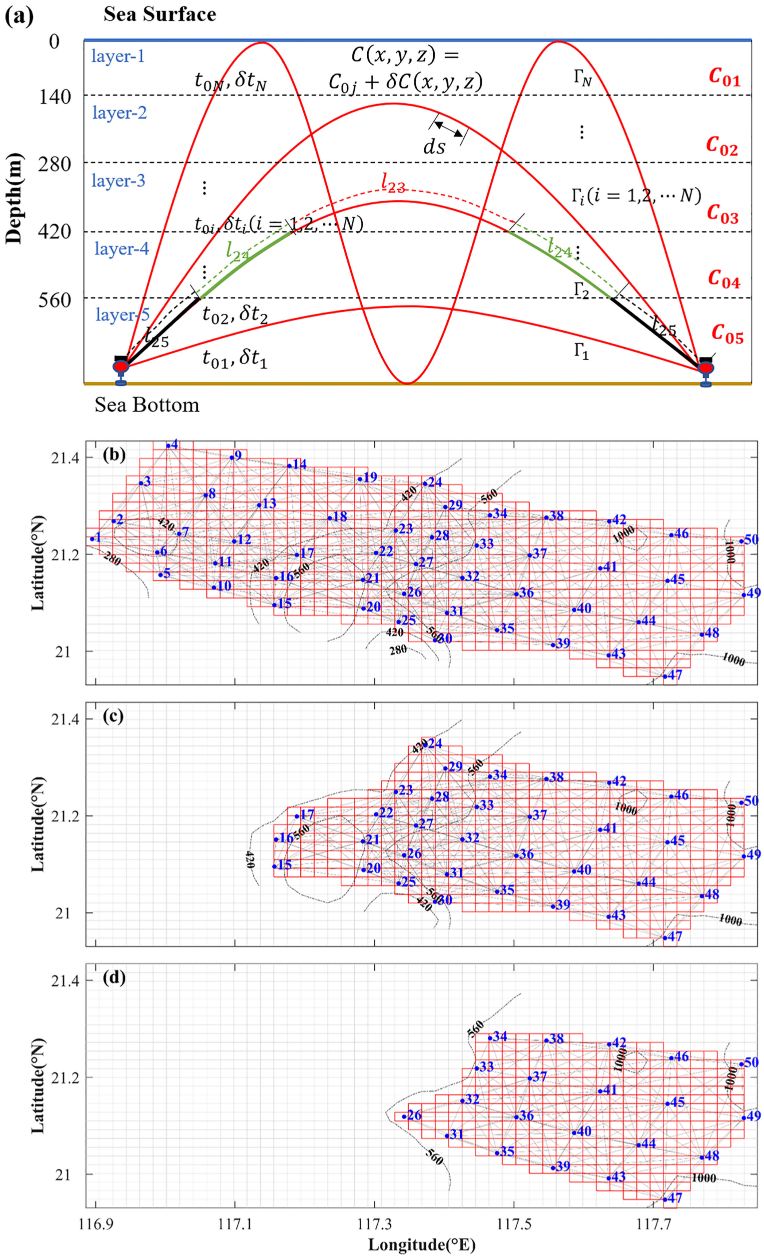


FIGURE 2

(A) Schematic diagram of the four refracted rays crossing the five layers in a vertical slice between two stations. C_{0j} and $\delta Cij(x,y,z)$ represent the reference sound speed and the sound speed deviation of the ith ray crossing jth layer, respectively. l_{ij} is the arc length of the ith ray crossing the jth layer. t_{0i} and δt_i represent the reference travel time and the travel time deviation, respectively. **(B)** Maps of the virtual MCAT stations and the projected rays and grids for the horizontal-slice inversion in the first, second, and third layers. **(C)** Maps of the virtual MCAT stations and the projected rays and grids for the horizontal-slice inversion in the fourth layer. **(D)** Maps of the MCAT stations and the projected rays and grids for the horizontal-slice inversion in the fifth layer. The blue number denotes the station number and the gray lines denote the transmission paths between individual station pairs.

fourth, and fifth layers, respectively. The number of layers was determined to harmonize the number of rays used in the inversion.

2.3 Ray simulation and determination of differential travel times

To simulate the acoustic transmission paths between each station pair, sound speed profiles were formed using temperature and salinity data from SUNTANS. Prior to vertical-slice inversion, it is needed to recognize two or more rays with different travel times at different depths. Thus, a vertical slice lying between the two stations was divided into multi-depth layers. A vertical slice with four rays and five depth layers is shown in Figure 2A.

According to the standard performance for sound propagation of the MCAT, it is assumed that reciprocal sound transmission is difficult between station pairs with a distance of > 25 km. Therefore, station pairs with distances of > 25 km were not considered in this simulation. As a result, 417 transmission paths were used in the 3D inversion of the sound speed. Figure 3 demonstrates an example of the simulated rays for one-way transmission from station C31 to C33 at hour 243 using the sound speed profiles between C31 and C33 (Figure 3A). In this case, there are four or five eigenray paths connecting two stations (Figure 3B). The rays arrived at the receiver within a time band of 10.793–11.198s (Figure 3C). Differential travel times were identified by the largest arrival peak method with the following steps: 1) because the time resolution for multi-arrival peaks of the M sequence ($t_r = Q/f$, where Q is 3 and f is 5 kHz here) is 0.6 ms, this value was set to the interval between the subsequent arrival peaks, 2) and signals with signal-to-noise ratios > 0.2 (Taniguchi and Huang, 2015) were considered (Figure 3D); and 3) during each interval, the largest arrival peak was selected (Figure 3D). Generally, 5 kHz was utilized in CAT studies. The primary objective of this paper is to discuss the feasibility of MCAT networking observations; the performances of different frequencies will be discussed in future work and are not tested here.

Finally, station pairs of 416, 416, 416, 258, and 116 were selected for the first, second, third, fourth, and fifth layers, respectively (Table 1). Based on the upper rules for selecting the ray between each station pair, 1,637, 1,944, 2,387, 1,579, and 744 rays were selected for the first, second, third, fourth, and fifth layers, respectively (Table 1). The spatial resolutions in the first, second, third, fourth, and fifth layers were estimated to be approximately 2.7, 2.7, 2,7, 2.8, and 3.4 km ($\sqrt{A/N}$, where A is the area and N is the number of stations), respectively (Table 1). For each station pair, at least one ray runs through and the selected rays are distributed in each layer.

3 Inverse problem

3.1 Inversion on a vertical slice

In the MCAT experiment, reciprocal travel time data can be directly observed between each station pair. However, in this work, they were simulated along the acoustic ray path using the results of the SUNTANS model.

Following the vertical-slice inversion in Kaneko et al. (2020), as shown in Figure 2A, we present an example to demonstrate the vertical-slice inversion for four rays and five horizontal layers. Range-average sound speed deviations are considered unknown variables between each station pair, which are need to be solved by the vertical-slice inversion. In this study, one-way travel times were simulated, and the effect of velocity on them was neglected as a suitable assumption. The one-way travel time along each ray path Γ_i is calculated as follows:

$$t_i = \int_{\Gamma_i} \frac{ds}{C_0(z) + \delta C(x, y, z)} \tag{1}$$

where $C_0(z)$ denotes the reference sound speed in each vertical layer, $\delta C(x,y,z)$ denotes the sound speed deviation in the 3D (x, y, z) coordinate frame, and ds represents the increment in the arc length along the ray.

The one-way travel time along each ray in Figure 2A was calculated using the following equation:

$$\begin{cases} \delta t_1 = -\frac{l_{1,1}}{C_{0,1}^2}\Delta C_1 - \frac{l_{1,2}}{C_{0,2}^2}\Delta C_2 - \frac{l_{1,3}}{C_{0,3}^2}\Delta C_3 - \frac{l_{1,4}}{C_{0,4}^2}\Delta C_4 - \frac{l_{1,5}}{C_{0,5}^2}\Delta C_5 \\ \delta t_2 = -\frac{l_{2,1}}{C_{0,1}^2}\Delta C_1 - \frac{l_{2,2}}{C_{0,2}^2}\Delta C_2 - \frac{l_{2,3}}{C_{0,3}^2}\Delta C_3 - \frac{l_{2,4}}{C_{0,4}^2}\Delta C_4 - \frac{l_{2,5}}{C_{0,5}^2}\Delta C_5 \\ \delta t_i = -\frac{l_{i,1}}{C_{0,1}^2}\Delta C_1 - \frac{l_{i,2}}{C_{0,2}^2}\Delta C_2 - \frac{l_{i,3}}{C_{0,3}^2}\Delta C_3 - \frac{l_{i,4}}{C_{0,4}^2}\Delta C_4 - \frac{l_{i,5}}{C_{0,5}^2}\Delta C_5 \\ \delta t_N = -\frac{l_{N,1}}{C_{0,1}^2}\Delta C_1 - \frac{l_{N,2}}{C_{0,2}^2}\Delta C_2 - \frac{l_{N,3}}{C_{0,3}^2}\Delta C_3 - \frac{l_{N,4}}{C_{0,4}^2}\Delta C_4 - \frac{l_{N,5}}{C_{0,5}^2}\Delta C_5 \end{cases} \tag{2}$$

where δt_i (i=1,2⋯N) represents the acoustic travel time along the ith ray and C_{0j} and ΔC_j represent the reference sound speed and sound speed deviation in the jth layer, respectively. l_{ij} is the arc length of the ith ray crossing the jth layer.

Eq. 2 can be written in the matrix form as follows:

$$y_v = \mathbf{E}_v x_v + n_{xv} \tag{3}$$

where y_v represents the vector of the sound speed deviation, $\mathbf{E}_v$ is the matrix obtained from the ray length and the reference sound speed, and n_{xv} represents the travel time error vector.

Eq. 4 constructs an ill-posed issue, and the regularized inversion scheme proposed by Rajan et al. (1987) was applied in this study. Then, the cost function is represented by

$$J_v = (y_v - \mathbf{E}_v x_v)^T (y_v - \mathbf{E}_v x_v) + \lambda_v x_v^T H_v^T H_v x_v \tag{4}$$

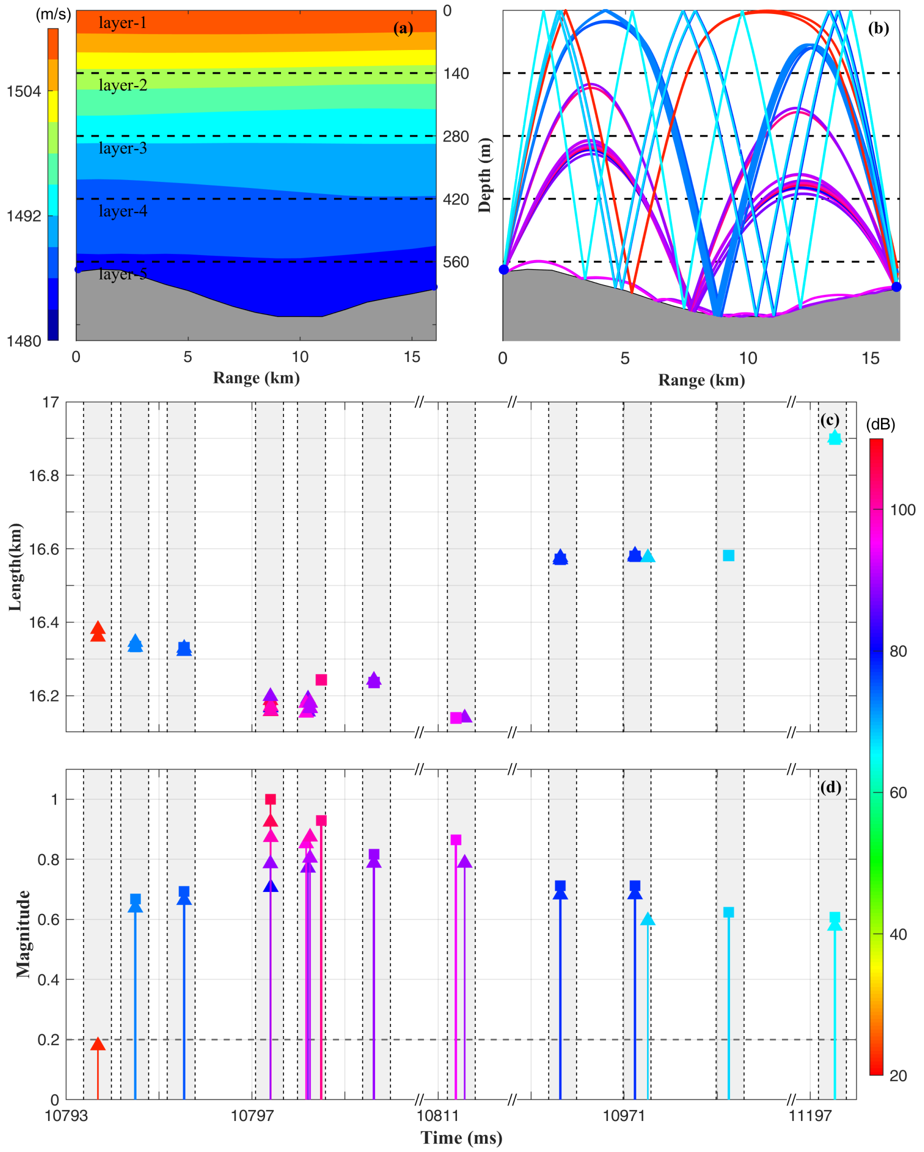


FIGURE 3
Ray simulation for the station pair C31–C33 at hour 243. **(A)** Contour map of the sound speed for the station pair C31–C33. **(B)** Eigenrays. **(C)** Ray arc length plotted against travel times. **(D)** Level of received sounds plotted against the travel time. The colors correspond to the sound level shown with color bar at the lower right. Shaded areas indicate the time bands for individual ray groups and the sound level is normalized using the maximum values.

where superscript T means the transpose of the matrix. H_v is the regularization matrix constructed from a finite difference approximation of the second-order derivative operator $s(x)$ and can be expressed as follows:

$$S(x) = \int_{-D}^{0} \left(\frac{\partial^2 x}{\partial z^2} \right)^2 dz = \sum_{i=1}^{M} \frac{(x_{i-1} - 2x_i + x_{i+1})^2}{(\Delta z)^2} = \frac{x_v^T H_v^T H_v x_v}{(\Delta z)^2} \tag{5}$$

In this work, $\Delta z = 1$ m, and then,

$$H_v = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix} \tag{6}$$

TABLE 1 Statistical factors of the eigenray between each station pair for each layer.

Layer	Number of rays				Total number of station pairs	Horizontal resolution	C_0
	Mean	Minimum	Maximum	Total		(km)	(m/s)
1st	4	1	10	1637	416	2.7	1531.8
2nd	6	1	10	1944	416	2.7	1513.9
3rd	6	1	12	2387	416	2.7	1501.8
4th	6	2	13	1579	258	2.8	1494.0
5th	7	3	12	744	116	3.4	1488.7

By minimizing J_v , the optimal solution $\vec{x}_v$ reduces to

$$\vec{x}_v = \left(\mathbf{E}_v^T\mathbf{E}_v + \lambda_v H_v^T H_v\right)^{-1}\mathbf{E}_v^T y_v \tag{7}$$

where λ_v is the Lagrange multiplier chosen using the L-curve method (O'Leary and Hansen, 1993).

3.2 Inversion on a horizontal slice

Horizontal-slice inversion rely on the results of the vertical-slice inversion. For the horizontal-slice inversion, in each layer, the tomography domain with 50 stations was segmented into numerous subdomains, as shown in Figures 2B–D. The rays and subdomains were numbered systematically for horizontal-slice inversion, the sound speed deviations are considered as unknown parameters. And the number of unknown parameters becomes $M \times N$, M represents the grid numbers and N represents the ray number in each layer.

In the first, second, and third layers, M and N were considered as 776 and 417, respectively. Since the bathymetry becomes shallow in the west of the experimental area, for the fourth layer, M was 508 and N was 258, while for the fifth layer, M was 335 and N was 116 (Figure 2B–D). The one-way sound speed deviation is calculated as follows:

$$\begin{cases} \Delta C_1 = \frac{l_{1,1}}{l_1}\delta C_1 + \frac{l_{1,2}}{l_1}\delta C_2 + \cdots + \frac{l_{1,j}}{l_1}\delta C_j + \cdots + \frac{l_{1,M-1}}{l_1}\delta C_{M-1} + \frac{l_{1,M}}{l_1}\delta C_M \\ \Delta C_2 = \frac{l_{2,1}}{l_2}\delta C_1 + \frac{l_{2,2}}{l_2}\delta C_2 + \cdots + \frac{l_{2,j}}{l_2}\delta C_j + \cdots + \frac{l_{2,M-1}}{l_2}\delta C_{M-1} + \frac{l_{2,M}}{l_2}\delta C_M \\ \Delta C_i = \frac{l_{i,1}}{l_i}\delta C_1 + \frac{l_{i,2}}{l_i}\delta C_2 + \cdots + \frac{l_{i,j}}{l_i}\delta C_j + \cdots + \frac{l_{i,M-1}}{l_i}\delta C_{M-1} + \frac{l_{i,M}}{l_i}\delta C_M \\ \Delta C_N = \frac{l_{N,1}}{l_N}\delta C_1 + \frac{l_{N,2}}{l_N}\delta C_2 + \cdots + \frac{l_{N,j}}{l_N}\delta C_j + \cdots + \frac{l_{N,M-1}}{l_N}\delta C_{M-1} + \frac{l_{N,M}}{l_N}\delta C_M \end{cases} \tag{8}$$

where δC_j denotes the sound speed deviation for the jth subdomain and $l_{i,j}$ denotes the length of the ith projected ray crossing the jth subdomain.

Similar to that in vertical inversion, equations 3, 4, 5, and 7 were also applied to horizontal inversion and were not introduced in detail here. The difference is H_h , which is the regularization matrix constructed from a finite difference approximation of the second-order derivative operator, that can be expressed as follows:

$$H_h \begin{bmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -4 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -4 & 1 & 0 & \cdots & 1 & \cdots & \cdots & 0 \\ 1 & 0 & 1 & -4 & 1 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 1 & -4 & 1 & \cdots & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 1 & -4 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & \cdots & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & \cdots & 1 & -4 \end{bmatrix} \tag{9}$$

4 Results of inverted 3D sound speed fields

4.1 Results of vertical inversion

The five-layer distributions of sound speed deviations between each station pair are shown in Figure 4. The values of sound speed deviation in the western area were higher than those in the eastern area (Figures 4A–F). The sound speed deviations in the first and second layers were almost uniform, and the difference between the western and eastern areas was< 0.5 m/s. The sound speed deviations in the third, fourth, and fifth layers showed an irregular distribution such that the sound speed deviations increased in the western part of the area. The results of the vertical-slice inversion were in good agreement with the results of the SUNTANS model (Figures 4A–F). Generally, the root-mean-square difference (RMSD) between the results of inversion and SUNTANS model are< 0.75 m/s (Figures 4K–O). Higher RMSDs (> 0.75 m/s) are mostly confined to the second layer and its edge. The RMSD was the lowest in the fifth layer, with a mean RMSD of 0.29 m/s.

Because the second layer exhibited the most significant variations in sound speed deviation, the results of the SUNTANS

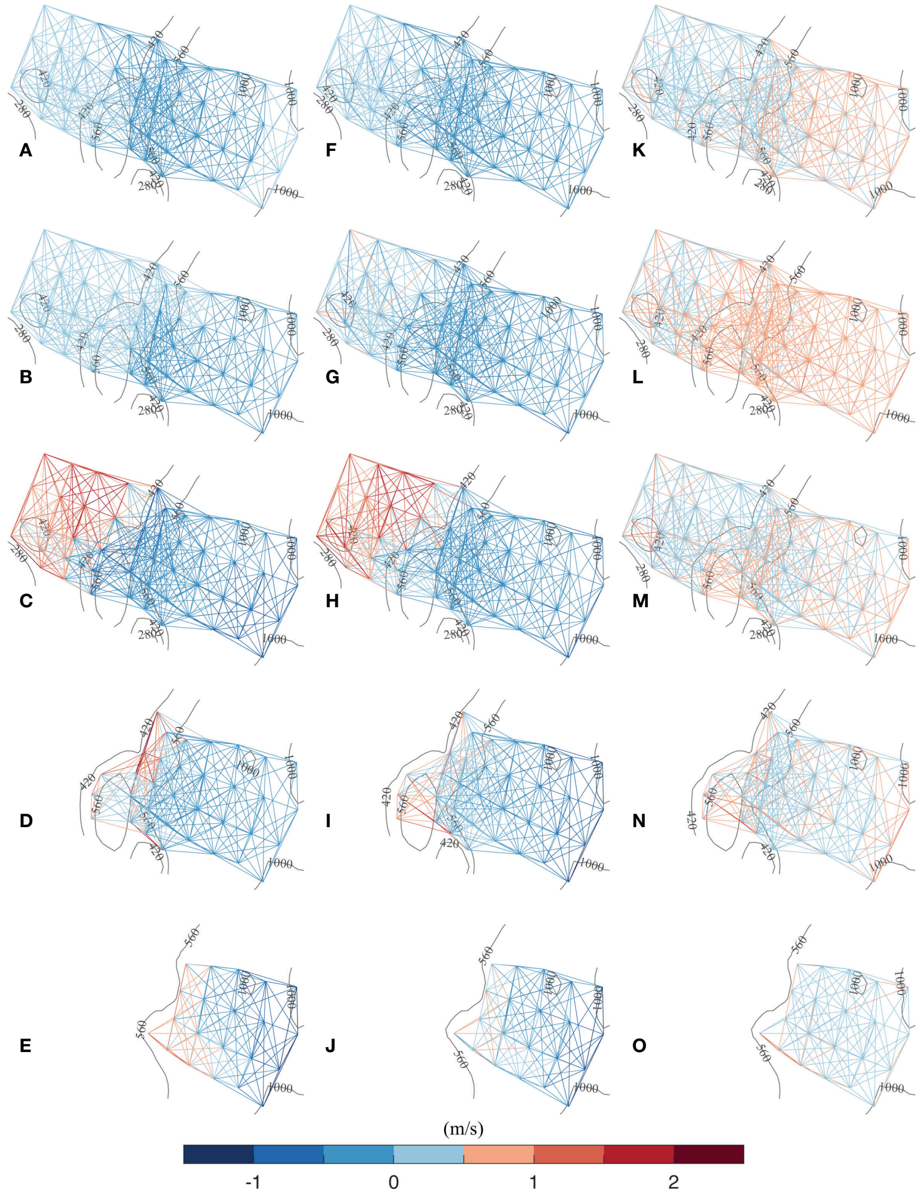


FIGURE 4
Results of the vertical-slice inversion. **(A–E)** Temporal plots of sound speed deviations calculated using the SUNTANS model along each transmission path. **(F–J)** Temporal plots of sound speed deviations derived from the vertical-slice inversion along each transmission path. **(K–O)** Temporal plots of RMSD calculated using the SUNTANS model and vertical-slice inversion. The first to fifth layers are from the top to bottom.

model and vertical-slice inversion in the second layer were compared during one tidal cycle from hour 84 to hour 97 (Figure 5). The results of inversion were in good agreement with the results of the SUNTANS model, such that the locations of the high (low) sound-speed zones were coincident. From the above data, the high sound speed zones of > 1516 m/s were considered an index of internal tides, which took approximately 12 h to propagate from the eastern to western boundaries of the observation area, which is a distance of 93.1 km (Figure 5). This value resulted in a wavelength of 82.9 km and a phase speed of 1.9 m/s.

4.2 Results of horizontal inversion

The five-layer variations of the sound speed deviations are shown with contour plots in Figures 6A–E for the SUNTANS model, Figures 6F–J for vertical-slice inversion, and Figures 6K–O for the RMSDs of both data. For both datasets, the values of sound speed deviations were higher in the western area than in the eastern area. In the first and second layers, the differences between the western and eastern areas were approximately 0.5 m/s, whereas in the third, fourth, and fifth layers, the

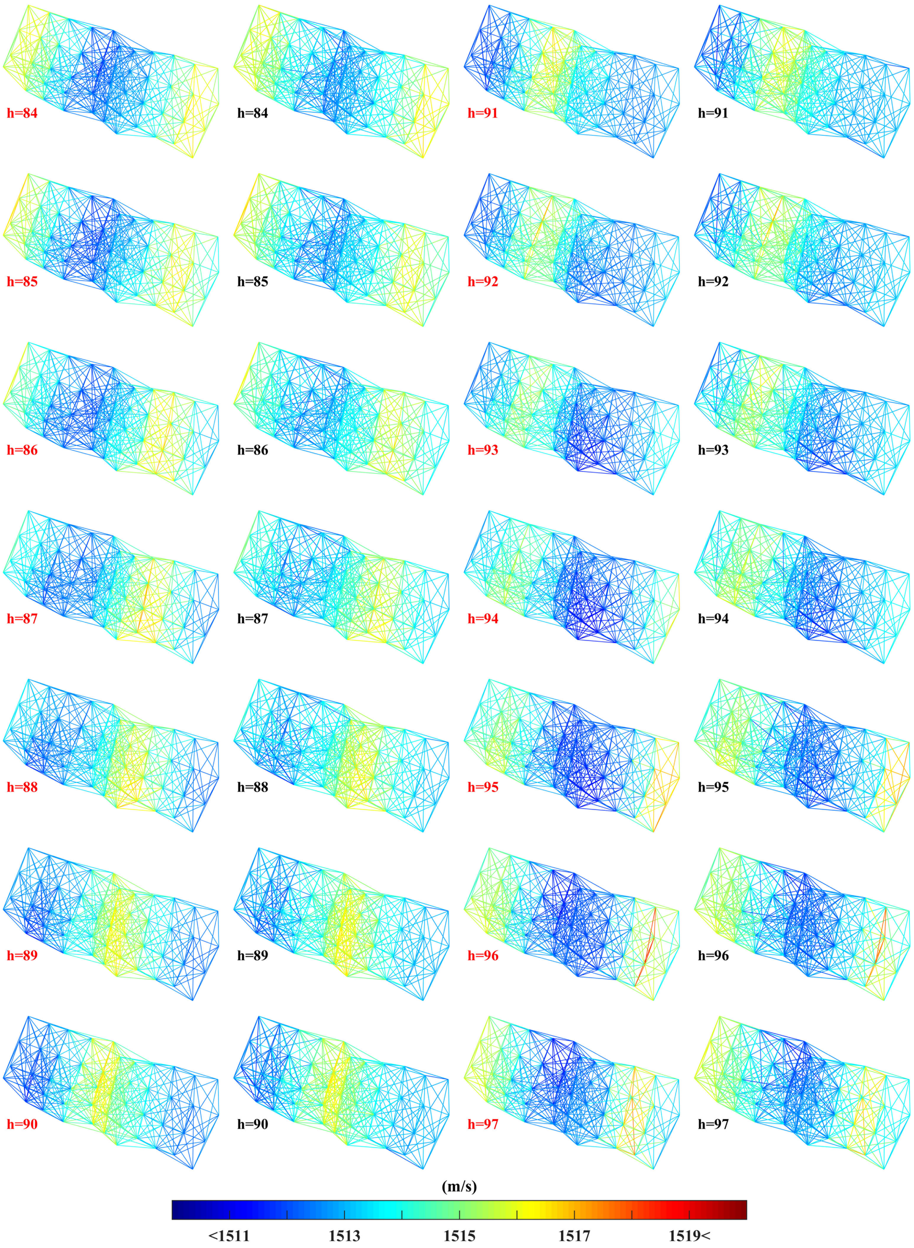


FIGURE 5
Results of vertical-slice inversion for the sound speed deviations calculated along each transmission path, during one tidal cycle from hour 84 to hour 97 in the second layer. The first and third columns are the results of the SUNTANS model, and the second and fourth columns are the results of vertical-slice inversion.

difference was approximately 1.5 m/s. The main reason for this increase is the variation in bathymetry in the observation area, wherein the effects of the continental slope became significant in the third, fourth, and fifth layers. The results of horizontal-slice inversion are in good agreement with those of the SUNTANS model. The RMSDs for each layer were< 1.0 m/s. Higher RMSDs (> 1.25 m/s) appeared at the western edge of the third and fourth layer, which were mostly caused by the sparse distribution of transmission paths and the complicated bathymetry.

The results of the SUNTANS model and horizontal-slice inversion for sound speed in the second layer are shown as contour plots during one tidal cycle from hour 84 to hour 97 in

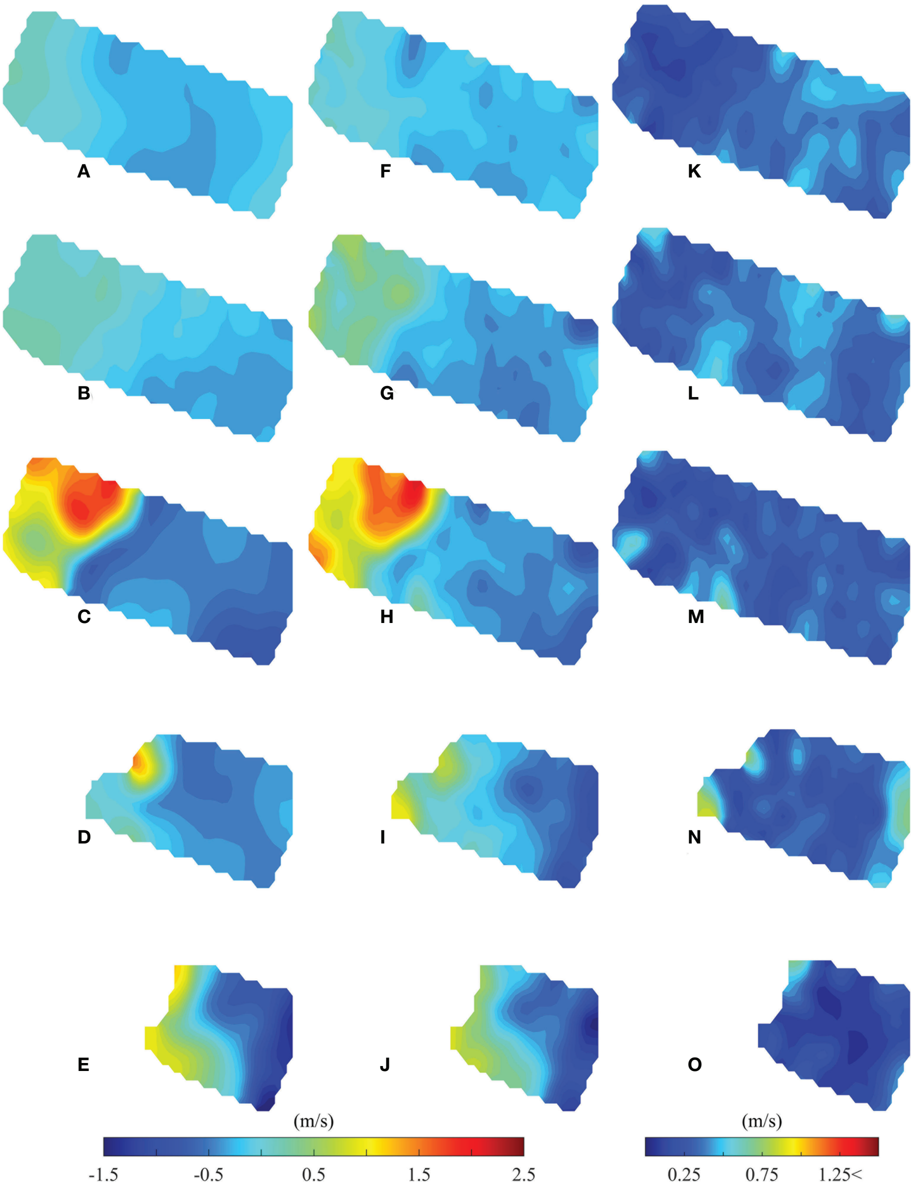


FIGURE 6
Results of the horizontal-slice inversion for sound speed deviations. **(A–E)** Temporal plots of sound speed deviation fields calculated using the SUNTANS model in each layer. **(F–J)** Temporal plots of sound speed deviation fields calculated from the horizontal-slice inversion in each layer. **(K–O)** Temporal plots of RMSDs calculated using the SUNTANS model and horizontal-slice inversions. The first to fifth layers are from the top to bottom.

Figure 7. The first and third columns of the figure are for the SUNTANS model, and the second and fourth columns are for the horizontal-slice inversion. The distributions of the large sound speed zones confined to the western area showed good agreement between the SUNTANS model and horizontal-slice inversion. The wavelength and phase speed values of internal tides calculated from the horizontal-slice inversion are similar to the vertical inversion results. The high sound speed zone transmits with a phase speed of 1.9 m/s from the east boundary to the west boundary.

5 Discussion on potential influencing factors

5.1 Swing of the anchor tether

Positioning errors are the most critical factors in sound-speed inversion. The swing of the moored MCAT system is a significant source of positioning error. Hence, a numerical experiment was conducted to evaluate the positioning error, which mostly results from the swing of the anchor tether. This study demonstrated the

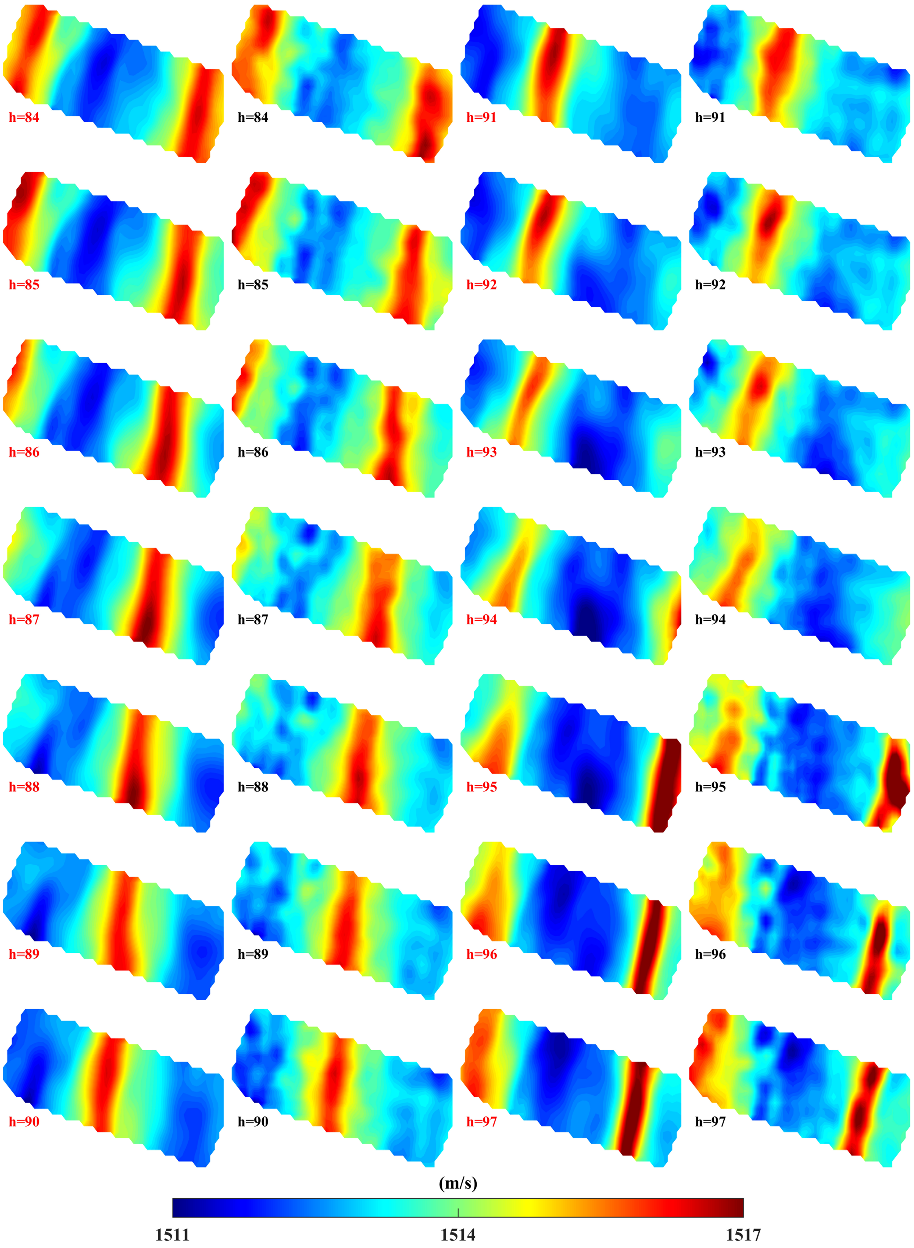


FIGURE 7
Results of the horizontal-slice inversion for sound speed during one tidal cycle from hour 84 to hour 97 in the second layer. The first and third columns are the results of the SUNTANS model and the second and fourth columns are results of the horizontal-slice inversion.

use of MCAT network observation in monitoring internal tides using sound speed. Generally, MCAT is deployed close to the bottom with an anchor, and it is set to approximately 5 m above the seafloor using a float. The results of the sensitivity experiment using a 5-m random swing are shown with the contour plot of sound speed deviations in Figure 8. Figure 8A represents the sound speed deviation obtained using the SUNTANS model. Figure 8B represents the results of the second-layer inversion with 0 m drift at hour 92. Both data showed good agreement, wherein the zone of higher values related to internal tides was coincident. Figure 8C shows the results of inversion, considering the 5-m random swing of the MCAT system, which is mostly generated by an oceanographic phenomenon (e.g., internal tide) (Figure 8D). The swing was random, although there are some differences in the results of inversion, the inversion errors are< 1.0 m/s (Figure 8C). Even though the moored MCAT would have some swing in the anchor tether, it did not affect its ability to capture the propagation processes of internal tides, suggesting the feasibility of using observations of the MCAT network to monitor internal tides.

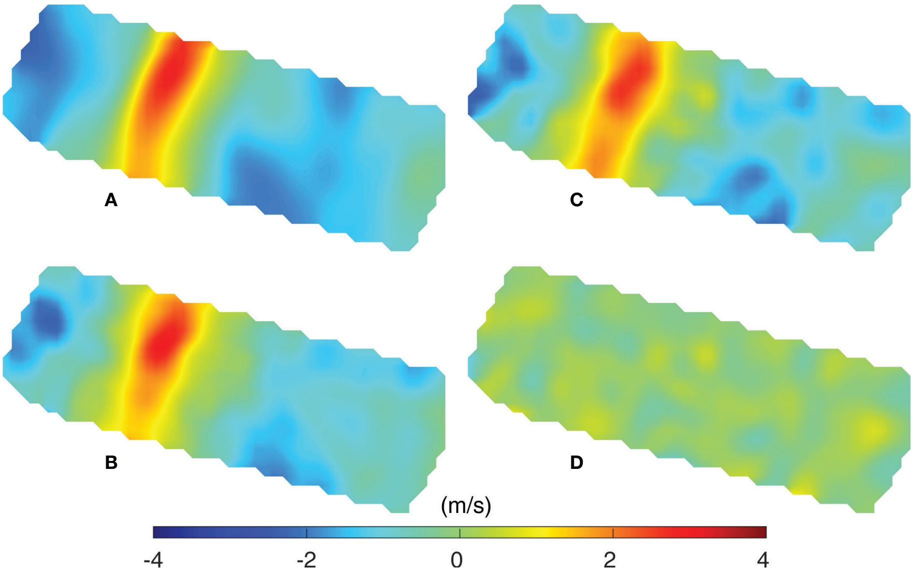


FIGURE 8
Distributions of the sound speed deviation at hour 92 from **(A)** SUNTANS, **(B)** inversion with no distance drift, **(C)** inversion with the addition of the random 5-m drift, and **(D)** distribution of sound speed deviations that resulted from the random 5-m drift.

5.2 Movement of the anchor mooring position

The accuracy of the results of inversion was significantly influenced by the positional movement of the acoustic stations. One sensitivity experiment was conducted to evaluate the effect of the movement of stations exerted by human activities on the results of inversion. Specifically, we suppose that a station is moved due to fishing activities and discuss how much the movement of one specific station affects the MCAT network observation of the entire area. As shown in Figure 9, the station movement was considered in various directions from the internal tide propagation (16° north to west). As the distance of movement increased in a specific direction, inversion error increased. Although a 10-m movement resulted in a maximum error of 1.1 m/s (Figure 9D), the effect on the network observation was weak and did not have a significant impact on the observation of internal tides (Figures 9A, D). When the distance of movement was 50 m, the inversion error reached 5.9 m/s and the area of influence was limited. Furthermore, a distance of movement of 100 m caused an error of approximately 11.9 m/s. As a result, internal tides could no longer be monitored well (Figures 9B, C, E, F).

5.3 System missing

Numerous MCATs were deployed to construct a network observation to confirm the possibility of monitoring internal tides. Hence, if unexpected troubles occur at one or several stations (e.g., lost or unable to transmit sound), we must ascertain their quantum of influence in monitoring internal tides *via* inversion. The effects generated by the missing MCAT stations are the focus of this study. As shown in Figure 10, four sensitivity experiments with different missing position arrangements were conducted to examine the effects on the results of inversion. When only one station located at the boundary is missing, the error was low with a maximum of 0.26 m/s. Besides, the associated missing rays lead to several subdomains without rays through the grids for horizontal inversion, thus decreasing the monitoring area (Figures 10A, E). Different from the aforementioned case, if the missing station was located inside the network with no missing subdomains, inversion errors were even lesser, with a maximum of 0.22 m/s (Figures 10B, F). In the case of two missing stations, the inversion error increased and the maximum error was< 0.50 m/s (Figures 10C, D, G, H), and the overall feature of internal tides was still reconstructed as their sound speed variations reached up to 6 m/s (see Figures 5, 7).

5.4 Clock drift

MCAT is equipped with a chip-scale atomic clock (CSAC), which is the lowest-power and lowest-profile atomic clock in the world, to capture the high-resolution of time. With a typical aging rate of 9×10^{-10}/s Hz/Hz (White Paper, 2018), the CSAC has a fairly good frequency drift (and corresponding time error). Taking the results of its derivation, the timing error accumulation over one month is 2.4 ms. Considering the time

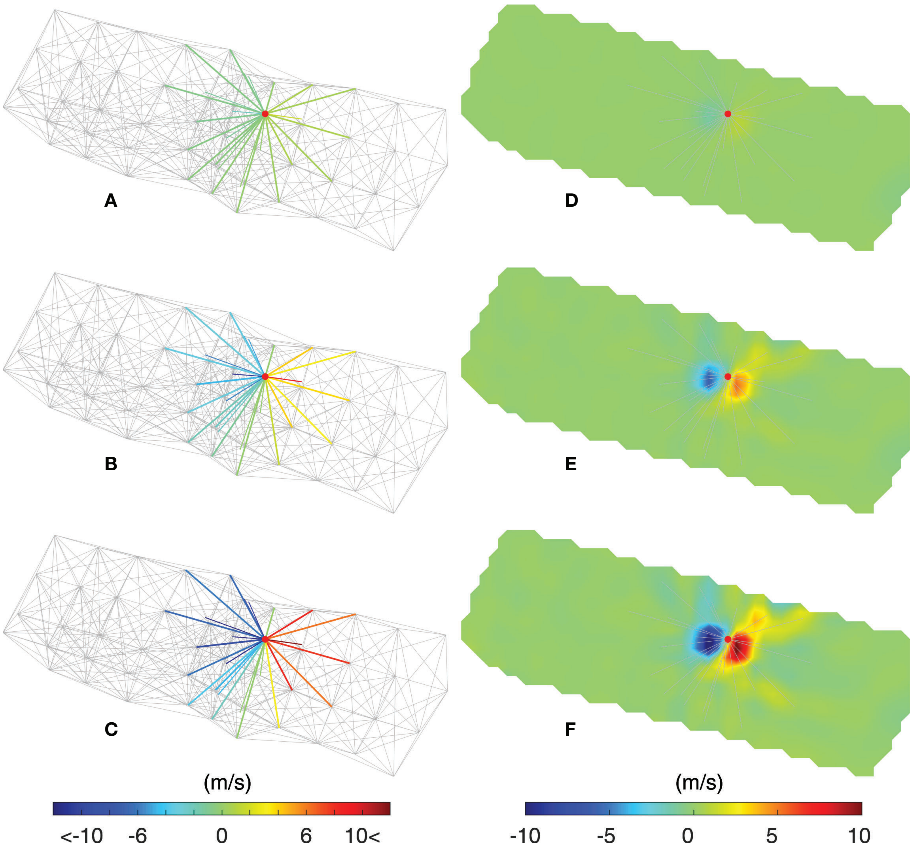


FIGURE 9
Results of the sensitivity experiments for station movement at hour 92. **(A–C)** Sound speed deviations generated by station movements of 10, 50, and 100 m at a specific station in the middle of the network area. The transmission lines marked with color represent the station pairs that are influenced by the movement of the station. The station is moved from east to west. **(D–F)** Inversion errors generated by the station movement of 10, 50, and 100 m at a specific station in the middle of the network area. The transmission lines represent the station pairs that are influenced by the movement of the station.

error generated by both drifts for one year of observation, the inversion errors resulting from the clock drift are as low as 0.5 m/s. The errors did not significantly influence the results of inversion because the sound speed variations related to the internal tides reached up to 6 m/s (see Figures 5, 7).

5 Conclusions

We examined the effectiveness of the network observation using 50 MCATs in a complicated deep-sea environment *via* numerical simulations using the synthetic ocean output obtained using the SUNTANS model. The considered cases were adequate to demonstrate the applicability of the network observation to tomographic mapping of sound speed fields in complicated deep-sea environments to capture the propagation of internal tides. The 3D fields of the sound speed inside the observation area were reproduced well in a sequence of vertical- and horizontal-slice inversions. Although some differences existed, the inverted sound speed fields showed good agreement with data of the SUNTANS model. The RMSD values of sound speeds obtained by the inversion and SUNTANS models were< 1.0 m/s. The locations of the higher sound speed zones generated by internal tides were coincident between the inversion result and the SUNTANS model. Sensitivity experiments involving the movement of the MCAT stations showed that a distance of movement of< 5 m, which is mostly caused by the swing of the anchor tether, had a low effect on the results of inversion. Large movements over 50 m, which resulted from the movement of stations, no longer served to observe sound speed variations due to internal tides. The sensitivity experiments of missing stations showed that although the error increased with two missing stations, the maximum error was still at an acceptable level of< 0.5 m/s.

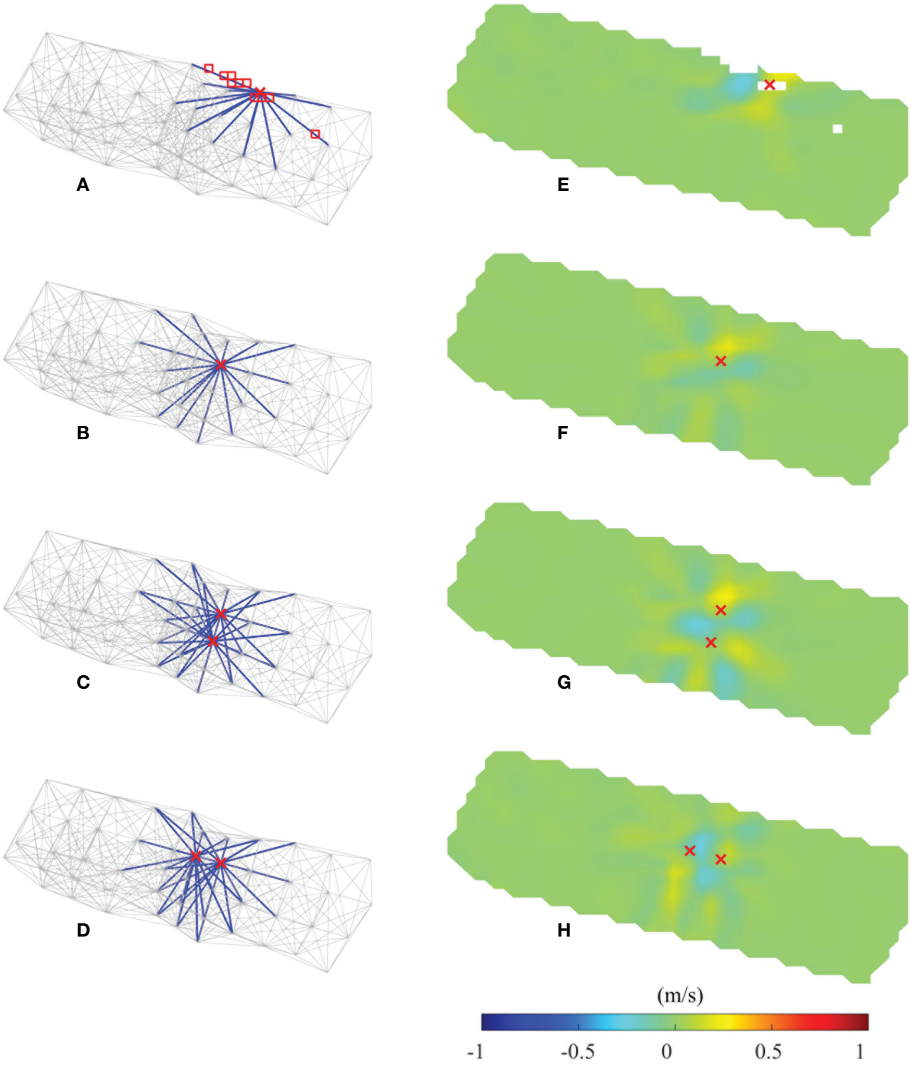


FIGURE 10
Results of the sensitivity experiments of missing stations in the network at hour 92. **(A–D)** The cross sign demonstrates the position of the missing station; the blue lines represent the transmission paths related to the missing station. **(E–H)** Inversion error generated by missing stations. The red cross marks indicate the positions of missing station. The red squares in **(A)** indicate the missing subdomains induced by the missing station, and the white squares in **(E)** match the red squares.

The higher sound speed zone generated by internal tides was captured well by the 3D inversion of the simulated MCAT data. MCAT is an advanced system of traditional CAT with a mirror forwarding function that transmits offshore observation data to the nearshore. These results suggest that MCAT network observation using numerous stations are the most efficient in capturing the propagation of internal tides. In this study, mirror function was not implemented in the SCS, and only simulation results were provided; therefore, in future studies, mirror function will be addressed, and the application of MCAT to monitor internal tides in the SCS will be determined.

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

LD and X-HZ contributed to conception and design of the study. LD and CX performed the statistical analysis. CX, X-HZ, Z-NZ and HZ wrote the first draft of the article. Z-NZ, CZ, and

QL organized the data. All authors contributed to article revision, read, and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (grants 41920104006, 52071293), the Scientific Research Fund of the Second Institute of Oceanography, MNR (JZ2001), the National Key Research and Development Program of China (2021YFC3101502), the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (SOEDZZ2207 and SOEDZZ2106), the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (project number SL2021MS021), and the Innovation Group Project of the Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai (No. 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

References

Buijsman, M. C., Kanarska, Y., and McWilliams, J. C. (2010). On the generation and evolution of nonlinear internal waves in the south China Sea. *J. Geophys. Res. Ocean.* 115, 1-17. doi: 10.1029/2009JC005275

Chen, M, Kaneko, A, Lin, J, and Zhang, C. (2017). Mapping of a Typhoon-Driven Coastal Upwelling by Assimilating Coastal Acoustic Tomography Data. *J. Geophys. Res. Ocean.* 122, 7822–7837. doi: 10.1002/2017JC012812

Chen, M., Syamsudin, F., Kaneko, A., Gohda, N., Howe, B. M., Mutsuda, H., et al. (2018). Real-time offshore coastal acoustic tomography enabled with mirror-transpond functionality. *IEEE J. Ocean. Eng.* 45, 645–655. doi: 10.1109/JOE.2018.2878260

Cornuelle, B., Munk, W., and Worcester, P. (1989). Ocean acoustic tomography from ships. *J. Geophys. Res. Ocean.* 94, 6232–6250. doi: 10.1029/JC094iC05p06232

Fringer, O. B., Gerritsen, M., and Street, R. L. (2006). An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. *Ocean Model.* 14, 139–173. doi: 10.1016/j.ocemod.2006.03.006

Kaneko, A., Zhu, X.-H., and Lin, J. (2020). *Coastal acoustic tomography* (Amsterdam, Netherlands: Elsevier). doi: 10.1016/B978-0-12-818507-0.00003-2

Li, Q., and Farmer, D. M. (2011). The generation and evolution of nonlinear internal waves in the deep basin of the south China Sea. *J. Phys. Oceanogr.* 41, 1345–1363. doi: 10.1175/2011JPO4587.1

O'Leary, D. P., and Hansen, P. C. (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

Park, JH, and Kaneko, A. (2001). Computer simulation of coastal acoustic tomography by a two-dimensional vortex model. *Journal of oceanography* 57, 593–602. doi: 10.1023/A:1021211820885

Porter, M. B. (2011). The bellhop manual and user's guide: Preliminary draft. heat. doi: 10.1038/299121a0

Rajan, S. D., Lynch, J. F., and Frisk, G. V. (1987). Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water. *J. Acoust. Soc Am.* 82, 998–1017. doi: 10.1121/1.395300

Send, U., Schott, F., Gaillard, F., and Desaubies, Y. (1995). Observation of a deep convection regime with acoustic tomography. *J. Geophys. Res. Ocean.* 100, 6927–6941. doi: 10.1029/94JC03311

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., and Huang, C. F. (2015). Simulated tomographic reconstruction of ocean current profiles in a bottom-limited sound channel. *J. Geophys. Res. Ocean.* 119, 4999–5016. doi: 10.1002/2014JC009885

The Ocean Tomography Group. (1982). A demonstration of ocean acoustic tomography. *Nature* 299, 121–125. doi: 10.1038/299121a0

White Paper (2018). *Chip-scale atomic clock (CSAC) performance during rapid chip-scale atomic clock (CSAC) performance during rapid temperature*. (Aliso Viejo, USA).

Worcester, P. F., Lynch, J. F., Morawitz, W. M. L., Pawlowicz, R., Sutton, P. J., Cornuelle, B. D., et al. (1993). Evolution of the large-scale temperature field in the Greenland Sea during 1988-89 from tomographic measurements. *Geophys. Res. Lett.* 20, 2211–2214. doi: 10.1029/93GL02373

Wunsch, M. C. (1979). Ocean acoustic tomography: A scheme for large scale monitoring. *Deep Sea Res. Part A. Oceanogr. Res. Pap* 26, 123–161. doi: 10.1016/0198-0149(79)90073-6

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

Yamoaka, H., Kaneko, A., Park, J. H., Hong, Z., 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

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, Z., And, O., and Ramp, S. R. (2011). Three-dimensional, nonhydrostatic numerical simulation of nonlinear internal wave generation and propagation in the south China Sea. *J. Geophys. Res. Ocean* 116, 1–26. doi: 10.1029/2010JC006424

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. Ocean.* 120, 4288–4305. doi: 10.1002/2014JC010676

Zhao, Z. (2015). Internal tide radiation from the Luzon strait. *J. Geophys. Res. Ocean.* 119, 5434–5448. doi: 10.1002/2014JC010014

Zhao, Z., Alford, M. H., Mackinnon, J. A., and Pinkel, R. (2010). Long-range propagation of the semidiurnal internal tide from the Hawaiian ridge. *J. Phys. Oceanogr.* 40, 713–736. doi: 10.1175/2009JPO4207.1

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, 117–128.

Zheng, H., Yamaoka, H., Gohda, N., Noguchi, H., and Kaneko, A. (1998). Design of the acoustic tomography system for velocity measurement with an application to the coastal sea. *J. Acoust. Soc Japan* 19, 199–210. doi: 10.1250/ast.19.199

Zhu, X. H., 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. Ocean. Eng.* 38, 285–296. doi: 10.1109/JOE.2012.2223911

Zhu, Z., Zhu, X., 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. Ocean.* 122, 7013–7030. doi: 10.1002/2017JC012715

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 20 January 2023
DOI 10.3389/fmars.2023.1104570

Inversion of shallow seabed structure and geoacoustic parameters with waveguide characteristic impedance based on Bayesian approach

Hanhao Zhu[1,2*†], Yangyang Xue[3†], Qunyan Ren[4,5*], Xu Liu[2], Jiahui Wang[6], Zhiqiang Cui[6], Shu Zhang[7] and Huili Fan[7]

[1]Donghai Laboratory, Zhoushan, China, [2]Institute of Marine Science and Technology, Zhejiang Ocean University, Zhoushan, China, [3]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, [4]Key Laboratory of Underwater Acoustic Environment, Chinese Academy of Sciences, Beijing, China, [5]Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, [6]School of Naval Architecture and Maritime, Zhejiang Ocean University, Zhoushan, China, [7]China Ship Development and Design Center, Wuhan, China

Check for updates

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Xiaoman Li,
Jiangsu University of Science and Technology, China
JianBo Zhou,
Northwestern Polytechnical University, China

*CORRESPONDENCE
Hanhao Zhu
✉ zhuhanhao@zjou.edu.cn
Qunyan Ren
✉ renqunyan@mail.ioa.ac.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 21 November 2022
ACCEPTED 03 January 2023
PUBLISHED 20 January 2023

CITATION
Zhu H, Xue Y, Ren Q, Liu X, Wang J, Cui Z, Zhang S and Fan H (2023) Inversion of shallow seabed structure and geoacoustic parameters with waveguide characteristic impedance based on Bayesian approach. *Front. Mar. Sci.* 10:1104570.
doi: 10.3389/fmars.2023.1104570

Underwater acoustic technology is essential for ocean observation, exploration and exploitation, and its development is based on an accurate predication of underwater acoustic wave propagation. In shallow sea environments, the geoacoustic parameters, such as the seabed structure, the sound speeds, the densities, and the sound speed attenuations in seabed layers, would significantly affect the acoustic wave propagation characteristics. To obtain more accurate inversion results for these parameters, this study presents an inversion method using the waveguide characteristic impedance based on the Bayesian approach. In the inversion, the vertical waveguide characteristic impedance, which is the ratio of the pressure over the vertical particle velocity, is set as the matching object. The nonlinear Bayesian theory is used to invert the above geoacoustic parameters and analysis the uncertainty of the inversion results. The numerical studies and the sea experiment processing haven shown the validity of this inversion method. The numerical studies also proved that the vertical waveguide characteristic impedance is more sensitive to the geoacoustic parameters than that of single acoustic pressure or single vertical particle velocity, and the error of simulation inversion is within 3%. The sea experiment processing showed that the seabed layered structure and geoacoustic parameters can be accurately determined by this method. The root mean square between the vertical waveguide characteristic impedance and the measured impedance is 0.38dB, and the inversion results accurately represent the seabed characteristics in the experimental sea area.

KEYWORDS

shallow sea, geoacoustic parameters inversion, Bayesian approach, waveguide characteristic impedance, fast filed method

1 Introduction

At present, using acoustic waves is the only method for detecting over long distances in seawater (Zhang et al., 2020; Miao et al., 2021a; Miao et al., 2021b). Underwater acoustic technology has become indispensable ocean observation, exploration, and exploitation (Xing et al., 2021; Zhang et al., 2021; Rong and Xu, 2022). The development of underwater acoustic technology is based on accurate prediction of underwater acoustic wave propagation (Zhou et al., 2021).

In the shallow seas, the structure and geoacoustic parameters of the seabed significantly affect underwater acoustic propagation (Guo et al., 2017; Yu et al., 2022). The geoacoustic parameters in each seabed layer, such as sound speeds, density and attenuations, all have significant effects on the underwater acoustic field prediction, the sonar performance prediction, the underwater acoustic detection and the underwater acoustic communication (Zhang et al., 2022; Zhao et al., 2022). Therefore, the acquisition of geoacoustic parameters is particularly important in shallow sea (Zhang et al., 2019; Yang et al., 2020). Although, the geoacoustic parameters can be obtained by direct core sample measurements, it is very time consuming and costly to obtain enough samples for a large area. Acoustic waves can propagate rapidly over a large area and carry amounts of information about the shallow seabed, requiring estimation of geoacoustic parameters.

In recent years, several studies have been conducted to derive geoacoustic parameters from shallow sea acoustic data. In these studies, the geoacoustic parameters could be inverted by matching the propagation characteristics of the acoustic waves with replicates from the acoustic computational model. As a results, the geoacoustic parameters inversion method was proposed (Yang et al., 2020). Hermand (Hermand, 1999) investigated an inversion method to rapidly estimate the distribution of seabed acoustic features using a controlled source and a single hydrophone. In the acoustic parameter inversion experiment conducted by the Chinese Academy of Sciences, the seabed sound speed and attenuation coefficient of the seabed were determined based on using the pulse waveform and transimission loss (Li et al., 2000). Park et al. (Park et al., 2005) used the time-domain waveform received from a towed array to invert the geoacoustic parameters, the inversion result is consistent with the previous measurementresults in the same sea area. By matching the acoustic pressure field excited with a point source, Zheng et al. (Zheng et al., 2021) studied a Bayesian inversion method for geoacoustic parameters in a shallow sea. However, the above inversion methods all belong to active inversion, which primarily uses the acoustic pressure field as the match object, which is actively excited by a point acoustic source. Recent research shows that, in a shallow sea, the acoustic pressure field lacks some acoustic field information, and resulting in its insensitivity to some geoacoustic parameters, which causes it to be insensitive to some geoacoustic parameters, reducing the accuracy of the inversion results. From the persoective of environmental protection, it is better to use passive acoustics inversion from the sources of opportunity in the propagation medium (Ren and Hermand, 2013; Li et al., 2019; Li et al., 2022). These sources of chance, such as ship noise, sea noise, and marine animal sounds, are widely distributed in shallow seas. However, their unknown propagation characteristics, appearances, and durations make passive inversion considerably more difficult than active inversion. With the advent of acoustic vector sensors, they have provided ideas for the development of passive inversion methods (Zhang et al., 2010).

Acoustic vector sensors can simultaneously measure the scalar acoustic field and the vector gradient field (first order as vector velocity field, or second order as acceleration field), which can contain more information about the acoustic field suitable for passive geoacoustic inversion (Sun et al., 2020). Geoacoustic parameters inversion using acoustic vector sensors has been a hot topic in the last decade.(Huang et al, 2010; Dahl et al, 2022). Santos et al. achieved a reliable estimation of seabed parameters using high-frequency signals and a small aperture vertical vector sensor array (Santos et al., 2009). Sun and Li (Sun et al., 2013) used a vector sensor array to invert the geoacoustic parameters, and effectively reduce the estimation range of seabed sound speed. Zhu et al. compared the acoustic pressure, the vertical particle velocity, and the vertical waveguide characteristic impedance for each geoacoustic parameter of a shallow seabed. The results showed that the waveguide characteristic impedance was more sensitive to changes in the geoacoustic parameters (Zhu et al., 2015). Dahl used automatic vector recorders to record ship noise and then estimated the seabed characteristics using them (Dahl and Dall'Osto, 2020). All of the above studies prove that the research on geoacoustic parameters inversion using vector sensors is beneficial, and the combinations of scalar and vector fields are more suitable for passive geoacoustic parameters inversion. However, in the current inversion research, most studies only considered the effect of seabed compression-wave (P-wave) speed, that is, the seabed is preset as a single-layer liquid seabed model, and the layered structure of the seabed is ignored. It is still a challenge to determine the seabed structure and the geoacoustic parameters of each seabed layer simultaneously (Ren et al., 2018; Kavoosi et al., 2021).

When analyzing inversion results, existing researches has focused on the optimal solution under specific conditions. Because geoacoustic parameter inversion is a complex, multidimensional, nonlinear optimization problem, the uncertainty analysis of inversion results is also vital. Uncertainty analysis of the results can be performed using the Bayesian approach to inversion (Tollefsen and Dosso, 2020). The Bayesian method combines prior information from the model with the measured dataset. It performs quantitative statistical analysis of the inversion results of parameters, and expresses the posterior probability density (PPD) of the inversion results. The Bayesian information criterion (BIC) can also determine the optimal model that fully explains the observed data using different parameterizations (Dosso et al., 2017).

To obtain more accurate inversion results, this study presents an inversion method with waveguide characteristic impedance based on the Bayesian approach. In an inversion, the vertical waveguide characteristic impedanceis set as the match object, and the Bayesian theory is used to invert the seabed structure, the sound speeds, the densities and the sound speed attenuations in seabed layers. We anticipate that these research results will enrich existing applications of acoustic vector sensor and can be used to develop an effective inversion method for geoacoustic parameters in complex shallow sea environments.

The remainder of this paper is organized as follows. Section 2 and Section 3 present the waveguide characteristic impedancein shallow sea and the proposed Bayesian inversion method are introduced in Section 2 and Section 3, respectively. Section 4 dicusses the advantages of the waveguide characteristic impedance in geoacoustic inversion and reviews the effectiveness of the proposed inversion method is verified also in this section. Section 5 describes the application of theproposed method using ship-radiated noise data measured in the Dalian Sea area. Finally, the conclusions are summarized in Section 6.

2 Modeling of waveguide characteristic impedance

Considering the influence of seabed shear-wave (S-wave) speed on acoustic propagation in a shallow sea cannot be ignored (Zhu et al., 2015). In this study, the characteristics of the shallow seabed were regarded as a semi-infinite elastic seabed covered with a uniform horizontal layered medium with multiple elastic sediments (Zheng et al., 2020). The modeling is performed assuming N×2D in a cylindrical coordinate system. Ignoring the mutual coupling of acoustic waves between the two-dimensional vertical planes in the θ direction, in the (r, z) plane, set z=0 to represents the sea surface, the downward sea surface is the direction of the positive value of the depth z-axis, and the positive r-axis represents the acoustic field propagation direction. A schematic representation of the model is shown in Figure 1.

As shown in Figure 1, the depth of the sea layer is set as h_1, the sound source with frequency f_0 is located at the depth of the seawater layer z_s, and the density and the sound speed in the layer are ρ_1 and c_1, respectively. The density, compression-wave (P-wave) speed, S-wave speed, P-wave speed attenuation, S-wave speed attenuation, and layer thickness of the seabed layers are represented by ρ_{bn}, c_{pn}, c_{sn}, α_{pn}, α_{sn} and h_n, respectively. These above parameters are the seabed geoacoustic parameters to be inverted in this study.

In the context of the acoustic wave theory, the physical quantities in the model can be represented by the displacement potential function. Setting thedisplacement potential function in the seawater layer is ϕ_1, it satisfies the wave equation as follows:

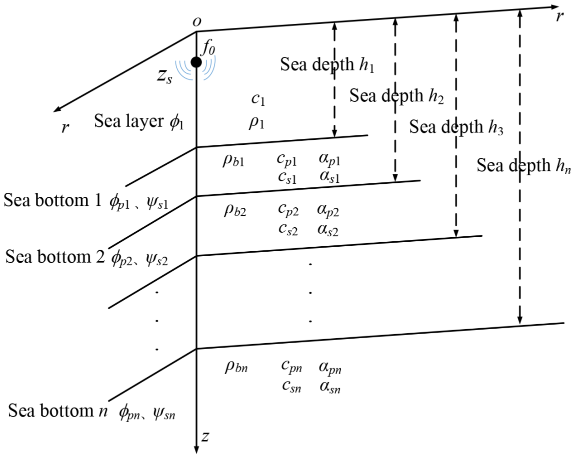


FIGURE 1
The shallow water environment model with an n-layered elastic seabed.

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \phi_1}{\partial r}\right)+\frac{\partial^2 \phi_1}{\partial z^2}+k_1^2\phi_1=-4\pi\delta(r,z-z_s)\ \ 0\le z\le h_1 \tag{1}$$

its formal solution is

$$\phi_1(r,z,\omega)=\int_0^{\infty}Z_1(z,\xi,\omega)J_0(\xi r)\xi\mathrm{d}\xi \tag{2}$$

where Z_1 is the ordinary differential equation of depth z and horizontal wavenumber ξ, and J_0 is the zero-order Bessel function. A detailed theoretical derivation can be found in the literature (Zhu et al., 2020).

The acoustic pressure field p, and the particle velocity field $\mathbf{v}$ can be expressed as follows:

$$p(r,z,\omega)=\rho_1\omega^2S(\omega)\phi_1=\rho_1\omega^2S(\omega)\int_0^{\infty}Z_1(z,\xi,\omega)J_0(\xi r)\xi d\xi \tag{3}$$

$$\mathbf{v}(r,z,\omega)=\begin{cases} v_r(r,z,\omega)=-i\omega S(\omega)\frac{\partial \phi_1}{\partial r}=i\omega S(\omega)\int_0^{\infty}Z_1(z,\xi,\omega)\mathrm{J}_0(\xi r)\xi^2\mathrm{d}\xi \\ v_z(r,z,\omega)=-i\omega S(\omega)\frac{\partial \phi_1}{\partial z}=-i\omega S(\omega)\int_0^{\infty}\frac{\partial[Z_1(z,\xi,\omega)]}{\partial z}J_0(\xi r)\xi d\xi \end{cases} \tag{4}$$

$S(\omega)$ represents the source level at f_0.

In general, the normal mode method (NMM) and the fast field method (FFM) can be used to solve Equation (3) and Equation (4) for the shallow sea environment, FFM converts the integral formula in the two equations into the form of Fourier transform and solved directly, it is more suitable for the rapid calculation of shallow sea acoustic field (Zhu et al., 2020). Therefore, in this study, the FFM is selected for solving the above-parameterized model. By discretizing the horizontal wavenumber ξ and the propagation distance r, the p and $\mathbf{v}$ can be calculated by Equation (5) and Equation (6).

$$p\left(r_j,z,\omega\right)=\frac{\rho_1\omega^2\Delta\xi}{\sqrt{2\pi r_j}}\mathrm{e}^{\mathrm{i}\left(\xi_{\min}r_j-\frac{\pi}{4}\right)}\sum_{l=0}^{N_S-1}\left[Z_1(z,\xi_l)\mathrm{e}^{\mathrm{i}r_{min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}} \tag{5}$$

$$\mathbf{v}(r,z,\omega)=\begin{cases} v_r\left(r_j,z,\omega\right)=\frac{\omega\Delta\xi}{\sqrt{2\pi r_j}}\mathrm{e}^{\mathrm{i}\left(\xi_{\min}r_j-\frac{\pi}{4}\right)}\sum_{l=0}^{N_S-1}\left[\xi_l\mathrm{Z}_1(z,\xi_l)e^{\mathrm{i}r_{\min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}} \\ v_z\left(r_j,z,\omega\right)=\frac{-\mathrm{i}\omega\Delta\xi}{\sqrt{2\pi r_j}}\mathrm{e}^{\mathrm{i}\left(\xi_{\min}r_j-\frac{\pi}{4}\right)}\sum_{l=0}^{N_S-1}\left[\frac{\partial[\mathrm{Z}_1(z,\xi_l)]}{\partial z}e^{\mathrm{i}r_{\min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}} \end{cases} \tag{6}$$

The waveguide characteristic impedance in the acoustic field (Zhu et al., 2015; Ren et al., 2018), is defined as the ratio of acoustic pressure over particle velocity, which can be generally expressed as Equation (7).

$$\begin{aligned}\mathbf{Z}(r,z,\omega)&=\frac{p(r,z,\omega)}{v(r,z,\omega)}\\ &=\begin{cases} Z_r(r,z,\omega)=\frac{\rho_1\omega^2S(\omega)\phi_1}{-i\omega S(\omega)\frac{\partial\phi_1}{\partial r}}=\rho_1\omega\frac{\sum_{l=0}^{N_S-1}\left[Z_1(z,\xi_l)e^{\mathrm{i}r_{min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}}}{\sum_{l=0}^{N_S-1}\left[\xi_lZ_1(z,\xi_l)e^{\mathrm{i}r_{\min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}}} \\ Z_z(r,z,\omega)=\frac{\rho_1\omega^2S(\omega)\phi_1}{-i\omega S(\omega)\frac{\partial\phi_1}{\partial z}}=\mathrm{i}\rho_1\omega\frac{\sum_{l=0}^{N_S-1}\left[Z_1(z,\xi_l)e^{\mathrm{i}r_{min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}}}{\sum_{l=0}^{N_S-1}\left[\frac{\partial[Z_1(z,\xi_l)]}{\partial z}e^{\mathrm{i}r_{\min}l\Delta\xi}\sqrt{\xi_l}\right]\mathrm{e}^{\mathrm{i}\frac{2\pi lj}{N_S}}} \end{cases}\end{aligned} \tag{7}$$

As illustrated in the above two equations, the source level $S(\omega)$, which intrinsically exists in p and $\mathbf{v}$ is absent in $\mathbf{Z}$. Consequently, the waveguide characteristic impedance is a unique feature of being $S(\omega)$ independent. Because it is difficult to measure the the source level accurately **Z** has a significant advantage for geoacoustic parameters inversion at-sea operations, especially for passive inversion.

For the l^{th} factorization in the waveguide characteristic impedance, the expression of **Z** can be simplified as follows:

$$\mathbf{Z}(r,z,\omega) = \begin{cases} Z_r(r,z,\omega) = \frac{\rho_1 \omega}{\xi_l} \\ Z_z(r,z,\omega) = \mathrm{i}\rho_1 \omega \frac{Z_1(z,\xi_l)}{\frac{\partial[Z_1(z,\xi_l)]}{\partial z}} \end{cases} \quad (8)$$

Compared with the horizontalwaveguide characteristic impedance Z_r, the vertical waveguide characteristic impedance Z_z depends crucially relies on the displacement potential function ϕ_1 and its derivative function. These functions have been shown to contain more information about the shallow sea environment, which making the Z_z more sensitive to the changes in shallow sea environmental parameters, as well as the seabed geoacoustic parameters (Zhu et al., 2015)

3 Bayesian inversion theory and derivation

3.1 Bayesian inversion theory

In Bayesian theory, the unknown parameters (environmental and model parameters) are considered as random variables (Jeon et al., 2022; Jiang and Zhang, 2022). Let **d** be a vector of experiment data, and let **m** be the vector of unknown parameters of a model I. Both **d** and **m** are related by Bayes' rule

$$P(\mathbf{m}|\mathbf{d},I) = \frac{P(\mathbf{d}|\mathbf{m},I)P(\mathbf{m}|I)}{P(\mathbf{d}|I)} \quad (9)$$

The dependence on model I is suppressed for simplicity, and includes I when considering model selection. $P(\mathbf{m})$ is the prior probability density function (PDF) and $P(\mathbf{d})$ is the PDF. $P(\mathbf{m}|\mathbf{d})$ is the PPD that solves the inversion problem in Bayesian theory. $P(\mathbf{d}|\mathbf{m})$ is the conditional PDF of the data **d** given parameters **m**. When **d** represents the experimental data, $P(\mathbf{d}|\mathbf{m})$ is interpreted as a function of **m**, known as the likelihood function $L(\mathbf{m})$, which can generally be written as follows:

$$L(\mathbf{m}) \propto \exp[-E(\mathbf{m})] \quad (10)$$

where $E(\mathbf{m})$ is the error function. Because the $P(\mathbf{d})$ is independent of **m**, after normalizing the Equation (9) can be written as follows:

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-E(\mathbf{m})]P(\mathbf{m})}{\int \exp[-E(\mathbf{m}')]P(\mathbf{m}')d\mathbf{m}'} \quad (11)$$

where the domain of integration spans the parameter space.

The interpretation of the PPD for multidimensional problems requires the estimation of the properties of the parameter value, uncertainties, and inter-parameters, such as the maximum posterior probability (MAP) model, mean model, and marginal probability distributions, which are defined, respectively, as follows:

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max}\{P(\mathbf{m}|\mathbf{d})\} \quad (12)$$

$$\overline{\mathbf{m}} = \int \mathbf{m}' P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \quad (13)$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i - m_i') P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \quad (14)$$

3.2 Error function and model selection

The data uncertainty distribution, which defined by the error function, is critical for Bayesian inversion (Zheng et al., 2020). However, it is difficult to determine the distribution that requires reasonable assumptions. In Bayes' rule, the error function $E(\mathbf{m})$ is established based on the likelihood function $L(\mathbf{m})$, assuming a Gaussian data errors assumption.

$$L(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{\pi^K \left|C_{\mathrm{m}}^f\right|} \exp\left\{-\left[Z_z^{f_{\mathrm{mea}}} - Z_z^{f_{\mathrm{pre}}}(\mathbf{m})\right]^{\mathrm{T}} \left(C_{\mathrm{m}}^f\right)^{-1} \left[Z_z^{f_{\mathrm{mea}}} - Z_z^{f_{\mathrm{pre}}}(\mathbf{m})\right]\right\} \quad (15)$$

where $\mathrm{Z}_{\mathrm{z}}^{\mathrm{f}_{\mathrm{mea}}}$ represents the measured vertical waveguide characteristic impedance data, $\mathrm{Z}_{\mathrm{z}}^{\mathrm{f}_{\mathrm{pre}}}$ $(\mathbf{m})$ and $\mathrm{C}_{\mathrm{m}}^{\mathrm{f}}$ represent the model prediction of the vertical waveguide characteristic impedance data and the covariance matrix.

If source information is not available, $\mathrm{Z}_{z}^{f_{\mathrm{pre}}}$ $(\mathbf{m})$ can be expressed as follows:

$$Z_z^{f_{\mathrm{pre}}}(\mathbf{m}) = A^f \mathrm{e}^{i\theta_f} Z_z^{f_{\mathrm{FFM}}}(\mathbf{m}) \quad (16)$$

where $\mathrm{Z}_{\mathrm{z}}^{\mathrm{f}_{\mathrm{FFM}}}$ $(\mathbf{m})$ is the vertical waveguide characteristic impedance predicted by the FFM, A^f and θ^f are the magnitude and phase of the unknown complex source at each frequency f, respectively. $L(\mathbf{m})$ can be maximized concerning the source by setting $\partial \mathrm{L}(\mathbf{m})/\partial \mathrm{A}^f = \partial \mathrm{L}(\mathbf{m})/\partial \theta^f$, resulting in

$$A^f \mathrm{e}^{i\theta_f} = \frac{\left[Z_z^{f_{\mathrm{FFM}}}(\mathbf{m})\right]^* Z_z^{f_{\mathrm{pre}}}(\mathbf{m})}{\left|Z_z^{f_{\mathrm{FFM}}}(\mathbf{m})\right|^2} \quad (17)$$

Then, $L(\mathbf{m})$ can be written as

$$L(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi \nu^f)^K} \exp\left[-\frac{B^f(\mathbf{m}) Z_z^{f_{\mathrm{pre}}}}{\nu^f}\right] \quad (18)$$

where $\mathrm{B}^f(\mathbf{m})$ is the normalized Bartlett disqualification.

To obtain a maximum-likelihood estimate of the data variance by setting the $\partial \mathrm{L}(\mathbf{m})/\partial \mathrm{v}^f = 0$, the maximum likelihood solution of ν^f is:

$$\hat{\nu}^f = \frac{B^f(\mathbf{m})\left|Z_z^{f_{\mathrm{pre}}}\right|^2}{K} \quad (19)$$

Using Equation(16) in Equation(15) and Equation(6), the error function $E(\mathbf{m})$ becomes

$$E(\mathbf{m}) = K \sum_{f=1}^{F} \ln\left[B^f(\mathbf{m})\left|Z_z^{f_{\mathrm{pre}}}\right|^2\right] \quad (20)$$

The MAP model for measured data can be found by minimizing the sum of the error function, indicating the correlation between measured and predicted data (Jiang and Zhang, 2022).

Owing to the large number of parameters that need to be inverted, an accurate selection of the seabed parameterized model that best matches the measured data is also extremely important for multi-layer seabed geoacoustic parameter inversion. In this study, based on the establishment of the error function $E(\mathbf{m})$, an accurate selection of the best-parameterized model is performed according to the Bayesian information criterion (BIC).

BIC is an asymptotic approximation of the Bayesian formula $P(\mathbf{d}|\mathbf{I})$ of the model **I**. The model **I** represents the multi-layer seabed model used in this paper, that is, given measurement data **d**, the likelihood function of model **I**, and its expression is as follows:

$$-2lnP(\mathrm{d}|I) \approx BIC \equiv -2\ln L(\mathbf{m}|I) + M\ln(N) \tag{21}$$

where M is the number of parameters in model **I**, and N is the number of observed data. The parameterization with the smallest BIC value was selected as the most appropriate model.

4 Numerical study

4.1 Sensitivity study for inversion parameters

To comprehensively investigate the feasibility of the vertical waveguide characteristic impedance in inversion, the sensitivity of the geoacoustic parameters of each seabed layer under this inversion method was investigated. In this study, a two-layer seabed shallow sea model was considered as an example. In this model, the density (ρ_{b1}, ρ_{b2}), P-wave speed (c_{p1}, c_{p2}), S-wave speed (c_{s1}, c_{s2}), P-wave speed attenuation in the semi-infinite elastic seabed and elastic sediment on layer (α_{p1}, α_{p2}), S-wave speed attenuation (α_{s1}, α_{s2}) and layer thickness (h_2) were the seabed geoacoustic parameters to be inverted.

Based on the model established in Section 2 and the error function in Section 3, the sensitivity of the geoacoustic parameters, which are inverted, to the error function $E(\mathbf{m})$ is investigated by numerical simulation. The fixed variable method is used to study the sensitivity of the geoacoustic parameters. In this technique, when studying the sensitivity of a particular parameter to the error function, i.e., with other parameters unchanged, only the change of the parameter in question in its previous interval is changed, and the statistical error function changes. For the reception of a single vector sensor in the simulation, the water depth is set as h_1 = 100m, the sound source depth z_s and the single sensor receiving depth z_r are set to 10 m and 20 m, respectively, and the sound source frequency f_0 = 150 Hz. The receiver range interval was set to 0m-1000m. The actual values and previous intervals of the geoacoustic parameters are listed in Table 1.

$$\begin{cases} \mathrm{TL}_p = -20\mathrm{lg}\left|\frac{p(r,z,\omega)}{p_{ref}|_{r-1m}}\right| p_{ref} = \frac{\mathrm{e}^{ik_0 r}}{r} \\ \mathrm{TL}_{v_z} = -20\mathrm{lg}\left|\frac{v_z(r,z,\omega)}{v_{ref}|_{r-1m}}\right| v_{ref} = \frac{k_0}{\rho\omega}\frac{\mathrm{e}^{ik_0 r}}{r} \\ ZL_{z_z} = 10\,\mathrm{lg}\left|\frac{Z_z(r,z,\omega)}{Z_{ref}}\right| Z_{ref} = 1.48 \times 10^6 \mathrm{Pa}\cdot\mathrm{s/m} \end{cases} \tag{22}$$

Figures 2A–C show the propagation characteristic curves in Table 1 for the p, v_z and Z_z, respectively. The transmission loss curves TL_p and TL_{vz} are used to describe the propagation characteristic of p and v_z, and the waveguide characteristic impedance curve ZL_{Zz} is used to describe the propagation characteristic of Z_z. These curves are defined by Equation (22) (Zhu, 2014).

From the comparison results, it can be seen that at some distances, that lead to notable peaks of the waveguide characteristic impedance Z_z, these peaks contain important information from the environment. This fluctuation is evident at short distances. As the propagation distance increases, the numerical value decreases at long distances decreases, and the change is relatively weakened. The fluctuation of the vertical waveguide characteristic impedance is more evident in the propagation distance because its definition cancels the influence of the source level and wave front expansion. The waveguide characteristic impedance also compensates for the transmission loss caused by the expansion of the wavefront, which is more conducive to the use of the waveguide information. Therefore, it is more suitable for the inversion of geoacoustic parameters as the match object.

The accuracy of the inversion result strongly depends on the sensitivity of the physical quantity to the geoacoustic parameters. The inversion results obtained with geoacoustic parameters with low sensitivity had larger uncertainty and lower inversion accuracy. When the sensitivity of a specific geoacoustic parameter is discussed in the inversion, the other parameters are fixed and only change the value of the discussed parameter within the range of interest. The value of the error function $E(\mathbf{m})$ for the discussed parameter reflects its sensitivity within the range of interest.

The sensitivities of p, v_z, and Z_z to the geoacoustic parameters in Table 1 are compared in Figures 3, 4. The two figures show the comparative results for the geoacoustic parameters in sediment and basement for a frequency of 150 Hz. In these figures, the blue curve represents p, the black curve represents v_z, and the red curve represents Z_z. The results show that Z_z is much more sensitive to these geoacoustic parameters than p or v_z, all the parameters will have an impact on the $E(\mathbf{m})$ for Z_z. Z_z shows higher sensitivity to wave speed and sediment layer thickness than to density and sound speed attenuation. The sensitivities of p and v_z were relatively close in the sediment and basement layers. Among the other parameters, the sensitivity of sound speed attenuation is relatively low, but Z_z has the most significant improvement in the sensitivity of these parameters. This means that using Z_z as the match object can improve the accuracy of the inversion results, especially for the sound speed attenuation.

According to the above analysis results, Z_z has been shown to have higher sensitivity to geoacoustic parameters than p and v_z. Z_z is expected to provide higher precision for geoacoustic parameters in geoacoustic inversion than when only p or v_z is used, especially for sound speed attenuations.

4.2 Simulation inversion results

In this section, the feasibility of the Bayesian inversion method using the characteristic impedance of a vertical waveguide is studied through simulations. In the simulation, the shallow sea model, listed in Table 1, was selected as the measurement environment. The true values of the geoacoustic parameters for the two-layered structure of the seabed are the inverse objects, and the vertical waveguide characteristic impedance data, as calculated in Figure 2C, are used as the measurement data $\mathrm{Z}_\mathrm{z}^{\mathrm{f}_\mathrm{mea}}$ for matching in the inversion. The

TABLE 1 Inversion results of seabed parameters in the simulation environment.

Stratification	Inversion parameters	Prior interval (Truth value)	Mean ± Standard deviation
Sediment	$c_{p1}/\mathrm{m \cdot s^{-1}}$	1800 – 2200 (2000)	2003.80 ± 51.07
	$c_{s1}/\mathrm{m \cdot s^{-1}}$	900 – 1100 (1000)	1001.13 ± 26.60
	$\rho_{b1}/\mathrm{g \cdot cm^{-3}}$	1.3 – 1.7 (1.5)	1.50 ± 0.05
	$\alpha_{p1}/\mathrm{dB \cdot \lambda^{-1}}$	0.3 – 0.7 (0.5)	0.50 ± 0.05
	$\alpha_{s1}/\mathrm{dB \cdot \lambda^{-1}}$	0.3 – 0.7 (0.5)	0.487 ± 0.068
	h_2/m	5 – 15 (10)	10.07 ± 1.32
Semi-infinite basement	$c_{p2}/\mathrm{m \cdot s^{-1}}$	2000 – 3000 (2500)	2519.71 ± 145.55
	$c_{s2}/\mathrm{m \cdot s^{-1}}$	1000 – 1400 (1200)	1192.91 ± 54.13
	$\rho_{b2}/\mathrm{g \cdot cm^{-3}}$	1.6 – 2.0 (1.75)	1.75 ± 0.06
	$\alpha_{p2}/\mathrm{dB \cdot \lambda^{-1}}$	0.3 – 0.7 (0.5)	0.51 ± 0.06
	$\alpha_{s2}/\mathrm{dB \cdot \lambda^{-1}}$	0.3 – 0.7 (0.5)	0.51 ± 0.05

parametric search optimization is performed by minimizing Equation (21) using the adjusted optimization annealing algorithm.

Model selection was performed by evaluating different layered models using BIC to determine the simplest parameterization that matches the measured data, and the results are shown in Figure 5. The BIC values of the three inversion models with different hierarchical structures are 36.52, 34.24, and 36.93, respectively. Based on the BIC, the 2-layer model was selected as the best-parameterized model, and the selected result agreed with the specification in Table 1.

Table 1 also shows the average and standard deviation of the maximum a posteriori probability estimate of the vertical waveguide characteristic impedance inversion results. Compared with the preset true value, it can be seen that the maximum error is within 0.3%. To further verify the accuracy of the inversion results, Figure 6 shows a comparison of the waveguide characteristic impedance curve ZL_{Zz} calculated using the true simulation value and inversion results. From the comparison of the results in Figure 6, it can be seen that the two curves are in agreement. The root mean square (RMS) of the

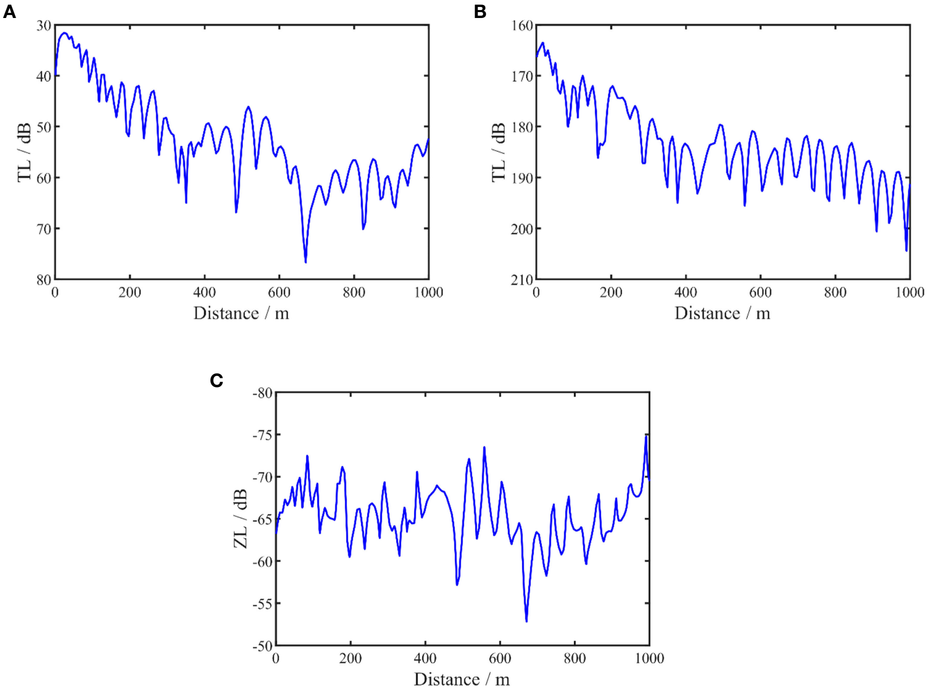


FIGURE 2
The calculated level for the p, v_z and Z_z, respectively. acoustic pressure **(A)**, vertical particle velocity **(B)** and waveguide characteristic impedance **(C)**.

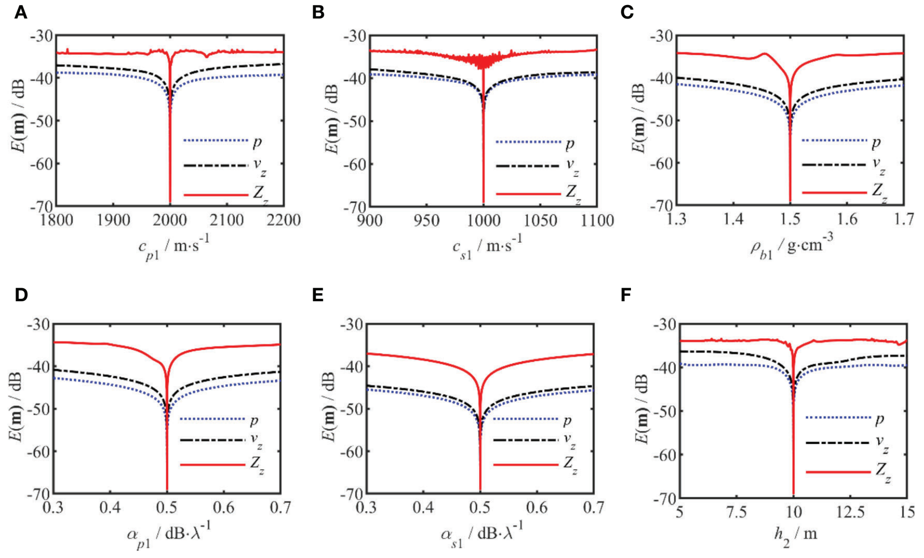


FIGURE 3

The sensitivity analysis of the error function $E(\mathbf{m})$) for the geoacoustic parameters of the sedimentary layers. **(A–F)** corresponds to the c_{p1}, c_{s1}, ρ_{b1}, α_{p1}, α_{s1}, and h_2, respectively. The red curve represents is for the vertical waveguide characteristic impedance, and the black curve represents the vertical particle velocity and blue curve is for the acoustic pressure.

inversion results was 0.38 dB, which proves the reliability of the results obtained with this inversion method.

Figures 7, 8 show the PPDs of the inverted geoacoustic parameters in the sediment layer and semi-infinite basement, respectively. In each figure, the vertical axis represents the probability of each parameter and the red line represents the true simulation value of each parameter. The average value and standard deviation of the inversion results are represented by the blue segment, with the average value shown in the center. The length of the blue segment reflects the standard deviation of each parameter. The PPD distribution range characterizes the uncertainty of each parameter. It can be seen from the results that the maximum probability density of each parameter is close to its true value. The uncertainty of the inversion results for each parameter is consistent with the results of the sensitivity analysis.

The marginal probability distribution between the two parameters is shown in Figure 9, which shows the uncertainty of the joint parameters. From the figures, it can be seen that the centralized probability density distributions characterize lower uncertainties, and all true values are close to the highest probability of PPD. The results combined with the sound speed attenuation have a broader distribution, the results combined with α_{s2} and α_{p2} are

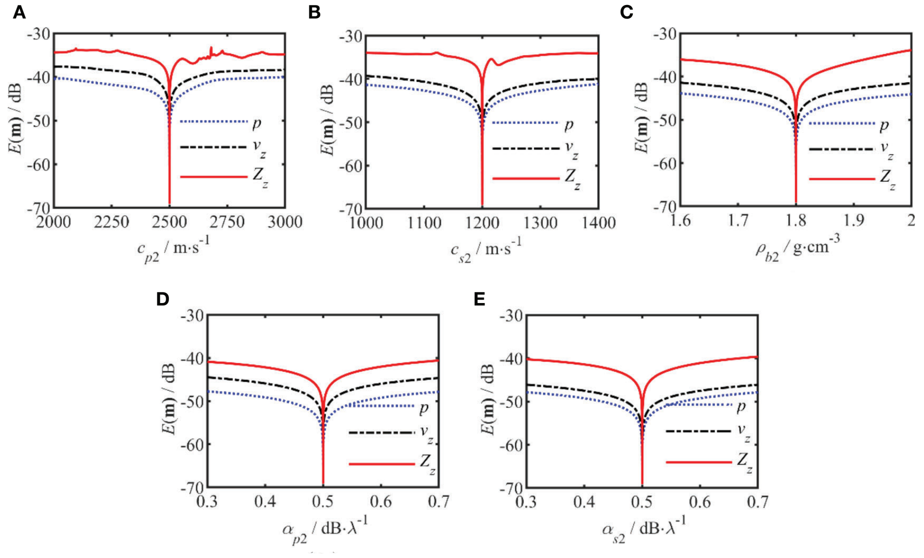


FIGURE 4

The sensitivity analysis of the error function $E(\mathbf{m})$) for the geoacoustic parameters of the semi-infinite seabed. **(A–E)** corresponds to the c_{p2}, c_{s2}, ρ_{b2}, α_{p2}, and α_{s2}, respectively. The red curve represents for the vertical waveguide characteristic impedance, the black curve represents the vertical particle velocity and blue curve is for the acoustic pressure.

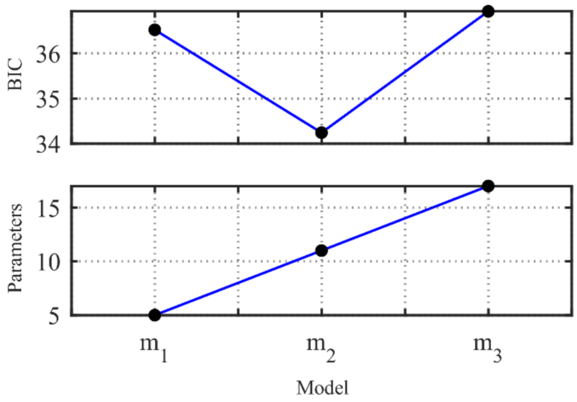


FIGURE 5
The result of model selection, the number of model parameters and the BIC value.

more distinct, and the span in the attenuation direction is larger. In conjunction with the analysis of the inversion accuracy of each parameter, it can be seen that, the sound speed attenuation has little effect on the theoretical acoustic pressure field in the process of matching inversion. Therefore, many results that deviate from the true value are retained, resulting in a more comprehensive distribution range.

5 Experiment results

5.1 Experiment description

To further illustrate the feasibility of the inversion method in practice, experimental data were processed and discussed. An experiment to measure sound-propagation was conducted in a shallow sea near Dalian. The water depth in the experimental area was approximately 25 m. According to previous information, the upper layer of the seabed in this sea area is sandy sediment and the lower layer is bedrock (Ren et al., 2018). During the measurement, a wideband signal was generated by the source ship engine along tracks at distance intervals of approximately 3 km, the depth of the sound source is equivalent to 1 m underwater. The signals were received by a vector hydrophone composed of acoustic pressure and vertical particle velocity in three directions, with a sampling frequency of 8192 Hz, and it is placed 5 m underwater by the receiver ship. The experimental configuration is shown in Figure 10. The deployment of the source and receiver ships are shown in Figure 10A. The motion trajectories of the two ships are shown in Figure 10B, where the solid black line represents the motion trajectory of the source ship and the solid red line represents the floating trajectory of the receiver ship during the experiment. The squares represent the recorded positions of the two ships at the beginning of the experiment and the circles represent the relative positions of the two ships at the end of the ship noise measurement. The motion trajectory of the yellow diamond represents the ambient noise measured after the experiment.

Before the start of the experiment, the line connecting the two ships was the x-coordinate, and the positive direction of the receiving ship was the y-coordinate. In this coordinate system, Figure 11A shows the distance between the source ship and receiving ship with time during the experiment. Figure 11B shows the relationship between the sound speed profile and the sea depth in the experimental area.

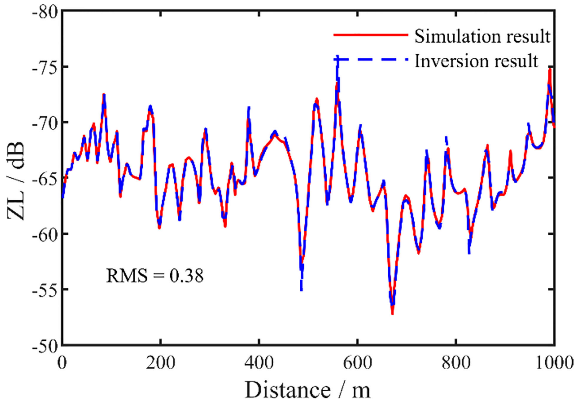


FIGURE 6
Comparison of the vertical waveguide characteristic impedance inversion results of simulation data. The inversion result is the dashed line and the experimental data is the solid line.

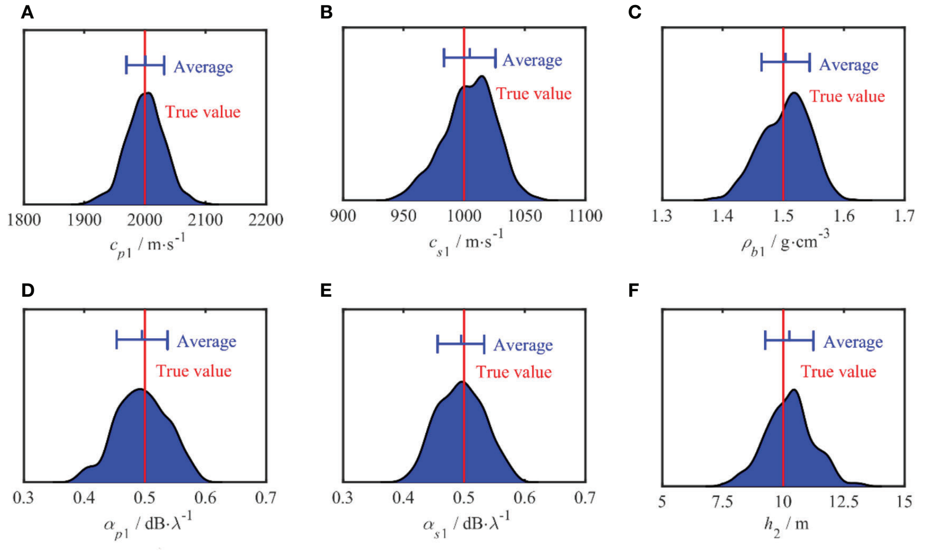


FIGURE 7
One-dimensional marginal probability distribution of sedimentary parameters. The red line means the true values of each parameter in simulation. The blue segments represent the average value and the standard deviation.

Figure 12 shows the acoustic pressure signal and the vertical particle velocity signal received from the vector hydrophone in the experiment. The recorded signals in the time domain and the corresponding low-frequency analysis recording (LOFAR) images are all given. Because the distance between the source ship and the receiving ship could satisfy the "distant-near-distant" relationship during the measurement, the amplitude of the two received signals in the time domain has the characteristic of "small-large-small." According to the LOFAR images corresponding to the signal received from each channel, the two received signals all show distinct interference striations in the range (time)–frequency domain.

To satisfy the propagation conditions of the point sources, the experimental data in the 0-700 s time period were selected for inversion. From the pressure p and the vertical particle velocity v_z measured by the vector hydrophone, the vertical waveguide characteristics impedance Z_z can be determined. To analyze the measurement results, we chose the 210 Hz signal, which can be seen more clearly in the time-frequency diagram, as the match object for the inversion. The depth of the source z_s

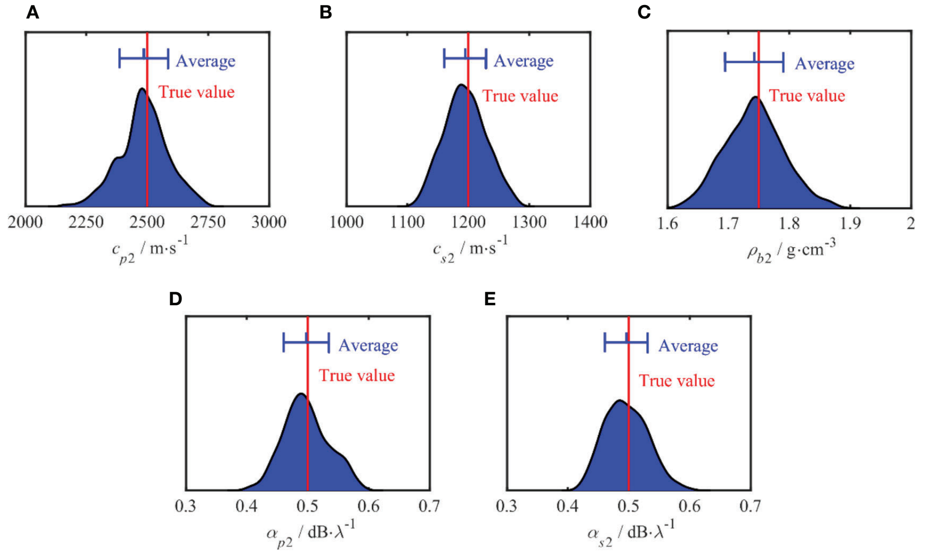


FIGURE 8
One-dimensional marginal probability distribution of semi-infinite parameters. The red line represents the true values of each parameter in the simulation. The blue segments represent the average value and the standard deviation.

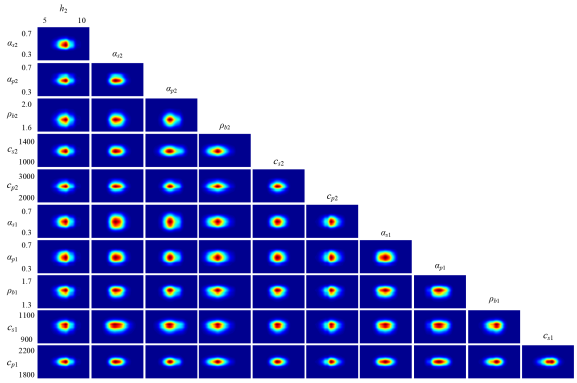


FIGURE 9
Two-dimensional joint marginal probability density distribution for 2-layer model.

was equal to 1 m, the depth of the vector sensor z_r=5 m, and the water depth of the experimental sea area was 25 m.

5.2 Inversion results and analysis

Model selection studies were performed for the three modes to determine the optimal parameterization for each parameter in each case. The unknown sound speed and density are assumed to depend on depth, as the acoustic pressure is mainly sensitive to these parameters. Therefore, each additional layer increases the number of parameters by six. Nonlinear optimizations for the impedance level inversion were performed to estimate the error function and hence the BIC for three different parameterizations with one to three uniform layers (including the basement layer). The results of the model selection are summarized in Figure 13 in terms of BIC values and number of layers. Note that, for intuitive presentation, the BIC values are normalized so that the minimum is unity. The results show that the lowest BIC value corresponds to the 2-layer inversion model. The inversion results (average + standard deviation) obtained with the 2-layer model are listed in Table 2, including the average and standard deviation with a confidence level of 70%. When comparing the inversion results of the same data (Ren, 2013), it was found that the speed and density of the sound were consistent within the error range.

The feasibility of the inversion method was further verified by comparing the inversion and measurement results. Figure 14 shows the comparison of the transmission loss curves TL_p (in Figure 14A) and the waveguide characteristic impedance curve ZL_{Zz} (in Figure 14B) obtained by the inversion of the two-layer model with the experimental data. The result of the inversion is the dashed line, and the experimental data is the solid line. Comparing the distribution trends of TL_p and ZL_{Zz}, it can be seen that ZL_{Zz} has a higher degree of consistency, especially at a larger distance. The ZL_{Zz} comparison results were better and showed higher stability. The RMS values of TL_p and ZL_{Zz} are 3.7 dB and 2.3 dB, respectively. A better fit for ZL_{Zz} was also indicated by the RMS values.

The marginal probability distributions for the parameters of the two-layer model inversion are estimated using Metropolis–Hastings sampling and are shown in Figures 15, 16, respectively. The red line represents the position with the highest probability density, i.e., the MAP value, and the blue line represents the average and standard deviation of the inversion results, with the average shown in the middle of the segment. Compared to the PPDs of the simulation inversion results, the distribution range of the PPDs obtained from the experimental data is more comprehensive, indicating that they are subject to larger uncertainties, and the uncertainties of the different parameters also vary. Compared to other parameters, the parameter h_2 has a narrow peak in the sediment layer, and the distribution of the sound speed attenuation (α_{p1}, α_{s1}) in the interval is relatively flat. The uncertainty of the parameters of the semi-infinite seabed was higher than that of the sedimentary layers.

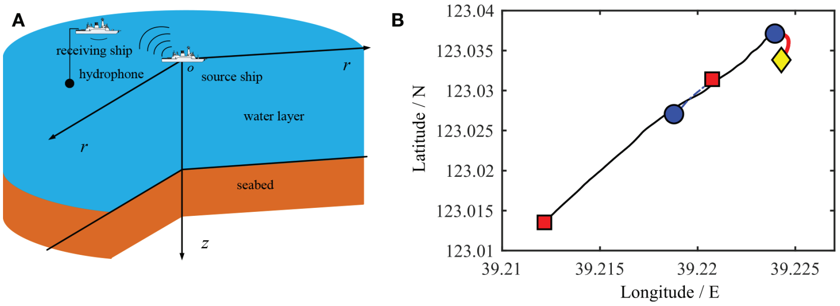


FIGURE 10
Schematic diagram of the experimental configuration, the deployment of the source ship and receiving ship **(A)** and the longitude and latitude of the distance between the two ships **(B)**.

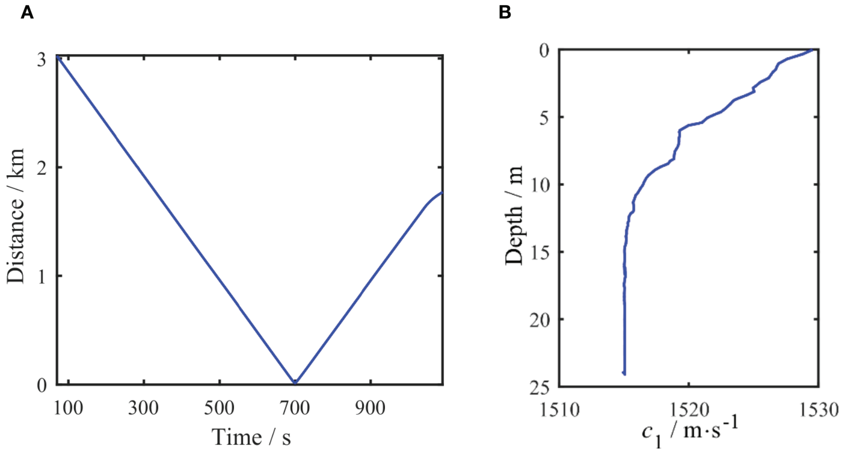


FIGURE 11
The distance between the source ship and the receiving ship with time during the experiment **(A)**, and the sound speed profile of the experiment sea area **(B)**.

The marginal probability distribution between the two parameters in the experimental data inversion is shown in Figure 17, which shows the uncertainty of the joint parameter. Compared with the two-dimensional probability density distribution in the simulation analysis, the distribution range obtained from the experimental data is more comprehensive, indicating that it has a higher uncertainty. This may be owing to the larger interference in the experiment. During the inversion process, more interference values affect the inversion results. This is consistent with the results of the one-dimensional probability density distribution. The probability density distribution of the individual joint parameters still exhibited a relatively obvious bright spot distribution, as shown in Figure 17. The parameters combined with h_2 are self-explanatory and have a narrower distribution range, indicating that the uncertainty of combining the two is small. With increasing depth, the fate of each parameter increased in the lower seabed compared to that in the upper layer. This feature is observed in the marginal probability distribution, and the inversion of the measured data is more evident.

6 Conclusions

To obtain more accurate inversion results for geoacoustic parameters in shallow seas, a new geocaoustic inversion method is presented in this study based on the nonlinear Bayesian approach and

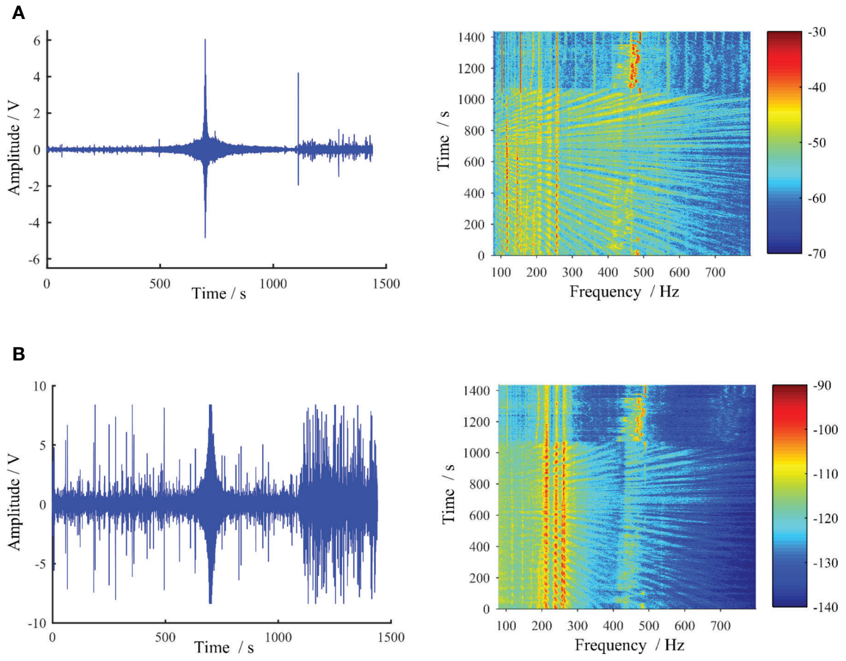


FIGURE 12
The recorded vector signals in time-domain and the corresponding LOFAR images, acoustic pressure **(A)** and vertical particle velocity **(B)**.

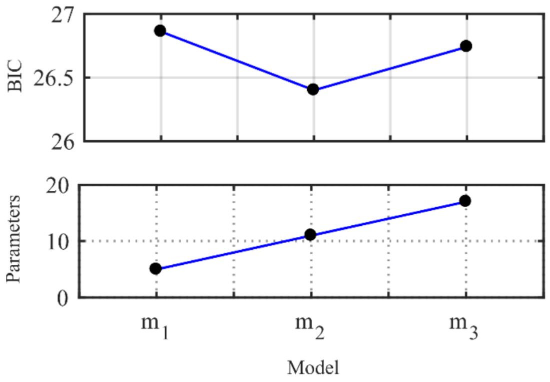


FIGURE 13
The result of model selection, the number of model parameters and the BIC value.

TABLE 2 Results of the measured data inversion.

Layer	Parameters	Search Bounds	Inversion Values (Mean ± Standard deviation)
Sedimentary	h_2/m	1 – 3	2.05 ± 0.21
	c_{p1}/m·s^{-1}	1500 – 1800	1651.91 ± 37.29
	c_{s1}/m·s^{-1}	700 – 800	748.95 ± 12.29
	ρ_{b1}/g·cm^{-3}	1.7 – 1.8	1.75 ± 0.01
	α_{p1}/dB·λ^{-1}	0.05 – 0.15	0.10 ± 0.01
	α_{s1}/dB·λ^{-1}	0.05 – 0.15	0.09 ± 0.01
Basement	c_{p2}/m·s^{-1}	1800 – 2200	1916.28 ± 59.25
	c_{s2}/m·s^{-1}	1100 – 1200	1120.31 ± 13.67
	ρ_{b2}/g·cm^{-3}	1.8 – 2.1	1.91 ± 0.04
	α_{p2}/dB·λ^{-1}	0.05 – 0.15	0.09 ± 0.01
	α_{s2}/dB·λ^{-1}	0.05 – 0.15	0.10 ± 0.01

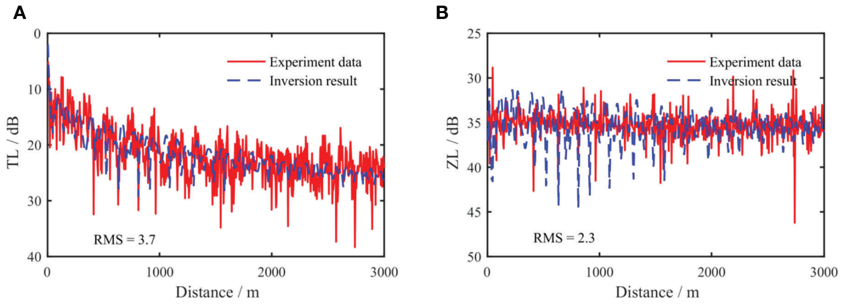


FIGURE 14
Comparison of inversion results of measured data. Acoustic pressure **(A)**, vertical waveguide characteristic impedance level **(B)**. The inversion result is the dashed line and Ex-periment data is the solid line.

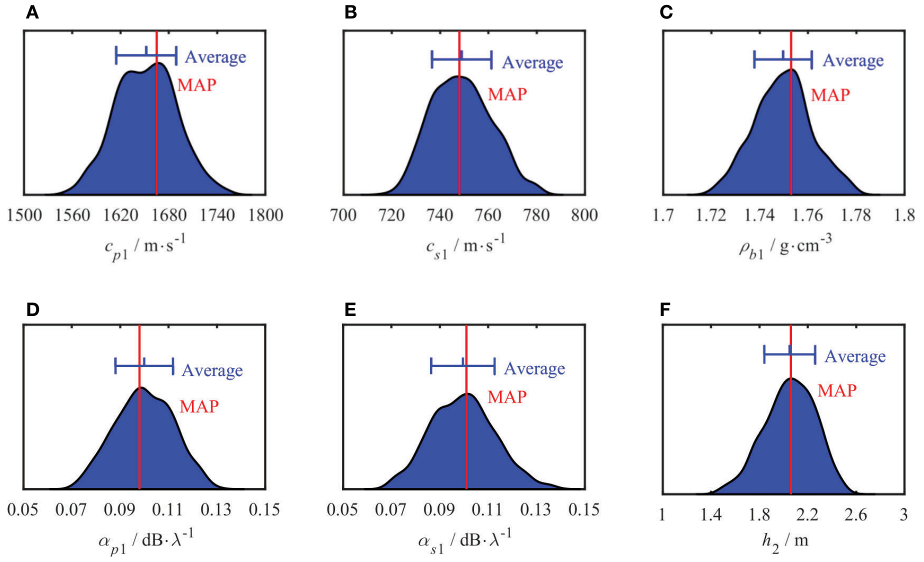


FIGURE 15
One-dimensional marginal probability distribution of sedimentary parameters. MAP value (red line), Average (blue vertical line in the middle) and standard deviation (blue vertical lines at both ends).

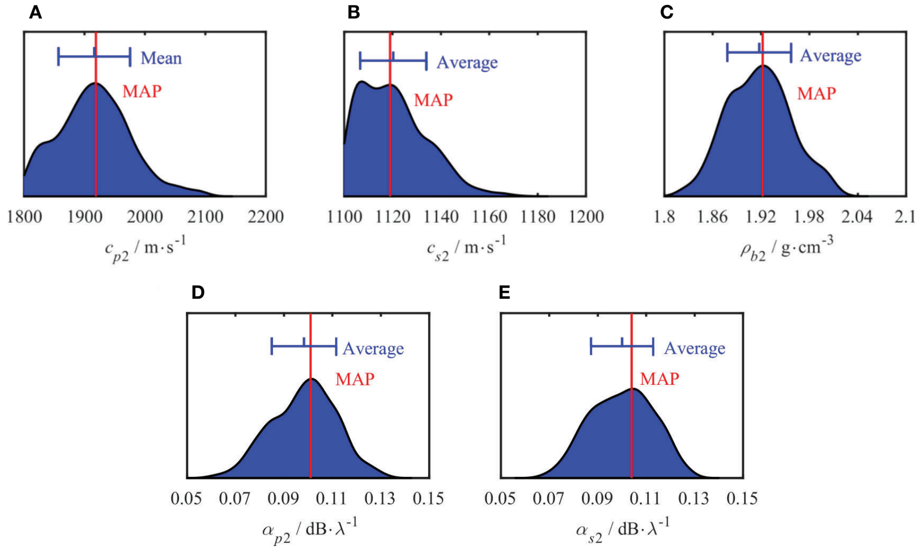


FIGURE 16
One-dimensional marginal probability distribution of semi-infinite parameters. MAP value (red line), Average (blue vertical line in the middle) and standard deviation (blue vertical lines at both ends).

set the vertical waveguide characteristic impedance Z_z as the match object,. The main conclusions are as follows:

(1) According to analysis results, in shallow sea environment with layered elastic seabed, the vertical waveguide characteristic impedance Z_z is proved has high sensitivity for geoacoustic parameters than that for p and v_z. Geoacoustic parameters include seabed structure, sound speed, density, and sound speed attenuation in seabed layers. Z_z is expected to provide higher precision for these geoacoustic parameters in geoacoustic inversion than the single acoustic pressure p or the single vertical particle velocity v_z, especially for sound speed attenuation, and Z_z is more suitable for geoacoustic parameter inversion.

(2) The Bayesian inversion approach can characterize the inversion result by the PPD from the statistical point of view. The best-parameterized model in inversion can be accurately selected based on the BIC. The Bayesian inversion approach can also measure the uncertainty of an inversion result. Because the inversion has a certain degree of randomness, it is more reasonable to determine the inversion result from the perspective of probability.

(3) Based on the experimental data of ship-radiated noise in the Dalian offshore area, the feasibility and reliability of the

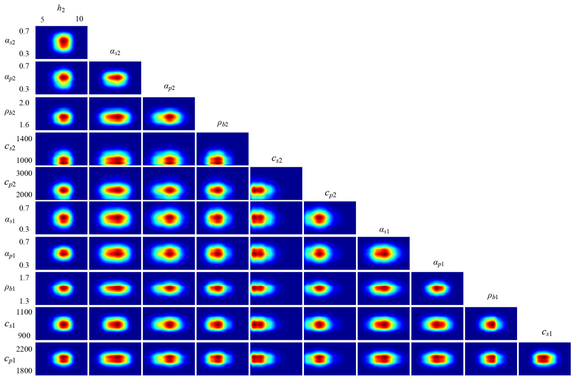


FIGURE 17
Two-dimensional marginal probability density distribution.

proposed inversion method were verified in practice. The sea experiment's inversion shows that the layered structure of the seabed and geoacoustic parameters can be accurately determined by this method. The RMS between the vertical waveguide characteristic impedance and the measured impedance is 0.38 dB, and the inversion results accurately reflect the characteristics of the seabed in the experimental area, and the inversion seabed structure is accordance with the previous information of experiment area.

We anticipate that these research results will enrich existing applications of acoustic vector sensors, and can be used to develop an effective inversion method for geoacoustic parameters in complex shallow sea environments, as well as to develop underwater acoustic technologies for long-range detection in seawater.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HZ and YX processed data, and co-authored wrote the original manuscript. QR supervised the study and reviewed the manuscript. XL and JW contributed to image format modification. ZC contributed to data. SZ and HF reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was found by the Zhejiang Province key research and development program (Grant No: 2020C02004), the Science Foundation of Donghai Laboratory (Grant No: DH-2022KF01018), the Marine defense Innovation Fund (Grant No: JJ-2020-701-09), the Science Foundation of Key Laboratory of Submarine Geosciences (Grant No: KLSG2201), Marine Science Zhejiang First-class Discipline Open Project Funding Project (Grant No: OFMS005).

Acknowledgments

We acknowledge the National Key Laboratory, College of Underwater Acoustic Engineering, Harbin Engineering University, China for supporting in this research.

Conflict of interest

Publisher's note

References

Dahl, P.-H., and Dall'Osto, D.-R. (2020). Estimation of seabed properties and range from vector acoustic observations of underwater ship noise. *J. Acoust. Soc Am.* 147, 345–350. doi: 10.1121/10.0001089

Dahl, P.-H., and Dall'Osto, D.-R. (2022). Range-dependent inversion for seabed parameters using vector acoustic measurements of underwater ship noise. *IEEE J. Ocean Eng.* 47, 682–689. doi: 10.1109/JOE.2021.3086880

Dosso, S.-E., Dong, H., and Duffaut, K. (2017). Model selection for profile structure in Bayesian geoacoustic inversion. *Acoust. Soc Am.J.* 14, 3470–3470. doi: 10.1121/1.4987215

Guo, X.-L., Yang, K.-D., and Ma, Y.-L. (2017). Geoacoustic inversion for bottom parameters *via* Bayesian theory in deep ocean. *Chin. Phys. Lett.* 34, 68–72. doi: 10.1088/0256-307X/34/3/034301

Hermand, J.-P. (1999). Broad-bandgeoacoustic inversion in shallow sea from waveguide impulse response measurements on a single hydrophone: theory and experimental results. *IEEE J. Ocean. Eng.* 24, 41–66. doi: 10.1109/48.740155

Huang, Y.-W., and Yang, S.-E. (2010). Spatial-temporal coherence of acoustic pressure and particle velocity in surface-generated noise. *J. Harbin. Eng. Univ.* 31, 137–143. doi: 10.3969/j.issn.1006-7043.2010.02.001

Jeon, Y., Lee, S.-J., and Kim, S. (2022). Estimation of shallow s-wave structures from the quasi-transfer spectrum and Bayesian inversion using microarray data at a deep drilling site at chungnam. *Geosci. J.* 26, 367–383. doi: 10.1007/s12303-021-0034-2

Jiang, X.-D., and Zhang, W. (2022). The research on Bayesian inference for geophysical inversion. *Rev. Geophys. Planet. Phys.* 53, 159–171. doi: 10.19975/j.dqyxx.2021-042

Kavoosi, V., Dehghani, M.-J., and Javidan, R. (2021). Underwater acoustic source positioning by isotropic and vector hydrophone combination. *J. Sound Vib.* 501, 116031. doi: 10.1016/J.JSV.2021.116031

Li, X.-M., Piao, S.-C., Zhang, M.-H., and Liu, Y. (2019). A passive source location method in a shallow water waveguide with a single sensor based on Bayesian theory. *Sensors* 19, 1452. doi: 10.3390/s19061452

Li, W.-L., Qiu, J.-S., Lei, P.-Y., Chen, X.-H., Fan, F., Deng, X.-J., et al. (2022). A real-time passive acoustic monitoring system to detect Yangtze finless porpoise clicks in ganjiang river, China. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.883774

Li, F.-H., and Zhang, R.-H. (2000). Bottom sound speed and attenuation inverted by using pulsed waveform and transmission loss. *Acta Acustica.* 25, 297–302. doi: 10.3321/j.issn:0371-0025.2000.04.002

Miao, Y.-J., Li, J.-H., and Sun, H.-X. (2021a). Multimodal sparse time-frequency representation for underwater acoustic signals. *IEEE J. Ocean. Eng.* 46, 642–653. doi: 10.1109/JOE.2020.2987674

Miao, Y.-J., Zakharov, V.-Z., Sun, H.-X., Li, J.-H., and Wang, J.-F. (2021b). Underwater acoustic signal classification based on sparse time-frequency representation and deep learning. *IEEE J. Ocean. Eng.* 46, 952–962. doi: 10.1109/JOE.2020.3039037

Park, C., Seong, W., and Gerstoft, P. (2005). Geoacoustic inversion in time domain using ship of opportunity noise recorded on a horizontal towed array. *J. Acoust. Soc Am.* 117, 1933–1941. doi: 10.1121/1.1862574

Ren, Q.-Y. (2013). *Using ship noise for ocean bottom geoacoustic parameters inversion* (Harbin. Eng. Univ: Harbin, China). doi: 10.7666/d.D430310

Ren, Q.-Y., and Hermand, J.-P. (2013). "Passive pressure and vector geoacoustic inversion offshore Amazon Rio mouth: A sensitivity study," in *Acoustics in Underwater Geosciences Symposium (RIO Acoustics)* (IEEE) Rio de Janeiro, Brazil, 1–4. doi: 10.1109/rioacoustics.2013.6683979

Ren, Q.-Y., Piao, S.-C., Ma, L., Guo, S.-M., and Liao, T.-J. (2018). Geoacoustic inversion using ship noise vector field. *J. Harbin Eng. Uni.* 39, 236–240. doi: 10.11990/jheu.201609031

Rong, S., and Xu, Y. (2022). Motion parameter estimation of AUV based on underwater acoustic doppler frequency measured by single hydrophone. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1019385

Santos, P., Rodríguez, O., Felisberto, P., and Jesus, S. (2009). "Geoacoustic matched-field inversion using a vertical vector sensor array," in *Int. Conf. Underwater. Acoust. Measurements.* (UAM), Nafplion: Greece, 1-3

Sun, D.-J., Ma, C., Yang, T.-C., Mei, J.-D., and Shi, W.-P. (2020). Improving the performance of a vector sensor line array by deconvolution. *IEEE J. Ocean Eng.* 45, 1063–1077. doi: 10.1109/JOE.2019.2912586

Sun, M., Li, F.-H., and Zhu, L.-M. (2013). Geoacoustic inversion with a vector sensor array in shallow sea water. *Tech. Acoust.* 32, 73–74. doi: 10.1063/1.4765947

Tollefsen, D., and Dosso, S.-E. (2020). Ship source level estimation and uncertainty quantification in shallow water *via* Bayesian marginalization. *J. Acoust. Soc Am.* 147, 339–344. doi: 10.1121/10.0001096

Xing, C.-X., Wu, Y.-W., Xie, L.-X., and Zhang, D.-Y. (2021). A sparse dictionary learning-based denoising method for underwater acoustic sensors. *Appl. Acoust.* 180, 108140.1–108140.13. doi: 10.1016/j.apacoust.2021.108140

Yang, H., Lee, K., Choo, Y., and Kim, K. (2020). Underwater acoustic research trends with machine learning: ocean parameter inversion applications. *J. Ocean. Eng. Technol.* 34, 371–376. doi: 10.26748/KSOE.2020.016

Yu, S.-W., Xiao, L., and Sun, W.-T. (2022). Parameter measurement of aircraft-radiated noise from a single acoustic sensor node in three-dimensional space. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.996493

Zhang, X.-B., Wu, H.-R., Sun, H.-X., and Ying, W.-W. (2021). Multireceiver SAS imagery based on monostatic conversion. *IEEE J-STARS* 14, 10835–10853. doi: 10.1109/JSTARS.2021.3121405

Zhang, X.-B., Yang, P.-X., and Dai, X.-T. (2019). Focusing the multireceiver SAS data based on the fourth order Legendre expansion. *Circ. Syst. Signal Pr.* 38, 2607–2629. doi: 10.1007/s00034-018-0982-6

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

Zhang, H.-G., Yang, S.-E., Piao, S.-C., Ren, Q.-Y., and Ma, S.-Q. (2010). A method for calculating an acoustic vector field. *J. Harbin. Eng. Univ.* 31, 470–475. doi: 10.3969/j.issn.1006-7043.2010.04.011

Zhang, X.-B., Ying, W.-W., Yang, P.-X., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Nav.* 14, 1055–1060. doi: 10.1049/iet-rsn.2019.0477

Zhao, Y.-P., Xie, S.-D., and Ma, C. (2022). Experimental investigations on the hydrodynamics of a multi-body floating aquaculture platform exposed to sloping seabed environment. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1049769

Zheng, G.-X., Piao, S.-C., Zhu, H.-H., Zhang, H.-Z., and Li, N.-S. (2021). Bayesian Inversion method of geoacoustic parameter in shallow sea using acoustic pressure field. *J. Harbin. Eng. Univ.* 42, 497–504. doi: 10.11990/jheu.201906062

Zheng, G.-X., Zhu, H.-H., Wang, X.-H., Sartaj, K., Li, N.-S., and Xue, Y.-Y. (2020). Bayesian Inversion for geoacoustic parameters in shallow Sea. *Sensors* 20, 2150. doi: 10.3390/s20072150

Zhou, J.-B., Tang, J., and Yang, Y.-X. (2021). A study on the estimation of source bearing in an ASA wedge: Diminishing the estimation error caused by horizontal refraction. *J. Mar. Sci. Eng.* 9, 1449. doi: 10.3390/jmse9121449

Zhu, H.-H. (2014). *Geoacoustic parameters inversion based on vertical waveguide characteristic impedance in acoustic vector field* (Harbin. Eng. Univ: Harbin, China). doi: 10.7666/d.D749708

Zhu, H.-H., Piao, S.-C., Zheng, H., Zhang, H.-G., and Tang, Y.-F. (2015). Research on geoacoustic parameters inversion in shallow sea from acoustic impedance. *Tech. Acoust.* 34, 204–206.

Zhu, H.-H., Zhu, J., Tang, J., Zhang, H.-G., and Zheng, G.-X. (2020). Effect of sediment acoustic characteristics on low-frequency acoustic field distribution in shallow water. *Acta Acustica* 45, 664–674. doi: 10.15949/j.cnki.0371-0025.2020.05.006

Frontiers in Marine Science

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fmars.2023.1111176

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Hiroyuki Matsumoto,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Ze-Nan Zhu,
Ministry of Natural Resources, China

*CORRESPONDENCE
Haocai Huang
✉ hchuang@zju.edu.cn
Guangming Li
✉ guangming_1224@hotmail.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 29 November 2022
ACCEPTED 10 January 2023
PUBLISHED 27 January 2023

CITATION
Xu S, Feng R, Xu P, Hu Z, Huang H and Li G (2023) Flow current field observation with underwater moving acoustic tomography. *Front. Mar. Sci.* 10:1111176.
doi: 10.3389/fmars.2023.1111176

Flow current field observation with underwater moving acoustic tomography

Shijie Xu[1,2], Rendong Feng[1,2], Pan Xu[3], Zhengliang Hu[3], Haocai Huang[1,4,2]* and Guangming Li[5]*

[1]Ocean College, Zhejiang University, Zhoushan, China, [2]Hainan Institute of Zhejiang University, Sanya, China, [3]College of Meteorology and Oceanography, National University of Defense Technology, Kaifu District, Changsha, China, [4]Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [5]National Innovation Institute of Defense Technology, Fengtai District, Beijing, China

Underwater environment observation with underwater acoustic tomography has been considerably developed in recent years. Moving sound transmission can obtain the observation of entire spatial area with sound station moving. Various internal structures, unique surface and submarine boundaries and changing environment constitutes a complex acoustic propagation channel. This paper focus on the inversion method and signal resampling for sound moving transmission. Also, the current field in three-dimensional (3D) scale is also studied. A five-station sound transmission experiment with four moored station and one moving station that conducted in range of 500m×500m at Huangcai reservoir, Changsha, China is presented. Signal resampling is performed to get correlation of received acoustic data. The vertical layer-averaged flow current results between moving station and moored station are inversed with 2D grid method. 3D flow current field result is composed by grid-averaged inversion current of vertical profile *via* moving station at different moment. The received results of reciprocal signal transmission between two moored stations and one moving station were used for layer-averaged current variations at vertical scale and grid-averaged current. The feasibility of the method in underwater moving acoustic tomography research is proved and its applicability is discussed. The proposed underwater acoustic tomography technology develops an innovative idea for the further development of temporal- spatial grided tomography observation.

KEYWORDS

underwater moving acoustic tomography, 3D flow current field, Doppler shift, coastal acoustic tomography, water temperature

1 Introduction

The development of underwater observation technology promotes the sensing and monitoring methods of marine environment (Munk & Wunsch, 1979). Coastal acoustic tomography (CAT) is an advanced technology for environment observation in the shallow waters, which is developed from ocean acoustic tomography (OAT) (Munk & Wunsch, 1979; Kaneko et al., 2020). The travel time delays obtained from the reciprocal transmission signals between sound stations are used to inverse the range-average water temperature and flow current in the observation area. In typical CAT observation experiments, inversion results are calculated using sound transmission information between sound stations in the multi-station network. This approach is of higher efficiency compared with single point measurement such as Acoustic Doppler Current Profiler (ADCP), Conductivity Temperature Depth system (CTD), Temperature Depth system (TD), etc. However, the previous method used travel time to inverse the mean variations of environment (temperature or flow) in observation area by layer-averaged or grid-averaged, without refining of all changing at each spatial position. The high precision variations at each area are necessary. In recent years, CAT technology is widely used for water parameter survey and observation in coastal water, islands region, river and so on. More advanced methods and experimental formats are proposed (Chen et al., 2021; Xiao et al., 2021; Xu et al., 2021; Chen et al., 2022; Mohamad Basel Al et al., 2022; Xu et al., 2022).

Sawaf et.al analyzed the inflow characteristics within a slow water-flow environment barrier lake *via* two pairs of fluvial acoustic tomography (FAT) systems, and developed an equation to estimate the flow direction (Al Sawaf et al., 2017; Al Sawaf & Kawanisi, 2018). Chen Minmo et.al conducted many studies *via* CAT in coastal areas for water temperature and flow current inversion (Chen et al., 2016; Chen et al., 2017; Chen et al., 2019; Chen et al., 2020; Chen et al., 2020; Chen et al., 2020; Chen et al., 2021). Recently, they found the standing-type waves by reconstructing the two-dimensional (2D) flow fields using a coast-fitting inversion model of five sound stations, which led to flood- and ebb-dominant tidal current asymmetries for the western and eastern halves in model domain, respectively (Chen et al., 2022). Besides, CAT technology was used for mapping tidal current and salinity. Kawanisi et. al continuously monitored two-dimensional distributions of current and salinity at a shallow tidal junction for 34.4 days and efficiently estimated the variations of the currents and salinity (Bahreinimotlagh et al., 2016; Kawanisi et al., 2016; Bahreinimotlagh et al., 2019; Chen et al., 2021; Xiao et al., 2021; Mohamad Basel Al et al., 2022). Future more, in order to obtain more spatially environment data, underwater moving acoustic tomography is proposed and developed. Chengfen Huang et.al achieved the first application of this method for the acoustic mapping of ocean currents in a shallow-water environment *via* four moored station and one ship-towed station (Huang, 2019; Huang et al., 2019). The errors of travel time difference caused by relative motion were corrected by Doppler shift estimated from the measured channel impulse response.

Unlike fix-station observations, moving sound transmission can obtain the observation of entire spatial area with sound station moving. Various internal structures, unique surface and submarine boundaries and changing environment constitutes a complex acoustic propagation channel. In particular, the problem of signal frequency drift, multi-path ray identifying need to be considered with sound station motion. This paper focus on the inversion method and signal resampling for sound moving transmission. Also, the current field in three-dimensional (3D) scale is also studied. A five-station sound transmission experiment with four moored station and one moving station that conducted at Huangcai reservoir, Changsha, China is presented. Signal resampling is performed to get correlation of received acoustic data. The vertical layer-averaged flow current results between moving station and moored station are inversed with 2D grid method (Huang et al., 2021; Xu et al., 2021; Xu et al., 2022). 3D flow current field is also reconstructed. Meanwhile, the advantage and disadvantage of underwater moving acoustic tomography are discussed.

This rest of paper is structured as follows: in Section 2, experimental settings of moving acoustic tomography is introduced. The inversion method of moving acoustic tomography and ray simulation is presented in Section 3. Section 4 focus on inversion results in 2D vertical profiles, 3D flow current field is also mapped in this section. Discussion and conclusion are given in Section 5 and Section 6, respectively.

2 Overview of experiment

An underwater moving acoustic tomography experiment was conducted with five high-frequency CAT systems at March 5th, 2022 in the Huangcai Reservoir, west of Changsha, China. Figure 1 shows the deployment of four moored stations and one moving station. Moored acoustic stations (S1-S4) were distributed at the corners of square flow confluence area, the moving acoustic station (S5) was towed with a fish boat using s 6m rope along circle and straight-line paths inside the observation area. Each moored station was fixed by a buoy and an anchored. All acoustic stations were equipped with high-frequency CAT system, 24V and 12V battery array, sound transceiver, GPS (less than 0.6μs time error) and TD (Temperature Depth Sensor). The reciprocal sound transmissions were successfully carried simultaneously between station pairs. GPS was used for time synchronisation of five stations and TD is used to record depth variations of each station.

In this experiment, sound transmissions between moving station and each moored station were analyzed. The high-frequency CAT systems that used in this experiment were developed by Hiroshima University (Kaneko et al., 2020). During the experiment, the distance between moving station and fix station is changing at range from 60m to 600m. The 9th order M-sequence phase-modulated sound signals were transmitted by transceivers; modulation depth is 2; each station was assigned a specific signal M-sequence code for code division multiple access. The signal-to-noise ratio (SNR) was improved by M-sequence of $20\log\sqrt{2^9 \times 2 - 1} = 30.098$ dB. The operation frequency range of transceiver cover 10kHz-90kHz. The center frequency of transmitted sound is 50kHz with 1-mintue transmission interval in this experiment. The length of one-period was 20.46ms (1023 bits). Table 1 shows the detail of acoustic signal and sound stations.

Due to the sensitivity of the high-frequency signal to the observation environment, undulate terrain and the station distance, the sound ray path structure is significantly fluctuated by the moving station. The terrain of experiment area (S1-S2, S2-S3, S3-S4 and S4-S1) were measured by depth gauge scanning. Unfortunately, S2 did not receive the acoustic signals because of the incorrect system setting,

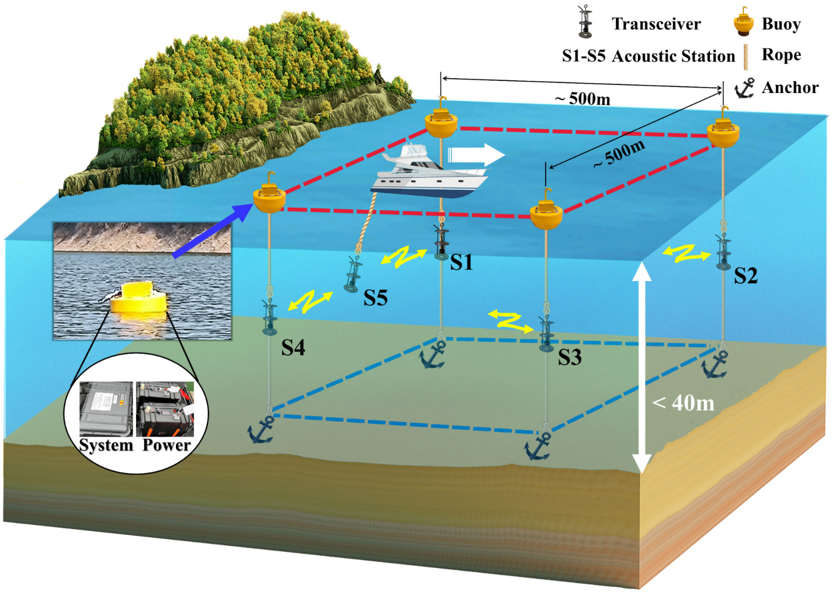


FIGURE 1
CAT station deployment in Huangcai Reservoir view from the north. Magnified figures with black circle show the layout of CAT system. Figure legends is shown at top right. White arrow shows the direction of motion station. The observation area is within 500m × 500m and the water depth is less than 40m.

and the acoustic signal between S4 to S5 is mostly blocked due to the terrain. Therefore, in this paper, only the received data of S1to S5 and S3 to S5 are analyzed to simplify the processing of acoustic signals.

3 Methodology

In this section, the inversion problem and the Doppler shift effect are analyzed to reconstruct the flow current field by sound reciprocal transmission under relative motion condition. Ray simulation and multi-peak matching are also demonstrated.

3.1 Forward problem under relative motion condition

Reconstructing range-averaged sound speed and flow current velocity in the observation area by sound reciprocal travel time is the basic principle of CAT technology. Moving transceivers have relative velocity under the relative motion condition between sound station pairs (Xu et al., 2022). Figure 2 shows reciprocal sound signal transmission between relative motion station of S_i and S_j. Without loss of generality, the transmission mode is assumed with a uniform sound speed (C_m) and a uniform current velocity (U_m) in each grid of the unbounded medium. Two transceivers are moving with speeds V_I and V_J, the v_i and v_j are the projection of real speeds in the vertical of station pairs. The acoustic signals are sent and receipted simultaneously, where sound rays travel along sound paths connecting the station pairs. Thus, the reciprocal travel time of S_i and S_j are not only influenced by the environment (temperature, current, salinity), but also transceiver motion. Each sound ray travel time from S_i and S_j, t_{ij}, and the reversed direction, t_{ji}, can be expressed as follow:

TABLE 1 Transmission signal and station parameters.

Central frequency (Hz)	50000				
Order of M-sequence	9				
Q[1]	2				
Transmit interval (min)	1				
Station	S1	S2	S3	S4	S5
Transceiver depth (m)	23.5	24.9	25.3	24.2	8.2

[1] Q value denotes the number of cycles per digit in M-sequence.

$$
\begin{aligned}
t_{ij} &= \int_{\Gamma^+} \frac{ds}{C_m + U_m - v_j} \approx \sum \frac{l_{ij}^k}{C_m^k + U_m^k - v_j} \\
t_{ji} &= \int_{\Gamma^-} \frac{ds}{C_m + U_m - v_i} \approx \sum \frac{l_{ij}^k}{C_m^k - U_m^k - v_j}
\end{aligned}
\tag{1}
$$

where l_{ij}^k is the length of sound ray pass through each gird in Figure 2. C_m^k and U_m^k are the grid-averaged sound speed and current, respectively. k is the grid number. The positive direction of v_i is away from the station j, similar with v_j.

By summing and subtracting the reciprocal transmission travel time in Eq. 1 with the reference travel time t_0, we obtained the travel time deviation (δt) and travel time difference (Δt):

$$
\begin{aligned}
\delta t &= t_{ij} + t_{ji} - 2t_0 = \sum \left(\frac{l_{ij}^k}{C_m^k + U_m^k - v_j} + \frac{l_{ij}^k}{C_m^k - U_m^k - v_i} - 2\frac{l_{ij}^k}{C_0^k} \right) \\
\Delta t &= t_{ij} - t_{ji} = \sum \left(\frac{l_{ij}^k}{C_m^k + U_m^k - v_j} - \frac{l_{ij}^k}{C_m^k - U_m^k - v_i} \right)
\end{aligned}
\tag{2}
$$

Simplifying the Eq. 2, uniform sound speed (C_m^k) in grid k decomposes into the sum of reference sound speed (C_0^k) and sound speed difference (δC^k).Both uniform current velocity ($U_m^k < 1m/s$) and

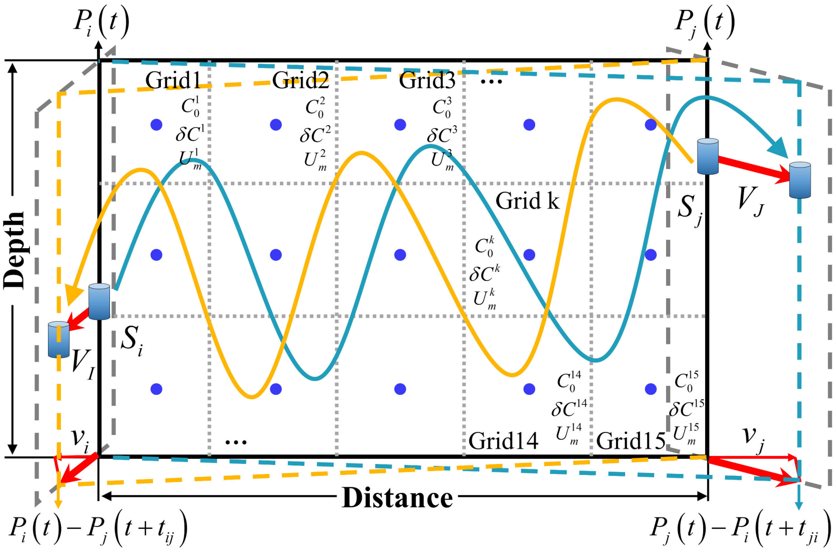


FIGURE 2
Reciprocal sound transmission between relative motion station S_i and S_j. Red arrows are the directions of station motion. The blue and yellow arrow lines denote the ray path and travel direction. The gray dotted lines are the grid division lines. The blue circles are the center point of girds for inversion calculation. Each grid assumes C_0^k, δC^k and U_m^k, respectively.

transceiver speed (v_i<1m/s) are far smaller than reference sound speed (about 1500m/s), the travel time deviation (δt) and travel time difference (Δt) can be simplified as Eq. 3 using Taylor expansions, the higher-order terms of Taylor expansion are omitted (Xu et al., 2021).

$$\delta t \approx -\sum\left(\frac{l_{ij}^k\left(2\delta C^k - v_i - v_j\right)}{\left(C_0^k\right)^2}\right) = -\sum\frac{2l_{ij}^k}{\left(C_0^k\right)^2}\left(\delta C^k - \frac{v_i + v_j}{2}\right)$$
$$\Delta t \approx -\sum\left(\frac{l_{ij}^k\left(2U_m^k + v_i - v_j\right)}{\left(C_0^k\right)^2}\right) = -\sum\frac{2l_{ij}^k}{\left(C_0^k\right)^2}\left(U_m^k - \frac{v_j - v_i}{2}\right) \quad (3)$$

From the Eq. (3), δt is mainly influenced by δC^k and the sum of v_i and v_j, where δC^k is dominated by water temperature fluctuations about 0~1°C. Normally, δC^k is less than $(v_i+v_j)/2$. Δt is mainly influenced by U_m^k and the sum of v_i and v_j, U_m^k is larger than $(v_i+v_j)/2$. Meanwhile, the error of δt is caused by the position drift of acoustic station and the error of Δt is caused by the clock drift (Zhu et al., 2021). Thus, it can be summarized that the large error exists in solving of δt (water temperature). In other words, the underwater moving acoustic tomography is appropriate to solve Δt (water current velocity).

In this experiment, four moored acoustic station and one moving acoustic station are used. Comparing with the tidal current velocity at each moored station, the speeds of moored stations are considered to be close to zero, which can be neglected. Therefore, during the reciprocal transmission the speed of moving station S_j can represent the relative speed between moored station S_i and moving station S_j, i.e., $v_i = 0$. The S_j relative speed v_j is positive when the two stations move closer (S_j away from S_i). Thus, the travel time difference and the current speed U_m are expressed as Eq. 4.

$$\Delta t = -\sum\frac{2l_{ij}^k}{\left(C_0^k\right)^2}\left(U_m^k - \frac{v_j}{2}\right) \Rightarrow U_m = -\frac{\left(C_0\right)^2}{2L_{ij}}\Delta t + \frac{v_j}{2} \quad (4)$$

where Δt is obtained by the sound transmission, C_0 is measured by CTD, L_{ij} is calculated by ray simulation and v_j is the travel speed of the boat using GPS. Eq. (4) can be expressed in a matrix form as:

$$\begin{bmatrix}\Delta t_1\\ \Delta t_2\\ \vdots\\ \Delta t_n\end{bmatrix}_{n\times 1} = -2\begin{bmatrix}\frac{l_{ij}^{11}}{\left(C_0^1\right)^2} & \frac{l_{ij}^{12}}{\left(C_0^2\right)^2} & \cdots & \frac{l_{ij}^{1K}}{\left(C_0^k\right)^2}\\ \frac{l_{ij}^{21}}{\left(C_0^1\right)^2} & \frac{l_{ij}^{22}}{\left(C_0^2\right)^2} & \cdots & \frac{l_{ij}^{2K}}{\left(C_0^k\right)^2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{l_{ij}^{n1}}{\left(C_0^1\right)^2} & \frac{l_{ij}^{n2}}{\left(C_0^2\right)^2} & \cdots & \frac{l_{ij}^{nk}}{\left(C_0^k\right)^2}\end{bmatrix}_{n\times k}\begin{bmatrix}U_m^1 - \frac{v_j}{2}\\ U_m^2 - \frac{v_j}{2}\\ \vdots\\ U_m^k - \frac{v_j}{2}\end{bmatrix}_{k\times 1} \quad (5)$$

where Eq. 5 represents the reciprocal transmission between a sound station pair. n is the number of sound ray paths, n = 1 corresponds to the 1st ray path. k is the number of girds, k = 1 corresponds to gird 1. l_{ij}^{nk} represents the length of nth ray path in kth grid between S_i and S_j.

The inversion method, the ray simulation of motion station pair and the travel time identification are presented in the following. Besides, the approximate distance of station pair at clock t is denoted by $P_i(t) - P_j(t)$. It should be noted that, due to the relative motion between S_i and S_j, the exact distance of the sound signal transmission from S_i to S_j is $P_i(t) - P_j(t+t_{ij})$ and S_j to S_i is $P_j(t) - P_i(t+t_{ji})$. The relative speed is much smaller than the sound speed, thus the error of the exact and the approximate distances can be ignored (Kaneko et al., 2020). However, the Doppler effort caused by the relative motion on the signal transmission need to be discussed, as shows in section 3.2.

3.2 Estimation and compensation of Doppler effect

The Doppler effect is caused by the motion of inhomogeneous water movement, such as: vehicle motions, surface waves, water turbulence, the relative motion of transceivers, etc. The time-varying effects of acoustic signal caused by Doppler effect in the water environment result in Doppler frequency shifts and additional frequency extensions. In this paper, for the underwater acoustic broadband transmission system, the Doppler effect is described by α (Doppler factor), which is defined as the ratio of the relative velocity V (v_j) between moored and moving transceivers to the sound speed C of observation area]. The Doppler factor is the ratio of frequency offset to signal frequency in the frequency domain as Eq. 7.

$$\alpha = \frac{C}{V} = \frac{\Delta f}{f} \tag{6}$$

where Δf is the frequency offset and f is the carrier frequency of acoustic signal.

Compared to the rate of electromagnetic wave in wireless communication (3×10^8 m/s), the sound wave in water is only about $1500m/s$, where exists significant Doppler effect for moving signal transmission. Thus, the Doppler effect need to be estimated and compensated during the reciprocal signal transmission of transceivers' relative motion.

The equivalent frequency of the moored station and moving station is different. If the receiving station (moving station S_j) is moving away from the sending station (moored station S_i) with speed v_j, for moving station, the travel time of one wavelength (λ) is $t_v \approx \lambda/(C+v_j) = (C/f)/(C+v_j)$, the equivalent frequency and Doppler frequency shift are as:

$$\begin{aligned} f_v &= \tfrac{1}{t_v} = \tfrac{C+v_j}{C/f} = f\left(1+\tfrac{v_j}{C}\right) = f(1+\alpha) \\ \Delta f &= f_v - f \approx f\cdot\alpha \end{aligned} \tag{7}$$

For moored station, the received signal wavelength of moored station is compressed v_j/f, thus, equivalent signal wavelength is $\lambda_v = \lambda - v_j/f = (C-v_j)/f$, the equivalent frequency and Doppler frequency shift are as:

$$\begin{aligned} f_v &= \tfrac{C}{\lambda_v} = \tfrac{C}{\left(C-v_j\right)/f} = f\left(\tfrac{1}{1-v_j/C}\right) = f\left(1+\tfrac{v_j}{C}+\left(\tfrac{v_j}{C}\right)^2\cdots\left(\tfrac{v_j}{C}\right)^n\right) \approx f(1+\alpha) \\ \Delta f &= f_v - f \approx f\cdot\alpha \end{aligned} \tag{8}$$

During the experiment, if single-frequency acoustic signal was used. Received signal usually consists of multipath arrivals with different travel times that caused different Doppler frequency shifts in an underwater acoustic channel (Huang, 2019; Huang et al., 2019). As Eq. 7 and Eq. 8 shows, the moored station has same Doppler frequency shift with moving station. The received acoustic signal $s(t) = e^{jwt}$ for multipath arrivals can be expressed as Eq. 9 when there is no Doppler frequency shift.

$$y(t) = H(\omega)e^{jwt}, \quad \left(H(\omega) = \sum_{n=1}^{N} A_n e^{-jw\tau_n}\right) \tag{9}$$

where n is the sound ray number, A_n and τ_n are the magnitude and travel time delay, respectively.

The received multipath signal with Doppler shift is express as follows:

$$y(t) = H(\omega,t)e^{j\omega t} = \sum_{n=1}^{N} A_n e^{-j\omega(\tau_n-\alpha_n t)} e^{j\omega t} \tag{10}$$

It can be concluded that received signal is no longer a single-frequency signal and the transmission channel is time-varying. Therefore, non-consistent frequency received signal is compensated by resampling (Eq. 11), where basically eliminate the signal compression in time domain due to Doppler effect.

$$y'(t) = y\left(\frac{t}{1+\overline{\alpha}}\right) \tag{11}$$

It should be noted that the moving boat (moving station) was at a constant speed ($0.6m/s$) during the experiment, and the real-time positions were recorded by GPS. In post-processing, the direction and speed of moving station were established and plotted at each reciprocal signal transmission moment, as shown in Figures 3A, D. The Doppler effect ($\overline{\alpha}$) in Eq. 11 at each moment is calculated by this way. However, system still exist a deviation between computational and real solutions as Eq. 12 shows, which is small enough to be negligible.

$$\ell = \frac{\alpha - \overline{\alpha}}{1+\overline{\alpha}} f \tag{12}$$

3.3 Inversion problem

The two-dimensional (2D) vertical depth-averaged flow current velocity can be calculated by the inverse method (Xu et al., 2021). Furthermore, three-dimensional (3D) current variations are reconstructed by multiple two-dimensional vertical current results through moving acoustic tomography model in a short period. Eq. 13 is the basic vertical solving matrix of S_i and S_j. The matrix expression is simplified to the following form:

$$\mathbf{y=Ex+n} \tag{13}$$

where y is the travel time difference data vector ($y = \{\Delta t_i\}$), E is the observation matrix ($E = -2\{l_{ij}^{nk}/(C_0^k)^2\}$) created by ray simulation, X is the unknown variable vector of current velocity ($X = U_m^k - \frac{v_j}{2}$), and n is the residual vector to measure the error terms.

The objective optimization function is defined as follows (Kaneko et al., 2020):

$$\mathbf{J=(y-Ex)^T(y-Ex)+\lambda x^T H^T Hx} \tag{14}$$

where H is a regularization matrix to smooth the grid solution by moving average of nearby grids. λ is the Lagrange multiplier to converge the results.

The expected solution is:

$$\mathbf{\hat{x}=(E^T E+\lambda H^T H)^{-1}E^T y} \tag{15}$$

The λ in Eq. 14 and Eq. 15 is obtained when the residual defined is less than the predetermined error value. The inversion error $\hat{n}$ and the iteration residual error $\hat{x}_{err}$ are expressed as:

$$\mathbf{\hat{n}=y-E\hat{x}=\left\{I-E\left(E^T E+\lambda H^T H\right)^{-1}E^T\right\}y} \tag{16}$$

$$\mathbf{\hat{x}}_{err} = \left(\mathbf{E}^T\mathbf{E}+\lambda\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{E}^T\mathbf{\hat{n}} \tag{17}$$

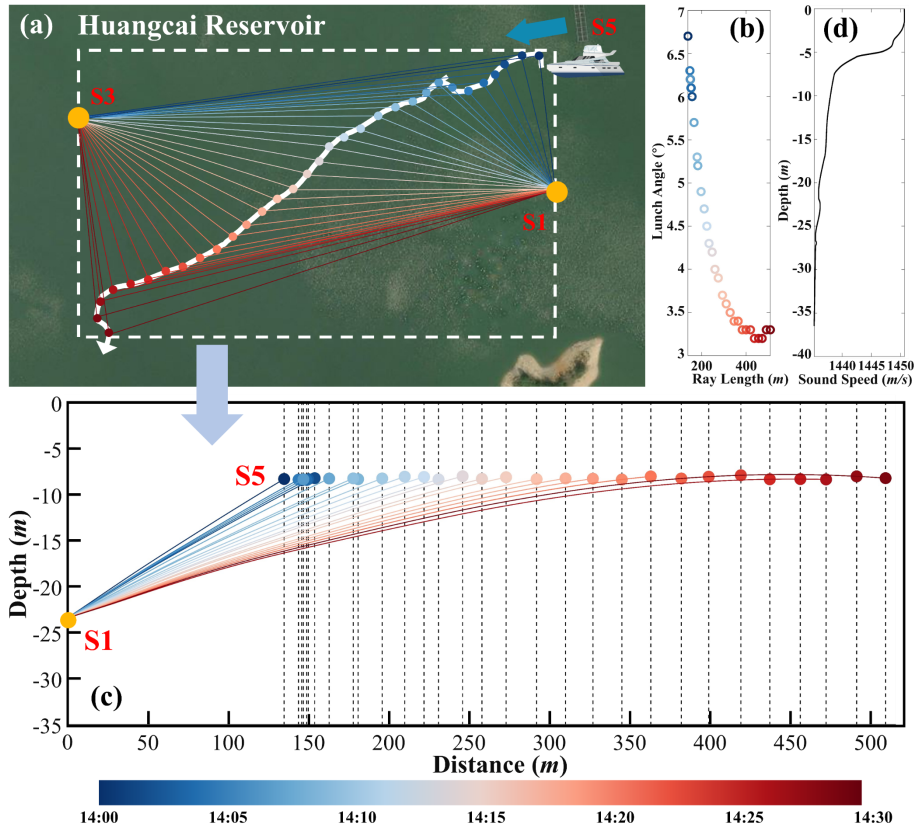


FIGURE 3
Reciprocal signal transmission during relative motion. **(A)** is the path of moving station at different moment. **(B)** is the relationship between launch angle and ray length. **(C)** is the ray simulation between station S1and S5 (Note that: only the direct rays are shown at **(C)**. **(D)** is the sound speed profile in observation area. The color bar below shows the color of different moment.

The grid-averaged current velocity in 2D vertical profile can be obtained *via* above process. Thus, 3D current field is reconstructed by the 2D current dataset at multiple moments.

3.4 Ray simulation and multi-peak matching

Ray simulation is used as a post-processing simulation to identify the multi-peak of received acoustic signal, E in Eq. 13 is formed by grid divided of ray structure after matching multi-peaks. In typical acoustic tomography experiments, all acoustic stations are fixed, and it only needs to use the obtained terrain for ray simulation between the station pairs. However, in this experiment, the ray structure is changing due to the varying terrain with moving station and the time-varying channel by Doppler effect.

The obtained resampled signal results after correlation in Figure 4 show that the peak is not very obvious. The red circles in top of Figure 4 are the identified peaks by matching with ray simulations. The process of peak identification has followed steps: acoustic signals are resampled as method in part 4.2 before correlation; the station distances at each moment are obtained by GPS recording between moving and moored station; the arrival signals are identified and matched with the largest SNR value peaks in the reference travel time window (obtained by station distances). Also, ray simulations at different distance are conducted to match peaks in receiving signals. Figures 3B, C show the relationship between launch angle with ray length and ray simulation results of S1 to S5, respectively. The sound ray structure stacked diagram (Figure 4C) is the simulation at different station distance based on sound speed profile measured by CTD.

The reciprocal resampled acoustic signals after correlation at 14:00 are shown in Figure 4 blow, where the results have significant multi ray path. Due to uncertainty of terrain caused by the moving stations, the multipath cannot be matched with ray simulation, and only the direct ray was simulated and identified. In addition, the low SNR values show that the Doppler effect caused by the moving station has a significant influence on signal attenuation. It should be noted that the sound speed profile is a negative gradient and has been introduced in the article (Xu et al., 2022).

In the inversion calculation, the vertical profiles of sound station pairs are used as gridded slice method in 2D (Xu et al., 2021). As only the direct ray was used, the horizontal stratification lines were set at 6m and 26m and the vertical stratification was divided into four layers evenly. The inverse results of the current field for a total of four grids at this layer were solved. The division method is designed to ensure that the ray length in each grid is close to reduce the inversion error.

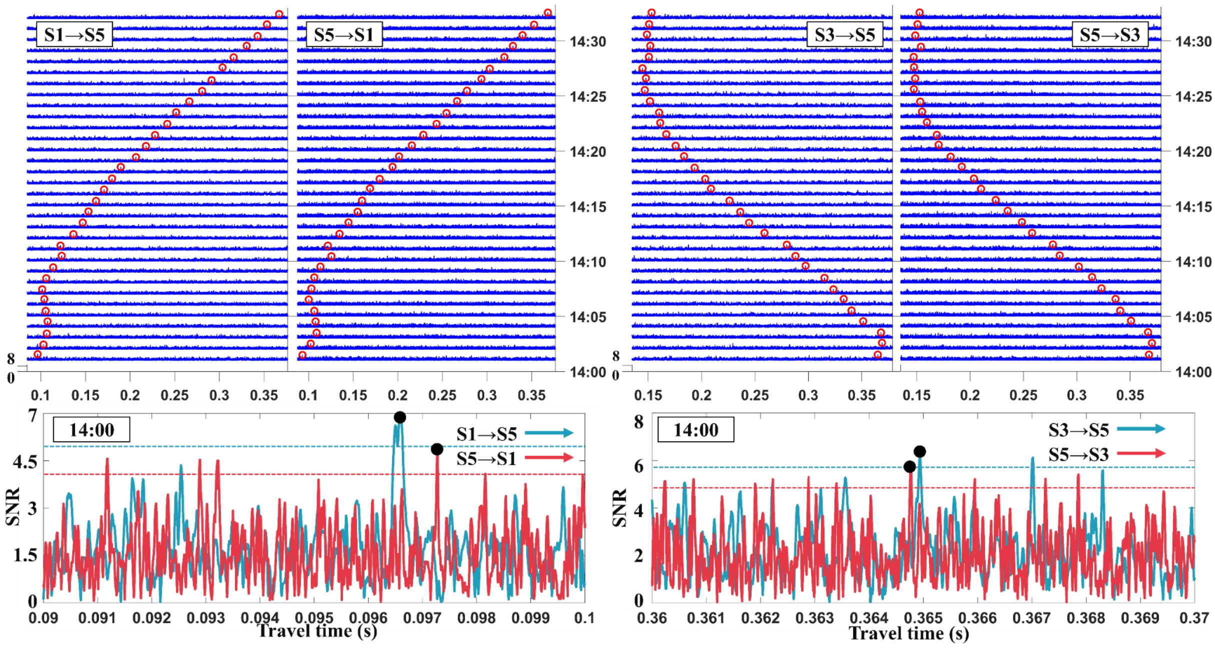


FIGURE 4

Correlation results of resampled signal (S1-S5 and S3-S5). The upper four figures are correlation results from 14:00 to 14:32. The abscissa axis above denotes the travel time of acoustic signal; ordinate axis denotes the date time of sending signals; the peaks' height denotes SNR value; and the red circles denote the identified peaks. The lower two figures are the reciprocal transmission patterns of S1-S5 and S3-S5, respectively. Red and blue dotted lines are the divide lines for high SNR.

4 Results

4.1 Travel time differences

Figure 5 shows the variations of travel time and travel time differences of direct arrival rays of S1 to S5 and S3 to S5, respectively. As shown in Figure 5, the variation trends of reciprocal transmissions at each station pair are basically the same. The travel time difference (TTD) calculated as Eq. 2 is the fluctuate obviously which corresponds to the current changing. The mean and standard travel time differences are displayed in Table 2. The mean travel time difference of S1-S5 is significantly higher than S3-S5, which means the flow current variations along S1-S5 path are stronger.

4.2 Inversion result

The layer-averaged flow current velocity of S1 to S5 and S3 to S5 are shown at Figures 6A, B, respectively. The result of layer-averaged

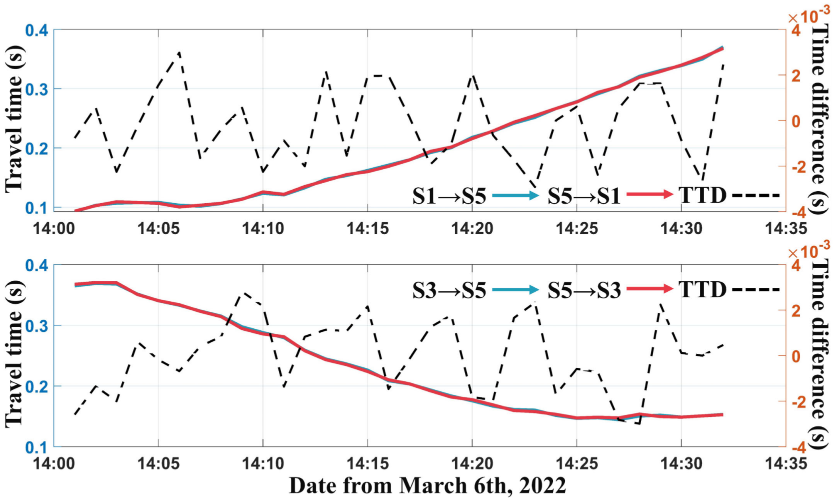


FIGURE 5

Travel time and travel time differences. Left axis is the range of travel time and right axis is the range of travel time differences.

TABLE 2 Mean and standard deviation of travel time differences.

Station pair	S1-S5	S3-S5
Mean of TTD (*ms*)	-0.1659	-0.0141
Standard deviation of TTD (*ms*)	1.7132	1.6521

current is the mean of all grid central point inversion results at same depth. The red lines display the current fluctuations and blue lines are the inversion results using 5-minute moving average. The moving average reduces the effect of anomalies and retains more information on variations through weighted distribution, which makes current trends more visible.

From Figure 6, variations of the layer-averaged current in the vertical profile is about -5 to 5 *cm/s*, positive and negative represent the direction of current (the pre-set range-layer error before inversion is 0.01*m/s*, and the inversion error is much smaller than the pre-set value). Although the current at this layer is not strong, the direction of current varies quickly. The variability of directions in the layer-averaged currents is caused by the turbulence current in the reservoir. The inversion method for average current ignores the current in the vertical direction, which normalises the changing of current *via* travel time differences in certain depth lines (layer-averaged method) or grid points (grid-averaged method).

Figure 7A is a 3D combination figure of flow current velocity and direction at each grid in vertical profiles. It shows the inversion current field results *via* the moving station at different locations. All 2D vertical profiles have four grid-averaged results at depth of 16*m*. Figures 7B-G is the 2D horizontal flow fields. The Figure 6 indicates the convergence results of inversion current field over a 30-minute observation period. The current along vertical profiles vary about 5-minute interval of each station pair. But the results of horizontal flow fields show the current directions is basically unchanged. It is speculated that the flow field is stability, but a small-scale dynamic progress and water exchange exist in observation area. However, only the matched direct ray paths are used for inversion, where more accurate changing of current field cannot be identified.

5 Discussion

Underwater moving acoustic tomography is an enhancement and application of CAT technique. It combines the traditional observation with the motion observation technique to cover the entire spatial area. By reconstructing 2D layer-averaged results and 3D spatial distribution in a feasibility experiment at Huangcai Reservoir, the method has been basically proved to facilitate the observation of current field.

The advantage of moving acoustic tomography compared with traditional observation *via* fixed stations is that the moving stations can compensate for areas of that sound ray in fixed observation do not pass through, reconstructing the environment variation over the whole area. However, the disadvantage of this technique is the drift and multipath uncertainty of reciprocal acoustic signals caused by Doppler effect and terrain variations. For example, in this paper, the experiment was carried out with four moored stations and one moving station, but due to the terrain and signal, only two moored stations and one moving station were effectively communicated with each other. Further, although recovering correlation signal *via* resampling techniques in moving acoustic tomography is initially proposed in this paper, the resolution of multipath identification is still a challenge (especially in shallow areas).

In addition, the 3D current results in 4.2 are reconstructed using grid-averaged current in vertical profile, instead of using stream function to create 2D current field (Syamsudin et al., 2016; Zhu et al., 2017a; Zhu et al., 2017b; Syamsudin et al., 2019). This approach is faster and more intuitive, but is unable to characterise the variation of current along vertical direction. This experiment used only three stations to reconstruct the current field, the errors in final results still exists. It is believed that the observation data and inversion results can be more accurate with more station reciprocal transmission data.

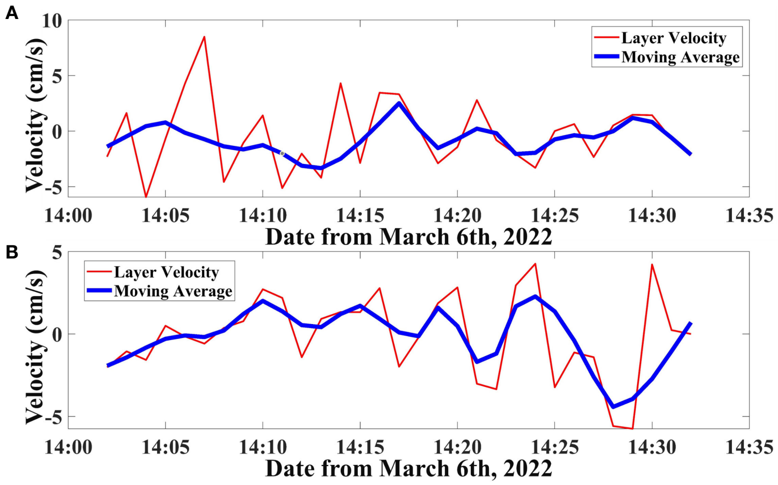


FIGURE 6
Inversion results of layer-averaged current. **(A)** is the layer-averaged current of S1 to S5. **(B)** is the layer-averaged current of S3 to S5. Red lines are the variations of inversion current results real-time at depth of 16*m*, and blues are the moving average of red lines.

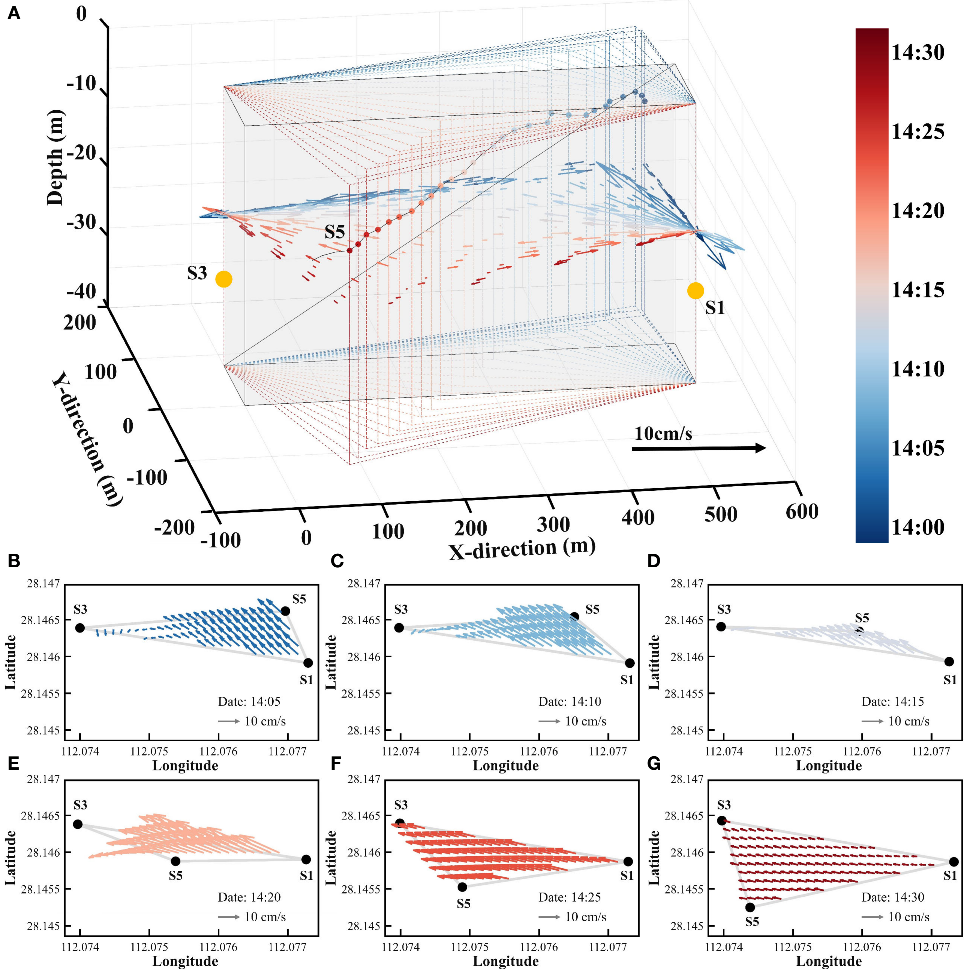


FIGURE 7
Inversion results of current field. **(A)** is the 3D combination figure. Different colour denotes different profile that same as the position at Figure 3. The arrows denote the velocity and direction of flow current at each grid. The grey area is critical observation area. **(B–G)** is the horizontal flow fields at six moments.

Huang et.al is the first developed moving acoustic tomography and applied in ocean flow field observation. The main differences between ours and theirs are the experiment scale and signal processing method. On the one hand, we used 50kHz transceiver for high precision moving tomography in a small-scale area and they used 18k transceiver in long-range (15km) observation. On the other hand, we resampled the signal to correct the signal travel time, while they corrected the arrival travel time *via* Doppler compensation. The inversion results have higher accuracy after high frequency transmission and signal resampled.

Moving acoustic tomography is more suitable for phenomena with slow fluctuations of current velocity and is not applicable to phenomena with temperature or quick variations of current. The method can be implemented similarly to a large long baseline array observation, where an accurate 3D environmental field is constructed by moving stations to observe the entire environment region.

To conclude, combining experiments and data analysis, the following empirical applications are given: moving acoustic tomography is more suitable for shallow areas with slow terrain fluctuations; using higher frequency signals (>3k) to obtain effective multipath sound structure; controlling the distance between moving and moored stations to reduce the signal travel time drift by the Doppler effect.

6 Conclusion

Underwater environment observation *via* advanced acoustic tomography has been considerably developed in recent years. In this paper, a method for reconstructing 3D layer-averaged flow current fields using high-frequency acoustic signal *via* short-term moving tomography is developed. The study introduces the basic idea of resampling acoustic signal and the calculation method for obtaining flow current. 3D flow current field result is composed by grid-averaged inversion current of vertical profile *via* moving station at different moment. Thus, a moving CAT observation experiment was carried out at range of 500m×500m in a reservoir to verify the feasibility of underwater moving acoustic tomography. The received results of reciprocal signal transmission between two moored stations and one moving station were used for layer-averaged current variations at vertical scale and grid-averaged current. The feasibility of the method is proved and its applicability is discussed.

The main conclusions of this research are as follows:

1. Underwater moving acoustic tomography can be used to complement areas that traditional fixed acoustic stations cannot observe and fuse the observation data. Experiments have initially proved the feasibility of this idea and the reliability of the inversion method process.
2. The acoustic signal needs to be re-sampled and correlated as the moving stations brings Doppler effect on the signal transmission. Particularly in shallow areas, the effect of real time variations in terrain causes signals difficult to be identified and matched. It is a problem that needs to be further analysed.
3. Moving acoustic tomography can be used in 3D observations. Rather than range averaging acoustic information over the path of sound rays passing in 2D planes, moving tomography is able to fuse 2D data to perceive and acquire 3D spatial observation, which has high application value.

The application of underwater moving acoustic tomography and signal processing *via* multiple stations in the ocean will be the focus of our future work. Data assimilation in 3D flow current field structures is also a research topic.

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

SX: Experimentation, methodology, formal analysis, data curation, investigation, validation, writing original draft, reviewing and editing. GL: Conceptualization, experimentation, formal analysis, data curation, validation, reviewing and Editing. RF: formal analysis, data curation, reviewing and editing. ZH and PX: supervision, validation, reviewing and editing. HH: Conceptualization, funding acquisition, project administration, supervision, reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 52071293).

Acknowledgments

We thank our colleagues in National University of Defense Technology for their experimental platform support and access to the experiment field sites. The raw data obtained from small-scale CAT system TD, CTD would be available upon request through Haocai Huang at Zhejiang University (hchuang@zju.edu.cn).

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

Al Sawaf, M. B., and Kawanisi, K. (2018). Novel high-frequency acoustic monitoring of streamflow-turbidity dynamics in a gravel-bed river during artificial dam flush. *Catena* 172, 738–752. doi: 10.1016/j.catena.2018.09.033

Al Sawaf, M. B., Kawanisi, K., Kagami, J., Bahreinimotlagh, M., and Danial, M. M. (2017). Scaling characteristics of mountainous river flow fluctuations determined using a shallow-water acoustic tomography system. *Physica A: Stat. Mech its Appl* 484, 11–20. doi: 10.1016/j.physa.2017.04.168

Bahreinimotlagh, M., Kawanisi, K., Al Sawaf, M. B., Roozbahani, R., Eftekhari, M., and Khoshuie, A. K. (2019). Continuous streamflow monitoring in shared watersheds using advanced underwater acoustic tomography system: a case study on zayanderud river. *Environ. Monit. Assess* 191 (11), 1–9. doi: 10.1007/s10661-019-7830-4

Bahreinimotlagh, M., Kawanisi, K., Danial, M. M., Al Sawaf, M. B., and Kagami, J. (2016). Application of shallow-water acoustic tomography to measure flow direction and river discharge. *Flow Meas Instrum* 51, 30–39. doi: 10.1016/j.flowmeasinst.2016.08.010

Chen, M., Hanifa, A. D., Taniguchi, N., Mutsuda, H., Zhu, X., Zhu, Z., et al. (2022). Coastal acoustic tomography of the neko-seto channel with a focus on the generation of nonlinear tidal currents–revisiting the first experiment. *Remote Sens* 14 (7), 1699. doi: 10.3390/rs14071699

Chen, K., Huang, C.-F., Huang, S.-W., Liu, J.-Y., and Guo, J. (2020). Mapping coastal circulations using moving vehicle acoustic tomography. *J. Acoust Soc. America* 148 (4), EL353–EL358. doi: 10.1121/10.0002031

Chen, M., Kaneko, A., Lin, J., and Zhang, C. (2017). Mapping of a typhoon-driven coastal upwelling by assimilating coastal acoustic tomography data. *J. Geophys Res: Ocean* 122 (10), 7822–7837. doi: 10.1002/2017jc012812

Chen, M., Kaneko, A., Zhang, C., and Lin, J. (2016). 3D assimilation of Hiroshima bay acoustic tomography data into a Princeton ocean circulation model. *J. Acoust Soc. America* 140 (4), 3183–3183. doi: 10.1121/1.4970008

Chen, M., Syamsudin, F., Kaneko, A., Gohda, N., Howe, B. M., Mutsuda, H., et al. (2020). Real-time offshore coastal acoustic tomography enabled with mirror-transpond functionality. *IEEE J. Ocean Eng* 45 (2), 645–655. doi: 10.1109/joe.2018.2878260

Chen, C., Yang, K., and Ma, Y. (2019). Sensitivity of sound speed fluctuation on acoustic arrival delay of middle range in deep water. *Appl. Acoust* 149, 68–73. doi: 10.1016/j.apacoust.2019.01.020

Chen, Z., Zheng, H., Tang, Y., and Chen, C. (2021). Measurement of Yangtze river flow based on coastal acoustic tomography. *J. Phys: Conf. Ser* 1739 (1), 012032. doi: 10.1088/1742-6596/1739/1/012032

Chen, M., Zhu, Z., Zhang, C., Zhu, X., Liu, Z., and Kaneko, A. (2021). Observation of internal tides in the qiongzhou strait by coastal acoustic tomography. *J. Ocean Univ. China* 20 (5), 1037–1045. doi: 10.1007/s11802-021-4590-x

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 (2), 558. doi: 10.3390/s20020558

Huang, C.-F. (2019). The development of acoustic mapping of ocean currents in coastal seas. *J. Acoust Soc. America* 146 (4), 2808–2808. doi: 10.1121/1.5136728

Huang, C. F., Li, Y. W., and Taniguchi, N. (2019). Mapping of ocean currents in shallow water using moving ship acoustic tomography. *J. Acoust Soc. America* 145 (2), 858–868. doi: 10.1121/1.5090496

Huang, H., Xu, S., Xie, X., Guo, Y., Meng, L., and Li, G. (2021). Continuous sensing of water temperature in a reservoir with grid inversion method based on acoustic tomography system. *Remote Sens* 13 (13), 2633. doi: 10.3390/rs13132633

Kaneko, A., Zhu, X. H., and Lin, J. (2020). *Coastal acoustic tomography* (Amsterdam: Elsevier Science Bv), WOS:000616677100016.

Kawanisi, K., Bahrainimotlagh, M., Al Sawaf, M. B., and Razaz, M. (2016). High-frequency streamflow acquisition and bed level/flow angle estimates in a mountainous river using shallow-water acoustic tomography. *Hydrol Process* 30 (13), 2247–2254. doi: 10.1002/hyp.10796

Mohamad Basel Al, S., Kiyosi, K., Cong, X., Gillang Noor Nugrahaning, G., and Faruq, K. (2022). Monitoring inflow dynamics in a multipurpose dam based on travel-time principle. *Water Resour. Manage*, 1–22. doi: 10.1007/s11269-022-03161-w

Munk, W., and Wunsch, C. (1979). Ocean acoustic tomography - scheme for Large-scale monitoring. *Deep-Sea Res. Part a-Oceanogr Res. Pap* 26 (2), 123–161. doi: 10.1016/0198-0149(79)90073-6

Syamsudin, F., Adityawarman, Y., Sulistyowati, R., Sutedjo, B., Kaneko, A., and Goda, N. (2016). Applicability and feasibility studies of coastal acoustic tomography for long-term monitoring of the Indonesian throughflow transport variability. *J. Acoust Soc. America* 140 (4), 3076–3076. doi: 10.1121/1.4969586

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 (17-18), 10475–10483. doi: 10.1029/2019gl084595

Xiao, C., Kawanisi, K., Torigoe, R., and Al Sawaf, M. B. (2021). Mapping tidal current and salinity at a shallow tidal channel junction using the fluvial acoustic tomography system. *Estuar Coast. Shelf Sci* 258, 107440. doi: 10.1016/j.ecss.2021.107440

Xu, S., Li, G., Feng, R., Hu, Z., Xu, P., and Huang, H. (2022). Tomographic mapping of water temperature and current in a reservoir by trust-region method based on CAT. *IEEE Trans. Geosci. Remote Sens* 60, 1–14. doi: 10.1109/TGRS.2022.3220648

Xu, S., Xue, Z., Xie, X., Huang, H., and Li, G. (2021). Ayer-averaged water temperature sensing in a lake by acoustic tomography with a focus on the inversion stratification mechanism. *Sensors* 21 (22), 7448. doi: 10.3390/s21227448

Zhu, Z.-N., Zhu, X.-H., and Guo, X. (2017a). Coastal tomographic mapping of nonlinear tidal currents and residual currents. *Cont Shelf Res* 143, 219–227. doi: 10.1016/j.csr.2016.06.014

Zhu, Z. N., Zhu, X. H., Guo, X., Fan, X., and Zhang, C. (2017b). Assimilation of coastal acoustic tomography data using an unstructured triangular grid ocean model for water with complex coastlines and islands. *J. Geophys Res: Ocean* 122 (9), 7013–7030. doi: 10.1002/2017jc012715

Zhu, Z.-N., Zhu, X.-H., Zhang, C., Chen, M., Zheng, H., Zhang, Z., et al. (2021). Monitoring of Yangtze river discharge at datong hydrometric station using acoustic tomography technology. *Front. Earth Sci* 9, 723123. doi: 10.3389/feart.2021.723123

Frontiers in Marine Science

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fmars.2023.1058542

Check for updates

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Hiroyuki Matsumoto,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Liwen Liu,
Northwestern Polytechnical University, China

*CORRESPONDENCE
Hanhao Zhu
zhuhanhao@zjou.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 September 2022
ACCEPTED 09 January 2023
PUBLISHED 27 January 2023

CITATION
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.* 10:1058542. doi: 10.3389/fmars.2023.1058542

Bayesian geoacoustic parameters inversion for multi-layer seabed in shallow sea using underwater acoustic field

Yangyang Xue[1†], Hanhao Zhu[2,3*†], Xiaohan Wang[1], Guangxue Zheng[1], Xu Liu[2] and Jiahui Wang[3]

[1]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, [2]Institute of Marine Science and Technology, Zhejiang Ocean University, Zhoushan, China, [3]School of Naval Architecture and Maritime, Zhejiang Ocean University, Donghai Laboratory, Zhoushan, China

Seabed geoacoustic parameters play an important role in underwater acoustic channel modeling. Traditional methods to determine these parameters, for example, drilling, are expensive and are being replaced by acoustic inverse technology. An inversion method based on Bayesian theory is presented to derive the structure and geoacoustic parameters of a layered seabed in a shallow sea. The seabed was considered a layered elastic medium. The objective of this research was to use the sound pressure detected by underwater acoustic sensors at different positions and to use nonlinear Bayesian inversion to estimate the geoacoustic parameters and their uncertainties in the multi-layer seabed. Specifically, the thickness, density, compression wave speed, shear wave speed, and the attenuation of these two wave speeds were determined. The maximum a posterior (MAP) model and posterior probability distribution of each parameter were estimated using the optimized simulated annealing (OSA) and Metropolis-Hastings sampling (MHS) methods. Model selection was carried out using the Bayesian information criterion (BIC) to determine the optimal model that thoroughly explained the experimental data for different parameterizations. The results showed that the OSA is much more capable of delivering high-accuracy results in multi-layer seabed models. The compression wave speed and shear wave speed were less uncertain than the other parameters, and the parameters in the upper layer had less uncertainty than those in the lower layer.

KEYWORDS

shallow sea, underwater acoustic channel, geoacoustic parameter inversion, nonlinear Bayesian theory, model selection, uncertainty analysis

1 Introduction

Acoustic waves are currently the only means of realizing underwater long-distance detection and communication. When acoustic waves propagating in shallow sea, they will be significantly affected by environment parameters. Moreover, the influence of seabed geoacoustic parameters on acoustic propagation is apparent. Therefore, the accurate inversion of shallow sea geoacoustic parameters is vital. (Benavente et al., 2019). Methods used to determine the seabed's geoacoustic parameters can be divided into those that rely on direct and indirect measurements (Yin and Hu, 2016). Direct methods mainly involve the acquisition of samples of the seabed substrate by drilling into the seabed. In contrast, acquiring of geoacoustic parameters relies more on the indirect measurement of acoustic inversion techniques. This technology has been widely used to obtain geoacoustic parameters because of its technical efficiency.

Past inversion studies for the acoustic properties of shallow seabed mainly focused on the geoacoustic parameters in the surface seabed, and assumed the seabed to be a liquid medium (Michalopoulou and Gerstoft, 2019). With the development of technology, simply treating the seabed as a liquid medium cannot meet the research needs, also it is urgent to obtain the stratified structure and parameters of the seabed simultaneously and accurately. Since Low-Frequency (LF)/Very-Low-Frequency (VLF) acoustic waves can penetration into the seabed deeply, they are more suitable to be used to obtain the structure and parameters (Yuan et al., 2019). However, there are few related studies at present. Therefore, this paper proposes an algorithm suitable for obtaining the geoacoustic parameters of the elastic seabed with multiple layers. (Zhang et al., 2020)

Geoacoustic parameter inversion represents a strongly nonlinear problem with multiparameter solutions for which an immediate answer is unavailable. Initially, optimization algorithms such as simulated annealing (SA) and genetic algorithm (GA) approaches have been widely used in the field of inversion to solve the error function (Sen and Stoffa, 2006; Zhang et al., 2021). However, the disadvantage of SA and GA algorithms is that they can only provide a set of MAP (Maximum A Posteriori, MAP) values, which is unrealistic for the multi-parameter coupling problem. In fact, because of the high randomness of the sea environment, the measured data inevitably contain systematic and random errors; thus, an uncertainty analysis of the inversion results is essential. It is unreasonable and insufficiently comprehensive to provide only a set of MAP values for optimal solution parameters to obtain the inversion result. The inversion method based on nonlinear Bayesian theory is a global optimization algorithm with roots in interdisciplinary subjects, such as probability theory, applied mathematics, and optimization theory. This algorithm can not only be used to effectively estimate the parameters of the MAP model, but it can also evaluate the uncertainty in the parameter inversion results from a statistical perspective, making it more suitable for meeting the current needs of the research on geoacoustic parameter inversion.

In general, the seabed of the shallow sea environment along the coast has a layered structure with multiple sedimentary layers, and the hardness of this seabed structure increases with the depth of the seabed. The research objective of this study was to determine the sound pressure detected by underwater acoustic sensors at different positions to extract the geoacoustic parameters of this environment, thus contributing to the modeling of underwater acoustic channels in the shallow sea. In addition, a nonlinear Bayesian inversion method was applied to estimate the seabed density, compression wave speed, shear wave speed, two types of wave speed attenuation, and their corresponding uncertainties. To improve the accuracy of the optimization results, optimized simulated annealing (OSA) is proposed to solve the error function, and Metropolis-Hastings sampling (MHS) was applied to estimate the MAP model and compute the marginal posterior probability distributions. The Bayesian information criterion (BIC) was used to determine the optimal parameterization model in terms of the likelihood function.

The paper is organized as follows: the inverse theory and algorithms are described in Section 2. Section 2.1 introduces the shallow sea model with an n-layered seabed and nonlinear Bayesian inversion theory. In addition, the BIC is described in this section. Section 2.2 presents a description of the OSA method. Section 3 provides the motivation for selecting a shallow sea model with a two-layer seabed as the discussion model and presents the analysis, which was conducted by way of simulation, of the feasibility of the result obtained using BIC, and the best model selected by BIC is analyzed by OSA. The inversion results for the experimental data obtained using the proposed method are provided in Section 4. Specifically, in Section 4.1, the laboratory experiments are described; Section 4.2 provides the results of the model selection analysis; Section 4.3 presents the inversion results and introduces the uncertainty analysis of the inversion results. Finally, the conclusions are presented in Section 5.

2 Inverse theory and algorithms

2.1 Nonlinear Bayesian inversion theory

This section describes the nonlinear Bayesian inversion theory as applicable to the pressure field generated by underwater sound. More general descriptions of Bayesian theory can be found elsewhere (Dosso et al., 2009; Dosso and Dettmer, 2011). In this study, the shallow sea environment was considered a layered waveguide model, and the seabed was considered a layered elastic medium, as illustrated in Figure 1. In the model, the r-axis represents the propagation direction of the acoustic signals, and the z-axis represents the depth of the sea. Furthermore, z_s and f_0 represent the depth and frequency of the sound source, respectively; ρ_1 and c_1 represent the density and speed of sound in the sea column; c_{pn}, c_{sn}, ρ_{bn}, α_{pn}, α_{sn}, and H_n are the parameters of the seabed model; c_{pn}, c_{sn}, and ρ_{bn} are the compression wave speed, shear wave speed, and density of the n-layered seabed, respectively; α_{pn} and α_{sn} represent the two types of wave speed attenuation, and H_n is the depth of the n-layered seabed (Li et al., 2019).

In the model shown in Figure 1, under the wave theory, the wave equation can be transformed into a form that conforms to the Fourier transform through discrete processing. This enables the model to be solved quickly by the fast field method (FFM) (Zhu et al., 2012), and to derive the potential functions of each layer included in the wave equation. The relationship between the sound pressure p in the seabed layer and the potential function φ, represented by $p=\rho_1\omega^2\varphi$, is then used to obtain the value of the sound pressure at each point in the seabed layer. Where ω represents the angular frequency corresponding to the source frequency f, that is $\omega=2\pi f$. The sound pressure field can be expressed as follows, where ξ represents the wave number.

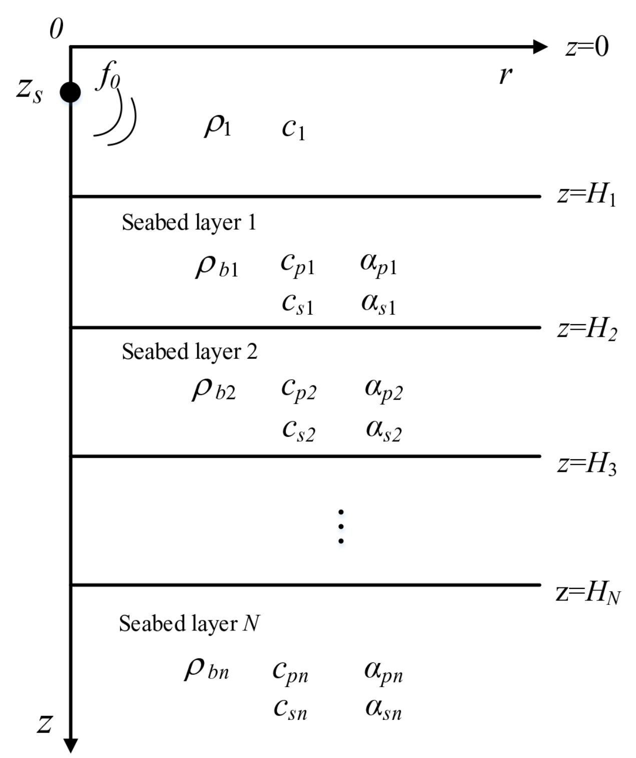


FIGURE 1
Shallow sea waveguide model with n-layered seabed.

$$p(r,z,\omega) = \rho_1\omega^2\int_0^\infty Z_1(z,\xi,\omega)J_0(\xi r)\xi d\xi \tag{1}$$

The random variables **d** and **m** represent the experimental data from the sound field and the awaiting inversion geoacoustic parameters in the seabed model (Zhu et al., 2019). The vectors **d** and **m** satisfy the Bayesian theorem.

$$\boldsymbol{P}(\mathbf{m}|\mathbf{d}) = \boldsymbol{P}(\mathbf{d}|\mathbf{m})\boldsymbol{P}(\mathbf{m})/\boldsymbol{P}(\mathbf{d}) \tag{2}$$

where $P(\mathbf{m}|\mathbf{d})$ is the posterior probability density (PPD), $P(\mathbf{d})$ is the probability density function (PDF) of **d**, $P(\mathbf{m})$ is the prior PDF, and represents the available parameter information independent of the data. $P(\mathbf{d}|\mathbf{m})$ is the conditional PDF of **m** under the given measured data **d**, which is expressed by the likelihood function $L(\mathbf{d}|\mathbf{m})$ for the measured data. Thus, Equation (2) can be written as

$$P(\mathbf{m}|\mathbf{d}) \propto L(\mathbf{m})P(\mathbf{m}) \tag{3}$$

The likelihood function is determined by the form of the measured data and the statistical distribution of the data errors. In practice, obtaining an independent estimate of the error statistics is often tricky; thus, the assumption of unbiased Gaussian errors is used in processing, and the likelihood function is given by

$$L(\mathbf{m}) = P(\mathbf{d}|\mathbf{m}) \propto \exp[-E(\mathbf{m})] \tag{4}$$

where $E(\mathbf{m})$ represents the error function. After normalization, we obtain

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-E(\mathbf{m})]P(\mathbf{m})}{\int \exp\left[-E\left(\mathbf{m}'\right)\right]P(\mathbf{m}')d\mathbf{m}'} \tag{5}$$

The domain of integration spans the M-dimensional parameter space.

In the Bayesian inversion method, the multidimensional PPD of **m** represents the most general solution to the inversion problem. However, interpreting the PPD for a multi-dimensional problem requires the properties to be estimated by defining the parameter values, uncertainties, and inter-relationships, such as the MAP model, mean model, and marginal probability distributions, which are defined, respectively, as

$$\hat{\mathbf{m}} = Arg_{\max}\{P(\mathbf{m}|\mathbf{d})\} \tag{6}$$

$$\bar{\mathbf{m}} = \int \mathbf{m}'P(\mathbf{m}')d\mathbf{m}' \tag{7}$$

$$P(\mathbf{m_i}|\mathbf{d}) = \int \delta\left(\mathbf{m_i} = \mathbf{m_i'}\right)P(\mathbf{m}'|\mathbf{d})d\mathbf{m}' \tag{8}$$

where δ is the Dirac delta function, and the range of the integral is the space of each parameter.

Inter-parameter relationships are quantified by normalizing the covariance matrix C_m to produce a correlation matrix with elements between +1 and −1.

$$R_{ij} = \frac{C_{m_{ij}}}{\sqrt{C_{m_{ii}}C_{m_{jj}}}} \tag{9}$$

where the parameters are completely positively correlated for R_{ij}= +1 and negatively correlated for $R_{ij} = -1$. Furthermore, $R_{ij} = 0$ indicates that the parameters were not relevant.

MAP estimates in this study were computed using the OSA. Marginal distributions were determined by numerical integration of the PPD, which was obtained using MHS. Although the MHS results do not depend on the initial model, the accuracy of the inversion results can be improved by selecting the correct initial value. In this study, the MAP estimate searched by the OSA was used as the initial model for the MHS (Zheng et al., 2020).

In Bayesian inversion theory, solving the PPD requires the likelihood function to be obtained. In addition, because the uncertainties in the data are often not well known, physically reasonable assumptions are required for the form of the uncertainty distribution; thus, assuming the data errors are independent Gaussian-distributed random variables, the likelihood function can be given by

$$L(\mathbf{m}) = \prod_{f=1}^{F}\frac{1}{\pi^k\left|C_\mathrm{m}^f\right|}\exp\left\{-\left[\mathrm{P}_{mea}^f - \mathrm{P}_{pre}^f(\mathbf{m})\right]^T\left(C_\mathrm{m}^f\right)^{-1}\left[\mathrm{P}_{mea}^f - \mathrm{P}_{pre}^f(m)\right]\right\} \tag{10}$$

where $\mathbf{p}_\mathrm{mea}^\mathbf{f}(\mathbf{m})$ represents the measured sound pressure data at K receiving positions for the f^th frequency, and $\mathbf{p}_\mathrm{pre}^\mathbf{f}(\mathbf{m})$ and $\mathbf{C}_m^f$, respectively, represent the sound pressure data predicted by the model and the covariance matrix under the same conditions (Dong and Dosso, 2011; Gao et al., 2017).

The sound pressure predicted by the model, $\mathrm{p}^{\mathrm{f}}_{\mathrm{pre}}(\mathbf{m})$, can be expressed as

$$\mathbf{p}^{\mathbf{f}}_{\mathrm{pre}}(\mathbf{m}) = A^{f} \mathrm{e}^{\mathrm{i}\theta^{f}} \mathbf{p}^{\mathbf{f}}_{\mathrm{FFM}}(\mathbf{m}) \tag{11}$$

where $\mathbf{p}^{\mathbf{f}}_{\mathrm{FFM}}(\mathbf{m})$ is the sound pressure computed by the FFM, and A^{f} and θ^{f} are the magnitude and phase of the unknown complex source at each frequency, respectively. Then, the likelihood can be maximized with respect to the source by setting $\partial L(\mathbf{m})/\partial \mathrm{A}^{\mathrm{f}}=\partial L(\mathbf{m})/\partial \theta^{\mathrm{f}}=0$, leading to

$$A^{f} \mathrm{e}^{\mathrm{i}\theta^{f}} = \frac{\left[\mathbf{p}^{f}_{\mathrm{FFM}}(\mathbf{m})\right]^{*} \mathbf{p}^{f}_{\mathrm{mea}}}{\left|\mathbf{p}^{f}_{\mathrm{FFM}}(\mathbf{m})\right|^{2}} \tag{12}$$

where "*" denotes the conjugate transpose.

For each frequency, the standard approximation of the diagonal covariance $\mathrm{C}^{f}_{m}= \mathrm{v}^{\mathrm{f}}\mathrm{I}$, where v^{f} is the unknown variance at the f^{th} frequency, and $\mathbf{I}$ is the identity matrix, in which case the likelihood function becomes

$$L(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi v^{f})^{K}} \exp\left[-\frac{B^{f}(\mathbf{m})\left|\mathbf{p}^{f}_{\mathrm{mea}}\right|^{2}}{v^{f}}\right] \tag{13}$$

where $B^{f}(\mathbf{m})$ represents the normalized Bartlett disqualification, which can be expressed as

$$B^{f}(\mathbf{m}) = 1 - \frac{\left|\left[\mathbf{p}^{f}_{\mathrm{FFM}}(\mathbf{m})\right]^{*} \mathbf{p}^{f}_{\mathrm{mea}}\right|^{2}}{\left|\mathbf{p}^{f}_{\mathrm{mea}}\right|^{2}\left|\mathbf{p}^{f}_{\mathrm{FFM}}(\mathbf{m})\right|^{2}} \tag{14}$$

To obtain a maximum likelihood estimate of the data variance by setting $\partial L(\mathbf{m})/\partial \mathrm{v}^{\mathrm{f}} = 0$, the maximum likelihood solution of the variance estimate can be expressed as

$$\hat{v}_{f} = \frac{B^{f}(\mathbf{m})\left|\left|\mathbf{p}^{f}_{\mathrm{mea}}\right|^{2}\right|}{K} \tag{15}$$

Substituting Equation (15) into Equation (13) and Equation (4), the error function $E(\mathbf{m})$ becomes

$$E(\mathbf{m}) = K\sum_{f=1}^{F} \ln\left[B^{f}(\mathbf{m})\left|\mathbf{p}^{f}_{\mathrm{mea}}\right|^{2}\right] \tag{16}$$

Determining an appropriate model parameterization is another essential aspect of geoacoustic parameter inversion. Model selection involves selecting a set of parameterized models that are most consistent with the measured data based on a certain target criterion. Models for which either too many or too few parameters are specified (over-parameterization and under-parameterization, respectively) are known to have a certain impact on the inversion results (Dettmer et al., 2009; Li et al., 2012). In this study, BIC is used to select the parameterized model that is most consistent with the measured data. BIC is an asymptotic approximation of the Bayesian theorem P(**d**|**I**) of model **I**, and **I** denotes the presupposition of different inversion models. Assuming the measurement data **d**, the likelihood function of model **I** is expressed as follows

$$-2\mathrm{In}P(\mathbf{d}|\mathrm{I}) \approx \mathrm{BIC} = -2\mathrm{In}L(\hat{\mathbf{m}}) + M\mathrm{In}N \tag{17}$$

The likelihood function is replaced with the error function to obtain, where M represents the number of model parameters, N represents the number of data

$$\mathrm{BIC} = 2E(\hat{\mathbf{m}}) + M\mathrm{In}N \tag{18}$$

The parameterization with the smallest BIC value was selected as the most appropriate model. According to Equation (18), the BIC value is determined by the error function, number of parameters, and number of measured data points. Therefore, minimizing the BIC provides a balance between the error value and number of parameters, identifying the simplest parameterization consistent with the resolving power of the data.

2.2 Optimised simulated annealing

Searching for the MAP values in the prior interval based on the error function is an important step in the Bayesian inversion method. Because the error function is nonlinear, multi-dimensional, and may have multiple peaks, the search process is highly challenging, and the correlation between the parameters complicates the search. Currently, most researchers apply global optimization algorithms, such as SA and GA, to search for the MAP (Dosso et al., 2001). In geoacoustic parameter inversion for a multi-layer seabed, the number of awaiting inversion parameters increases as the number of layers in the seabed increases, and the choice of the parameter prior interval also becomes more difficult. An interval range that is too large increases the difficulty, at the same time, reduces the accuracy of the results, whereas a range that is too small may not include the true value of the geoacoustic parameters in the interval. The use of traditional global optimization algorithms, owing to the lack of effective limiting measures, usually involves setting large prior intervals for the inversion, resulting in low inversion efficiency and poor results. The effective reduction of the range of the previous interval in the optimization process would enable the accuracy and efficiency of the inversion result to be effectively improved (Ohta et al., 2008).

Previous research showed that the solution to multi-layer sea environment problems led to the emergence of an objective law according to which the value of the acoustic impedance in the seabed generally increases with the depth of the seabed (Yang, 2009; Li, 2012). This law provides the possibility of narrowing the prior interval in the inversion. Therefore, the OSA we proposed is based on the SA algorithm, which is an improved SA that can be used to correct the prior interval in the process of optimizing and solving the parameters according to the acoustic impedance law of the upper and lower layers of the seabed (Xue et al., 2021). The algorithm first optimizes and solves the parameters of the lower layer, and adds constraints on the basis of the existing results to realize the accurate solution of the parameters of the upper layer. The inversion flow chart that includes the additional correction process is shown in Figure 2.

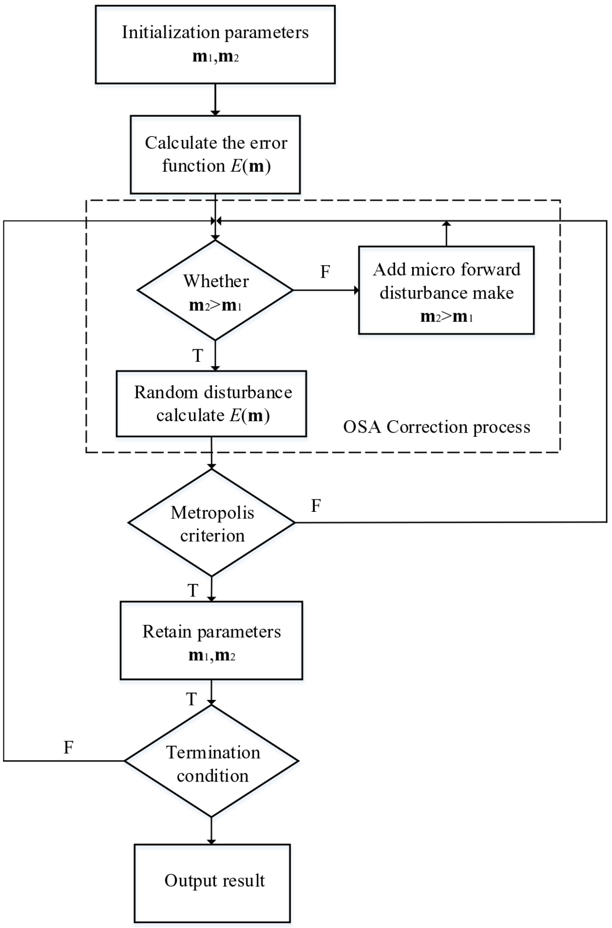


FIGURE 2
Inversion flow chart reflecting the additional correction process.

3 Simulation analysis

Selecting an appropriate forward model is crucial for improving the accuracy of the inversion (Seongryong et al., 2016; Enming et al., 2018). For shallow sea environments, the seabed is regarded as a uniform and isotropic multi-layer medium, and the density ρ_b, compression wave speed c_p, shear wave speed c_s, and the two types of wave speed attenuation α_p and α_s are taken as the inversion objects in this study. To verify the reliability of the above-mentioned theoretical research pertaining to inversion, a simulation example was used for verification. In this example, the seabed is modelled as a two-layer elastic medium. The true values are listed in Table 1, where the source frequency is 155 Hz, the depth of the sound source z_s is 20 m, the receiving depth z_r is 10 m, and the depth of the sea H_1 is 100 m. The transmission loss (TL) is calculated by FFM under the true value condition (Zhu et al., 2012), as shown in Figure 3. The sound waves propagating in the sound field can be regarded as the superposition of normal waves of different orders. The superposition of normal waves is related to the phase and will show interference structure. Thus, the TL at different locations is up and down, and the TL increases significantly at the 500m location due to the reverse superposition of a large number of normal positive waves.

3.1 Model selection

The main discussion in this section is the feasibility of BIC. The sound pressure data obtained from the 2-layer model simulation is used as the validation object. Different inversion models are used to invert this set of sound pressures, When the BIC value is smallest, if the inversion results of the corresponding model best match. We considered three different versions of the inversion model: a parameterization model with one to three layers of the isotropic uniformly distributed seabed. Based on the sound pressure values obtained under the simulation conditions listed in Table 1, we used the OSA to get the BIC values corresponding to the inversion results for the different versions of the model for model selection.

The results of the study designed to select the model were obtained from the simulation analysis and are shown in Figure 4. Here, m_1, m_2, and m_3 represent the inversion model of the one-, two-, and three-layer uniformly distributed seabed, respectively. Figure 4A shows that the simulation inversion of different models lowers the error value as the number of layers in the model increases, with the error value of the one-layered model being the largest. Figure 4B shows that the number of parameters that need to be also considered increases, which increases the uncertainty of the inversion results. BIC fully considers the balance among the error values of the model, the number of parameters, and the number of data points and selects the parameterized model most consistent with the measured data. Figure 4C shows that the model based on the initial settings determined by BIC with the best fit is the two-layer model of the uniformly distributed seabed. These results show that the BIC can be employed to select the layered model of the seabed that best fits the parameters that were initially specified.

The TL values obtained by the different inversion models are compared in Figure 5. Figures 5A–C Shows the TL comparison between the inversion model using different layered structures and the two-layer structure simulation data inversion results. The results qualitatively indicate that the inversion model of the two-layer structure is the most consistent with the initially specified model and demonstrates that BIC can select the model with the best parameterization.

3.2 Optimisation error value by OSA

Many optimization algorithms, such as SA and GA, can optimize the error function (Bevans and Buckingham, 2016; Yang et al., 2017;

TABLE 1 Parameters used for the simulation of the sea environment.

Parameters		c_p/m·s^{-1}	c_S/m·s^{-1}	ρ_b/g·cm^{-3}	α_p/dB·λ^{-1}	α_s/dB·λ^{-1}	H_1/m
True value	Sedimentary	2000	1000	1.5	0.1	0.1	100
	Semi-infinite	2500	1200	1.7	0.1	0.1	/

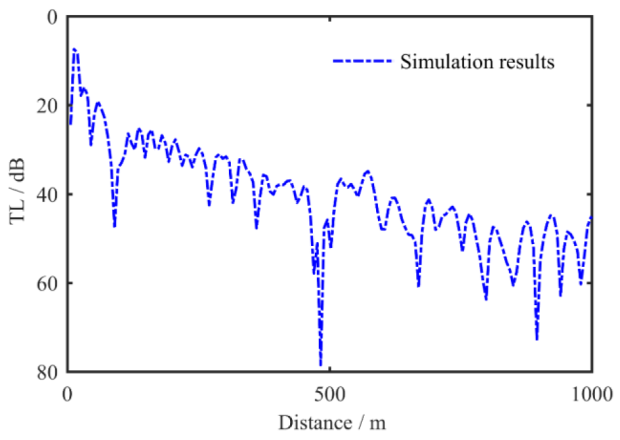


FIGURE 3
TL obtained by using the true value set in the simulation.

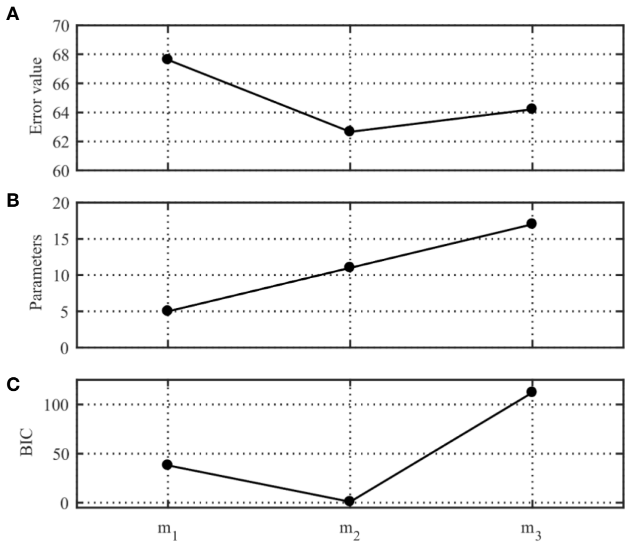


FIGURE 4
Results of the model selection study in terms of the **(A)** error value, **(B)** parameters, and **(C)** BIC, where m_1, m_2, and m_3 represent the inversion model of the one-, two-, and three-layer uniformly distributed seabed, respectively.

Fu et al., 2018). However, application of the traditional optimization algorithm to the error function established in this study for the multi-layer uniformly layered seabed model often produces optimization results that do not conform to the actual physical laws. In addition, considering seabed stratification, the prior interval inevitably needs to be expanded. This means that the accuracy of the results obtained when the traditional optimization algorithm is optimized for the above model is greatly reduced. Because the conventional global optimization algorithms SA and GA have certain difficulties with the application of multi-layer seabed inversion, we proposed an improved algorithm (Julien et al., 2018; Li et al., 2018).

Table 2 compares the mean values and standard deviation (std) before and after the application of OSA. The results in the table indicate that the mean value of the optimization result after the application of OSA is closer to the true value obtained in the simulation, and the corresponding standard deviation is reduced; in particular, the parameters of the second seabed improved significantly. Figures 6, 7 show the PPD of the parameters of the first and second layers of the seabed, respectively. The red line represents the true value, and the blue line represents the distribution of the difference between the mean and standard deviation of the data results in the interval. These results show that the PPD distribution of the parameter inversion obtained after applying OSA is narrower. Figure 7 shows that the PPD distribution after application of the OSA is further concentrated near the true value; in particular, the improvement in the speed and density of the acoustic wave of the second seabed is more pronounced.

Figure 8 compares the sound pressure TL curve calculated by the average value of the inversion results of each parameter before and after the algorithm was improved with the TL curve calculated using the true value. This comparison reveals that the TL curve obtained by the improved optimization algorithm is in closer agreement with the TL curve obtained with the true value, which further demonstrates the feasibility of the improved algorithm. From the interpretation of the interference phenomenon, using the normal wave theory, it can be known that the rapid increase in TL here is caused by the interference and superposition of different order normal waves at this position, and the appearance of the interference structure makes the sound field more complicated.

4 Measurements and results

4.1 Introduction to the experiment

The propagation characteristics of high-frequency underwater acoustic signals in small-scale environments can be used to simulate the propagation characteristics of low-frequency signals in actual large-scale sea environments (Zhang et al., 2018; Zhang et al., 2022). Compared with the original sound field, the expression of the sound field calculated by the reduced-scale experiment is only increased in equal proportion, and the characteristics of the undulation and distribution of the sound pressure field remain the same. Detailed introduction to this topic can be found elsewhere (Zhu, 2014; Zheng et al., 2019).

The accuracy of the inversion method was verified by conducting a scaling experiment. The experiment was carried out in an anechoic tank in the laboratory using polyvinyl chloride (PVC) slab (size: 153 mm × 110 mm × 10.5 mm) to simulate the elastic semi-infinite seabed. A layer of fine sand was placed on the PVC slab to simulate a shallow sea waveguide environment with elastic sediment and an elastic semi-infinite seabed with a sand thickness of 250 mm. The experimental setup is shown in Figure 9, where the sound source sensor frequency is 155 kHz, the depth of the sound source sensors z_s is 200 mm, the receiving depth z_r is 200 mm, and the depth of the freshwater z_1 is 300 mm. The acoustic speed of the sound wave in the sea, calculated at room temperature (11.15°C) in the laboratory, was 1450.212 $m \cdot s^{-1}$. A

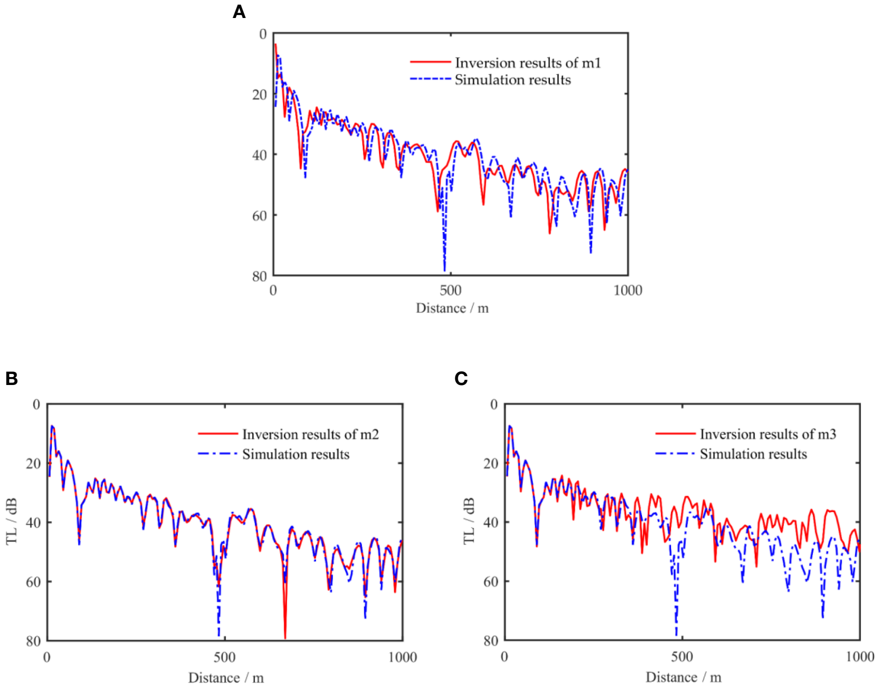


FIGURE 5
Comparison of transmission loss obtained by the different inversion models **(A)** m1, **(B)** m2, **(C)** m3. Inversion results of the simulation data of the two-layer structure using different layered structure inversion models. The blue dotted line indicates the transmission loss resulting from simulation with the true value set.

high-frequency underwater acoustic wave was transmitted by a source (a custom-built transmitting transducer), which was installed at a fixed position at one end of the equipment. The sound wave was received by a single acoustic sensor (a standard TC4038 underwater sensor), which was attached to the mobile micro-worktable to enable the position of the sensor to be varied at equal distances. This arrangement allowed the underwater sensor to move 2 mm every time, with an error of less than 20 μm. The movement of the mobile workbench was computer-controlled for data measurement and collection. Upon completion of the measurement in one position, the worktable automatically moved to the next position, and the measurement at each position was recorded ten times to obtain the average value for each position. A total of 500 locations were selected for measurement during the experiment (Ballard et al., 2009; Zheng, 2019).

TABLE 2 Comparison of the mean and standard deviation of inversion parameters before and after improvement of the simulated algorithm.

Layered	Parameters	True value	Prior range	Before applying OSA mean ± std	After applying OSA mean ± std
water	H_1/m	100	/	/	/
	c_1/m·s^{-1}	1500	/	/	/
	ρ_1/g·cm^{-3}	1.00	/	/	/
Sedimentary	c_{p2}/m·s^{-1}	2000	1800-2200	1998.71 ± 74.28	2001.78 ± 21.34
	c_{s2}/m·s^{-1}	1000	900-1100	995.77 ± 40.05	1003.15 ± 17.62
	ρ_{b2}/g·cm^{-3}	1.50	1.00-2.00	1.55 ± 0.18	1.52 ± 0.07
	α_{p2}/dB·λ^{-1}	0.10	0.10-0.11	0.09 ± 0.004	0.10 ± 0.001
	α_{s2}/dB·λ^{-1}	0.10	0.09-0.11	0.10 ± 0.004	0.10 ± 0.001
	H_2/m	20	15-25	20.11 ± 1.49	19.97 ± 0.74
Semi-infinite seabed	c_{p3}/m·s^{-1}	2500	1800-2800	2296.28 ± 198.42	2429.02 ± 59.65
	c_{s3}/m·s^{-1}	1200	900-1400	1130.89 ± 101.27	1205.55 ± 28.32
	ρ_{b3}/g·cm^{-3}	1.70	1.00-2.00	1.52 ± 0.19	1.67 ± 0.09
	α_{p3}/dB·λ^{-1}	0.10	0.09-0.11	0.10 ± 0.004	0.10 ± 0.001
	α_{s3}/dB·λ^{-1}	0.10	0.09-0.11	0.10 ± 0.004	0.10 ± 0.001

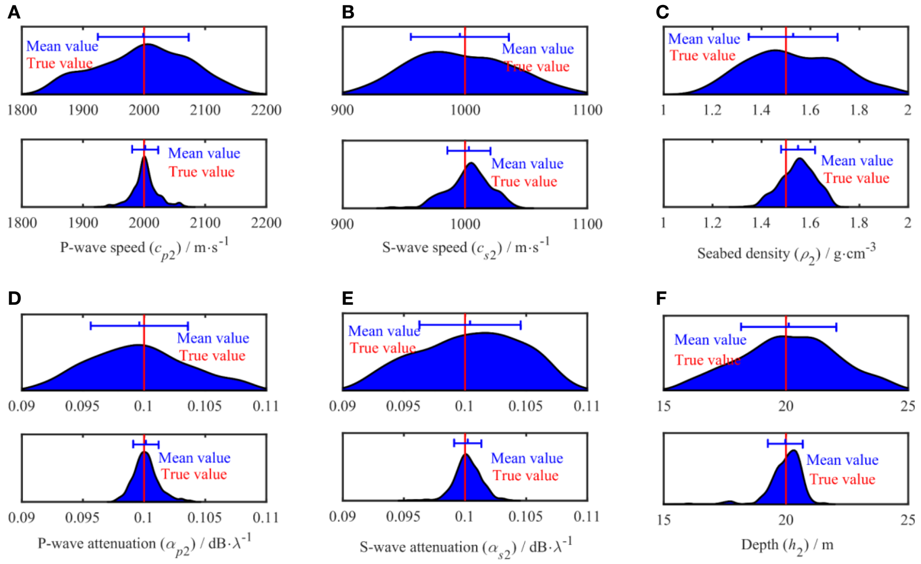


FIGURE 6
One-dimensional marginal probability distribution of sedimentary parameters.

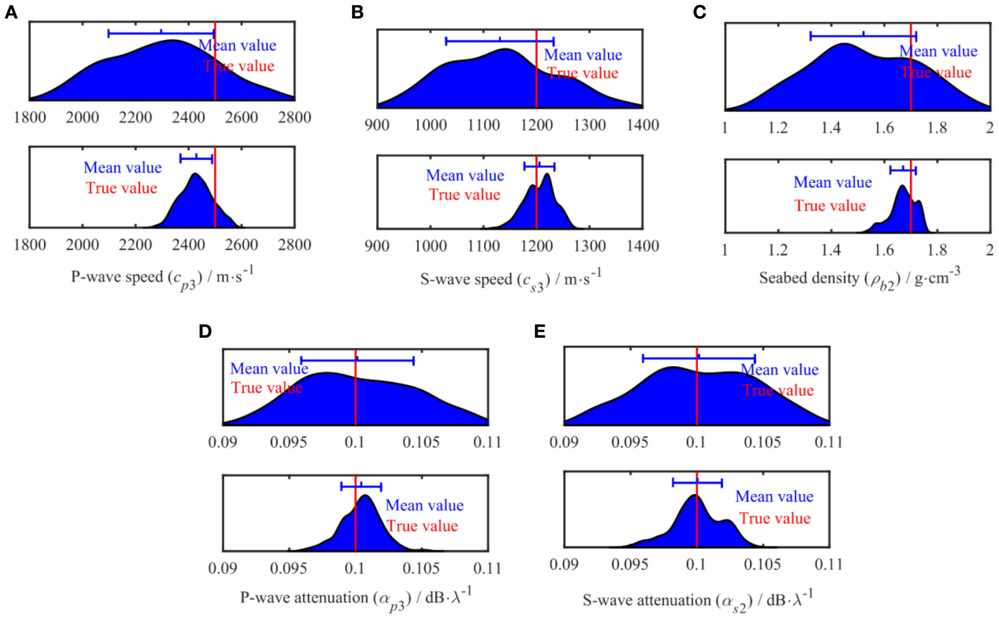


FIGURE 7
One-dimensional marginal probability distribution of semi-infinite parameters.

Figure 9C gives the arrival times of the 50-150 channel time domain waveforms measured by the above experimental equipment, from which the direct and reflected waves can be observed more clearly. Figure 9D shows the time domain waveform of the signal received by the hydrophone at the 50th position. Since the experiments were conducted in an anechoic environment and the noise effect is relatively small, the arrival time of the signal in the time domain waveform can be used to distinguish between direct and reflected waves. We choose the bottom reflected wave as the actual signal used in the inversion.

4.2 Model selection

The feasibility of BIC has been verified in the simulation, and this section uses the BIC criterion to select the best parameterized model for the experimental data. Three models with different hierarchical structures are still applied to invert the experimental data, and the BIC values can be calculated by combining the error values and the number of model parameters obtained during the inversion of the different models. The BIC values are solved by Equation (17), the model with the smallest BIC value is the best parameterized model.

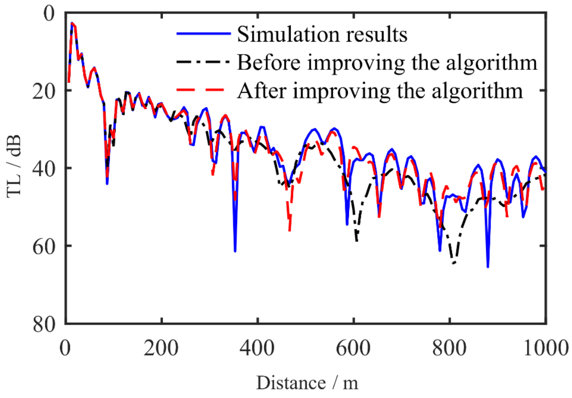


FIGURE 8
Comparison of transmission loss of inversion parameters before and after improvement of the algorithm.

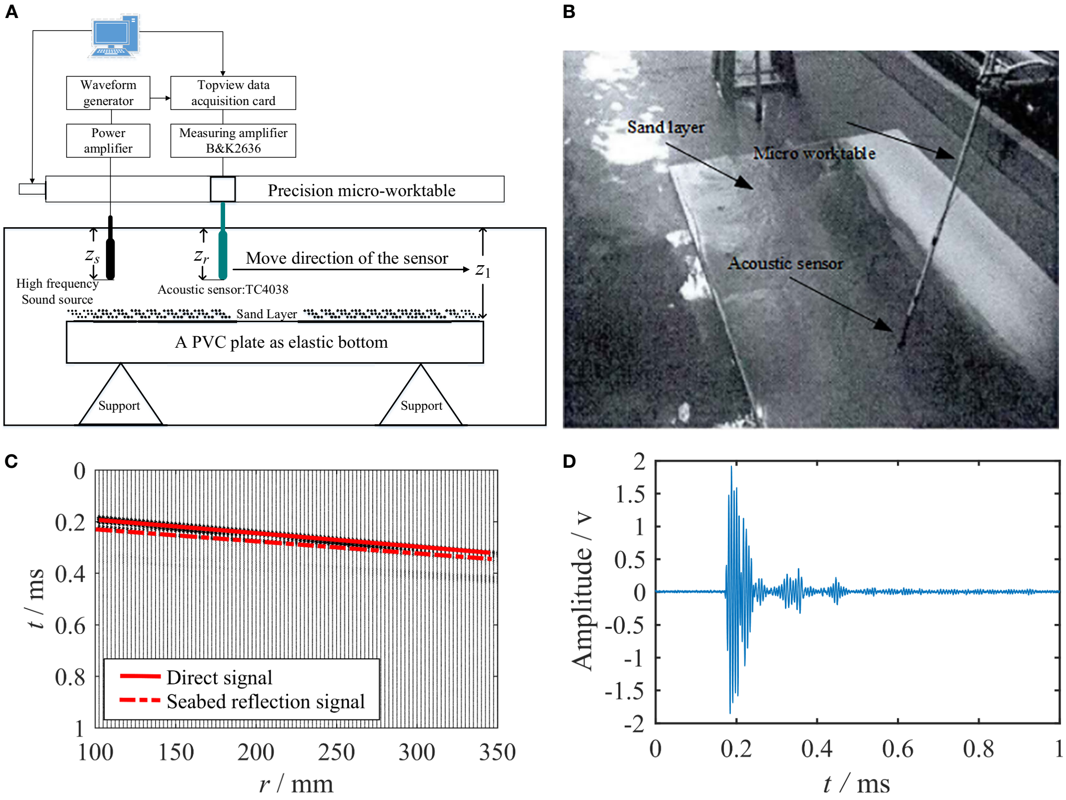


FIGURE 9
Experimental setup with the measuring instruments and equipment and time domain waveform. **(A)** Diagram of the layout of the experimental measurement system, **(B)** Photographic image of the movable micro-worktable for the measurement equipment, **(C)** Time domain waveform arrival time, **(D)** Time domain waveform at position 50.

The results of the model selection are shown in Figure 10. Figure 10A shows that the error decreases substantially when proceeding from one to two seabed layers and that this reduction is significantly less when the number of layers is increased from two to three. Figure 10B shows that the number of parameters to be considered also increases, which raises the uncertainty of the inversion results. According to Figure 10C, the optimal model determined by BIC is a two-layer uniformly distributed seabed model.

Figure 11 shows the profile structure of the compression wave speed, shear wave speed, and variation in density with depth. The acoustic impedance at the bottom is proportional to the deep; that is, in general marine environment, the acoustic impedance increases with depth. The results indicate that the trends observed for the three different inversion models are consistent with this physical law. Moreover, increasing the stratification of the seabed does not include the “ reversal “ phenomenon in the distribution of parameters; this demonstrates the significance of OSA in searching for parameters in the prior interval. The structure of the section, with 1000 mm as the boundary, reveals an obvious structural division of the density, compression wave speed, and shear wave speed within this value. As the depth increased, each parameter tended to a certain value, and the change in the compression wave was particularly obvious.

4.3 Inversion results and uncertainty analysis

The BIC optimization results demonstrate that the two-layer uniformly layered seabed model is the most consistent with the experimental results. The inversion results of this model are therefore analyzed and discussed in this section. Figure 12 compares the TL

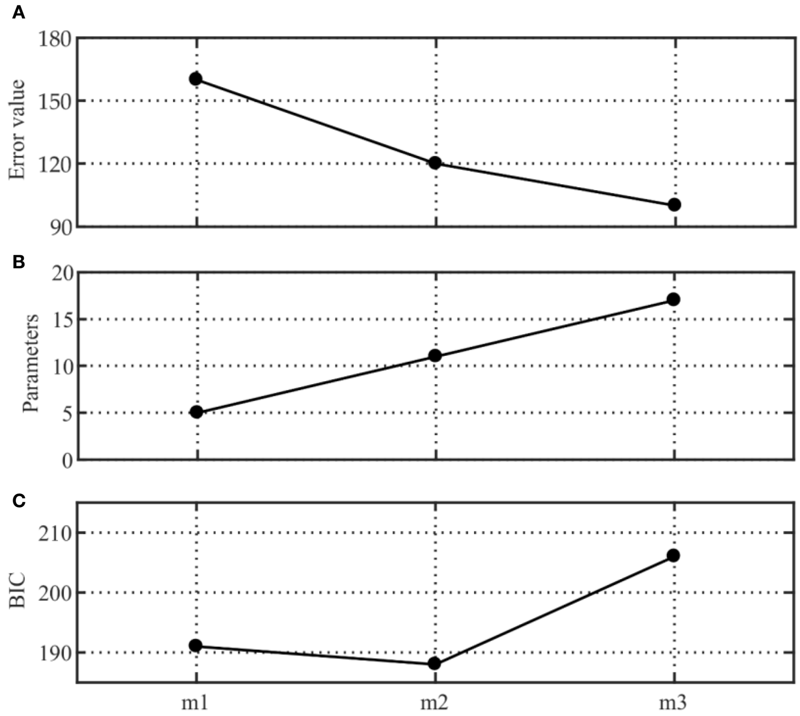


FIGURE 10
Results of the model selection study: **(A)** Error value, **(B)** parameters, and **(C)** BIC, where m_1, m_2, and m_3 represent the inversion model of the one-, two-, and three-layer uniformly distributed seabed, respectively.

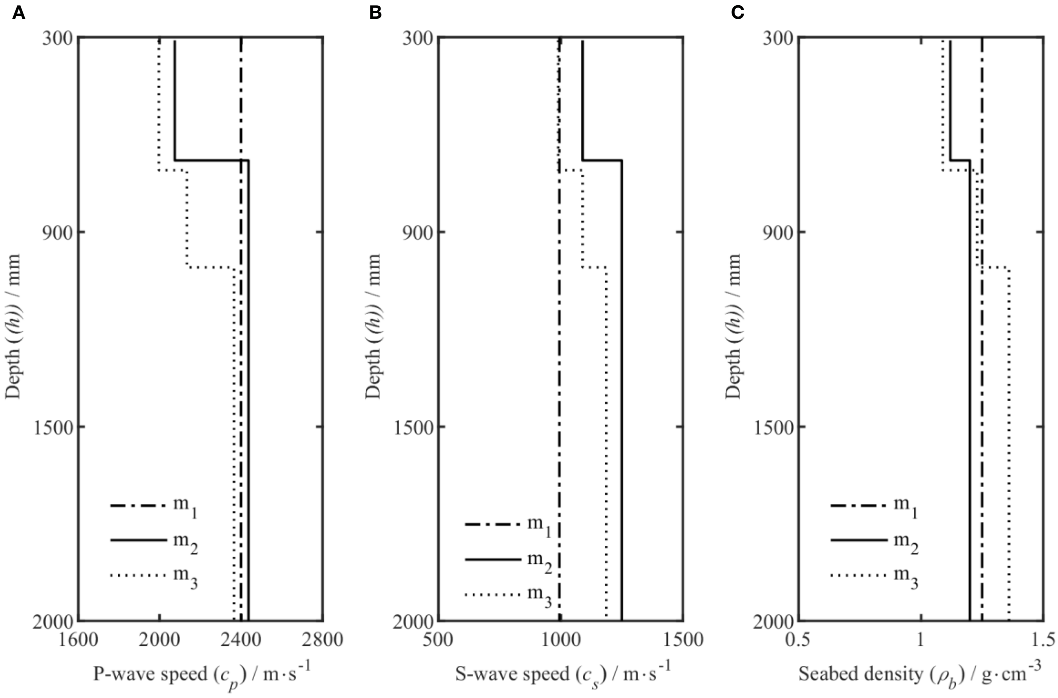


FIGURE 11
Density, compression wave speed, and shear wave speed profiles from the model selection study, where m_1, m_2, and m_3 represent the inversion model of the 1-layer, 2-layer, and 3-layer uniformly distributed seabed, respectively. **(A–C)** Relationship of c_p, c_s, and ρ with depth, respectively.

TABLE 3 Results of the measured inversion data.

Layer	Parameters	Search Bounds	Inversion Values
Sedimentary	$c_{p2}/\mathrm{m\cdot s^{-1}}$	2000-2200	2073.65
	$c_{s2}/\mathrm{m\cdot s^{-1}}$	1000-1200	1093.76
	$\rho_{b2}/\mathrm{g\cdot cm^{-3}}$	1-1.3	1.11
	$\alpha_{p2}/\mathrm{dB\cdot\lambda^{-1}}$	0.09-0.11	0.09
	$\alpha_{s2}/\mathrm{dB\cdot\lambda^{-1}}$	0.09-0.11	0.10
	H_2/mm	230-270	249.69
Semi-infinite seabed	$c_{p3}/\mathrm{m\cdot s^{-1}}$	2000-2900	2440.17
	$c_{s3}/\mathrm{m\cdot s^{-1}}$	1000-1500	1255.82
	$\rho_{b3}/\mathrm{g\cdot cm^{-3}}$	1-1.5	1.21
	$\alpha_{p3}/\mathrm{dB\cdot\lambda^{-1}}$	0.09-0.11	0.10
	$\alpha_{s3}/\mathrm{dB\cdot\lambda^{-1}}$	0.09-0.11	0.09

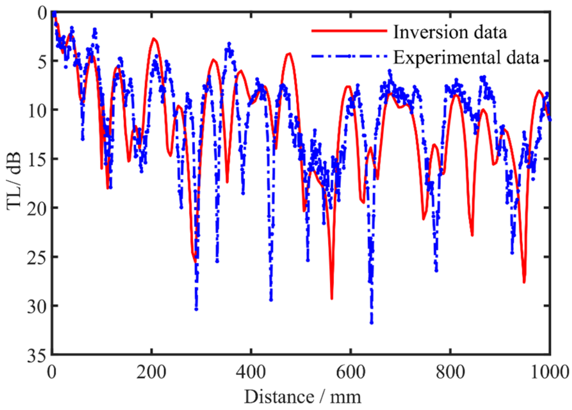


FIGURE 12
Comparison of the transmission loss between the measured (dashed blue trace) and the inversion data (solid red trace).

between the inversion data of the MAP model and the data acquired by conducting experimental measurements, the TL obtained by inversion uses the same source frequency as the experimental data. Figure 12 shows that Comparison of the transmission loss between the measured data and the inversion data, Table 3 lists the prior inversion intervals and the mean of the inversion results. The density of the polyvinyl chloride (PVC) slab that was used to simulate the semi-infinite seabed is known to be 1.20 g·cm^{-3}, and the thickness of the sediment, simulated by sand in our experiments, is approximately 250 mm. The average density of the semi-infinite seabed obtained by inversion was 1.21 g·cm^{-3}, and the average thickness of the sedimentary layer was 249.69 mm. The results show that the inversion results were consistent with the experimental results.

In addition, we compared the inversion results with those obtained in the simulation of a semi-infinite seabed with a plate of the same material. The inversion results showed that the speed of the compression wave and the shear wave was 2399.36 m·s^{-1} and 1242.97 m·s^{-1}, respectively. Considering that the sediment layer was replaced by fine sand and that there were many uncertain factors, such as the uniformity of the fine sand, errors would inevitably exist. Thus, the inversion results obtained for the speed of the compression wave and the shear wave of the semi-infinite seabed in this study were basically the same. It further verified the correctness of the inversion method proposed in this paper.

The uncertainty analysis of each parameter obtained by the inversion is helpful to judge the difficulty of accurate inversion of each parameter. The estimated value obtained by the OSA method was used as the initial model for sampling, and the PPD numerical integration solution of the parameters was obtained using the MHS method. The one-dimensional probability distribution map of the parameters corresponding to the two-layer uniform seabed distribution model is shown in Figures 13, 14, where the red solid line represents the MAP value of each parameter, and the blue line represents the distribution of the difference between the mean and standard deviation of the data result in the interval. In the figure, the one-dimensional marginal probability distribution of the parameters in the sedimentary layer and the semi-infinite seabed follows a normal distribution in the prior interval. The distribution of the sedimentary parameters is narrower than that of the semi-infinite seabed, indicating that these parameters are more sensitive. From the point of view of a single parameter, the distributions of c_{p2}, c_{s2}, and c_{s3} are narrow. This shows that the parameter can be well distinguished, and the uncertainty is small; in contrast, c_{p3} and the results of the wave speed attenuation distribution in each layer are poor, indicating that the uncertainty of this parameter is relatively large and that the parameter is not easy to determine.

The inversion of the geoacoustic parameters is a multi-parameter optimization problem; the parameters are coupled with each other, and the correlation between the parameters directly affects the inversion results and increases the uncertainty in the parameters that need to undergo inversion. Therefore, we considered it necessary to analyze the extent to which the inversion parameters are correlated. The parameter correlation matrix (Figure 15) shows a strong positive correlation between α_{s2} and c_{p3}. In addition, c_{s3} and α_{p3} are also strongly positively correlated, and this also holds for c_{p2} and ρ_{b2}. In contrast, α_{s2} and ρ_{b2} have a negative correlation, and a strong negative correlation exists between h_2 and ρ_{b2} and ρ_{b3}.

Figure 11 shows the profile structure of parameters, which is intended to represent the depth distribution of parameters

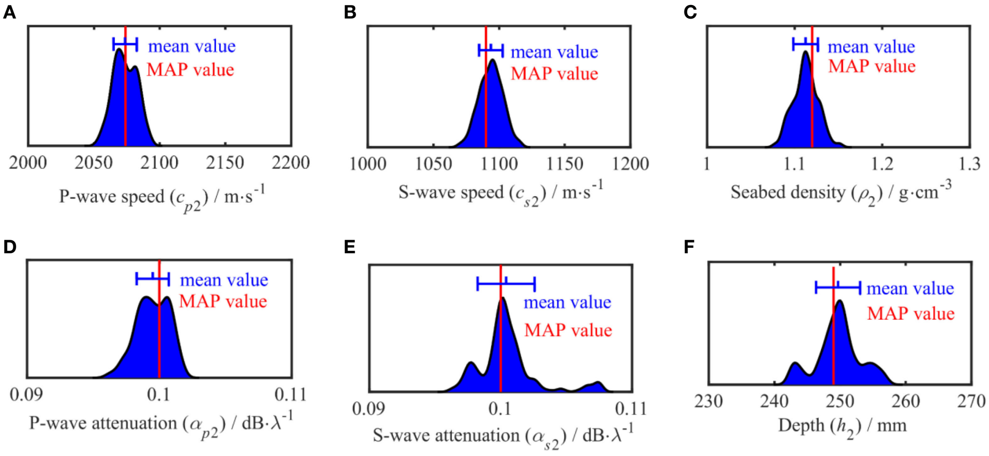


FIGURE 13
One-dimensional marginal probability distribution of sedimentary parameters.

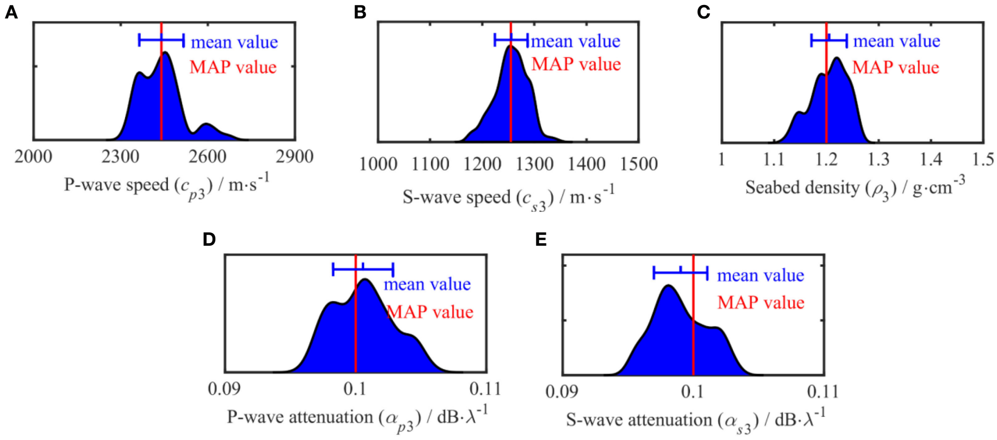


FIGURE 14
One-dimensional marginal probability distribution of semi-infinite parameters.

corresponding to the MAP, in which each parameter is a specific value. However, what we want to show in Figure 16 is the two-dimensional edge probability density distribution of parameters changing with depth, highlighting that the two-dimensional edge probability density of different layers and parameters is different, while the distribution range of color represents the probability density distribution, that is, the uncertainty of parameters. Figures 16A–C shows the two-dimensional edge probability distribution of the compression wave speed, shear wave speed, and density with depth to express the relationship among these parameters more intuitively and accurately. Because the inversion parameters are related to those of the sedimentary layer and lower seabed, the two-dimensional edge probability density distribution starts from the sedimentary layer. The results showed that, compared with the semi-infinite seabed, the probability distribution of the compression wave speed, shear wave speed, and density of the sedimentary layer were narrower and less uncertain, which further indicated that the parameters of the sedimentary layer were more sensitive.

5 Conclusions

The nonlinear Bayesian inversion method is not only valuable for effectively estimate the MAP model of the seabed parameters and analyzing the uncertainty in these parameters from a statistical point of view. In this study, our method based on nonlinear Bayesian inversion theory was employed for the inversion of geoacoustic parameters, which enriches and improves the related fields of geophysics and sea acoustics. Furthermore, the proposed method has essential prospects for practical application to seabed resource exploration, sea environment detection, sea engineering, and sea development. Specifically, in this study, a nonlinear Bayesian inversion method was used to invert the pressure field created by underwater sound wave as detected by underwater sensors. We varied the number of layers in the seabed in the parameterized models that we considered. Then, the optimal model was selected using the BIC, and the uncertainty in the inversion results was analyzed. The following conclusions were drawn:

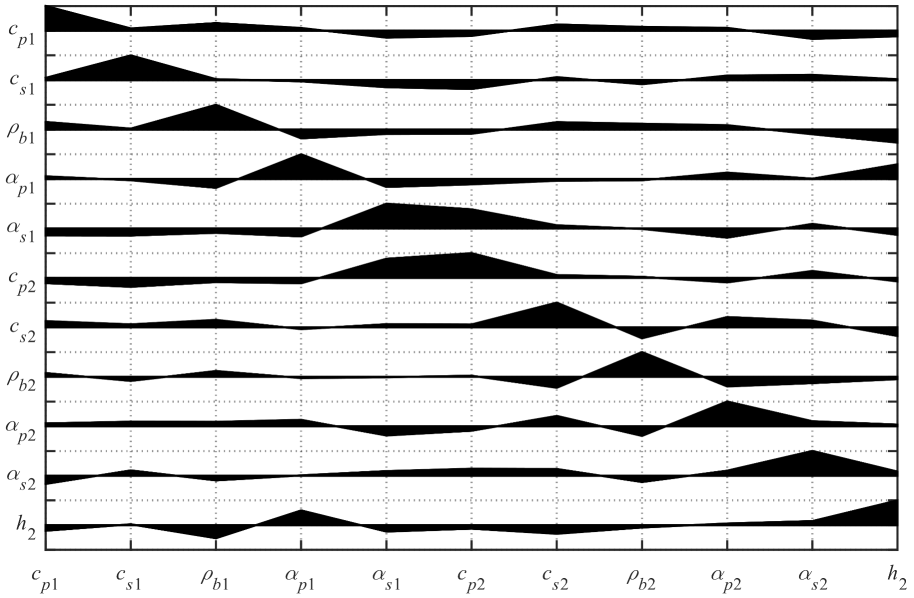


FIGURE 15
Correlation matrix diagram between inversion parameters.

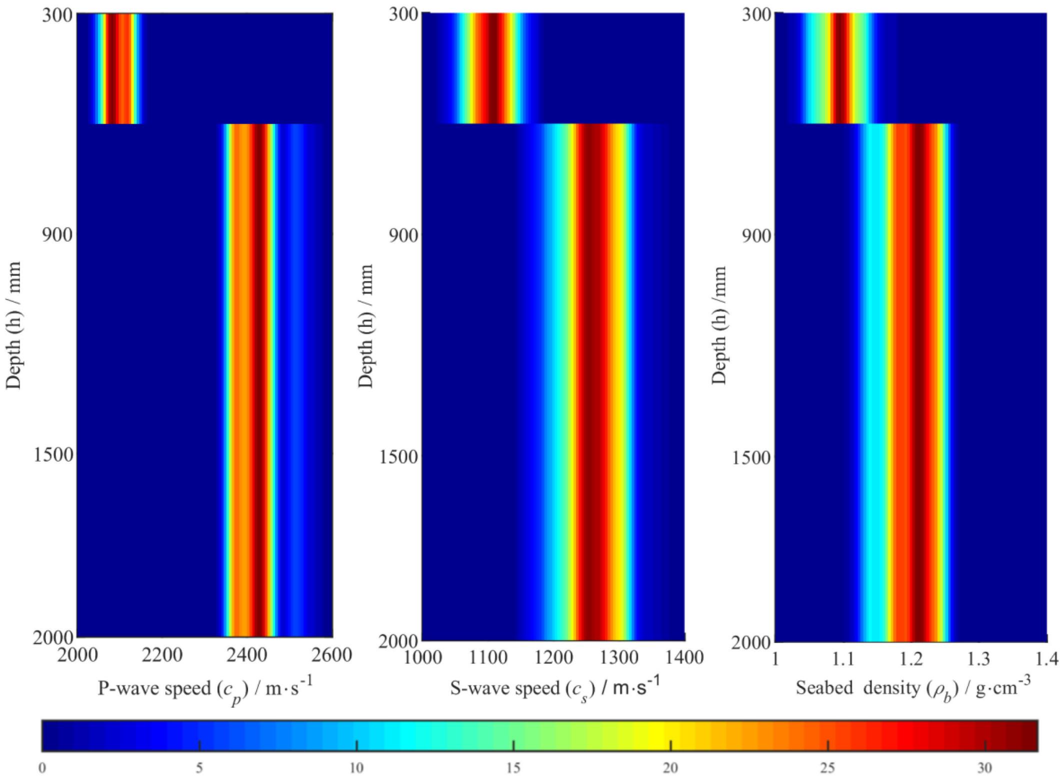


FIGURE 16
Two-dimensional marginal probability distribution. **(A-C)** two-dimensional edge probability distribution of compression wave speed, shear wave speed and density with depth.

(1) Bayesian inversion of the pressure field of sound can be used to obtain the parameters that define the seabed. The optimal model determined by BIC is a two-layer uniform seabed distribution model, which is consistent with the experimental results, The TL comparison results also prove the feasibility of the inversion method.

(2) With the aim of developing a nonlinear Bayesian inversion method based on the multi-layer seabed model, in the context

of the research discussed in this article, the OSA was able to compute the estimated value of MAP efficiently and accurately, and avoided the phenomenon of parameter "reversal" in the multi-layer seabed model. OSA can improve the accuracy of the inversion results, which is verified in the simulation.

(3) The uncertainty analysis of the inversion results enabled us to conclude that, compared with the compression wave speed attenuation, shear wave speed attenuation, and density, the uncertainty in the compression wave speed and shear wave speed is smaller when the inversion of the sound pressure field is used for the calculation, and these parameters have greater sensitivity. Compared with those in the case of semi-infinite seabed, the sediment layer parameters are more sensitive and less uncertain.

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

YX processed data, reviewed literature and wrote the original manuscript. HZ co-authored the "Abstract," "Introduction," and supervised the study and reviewed the manuscript. XW contributed to data. GZ provides support in optimization algorithm. XL and JW contributed to image format modification. All authors contributed to the article and approved the submitted version.

Funding

This research was found by the National Natural Science Foundation of China (Grant No: 11704337), the Stable Supporting Fund of Acoustic Science and Technology Laboratory (Grant No: JCKYS2020604SSJS011), the Open Foundation from Marine Sciences in the First-Class Subjects of Zhejiang (20200101) and the Youth Innovation Promotion Association CAS (Grant No. 2020023). The Science Foundation of Donghai Laboratory (Grant No: DH-2022KF01018).

Acknowledgments

We acknowledge the National Key Laboratory, College of Underwater Acoustic Engineering, Harbin Engineering University, China for supporting in 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

Ballard, M.-S., Becker, K.-M., and Goff, J.-A. (2009). Geoacoustic inversion for the new Jersey shelf: 3-d sediment model. *IEEE J. Oceanic. Eng.* 35, 28–42. doi: 10.1109/JOE.2009.2034490

enavente, R., Dittmer, J., Cummins, P.-R., and Sambridge, M. (2019). Efficient Bayesian uncertainty estimation in linear finite fault inversion with positivity constraints by employing a log-normal prior. *Geophys. J. Int.* 217, 469–484. doi: 10.1093/gji/ggz044

Bevans, D.-A., and Buckingham, M.-J. (2016). A geoacoustic inversion technique using the low-frequency sound from the main rotor of a Robinson R44 helicopter. *J. Acoust. Soc Am.* 140, 3169–3169. doi: 10.1121/1.4969952

Dettmer, J., Dosso, S.-E., and Holland, C.-W. (2009). Model selection and Bayesian inference for high-resolution seabed reflection inversion. *J. Acoust. Soc Am.* 125, 706–716. doi: 10.1121/1.3056553

Dong, H.-F., and Dosso, S.-E. (2011). Bayesian Inversion of interface-wave dispersion for seabed shear-wave acoustic profiles. *IEEE J. Oceanic. Eng.* 36, 1–11. doi: 10.1109/JOE.2010.2100490

Dosso, S.-E., and Dettmer, J. (2011). Bayesian Matched-field geoacoustic inversion. *Inverse. Problems.* 27, 55009. doi: 10.1088/0266-5611/27/5/055009/meta

Dosso, S.-E., Nielsen, P.-L., and Harrison, C.-H. (2009). Bayesian Inversion of reverberation and propagation data for geoacoustic and scattering parameters. *J. Acoust. Soc Am.* 125, 2867. doi: 10.1121/1.3106524

Dosso, S.-E., Wilmut, M.-J., and Lapinski, A.-L. S. (2001). An adaptive-hybrid algorithm for geoacoustic inversion. *IEEE J. Oceanic. Eng.* 26, 324–336. doi: 10.1109/48.946507

Enming, X., Rongwen, G., Dosso, S.-E., Jianxin, L., and Hao, D. (2018). Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluricdata. *Geophys. J. Int.* 213, 1751–1767. doi: 10.1093/gji/ggy071

Fu, D., Xiao, G., and Zhou, L. (2018). Nonlinear Bayesian theory and BIC criterion-based on high precision Rayleigh wave inversion of cutoff-wall. *Water Resour. Hydropower. Eng.* 49, 64–70. doi: 10.13928/j.cnki.wrahe.2018.08.008

Gao, F., Pan, C., and Sun, L. (2017). Geoacoustic parameters inversion of bayes matched-field: A multi-annealing Gibbs sampling algorithm. *Acta Armamentarii.* 38, 1385–1394. doi: 10.3969/j.issn.1000-1093.2017.07.017

Julien, B., Ying, T.-L., Dimitrios, E., Goff, J.-A., and Dosso, S.-E. (2018). Geoacoustic inversion on the new England mud patch using warping and dispersion curves of high-order modes. *J. Acoust. Soc Am.* 143, 405–411. doi: 10.1121/1.5039769

Li, C. (2012). Study of multi-mode interface-wave dispersion curves inversion based on nonlinear Bayesian theory. *Ocean. Univ. China* 5, p74–p76. doi: 10.7666/d.y1928402

Li, C., Dosso, S.-E., and Dong, H. (2012). Interface-wave dispersion curves inversion based on nonlinear Bayesian theory. *Acta Acustica.* 37, 225–231. doi: 10.15949/j.cnki.0371-0025.2012.03.018

Li, X., Piao, S., Zhang, M., and Liu, Y. (2019). A passive source location method in a shallow sea waveguide with a single sensor based bayesian theory. *Sensors* 19, 1452. doi: 10.3390/s19061452

Li, C., Yang, Y., Wang, R., and Yan, X. (2018). Acoustic parameters inversion and sediment properties in the yellow river reservoir. *Appl. Geophys.* 15, 78–90. doi: 10.1007/s11770-018-0663-z

Michalopoulou, Z., and Gerstoft, P. (2019). Multipath broadband localization, bathymetry, and sediment inversion. *IEEE J. Ocean. Eng.* 45, 92–102. doi: 10.1109/JOE.2019.2896681

Ohta, K., Matsumoto, S., Okabe, K., Asano, K., and Kanamori, Y. (2008). Estimation of shear wave acoustic in ocean bottom sediment using electromagnetic introduction source. *IEEE J. Oceanic. Eng.* 33, 233–239. doi: 10.1109/JOE.2008.926108

Sen, M.-K., and Stoffa, P.-L. (2006). Bayesian Inference, gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophys. Prospecting.* 44, 313–350. doi: 10.1111/j.1365-2478.1996.tb00152.x

Seongryong, K., Jan, D., and Junkee, R. (2016). Highly efficient Bayesian joint inversion for receiver-based data and its application to lithospheric structure beneath the southern Korean peninsula. *Geophys. J. Int.* 206, 328–344. doi: 10.1093/gji/ggw149

Xue, Y., Lei, F., Zhu, H., Xiao, R., Chen, C., and Cui, Z. (2021). An inversion method for geoacoustic parameters of multilayer seabed in shallow water. *J. Phys. Conf. Series.* 1739, 12019. doi: 10.1088/1742-6596/1739/1/012019/met

Yang, Z. (2009). *Geological oceanography* (Shandong Education Press). p125-p135.

Yang, K., Xiao, P., Duan, R., and Ma, Y. (2017). Bayesian Inversion for geoacoustic parameters from ocean bottom reflection loss. *J. Comput. Acoustics.* 25, 1750019. doi: 10.1142/S0218396X17500187

Yin, B., and Hu, X.-Y. (2016). Overview of nonlinear inversion using Bayesian method. *Prog. Geophys.* 31, 1027–1032. doi: 10.6038/pg20160313

Yuan, S., Wang, S., Lou, Y., Wei, W., and Wang, G. (2019). Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model. *Geophysics* 84, 149–164. doi: 10.1190/geo2017-0643.1

Zhang, X., Dai, X., and Yang, B. (2018). Fast imaging algorithm for the multiple receiver synthetic aperture sonars. *IET. Radar. Sonar. Navigation.* 12, 1276–1284. doi: 10.1049/iet-rsn.2018.5040

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

Zhang, X., Yang, P., and Sun, M. (2022). 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., Ying, W., Yang, P., and Sun, M. (2020). Parameter estimation of underwater impulsive noise with the class b model. *IET. Radar. Sonar. Navigation.* 14, 1055–1060. doi: 10.1049/iet-rsn.2019.0477

Zheng, G. (2019). *Geoacoustic parameter inversion based on sound field propagation characteristics in the shallow water* (Zhoushan, China: Zhejiang Ocean University).

Zheng, G., Zhu, H., and Wang, X. (2020). Bayesian Inversion for geoacoustic parameters in shallow Sea. *Sensors* 7, 2150. doi: 10.3390/s20072150

Zheng, G., Zhu, H., and Zhu, J. (2019). A method of geo-acoustic parameter inversion in shallow sea by the Bayesian theory and the acoustic pressure field. *2nd. Franco-Chinese. Acoustic. Conference.* 283, 6003. doi: 10.1051/matecconf/201928306003

Zhu, H. (2014). *Geoacoustic parameters inversion based on waveguide impedance in acoustic vector field* (Harbin, China: Harbin Engineering University). doi: 10.7666/d.D749708

Zhu, 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. University.* 33, 648–652+659. doi: 10.3969/j.issn.1006-7043.201105075

Zhu, H., Zhu, J., and Zheng, G. (2019). A separation method for normal modes in shallow sea under near field. *Acta Acust.* 44, 39–48. doi: 10.15949/j.cnki.0371-0025.2019.01.005

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fmars.2023.1112592

Reconstruction of horizontal tidal current fields in a shallow water with model-oriented coastal acoustic tomography

Naokazu Taniguchi[1]*, Hidemi Mutsuda[1], Masazumi Arai[1], Yuji Sakuno[1], Kunihiro Hamada[1], Toshiyuki Takahashi[2], Kengo Yoshiki[2] and Hironori Yamamoto[2]

[1]Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Japan,
[2]Fukken Co., LTD, Hiroshima, Japan

OPEN ACCESS

EDITED BY
Haixin Sun,
Xiamen University, China

REVIEWED BY
Xiao-Hua Zhu,
Ministry of Natural Resources, China
Yasumasa Miyazawa,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan

*CORRESPONDENCE
Naokazu Taniguchi
✉ ntaniguchi@hiroshima-u.ac.jp

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 November 2022
ACCEPTED 03 January 2023
PUBLISHED 03 February 2023

CITATION
Taniguchi N, Mutsuda H, Arai M, Sakuno Y, Hamada K, Takahashi T, Yoshiki K and Yamamoto H (2023) Reconstruction of horizontal tidal current fields in a shallow water with model-oriented coastal acoustic tomography. *Front. Mar. Sci.* 10:1112592. doi: 10.3389/fmars.2023.1112592

Reciprocal acoustic transmission and coastal acoustic tomography (CAT) is a powerful tool for measuring tidal currents in shallow coastal water, especially if data assimilation is employed. In previous CAT data assimilation studies, the ensemble Kalman filter (EnKF) has been implemented to assimilate observed path-averaged velocity, but *ad-hoc* procedures called localization and inflation, which compensate for issues associated with using ensemble approximation, were not always implemented. In this study, EnKF is applied to assimilate the path-averaged currents obtained from a reciprocal acoustic transmission experiment conducted at Mihara-Seto in the Seto Inland Sea, Japan, with four acoustic stations in 2020 to reconstruct spatiotemporal variations of tidal currents at the observation site. We executed EnKF with several combinations of different values of the inflation, localization, and the number of ensemble members. The resulting data assimilated velocity reconstructions are compared with acoustic Doppler current profiling (ADCP) results. The results show that data assimilation with EnKF improved the velocity reproduction compared with the model prediction and that implementing covariance inflation contributed to additional improvements. The covariance localization did not improve the results in our case. The best result in terms of fractional error variance (FEV) between the ADCP velocity was obtained from the case with 980 ensemble members with a covariance inflation of 1.01; the FEV was 7.9%. The case of 98 ensemble members with a covariance inflation of 1.01 resulted in similar performance; the FEV value was 8.2%. Thus, with the covariance inflation, the number of ensemble member used in previous CAT studies are reasonable. In the study, we also clarified the reason for the high-frequency variation in the observed path-averaged currents in a preliminary experiment; the path-averaged currents had captured the spatiotemporal variation of vortex generation associated with island wakes. The reciprocal acoustic transmission with EnKF can capture short-period variation over a long period; thus, it can be used in studies of coastal physical processes with various time scales.

KEYWORDS

tidal currents, reciprocal acoustic transmission, path-averaged velocity, coastal acoustic tomography, data assimilation, ensemble Kalman filter, island wake, covariance inflation

1 Introduction

Ocean acoustic tomography (OAT) is a unique method of acoustical oceanography for remotely sensing the sound speed and current in the ocean interior (Munk and Wunsch, 1979; Munk et al., 1995). In OAT, one transmits and receives acoustic signals between multiple sources/receivers through the ocean, precisely determines travel times of acoustic signals, and then, applies an inverse method to the determined travel times to estimate the ocean interior states (sound speed and ocean currents) traversed by the acoustic paths. The effects of sound speed and ocean currents on the travel times can be separated by computing the sum and difference of the travel times obtained from reciprocal acoustic transmission. This OAT and reciprocal acoustic transmission methods have been applied to coastal shallow water, which may be referred to as coastal acoustic tomography (CAT). Readers can refer to Kaneko et al. (2020) and references therein for the application of reciprocal acoustic transmission (and CAT) in shallow water around Japan. Because of its usefulness, the concept of OAT and reciprocal acoustic transmission have further been applied to a very small port (e.g., Ogasawara and Mori, 2016) and/or rivers (e.g., Kawanisi et al., 2017). Additionally, multiple new approaches have been proposed, such as deploying networked array (Huang et al., 2013; Chen et al., 2021), moving ship tomography (Huang et al., 2019) originating from the work by Cornuelle et al. (1989), or physics based approach (Wang et al., 2018).

A promising method of OAT/CAT is a model-oriented inverse method (state estimation or data assimilation; Munk et al., 1995). Data method assimilation have now been widely used in science and engineering fields and there are multiple assimilation schemes (e.g., Evensen, 2009, and references therein). Among various assimilation schemes, the data assimilation with ensemble Kalman filter (EnKF; Evensen, 1994) is ease to implement yet powerful enough to reproduce tidal currents in shallow water by assimilating the reciprocal acoustic transmission data (Park and Kaneko, 2000; Lin et al., 2005). The CAT with EnKF data assimilation have been used to discuss physical process in coastal seas (Zhu et al., 2017; Chen et al., 2017; Zhu et al., 2021). Recently, CAT data assimilation is implemented with velocity fields obtained from an ocean radar observation (Zhu et al., 2022).

EnKF approximates model covariance by ensemble covariance, which enables the application of Kalman filter theory to nonlinear models (like a nonlinear numerical ocean model) and large state dimensions. However, approximating model covariance by ensemble covariance with a finite ensemble size causes several issues, such as long-range spurious or unphysical correlations in covariances (sampling error) and the underestimation of the ensemble variance. Thus, EnKF implementation requires somewhat *ad-hoc* tuning such as localization and inflation (e.g., Evensen, 2009). The common approach of localization is covariance localization, which damps long-range spurious correlation and localize the effect of data assimilation near the observation location (e.g., Hamill et al., 2001). The inflation is to broaden the updated analysis in order to counteract the excessive variance reduction caused by spurious correlations in the update (e.g., Anderson and Anderson, 1999).

The above mentioned localization and inflation have not been implemented commonly in CAT data assimilation studies. An advantage of CAT method in coastal shallow sea is that path-averaged currents can be measured and assimilated into ocean models with a short time interval (e.g., every few minutes). However, frequent Kalman filter updates with an underestimated ensemble variance (due to small ensemble sizes) leads to filter divergence. Filter divergence might happen in previous CAT studies. For example, there was a case that ensemble spreads of predicted path-averaged currents over some paths were not large enough for some period, and analysis steps of assimilation did not improve effectively (e.g., Figure 11 in Zhu et al., 2017). Therefore, it seems that the covariance inflation has possibility to improve the CAT data assimilation with EnKF, in particular when the path-averaged currents are assimilated frequently. An example of the use of covariance localization is that by Chen et al. (2017), which implemented the covariance localization to the problem of upwelling associated with typhoon passage (with relatively strong wind). On the one hand, tidal currents will be spatially correlated in a small study area, and covariance localization may not be required; on the other hand, if EnKF is used with small ensemble members, there will arise spurious correlation and localization may be required. Chen et al. (2017) implemented the covariance localization, but improvements associated with the localization is not clear.

We conducted a reciprocal acoustic transmission experiment between a pair of transceivers (i.e., acoustic stations) deployed in an area of complex coastlines with multiple islands and channels around Mihara-Seto in the Seto Inland Sea, Japan (Figure 1) in 2019 and demonstrated the capability of reciprocal acoustic transmission to capture high frequency variations of small magnitude in tidal currents (Taniguchi et al., 2021a). In 2020, we conducted a reciprocal acoustic transmission experiment using four acoustic stations deployed at the same observation site to determine the capability of data assimilation with reciprocal acoustic transmission to reproduce complex current velocity fields in the observation site. We developed a numerical ocean model for simulating depth-averaged flow around the observation site (i.e., two dimensional model) and applied EnKF to assimilating the path-averaged currents to the numerical ocean model.

The specific purpose of this paper is twofold. First, we clarify the performances of EnKF with path averaged currents in regard to the tuning parameters of covariance localization and covariance inflation as well as the number of ensemble members. We executed the sequential data assimilation of EnKF with several combinations for these values and compared the resulting velocity fields with the velocity from acoustic Doppler current profiling (ADCP) observation. Clarified results for the method performance can be a guidance for parameter choices and contribute to future study using the same method. The second purpose is to clarify the reasons for high-frequency variation found in the path-averaged currents in the preliminary experiment (Taniguchi et al., 2021a). Reciprocal acoustic transmission can be performed over a long period (or operationally) with sufficiently short time intervals. Although the data with short time interval may not contribute directory to reproducing spatially precise velocity field of tidal currents, the estimation can be precise for both spatially and temporally by combining with a numerical ocean model (thus, data assimilation method). As we shall see, the data

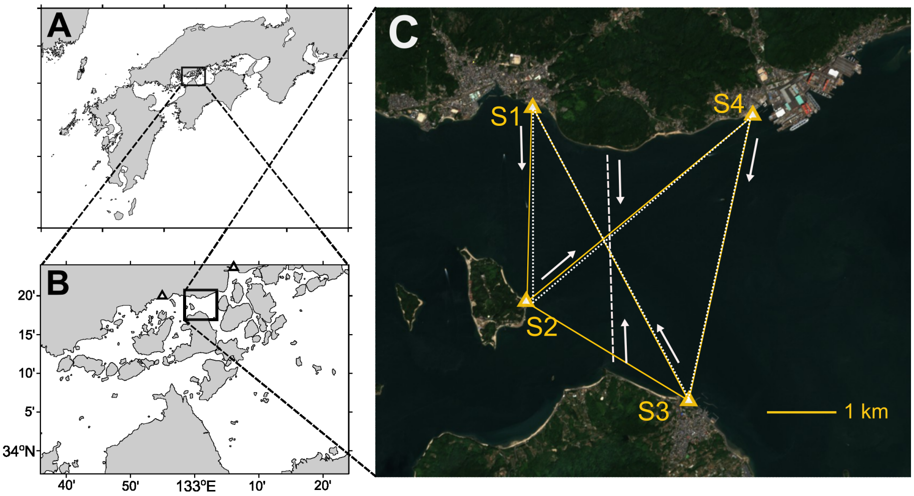


FIGURE 1
(A, B) Geographical location of observation site (Mihara-Seto) in the Seto Inland Sea, Japan; **(C)** satellite (Sentinel-2) image of observation site and locations of acoustic stations (four triangles with labels S1, S2, S3, and S4). The two black triangles in panel **(B)** is the location where astronomical tide information is obtained: Takehara for the west and Itozaki for the east. In panel **(C)**, the five yellow lines are the acoustic transmission paths over each station pair; the white dashed and dotted lines indicate the tracks of shipboard acoustic Doppler current profiling (ADCP) observation performed on Oct. 30 and 31, respectively, with the arrows for the route direction.

assimilation results reveals that the path-averaged current had captured vortex generation behind an island (i.e., island wake). Such capability to reproduce the transient vortexes can advance the studies on physical processes on coastal shallow water.

2 Materials and methods

2.1 Field observation

An experiment on reciprocal acoustic transmission between four acoustic stations was conducted in a complex sea area of Mihara-Seto with multiple islands and channels 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 (S1, S2, S3, and S4). The coast blocks the transmission between the S1 and S4 stations, and thus the travel time is not observable between them. The distances between the five station pairs (S1 and S2, S1 and S3, S2 and S3, S2 and S4, and S3 and S4) were 2842, 4895, 2805, 4300, and 4280 m, respectively. The travel-time measurement system used for this experiment was nearly the same as that used in a preliminary experiment in the same area in 2019 (Taniguchi et al., 2021a), which was initially developed for a moving-ship tomography study (Huang et al., 2019), with some changes and modifications. For this experiment, we used an electroacoustic transducer Model T235 (which can operate over a frequency band of 10–25 kHz) of Neptune Sonar as a transceiver. The ways of deploying the transceivers for stations S1 and S2 were the same as those in the preliminary experiment (Taniguchi et al., 2021a). The deployment of the S3 station was identical to that of the S1 station. For station S4, the transceiver was deployed about 1 m ahead of a pier using iron flames because there was rubble right next to the pier.

The reciprocal acoustic transmissions between the four stations were performed repeatedly every two minutes. A pulse compression method was implemented to increase the signal-to-noise ratio (SNR) without decreasing the time resolution. The transmitted signal in the experiment was a binary phase-shift keyed signal encoded by a pseudo-random binary sequence called a maximal length sequence (often referred to as an m-sequence).

Table 1 summarizes the transmission signal parameters for this experiment. In the experiment, we introduced two ideas on signal transmission (Taniguchi et al., 2021b). The first idea was about the periodicity of m-sequences: the transmission signal in the experiment was one period of an m-sequence of 2,047 digits (i.e., 11th order m-sequence; Table 1), but the last 63 and first 64 digits of the m-sequence were respectively added to the head and tail of the m-sequence to achieve the original auto-correlation property of repeated m-sequences. The other idea is about the transmission timing of the signal: the start time of each pulse was synchronized among all stations by using the one pulse-per-second signal of GPS, but time-shifted slightly so that the arrival signals do not overlap at all stations. Some of the results of the experiment are found in a report by Taniguchi et al. (2021b).

Travel times of acoustic signals were determined by detecting the time at which the received arrival pattern (i.e., the impulse response or cross-correlogram) first exceeded 14 dB. 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 c_0 is the reference sound speed value, and L is the transceiver-to-transceiver distances. Here we assumed straight acoustic ray paths

TABLE 1 Transmission signal parameters.

Order (length) of m-sequence	11 (2047 digits)
Carrier cycles per digit	4
Carrier frequency	18.018 kHz
Signal bandwidth	≈4.5 kHz
Digit duration	0.222 ms
Period duration	454 ms
Number of periods	1
Number of additional digits	63 & 64
Signal duration	482 ms

because the range-to-depth ratio is about 100:1, and the acoustic ray paths are nearly horizontal. Also, since the transceivers were nearly fixed in the experiment, the transceiver motion was not required to consider. The erroneous estimates of the path-averaged velocity were removed and linearly interpolated if the data gaps were less than 10 minutes. The obtained currents are path-averaged values but can be regarded as section-averaged currents because the tidal currents at the observation site are almost vertically uniform, as shown by an ADCP observation in the preliminary experiments (Taniguchi et al., 2021a). The path-averaged currents for each station pair were summarized as time series data starting from 16:00 on October 28, 2020. These time series of path-averaged (section-averaged) currents of the five station-pair were the data assimilated into a numerical ocean model.

We conducted shipboard ADCP observations on October 30 and 31 to obtain velocity data for comparison with data assimilation results. The ADCP used in the experiment was a WH-ADCP (Teledyne RD Instruments, Inc.). The details of the operating parameters for the ADCP observations are summarized in Table 2. The shipboard ADCP observations were made hourly from 07:00 to 17:00 along the north-south section on October 30 (dashed line in Figure 1C) and from 07:00 to 16:00 along the acoustic transmission lines (without the line connecting the S2 and S3 stations) on October 31 (dotted lines in Figure 1C). The fishing boat that had the ADCP installed traveled along each transect at an average ship speed of about 3–4 knots, taking about 20–45 min to complete the journeys. The ship track and speed were recorded using a differential GPS, which was used to correct the ADCP's compass error (Trump and Marmorino, 1997). When the ship approached the S2 station, we allowed the ADCP's transect to deviate from that of the acoustic transmission path for safety reasons because the location for the S2 station was continually used by ferry boats for passengers. In all, 21 cross-sections of velocity data were obtained along the north-south transect and along the transmission lines from these ADCP observations. After the shipboard ADCP observation, the recorded ADCP velocity data with the compass correction were averaged over the depth and then 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 used in the data assimilation (see Sec. 2.2). This spatial average also corresponds to the time average over about 30 s.

TABLE 2 ADCP operation parameters.

Frequency	600 kHz
Bin length	2 m
Bin number	30
First bin depth	3.58 m
Ensemble period	4.65 s
Pings in each ensemble	5 pings
Ping interval	0.72 s
S.D. of each ping	3.35 cm s^{-1}

2.2 Numerical ocean model

A numerical ocean model used in this study is based on the depth-averaged shallow water equations of the following forms:

$$\frac{\partial \eta}{\partial t}+\frac{\partial UD}{\partial x}+\frac{\partial VD}{\partial y}=0, \tag{2}$$

$$\frac{\partial UD}{\partial t}+\frac{\partial U^2D}{\partial x}+\frac{\partial UVD}{\partial y}-fVD = -gD\frac{\partial \eta}{\partial x}+F_x-C_D\sqrt{U^2+V^2}U, \tag{3}$$

$$\frac{\partial VD}{\partial t}+\frac{\partial UVD}{\partial x}+\frac{\partial V^2D}{\partial y}+fUD = -gD\frac{\partial \eta}{\partial y}+F_y-C_D\sqrt{U^2+V^2}V, \tag{4}$$

where (x,y) are the east-west and north-south coodinates; (U,V) are the eastward and northward components of the depth-averaged velocity; η is the free surface elevation; H is the water depth at rest; D is the total water depth (thus, $D=H+\eta$); f is the Coriolis parameter; g is the gravitational acceleration. The terms F_x and F_y are the lateral eddy viscosity and are defined according to

$$F_x=\frac{\partial}{\partial x}\left(2HA_H\frac{\partial U}{\partial x}\right)+\frac{\partial}{\partial y}\left[HA_H\left(\frac{\partial U}{\partial y}+\frac{\partial V}{\partial x}\right)\right], \tag{5}$$

$$F_y=\frac{\partial}{\partial y}\left(2HA_H\frac{\partial V}{\partial y}\right)+\frac{\partial}{\partial x}\left[HA_H\left(\frac{\partial U}{\partial y}+\frac{\partial V}{\partial x}\right)\right], \tag{6}$$

where A_H is the lateral kinematic eddy viscosity coefficient and was assigned a constant value of 4.8 m s^{-2} in the present model. The last term in Eqs. (3) and (4) are the bottom stress, and the bottom drag coefficient C_D is given by the larger of the two values between 0.0025 and $gn^2/H^{4/3}$ with Manning's roughness coefficient n=0.0029.

The above shallow water equations are solved numerically using the finite difference method. The model domain is shown in Figure 2. The numerical discretization and integration schemes followed those used in Princeton Ocean Model (POM; Blumberg and Mellor, 1987). The grid space is 100 m. 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). Tidal forcing at each open boundary drives the model interior. In the present model, a Flather condition, which is an extension of a radiation boundary condition, was implemented in

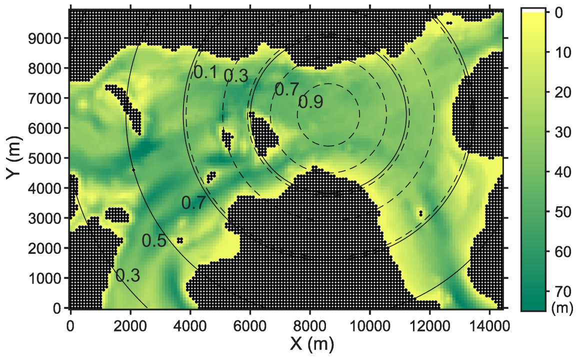


FIGURE 2
Domain of a numerical ocean model used in this study. The color indicates the water depth, and black dots correspond to land grids. The boundaries without dots (i.e., not land) are open boundaries where prescribed tidal forcing is given to drive the flow in the model domain. The dashed and solid circles are contours of the two covariance localization radii (4 km and 10 km) with a contour interval of 0.2.

the following way (e.g., Carter and Merrifield, 2007):

$$U_{n,b} = U_n^{ext} \pm \sqrt{\frac{g}{H}}(\eta_{b\pm 1} - \eta^{ext}), \tag{7}$$

where the subscripts n,b indicate the velocity normal to the boundary at the boundary grids; the superscript ext indicates using externally prescribed values. This Flather condition is less sensitive to errors in the prescribed boundary values compared with other boundary conditions, such as clamped surface elevations (Carter and Merrifield, 2007). The above Flather condition is provided for U at the west and east boundaries and V at the southern boundaries. The tidal elevations and normal velocities were constant along each open boundary in the present model. The tangential velocities at the boundary grids were given by the values of neighboring interior grids (i.e., a zero-gradient condition).

The prescribed tidal elevations at the west and east boundaries were derived from astronomical tides, i.e., the amplitudes and phases of five constituents (namely, M2, S2, K1, O1, and N2) at the nearest tide stations: Takehara for the west and Itozaki for the east (Figure 1B). These prescribed tidal elevations at the west and east boundaries are shown in Figure 3. 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 elevations as the western and eastern boundaries, respectively. The tidal currents normal to the open boundaries were determined by using the information on path-averaged current *via* an *ad-hoc* way as follows. First, we applied a harmonic analysis to the path-averaged current observed between the S1 and S3 stations and estimated the amplitudes and phases of each tidal constituent at the observation site. The estimated phases for each tidal constituent were used to prescribe the normal velocity at the open boundaries. Second, we obtained the maximum tidal current at the east 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 convert it to section averaged current speed, resulting in a maximum section-averaged current of 2.2 m s^{-1}. Third, the amplitudes of the estimated five tidal constituents for the normal velocity at the east boundary 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. 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 elevations equal to that at eastern boundaries.

2.3 Data assimilation

As is the case with previous CAT data assimilation studies (Park and Kaneko, 2000; Lin et al., 2005; Zhu et al., 2017), the EnKF scheme (Evensen, 1994; Evensen, 2003; Evensen, 2009) accompanied Monte Carlo method is used to assimilate the path-averaged current data into the ocean model. In the present numerical ocean model, state variables are the east-west and north-south components of the depth-averaged velocity ($U(x,y)$, $V(x,y)$) and sea surface elevation (tidal elevation; η). These state variables on all model grids are aligned and expressed as model state vector $\boldsymbol{x}_k$ where subscript k indicates the state vector at time k. The ocean model updates the model state vector $\boldsymbol{x}$ according to the governing equations and implemented numerical scheme. The relationship between the observed path-averaged current at time k and the model state vector $\boldsymbol{x}_k$ can be described as:

$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k, \tag{8}$$

where vector $\boldsymbol{y}_k$ contain the five path-averaged currents. Note that the observed path-averaged current cannot be assigned to a specific time step k of the model time step (1 s in the present model) because the acoustic signal travels and samples the ocean over a few seconds; thus, we assumed that the observed path-averaged currents are obtained at the scheduled time for each transmission (e.g., 00:00:00, 00:02:00, 00:04:00, and so on). The observation matrix $\boldsymbol{H}$ maps the model space (i.e., velocity fields) into observed space (i.e., path-averaged currents) and corresponds to the procedure that interpolates ($U(x,y)$, $V(x,y)$) along the transmission paths, projects it onto the direction parallel to the paths, and averages over the paths. In practical computation, we did not define (or use) observation matrix $\boldsymbol{H}$ but computed $\boldsymbol{H}\boldsymbol{x}_k$ by interpolating, projecting, and path-averaging the velocity field at time k. The term $\boldsymbol{v}_k$ is observation noise and is assumed to be Gaussian noise with covariance $\boldsymbol{R}_k$. The path-averaged currents are assumed to be independent over the station pairs, and thus, $\boldsymbol{R}_k$ is a diagonal matrix. We also assumed constant $\boldsymbol{R}_k$ over time in this study; thus, $\boldsymbol{R}_k$=$\boldsymbol{R}$. Considering the results of the preliminary experiment (Taniguchi et al., 2021a), we assigned 0.025^2 m^2 s^{-2} for all the diagonal elements.

We assume here that the observed path-averaged currents are available at time k. The numerical ocean model predicts $\boldsymbol{x}_k$ of each N ensemble member to apply EnKF. The N state vectors $\boldsymbol{x}_k$ corresponding to N ensemble members forms a matrix:

$$\boldsymbol{X}_k = (\boldsymbol{x}_k^{(1)}, \boldsymbol{x}_k^{(2)}, \ldots, \boldsymbol{x}_k^{(N)}), \tag{9}$$

where $\boldsymbol{x}_k^{(i)}$ is i-th ensemble member's state vector. The ensemble mean and ensemble anomaly are computed as follows:

$$\bar{\boldsymbol{x}}_k = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_k^{(i)} \tag{10}$$

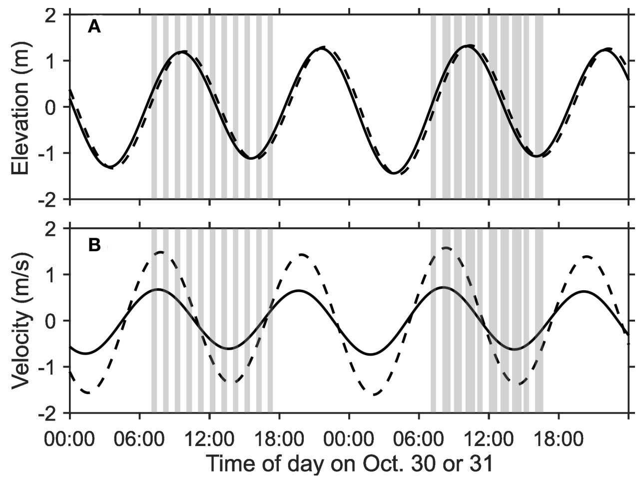


FIGURE 3
Time series of **(A)** tidal elevations and **(B)** tidal currents that were given as the open boundary conditions at the west (solid) and east (dashed) boundaries. The gray vertical bars indicate the time of the shipboard ADCP observation.

and

$$\boldsymbol{E}_k = (\boldsymbol{x}_k^{(1)} - \overline{\boldsymbol{x}}_k, \boldsymbol{x}_k^{(2)} - \overline{\boldsymbol{x}}_k, \ldots, \boldsymbol{x}_k^{(N)} - \overline{\boldsymbol{x}}_k). \tag{11}$$

With $\widetilde{\boldsymbol{E}}_k = \boldsymbol{E}_k / \sqrt{N-1}$, the ensemble covariance is formed as $\widetilde{\boldsymbol{E}}_k \widetilde{\boldsymbol{E}}_k^T$ where superscript T indicates matrix transpose. Correspondingly, a matrix of an ensemble of perturbed measurements (Burgers et al., 1998) are formed as:

$$\boldsymbol{D}_k = (\boldsymbol{y}_k, \boldsymbol{y}_k, \ldots, \boldsymbol{y}_k) + \sqrt{N-1}\boldsymbol{S}, \tag{12}$$

where $\boldsymbol{S}$ is the observation noise matrix whose columns are sampled from normal distribution with zero mean and covariance $\boldsymbol{R}$ and divided by $\sqrt{N-1}$. The ensemble covariance matrix for the perturbed measurements is formed as $\boldsymbol{S}\boldsymbol{S}^T$. With above $\boldsymbol{D}_k$ and $\boldsymbol{S}\boldsymbol{S}^T$, the ensemble of model states are updated as follows (Evensen et al., 2022):

$$\boldsymbol{X}_k^a = \boldsymbol{X}_k + \widetilde{\boldsymbol{E}}_k(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k)^{\mathrm{T}}(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k)^{\mathrm{T}} + \boldsymbol{S}\boldsymbol{S}^T)^{-1}(\boldsymbol{D}_k - \boldsymbol{H}\boldsymbol{X}_k), \tag{13}$$

where $\boldsymbol{X}_k^a$ indicates the updated ensembles of state vectors (i.e., ensemble of the assimilated model state vectors).

Approximating model covariance matrix by ensemble covariance matrix with a finite ensemble size causes several issues like long-range spurious or unphysical correlations in covariances. Covariance localization and inflation are commonly used to mitigate the impact of ensemble approximation (e.g., Evensen, 2009, and references therein). The covariance localization in this study was implemented as Kalman-gain localization (Chen and Oliver, 2017), which is given

$$\boldsymbol{X}_k^a = \boldsymbol{X}_k + \boldsymbol{\rho}_1 \circ \widetilde{\boldsymbol{E}}_k(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k)^{\mathrm{T}}(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k)^{\mathrm{T}} + \boldsymbol{S}\boldsymbol{S}^T)^{-1}(\boldsymbol{D}_k - \boldsymbol{H}\boldsymbol{X}_k), \tag{14}$$

where ∘ in Eq. (14) indicates Schur (element-wise) product, and $\boldsymbol{\rho}_1$ is a matrix of localization (damping) function. As is the case with Chen et al. (2017), the localization is implemented with an artificial point observation at the center of the tomographic array. The damping function is a fifth order polynomial (Gaspari and Cohn, 1999) that approaches zero far from the artificial point observation location. We tested two cases with localization radii of 4 and 10 km. The contour of the two damping functions are shown in Figure 2. When a localization radius of 4 km was chosen, the value of the damping function at the most outside station was about 0.5, which was the same condition as Chen et al. (2017). In this case, only state variables on the grids around the center of the tomographic array were fully updated. We tested a case for a weak localization effect with a localization radius of 10 km. In this case, the value of the damping function at the most outside station was about 0.9.

A covariance inflation (Anderson and Anderson, 1999) is a procedure replacing the updated ensemble analysis according to:

$$\boldsymbol{X}_k^a \leftarrow \rho_i\left(\boldsymbol{X}_k^a - \overline{\boldsymbol{X}}_k^a\right) + \overline{\boldsymbol{X}}_k^a, \tag{15}$$

where ρ_{i} is the inflation factor and $\overline{\boldsymbol{X}}^a$ indicates the ensemble average of $\boldsymbol{X}^a$. This inflation procedure is performed after the model state vector is updated. The covariance inflation suppresses filter divergence because it compensates for the variance reduction due to spurious correlations and other effects, which leads to underestimation of the ensemble variance. We tested the 1.01 and 1.02 as ρ_{i} for the covariance inflation in this study.

2.4 Investigation of data assimilation

In this study, we investigate the performance difference in terms of ensemble members, covariance inflation, and localization. We performed the data assimilation with several combinations for these values. The cases we investigated are listed in Table 3. Particularly, we were interested in the performance degradation of the cases with 42 and 98 members compared with the cases with 980 members. Ensemble members of 42, 98, and 980 are nearly the half, equal to and 10 times the maximum number of ensemble members used in previous CAT studies (100; Park and Kaneko, 2000; Lin et al., 2005). For the cases with 980 members, we assumed that the number of ensemble member is enough and that covariance localization was not required; thus, we considered only the three cases regarding covariance inflation.

The ensemble members were generated by adding noises to the amplitudes and phases of the five tidal constituents for both the tidal elevation and normal velocity at all open boundaries, following previous studies (Park and Kaneko, 2000; Lin et al., 2005; Zhu et al., 2017). The noises to the tidal elevation's five constituents were generated from the normal distributions with zero mean and standard deviation of 10% (15% or 20%) of M2 (S2 or K1, O1, and N2) constituent's amplitude and 2°, respectively. The corresponding amplitude of tidal elevation were 10.7, 7.1, 6.8, 4.4, and 4.4 cm for M2, N2, K1, O1, and N2 tidal constituents, respectively. We assumed that the normal velocities at open boundaries were more uncertain than the tidal elevation at the open boundaries. Thus, relatively large noises were added to the normal velocities at the open boundaries to generate ensemble members. The noises to the five tidal constituents of the normal velocity were generated from the normal distributions with zero mean and standard deviation of 10% (15% or 20%) of M2 (S2 or K1, O1, and N2) constituent's amplitude and 20° for semi-diurnal tides and 10° for diurnal tides, respectively. The corresponding amplitudes of the normal velocity were 12.6, 7.4, 4.2, 1.8, and 5.2 cm s^{-1} for M2, N2, K1, O1, and N2 tidal constituents, respectively. We also added

TABLE 3 Test cases of data assimilation.

	Number of ensemble	Covariance inflation	Covariance localization
case 1	42, 98, 980	1 (no inflation)	no
case 2	42, 98, 980	1.01	no
case 3	42, 98, 980	1.02	no
case 4	42, 98	1	4 km
case 5	42, 98	1	10 km
case 6	42, 98	1.01	4 km
case 7	42, 98	1.01	10 km

noises to Manning's roughness coefficient (n) and the lateral kinematic eddy viscosity coefficient (A_H). The noises to n and A_H were generated from the normal distributions with zero mean and standard deviation of 0.0002 and 0.4 m s^{-2}, respectively.

The velocity fields obtained from the data assimilation with the test cases in Table 3 were compared with the velocities obtained from the ADCP observations performed on October 30 and 31. Thus, we focused on assimilation results on October 30 and 31 in this paper. The model integration started at 00:00 on October 25, 2020, with no motion as the initial condition. The data assimilation step started at 16:00 on October 28, and the model state update was repeated every two minutes until 0:00 on November 1. The numerical model's time integration in the first time step after updating state vector was replaced with the Euler method.

The comparison with the ADCP results was evaluated by correlation coefficient, root-mean-squared difference (RMSD), and fraction error variance (FEV). The definitions of the RMSD and fractional error variance are

$$\mathrm{RMSD}_U = \sqrt{\frac{1}{K}\sum_i^K (U_i^{\mathrm{DA}} - U_i^{\mathrm{ADCP}})^2}, \tag{16}$$

$$\mathrm{RMSD}_V = \sqrt{\frac{1}{K}\sum_i^K (V_i^{\mathrm{DA}} - V_i^{\mathrm{ADCP}})^2}, \tag{17}$$

$$\mathrm{FEV} = \frac{\sum_i^K |(U_i^{DA} - U_i^{ADCP}) + j(V_i^{DA} - V_i^{ADCP})|^2}{\sum_i^K |U_i^{ADCP} + jV_i^{ADCP}|^2}, \tag{18}$$

where $j \times j = -1$, K is the number of ADCP observation data, and $|\cdot|$ indicates computing the absolute value. For these comparisons, the data assimilation results were spatiotemporaly interpolated to obtain the velocity at the same locations and times as each ADCP velocity.

3 Results

The EnKF updates were performed with the cases listed in Table 3, and the resulting time series of velocity field was compared with the ADCP observed velocity. The results of the model prediction without the data assimilation was also compared with the ADCP results. Table 4 summarizes the results of the comparisons. The correlation coefficients of eastward and northward velocity between the model prediction and the ADCP results are 0.95 and 0.68, respectively. The RMSDs (RMSD$_U$ and RMSD$_V$) are 0.29 and 0.20 m s^{-1}, and the fractional error variance is 19.6%. Although we implemented the information obtained from the observed path-averaged current into the open boundary condition in the numerical ocean model, there still were considerable differences from the ADCP results; thus, data assimilation is necessary.

The correlation coefficients of current velocity between the data assimilation and ADCP results are high. In particular, the correlation coefficients of the eastward current (r_U) are all 0.97. The correlation coefficients of the northward current is slightly lower than that of the eastward current but are still over 0.75. The RMSD is below 0.2 m s^{-1} and the FEVs are about 8–11%. Implementing the covariance inflation improve the results compared with the results without the covariance inflation for all the tested ensemble member (test cases 2 and 3). On the other hand, implementing the covariance localization did not always reduce the RMSD or FEV. Implementing both covariance inflation and localization improved but was not always better than the results with covariance inflation only. There seems no clear difference between the results with covariance localization with different localization radii (cases 4 and 5, or 6 and 7). The test cases 2 and 3 with 980 ensemble members resulted in the smallest difference with reference to the ADCP results, but some cases with smaller ensemble member resulted in the similarly small differences (case 3 with 42 members and case 2 and 3 with 98 members).

Figure 4 shows the results with 980 ensemble members: the path-averaged currents predicted from the data assimilation results (left column) and the standard deviation of the path-averaged current ensembles (i.e., forecast ensemble spread; right column) for the five transmission paths. In general, the data assimilation results reproduced the observed path-averaged current patterns (black). However, the levels of fit are different between the test cases. Implementing the covariance inflation (red and yellow) and using the larger values (yellow) reproduced the path-averaged currents closer to the observed values. The case 2 with ρ_i=1.01 resulted in smaller FEV (7.9%) than that of case 3 (8.2%); however, in terms of the reproduced path-averaged currents, using ρ_i=1.02 may be desirable if one is interested in reproducing such high-frequency variation. The results without the covariance inflation (blue) could not reproduce some of the high-frequency variation, which might be under a filter divergence condition. The ensemble spread in the results with ρ_i=1 is nearly equal or somewhat larger than the assumed observation noise covariance (diagonal matrix with 0.025^2 m^2 s^{-2}). This indicates that even using large number of ensemble members can results in filter divergence if updating the state vector so often (2 min in this study); that is, the covariance inflation is necessary. The ensemble spread effectively enlarged by using the covariance

TABLE 4 Summary of the comparison with ADCP observed currents for all the test cases listed in Table 3.

		r_U	r_V	RMSD_U (m/s)	RMSD_V (m/s)	FEV (%)
model		0.95	0.68	0.29	0.20	19.6
42 members	case 1	0.97	0.75	0.20	0.17	11.3
	case 2	0.97	0.82	0.19	0.15	9.4
	case 3	0.97	0.86	0.19	0.13	8.5
	case 4	0.97	0.76	0.20	0.17	10.9
	case 5	0.97	0.78	0.20	0.16	10.9
	case 6	0.97	0.83	0.19	0.14	8.8
	case 7	0.97	0.82	0.19	0.15	9.3
98 members	case 1	0.97	0.79	0.20	0.16	10.3
	case 2	0.97	0.85	0.18	0.14	8.2
	case 3	0.97	0.88	0.20	0.12	8.5
	case 4	0.97	0.78	0.20	0.16	10.2
	case 5	0.97	0.76	0.20	0.17	10.8
	case 6	0.97	0.85	0.19	0.13	8.5
	case 7	0.97	0.85	0.19	0.14	8.5
	case 1	0.97	0.78	0.19	0.16	10.0
980 members	case 2	0.97	0.86	0.18	0.13	7.9
	case 3	0.97	0.88	0.19	0.12	8.2

inflation, in particular for the paths S1S2 and S3S4 (panel B1 and B5). For the path between S2 and S3, the path-averaged current observation failed during low tides because of the shallow bank near station S3 (Taniguchi et al., 2021b). During no observation, using ρ_i=1.02 caused relatively large ensemble spread (panel B3).

Figures 5, 6 shows the same as Figure 4 but for the selected results (cases 1, 3, and 4) obtained from the data assimilation with 42 and 98 ensemble members. In the results with the 42 and 98 ensemble members, the path-averaged current reproduced from the data assimilation results agreed with the observed currents as is the case with the results of 980 ensemble members. The cases with covariance inflation (red) closely fit to the observed currents. The results with the covariance localization (yellow) did not improve the results much. It is also found that the difference between the data assimilation results and observed path-averaged current is smaller in the results with 980 ensemble members (gray in Figures 5A, 6A) than in 42 or 98 members.

The ensemble spread in the results with 42 and 98 members shows that covariance inflation is also effective to enlarge the ensemble spread and to suppress filter divergence. The covariance localization does not always enlarge the ensemble spread; it is effective for the S2S3 and S2S4 paths (panels B3 and B4 in Figures 5, 6). For these paths, the ensemble spread of the test case 4 becomes sufficiently large during some specific periods. On the other hand, for the S1S2 and S3S4 paths, the ensemble spread is nearly equal to that of the results without both covariance localization and inflation. In the results with 42 and 98 members, it founds that the ensemble spread more fluctuates compared with the results with 980 members and even with the same covariance inflation value. The resulting ensemble spreads vary differently between the 98 (or 42) and 980 members.

The ADCP observed velocities and data assimilation results (case 3 with 98 ensemble members) are compared in Figure 7. Although only the results from one test case are shown in the figure, the results of other cases such as cases 2 and 3 with 980 ensemble members reproduced the nearly equivalent patterns of velocity fields. In the figure, the model predicted velocities are also compared with those of ADCP. While the model predicted velocities seems to reproduce the velocity patterns, but the discrepancy from the ADCP results are clearly found. In particular, the model prediction overestimates the velocity magnitude during the period with strong current. On the other hand, the data assimilation can reproduce the ADCP observed velocity patterns even the sharp spatial variation (e.g., the transects at 08:00 and 09:00 in panel B1). However, the velocity magnitude of data assimilation still slightly larger than that from ADCP observation during strong currents; the model's overestimation was not perfectly corrected in the data assimilation results. The discrepancies between the data assimilation and ADCP results are also found near the coastlines (e.g., near the S4 station in the transects at 08:00 and 09:00 in panel B2) or around the curved flow during the period with weak currents (e.g., the transect at 12:00 in panel B3). There will be several reasons for this discrepancies between the data assimilation results and ADCP results during the weak currents. First, such weak current appears near the coastline, and the performance of the model (and assimilation too) may be restricted near the coastline because of the limited spatial resolution of the model. Also, there is no observation of path-averaged current along the coasts. Second, signal-to-noise ratio between the magnitude of tidal currents and assigned measurement noise become small during weak currents. Finally, tomographic path-averaged data are good at regularizing the red spectrum or fields with low wave number (e.g., Cornuelle and Worcester, 1996) like a nearly uniform flow during the strong current in the observation site, while the fields might be complex with vortex-shaped structures during weak currents (at slack water).

In the results of our preliminary experiment along the S1 and S2 stations (Taniguchi et al., 2021a), the observed path-averaged current exhibited high-frequency oscillation with a period of about 1.5 h during the flood current. We can clarify the velocity fields causing those high-frequency variations with the data assimilation results.

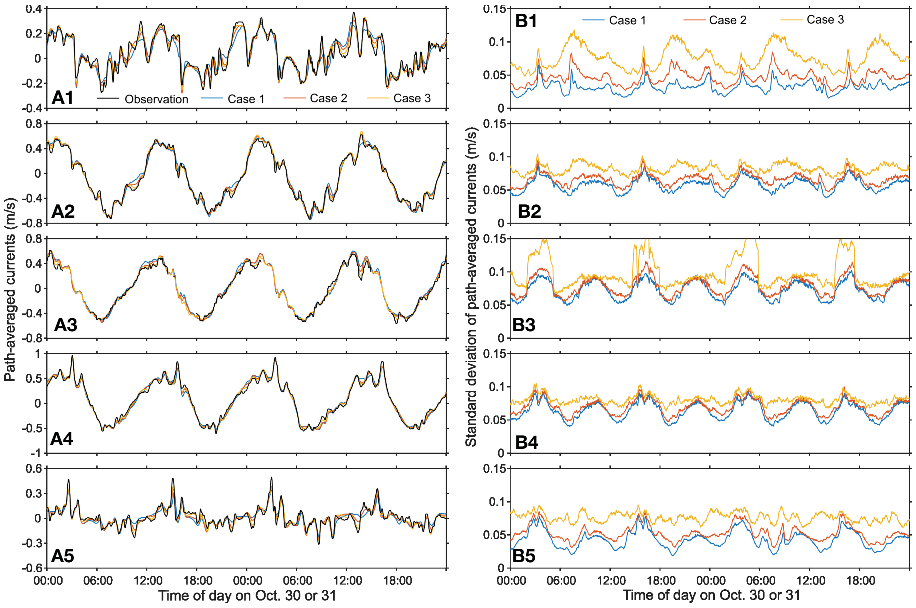


FIGURE 4

Results of data assimilation with 980 ensemble members. **(A17–A5)** Path-averaged currents along the station pairs S1S2, S1S3, S2S3, S2S4, and S3S4; **(B1–B5)** The standard deviation of path-averaged current ensembles (i.e., the forecast ensemble spread) just before the Kalman filter update in each transmission time for the same station pairs as of **(A1–A5)**. The blue, red, and yellow lines are the results of cases 1, 2, and 3, respectively, and the black line in **(A1–A5)** indicates the observed path-averaged current.

Figure 8 shows the time series of path-averaged currents along the S1 and S2 stations and six snapshots of the velocity fields during the periods with high frequency oscillation. The result used in the figure is from case 3 with 98 members. The observed and modeled results are also drawn in the path-averaged current's time series plot. The data assimilation results reproduced the high-frequency variation in the observed path-averaged current (red and blue). The reproduced velocity snapshots exhibit the occurrence of a vortex at the island's downstream (east) side. When the clockwise vortex exists behind the island, the path-averaged current exhibits a markedly sharp trend toward the S1 direction (at 7:30 and 9:00; panels C and F). On the other hand, when there is no clockwise vortex or an anticlockwise vortex behind the island, the path-averaged current has a trend in the direction of the S2 station (panels D, E, and G). That is, the high-frequency variation in the path-averaged current along the S1 and S2 stations during the flood tide is associated with the two vortex

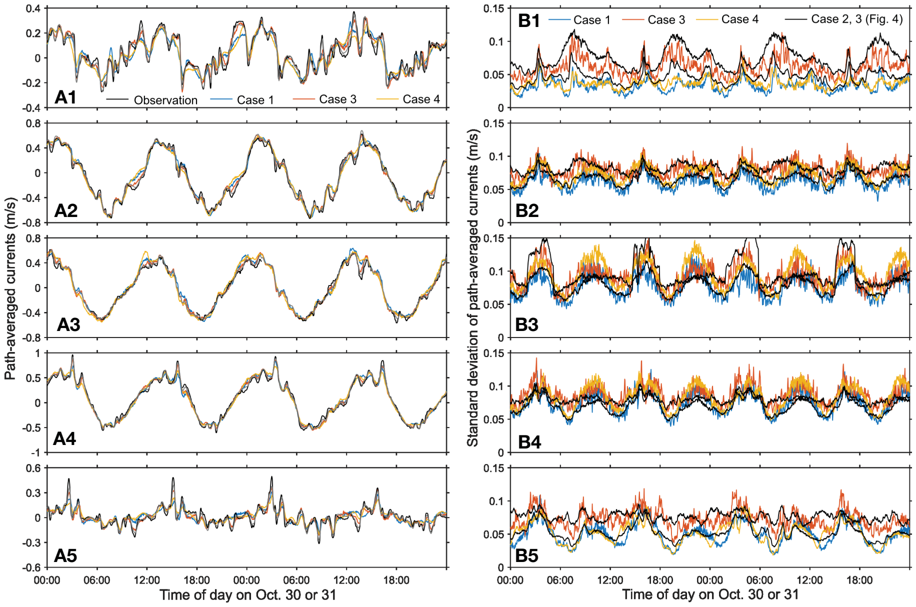


FIGURE 5

Same as Figure 3 but for the results of case 1 (blue), 3 (red), and 4 (yellow) with 42 ensemble members. The gray lines in panels **(A1-A5)** are the results of case 3 with 980 ensemble members. In panels **(B1-B5)**, the two black lines are the results of case 2 and 3 with 980 ensemble members.

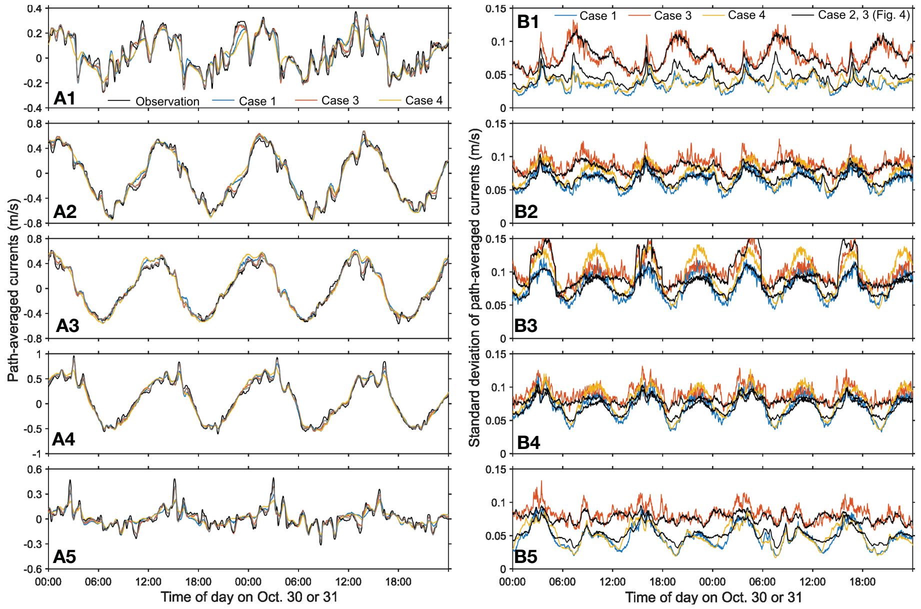


FIGURE 6
Same as Figure 4 but for the results of 98 ensemble members.

generations behind the island, i.e., due to the island wake. Note that the velocity fields at a northern part of the S1S2 path is nearly eastward and perpendicular to the path direction; thus, background eastward current less affect the variation in the S1S2 path-averaged current during the period. The numerical ocean model without data assimilation partly reproduced such island wake (the velocity fields are not shown here), but the occurrence timing and extent of the vortex were not well reproduced as we can expect from the model predicted path-averaged currents (the gray line in panel A).

In our previous report on this experiment, we showed that there were spike-like variations in the path-averaged currents between the S2 and S4 stations and the S3 and S4 stations with a slight time lag (Figure 4 in Taniguchi et al., 2021b). Such spike-like variation is found in almost every ebb tide. The velocity fields causing those spike-like variations are focused on in Figure 9. The model prediction did not reproduce the spike-like variation in the path-averaged current in the period shown in the figure but predicted several spike-like variation in other periods. The data assimilation result (case 3 with 98 ensemble members) reproduced the spike-like variation in the path-averaged currents (panels A1 and A2). Those spike-like variation may be associated with a strong current from the northeastern channel. During the period shown in Figures 9B, C, the relatively strong flow from the northeastern channel cross the S3S4 path with a slight angle; thus, the S3S4 path-averaged current shows rapid velocity variation toward negative value (toward the S3 station) at that time. During the same period, the flow is almost westward around the S2S4 path; thus, the path-averaged current between the S2 and S4 stations exhibits relatively large values. Then the area of the flow from the southeastern channel broadens northward (panels D and E). This variation cancels the contribution of the flow from the northeastern and southeastern channels to the path-averaged current along the S3S4 path, resulting in a path-averaged current with nearly zero. At the same time, strong flow from the northeastern channel reaches near the S2S4 path with the flow direction parallel to the path, causing further large (in the direction to S2 station) path-averaged current along the S2 and S4 stations. After that, the path-averaged current along both paths becomes nearly zero due to slack water (panel G). The path-averaged currents contain velocity information on both the magnitude and direction, and thus, the path-averaged current express rapid variation when both the magnitude and direction simultaneously cause the variation in the same sign (positively or negatively). The reproduced current velocity fields also exhibit markedly curved flow around the path between the S2 and S3 stations during the last half period (panels E and F); however, the period corresponds to low tide (Figure 3), and the transmission between the S2 and S3 station failed due to shallow bank near the S3 station. Thus, the velocity patterns near the S2S3 path can be somewhat uncertain. Applying data assimilation with successful five path-averaged currents will better clarify the velocity field time variation related to the spike-like variation in the path-averaged currents.

4 Discussion

The performance of EnKF with the several combination of covariance localization and inflation, and the number of ensemble member are investigated *via* the comparison with the ADCP results. These factors relate with the issues introduced by ensemble approximation of covariance matrix in Kalman filter: spurious correlation and filter divergence. These factors have not always been considered in the previous CAT studies (Park and Kaneko, 2000; Lin et al., 2005; Zhu et al., 2017). This may be because, in these studies (and in this study), the ensemble members are generated by adding noises to the open boundary forcing. Velocities around the open boundary are always different over ensemble members, and thus, permanent filter divergence do not occur, leading to somewhat untouched theme in the CAT data assimilation. However, as shown in this paper, covariance inflation, which spread ensembles after each updates, improves both the level of fit to the observed path-averaged currents and agreement with the ADCP

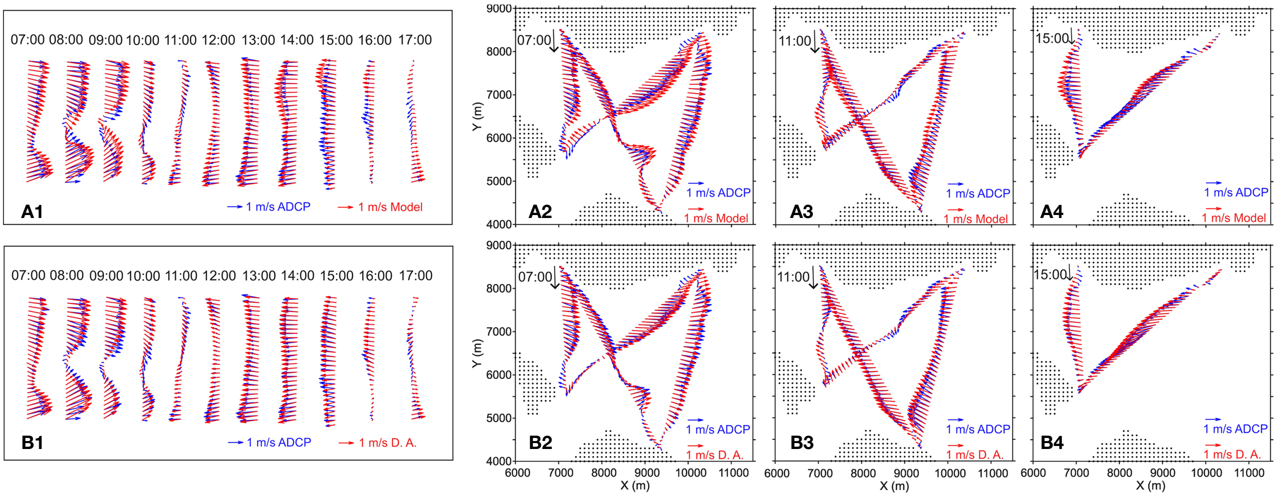


FIGURE 7
Comparison results of model prediction (panels **A**; first row) and data assimilation (panels **B**; second row) with the ADCP results. Panels **(A1, B1)** are the comparisons with the ADCP results on Oct. 30 (along the dashed line in Figure 1C), and panels **(A2–A4, B2–B4)** are the comparisons with the ADCP results on Oct. 31 (along the dotted lines in Figure 1C). The data assimilation result is obtained from case 3 with 98 ensembles.

result. In the present study, the constant inflation value of 1.01 or 1.02 worked and improved the results. With the covariance inflation, the difference of the results was minor between 98 and 980 ensemble members regarding the comparisons with the ADCP results. One note is that the numerical experiments shown in Table 4 were from the results of one realization for each test case. As the results (Figures 4–6) indicates, the results may depend on the noise added to the observed data (i.e., to make perturbed observation) and boundary condition, and the sensitivity to the added noise are larger with smaller ensemble members. Thus, the results in Table 4 may slightly change with different ensemble realizations; however, the superiority with covariance inflation and with larger ensemble members will be retained.

Low quality of assimilation results without covariance inflation is due to frequent update by observation with relatively small observational error (0.025 m s^{-1}). With frequent Kalman updates, ensembles may not spread sufficiently between updates and the next updates, leading to temporary filter divergence in some period. Thus, covariance inflation is required to inflate the analysis ensemble even with many ensembles. This indicates that the time interval between the Kalman updates may be extended although we conducted the acoustic transmission and the Kalman updates every two minutes. Note that however, there is flow patterns that cause spike-like or triangle-shape variation in the path-averaged currents (Taniguchi et al., 2021a; Taniguchi et al., 2021b); thus, one still needs high-frequent transmissions to measure (record) the path-averaged currents as references. Also, sufficiently frequent measurements and updates can recursively pull the ensembles toward the measured solution and do not allow the ensembles to diverge (to bimodal) even in strongly nonlinear dynamics (Evensen et al., 2022).

The comparisons with ADCP results suggest that covariance localization is not required if one focuses only inside the tomographic array. Here, we consider the performance differences in terms of spatial structure of covariance. Figure 10 shows the spatial map of ensemble correlation associated with cross-covariance $\widetilde{\boldsymbol{E}}_k(\boldsymbol{H}\widetilde{\boldsymbol{E}}_k)^{\mathrm{T}}$ in Eq. (13) (i.e., cross-covariance between the velocity field ensembles and path-averaged current ensembles) for eastward (*U*) and northward (*V*) components at a specific time of 9:00 on October 31, 2020. The results of the test case 3 with ensemble members of 980, 98, and 42, and the test case 4 with 42 ensemble members are shown. The results with 980 ensemble members (the first rows of panels A and B) shows high-correlation patches along the paths. In addition, there are weak correlations between each path-averaged current and velocity fields throughout the model domain. The results with 980 ensemble members had less spurious correlation because of sufficient number of ensemble members; thus, these weak spatial-correlations are physical correlations of tidal currents. Also, some of these spatial-correlation is partly explained by the characteristic of integral observations (path-averaged currents). As van Leeuwen (2019) demonstrated, non-local measurements (i.e., path-averaged currents) can influence distant state variables, which are not physically connected in the prior covariance but become connected *via* the non-local observation. The velocity fields throughout the model domain including those at open boundaries are corrected by the data assimilation *via* these correlation, and it is reasonable that the covariance localization is not required for this case. The results with 98 ensemble members (the second rows of panels A and B) show similar spatial structures of correlation but with slightly complex and higher (positively or negatively) correlation values, in particular, at outside the area covered by the reciprocal acoustic transmission. The results with 42 ensemble members (the third rows of panels A and B) show further spatially high-correlations. These correlations higher than that in the results of 980 ensemble members would be spurious correlation associated with insufficient ensemble members and sampling errors. However, if we focus on the area covered by reciprocal acoustic transmissions, the spatial patterns are retained even in the results with 42 ensemble members, resulting in similar performances between them (FEVs of 8.2, 8.5, and 8.5, respectively, when 980, 98, and 42 ensemble member were used). Short interval of Kalman update (2 minutes) is also responsible for the nearly equal performances.

Figure 10 also shows the results of case 4 with 42 ensemble members (the fourth rows of panels A and B), which implemented the covariance localization with a localization radius of 4 km. The results with 42 ensemble members with covariance localization shows further high-correlation. This is because Kalman updates are limited to the

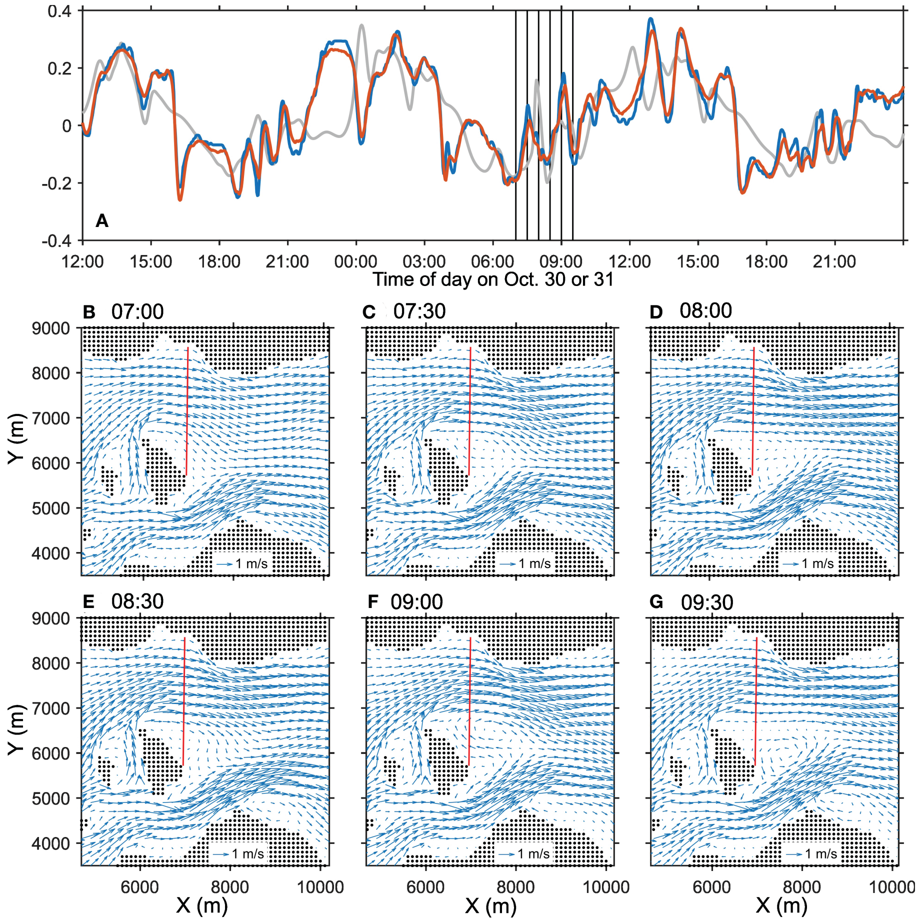


FIGURE 8
(A) Time series of path-averaged current between the S1 and S2 stations. The blue, red, and gray lines indicate the reciprocal acoustic transmission results, data assimilation results (case 3 with 98 members), and the model predictions, respectively. The positive value of the path-averaged currents corresponds to the flow toward the S1 station. Panels **(B-G)** show the snapshots of velocity fields at times shown as the vertical bars in **(A)**. The red line in panels **(B-G)** is the path connecting the S1 and S2 stations.

center of the observation area and the model ensemble predictions are somewhat retained even along the reciprocal acoustic transmission paths. It is also noted that the small spatial structures of correlation found in the results without covariance localization (from the first to third rows) are weaken in the results with 42 ensemble members with covariance localization. It is naturally expected that EnKF does not properly updates the velocity fields with the structures found in the results without the covariance localization. In fact, Cornuelle and Worcester (1996) used simple numerical experiments and showed that the data assimilation with integral observations requires retaining full information in a model covariance matrix, i.e., off-diagonal components in model covariance matrix is important. Since the covariance localization, with a short localization radius in particular, quickly decreases the off-diagonal terms (as well as diagonal terms far from the observation location) to zeros, it makes Kalman filter not optimal scheme for integral data such as path-averaged currents. However, if we need to run the EnKF with further smaller number of ensemble members such as 20 (Zhu et al., 2017), spurious correlation will be more significant and affect the assimilation results. For such case, the covariance localization can be used to mitigate spurious correlation. When we performed the same numerical experiments with 20 ensemble members, the fractional error variance (FEV) were 15.6%, 10.2%, 10.0%, 11.6%, 12.0%, 9.9%, and 11.2% for the Case 1 to Case 7, respectively.

The velocity fields reproduced by the data assimilation revealed the reasons for the high-frequency or spike-like variations in the observed path-averaged currents (Taniguchi et al., 2021a; Taniguchi et al., 2021b). The high-frequency variation in the path-averaged current along the S1 and S2 stations is associated with the island wake. The data assimilation results revealed the repeated occurrences of clockwise and anticlockwise vortexes behind the island. It is noted that the numerical model potentially included the responsible mechanisms to reproduce the island wakes (the gray line in Figure 7). However, it would be unable to reproduce the island wake precisely by the numerical model alone with limited open boundary condition (only five tidal constituents with constant velocity along the boundary and no tangential velocity), because the extents and timings of wake generations are complex and vary with every flood tides, as expected by the time series plots of the S1S2 path-averaged current (Figure 8 and Figure 8 in Taniguchi et al., 2021a). The island wake is a research topic in the geophysical fluid dynamics over a long period, and even today, there have been studies associated with the island wake (e.g., Chang et al., 2013; Chang et al., 2019). The period of wake evolution T is related to the Strouhal number $S_t=D/UT$

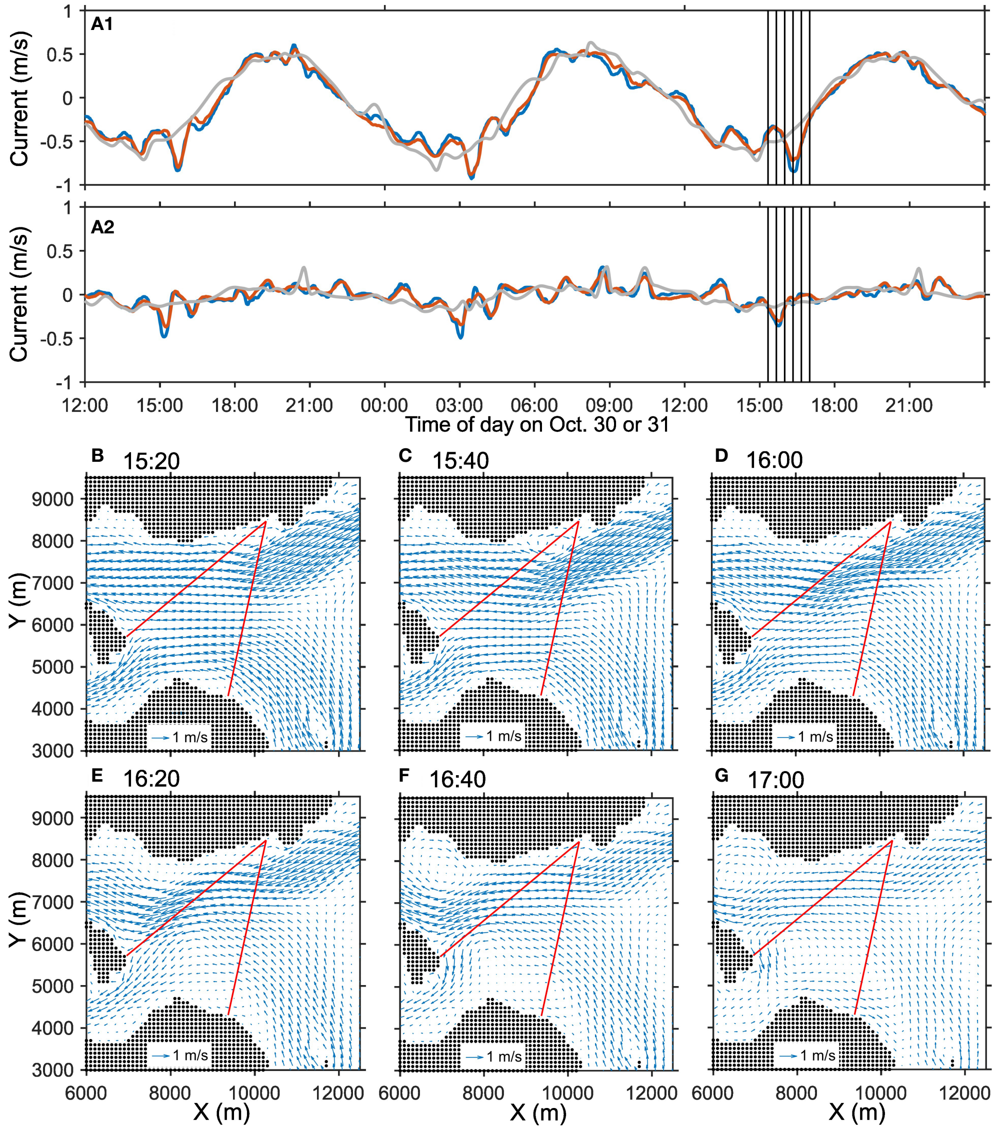


FIGURE 9
(A1, A2) time series of path-averaged current between the S2 and S4 stations and the S3 and S4 stations. The blue, red, and gray lines indicate the reciprocal acoustic transmission results, data assimilation results (case 3 with 98 members), and the model predictions, respectively. The positive value of the path-averaged currents corresponds to the flow toward the S4 station for both station pairs. Panels **(B-G)** show the snapshots of velocity fields at times shown as the vertical bars in **(A)**. The red lines in panels **(B-G)** are the paths connecting the S2 and S4 stations and the S3 and S4 stations.

where D and U are the characteristic length of the island and the reference velocity. With approximate values of D=1.5 km (the north-south length), U=1.5 m s^{-1}, and T=1.5 h, we obtain S_t=0.185, which is close to the rough value of S_t (about 0.2) with a range of Reynolds number 200–2×10^5. If we measure the vortex shedding period T and estimate the flow velocity, it would be regarded as a real-field application of the vortex flowmeter. The high-frequency variation in the S1S2 path-averaged current is not clear in the path-averaged current along the S3S4 path, which contain the northward velocity component same as the S1S2 path; thus, at this moment, it is expected that the vortexes quickly decay or corrupt. The data assimilation results also do not contain clear vortex street. Additional observation to sample the vortex in various angles will be required to reproduce the vortex generation, shape, downstream movement, and/or corruption more precisely. The island wake occurs almost every flood tide but only during a short period. For such a short observation period, a moving ship acoustic tomography (Huang et al., 2019) may be effectively used to improve the reproduction performance.

5 Summary and concluding remarks

In this paper, we investigated the performance of CAT data assimilation by applying EnKF to the path-averaged currents obtained from the 2020 reciprocal acoustic transmission experiment with four acoustic stations deployed at Mihara-Seto in the Seto Inland Sea, Japan. The results of EnKF with several combinations of the values of covariance localization, inflation, and the number of ensemble members were compared with the ADCP results. The results showed that data assimilation with EnKF improved the velocity reproduction compared with the model prediction and that

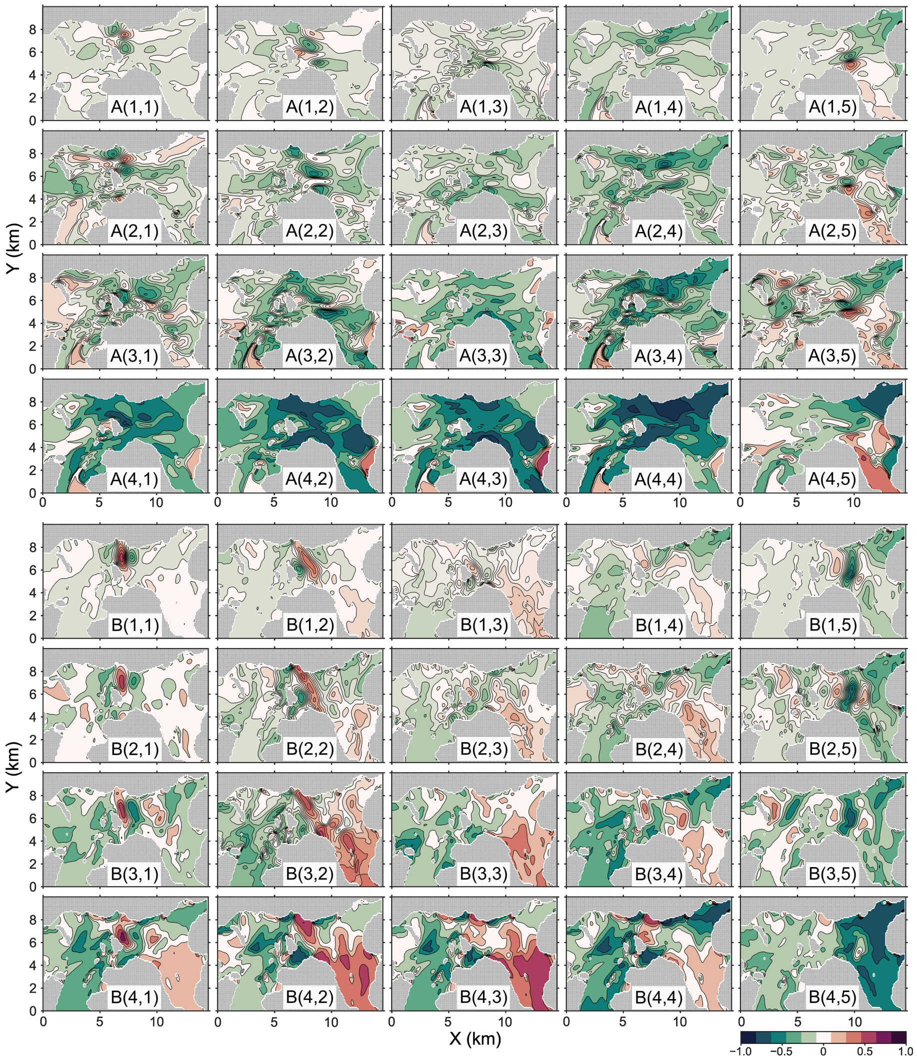


FIGURE 10
Spatial map of ensemble correlation associated with cross-covariance $\tilde{\boldsymbol{E}}_k(\boldsymbol{H}\tilde{\boldsymbol{E}}_k)^{\mathrm{T}}$ in Eq. (13) for eastward (*U*; panels **A**) and northward (*V*; panels **B**) components at a specific time of 9:00 on October 31, 2020. In each panel **(A, B)**, the first, second, and third rows show the results of case 3 with 980, 98, and 42 ensemble members, and the fourth row shows the result of case 4 with 42 ensemble members (see Table 3 for the test case). The column number indicates the station pair of the path-averaged currents: S1S2, S1S3, S2S3, S2S4, and S3S4, from the left to right column.

implementing covariance inflation contributed to additional improvements for all the tested ensemble members. Thus, it is suggested that the covariance inflation should be implemented in CAT data assimilation with EnKF, in particular, when EnKF updates are performed frequently. Focusing on the velocity fields inside the tomographic array, we obtained nearly equivalent performance of EnKF over 98 and 980 ensemble members with covariance inflation. The covariance localization did not improve the results for the number of ensemble members of 42, 98 and 980.

In this study, we implemented EnKF as the data assimilation scheme for path-averaged currents, but this does not mean that EnKF always provides the best estimates. Rather, in our study, we purpose developing a near-real-time (nowcasting) tidal current reproducing system, and EnKF is suitable to implement as the term *filter* indicates (i.e., not the smoother). If the purpose is hindcasting, then ensemble Kalman smoother may provide better estimates (Evensen, 2009). Also, we did not explore the applicability of other data assimilation schemes such as square-root-filter, which would provide better performance when the ensemble member is small.

With the above-mentioned limitation, the obtained comparison results provide some guides when CAT data assimilation with EnKF is applied. The reproduced velocity fields agreed with the ADCP results; the fractional error variance is about 8%. Additional improvements would be possible by successful transmissions over all the five paths. The data assimilation results also revealed the reason for the high-frequency variation in the path-averaged currents found in the previous reciprocal acoustic transmission experiment conducted in the same area (Taniguchi et al., 2021a). The repeated generation (but only two or three) of vortexes during the flood tide at the downstream side of the island (i.e., island wakes) caused high-frequency variation in the path-averaged current. This, in turn, indicates that the reciprocal acoustic transmission and the resulting path-averaged current can be used to effectively detect temporal variation in the velocity fields associated with island wakes. Thus, while previous CAT studies have showed its usefulness, this study is an additional example demonstrating the usefulness of CAT as an observational method for the studies on coastal physical processes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

NT, HM, and TT contributed to conception and design of the study. NT, TT, KY, HY, YS, and KH provided instrument resources and/or performed field observation, contributed to data acquisition. NT, MA, and HM contributed model development and validation. NT, TT, and KH contributed to data analysis methodology. NT wrote the first draft of the manuscript. HM, MA, YS, and TT wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

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. The Sentinel-2 image in Figure 1 was downloaded from the EO browser (https://www.sentinel-hub.com/explore/eobrowser/).

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

Anderson, J. L., and Anderson, S. L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Rev.* 127, 2741–2758. doi: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Blumberg, A. F., and Mellor, G. L. (1987). A description of a three-dimensional coastal ocean circulation model. In *Three-Dimensional Coastal Ocean Models*. N. S. Heaps Ed. 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)126<1719:ASITEK>2.0.CO;2

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

Chang, M.-H., Jan, S., Liu, C.-L., Cheng, Y.-H., and Mensah, V. (2019). Observations of island wakes at high rossby numbers: Evolution of submesoscale vortices and free shear layers. *J. Phys. Oceanogr.* 49, 2997–3016. doi: 10.1175/JPO-D-19-0035.1

Chang, M.-H., Tang, T. Y., Ho, C.-R., and Chao, S.-Y. (2013). Kuroshio-induced wake in the lee of green island off taiwan. *J. Geophys. Res.: Oceans* 118, 1508–1519. doi: 10.1002/jgrc.20151

Chen, H., Chen, H., Zhang, Y., and Xu, W. (2021). Decentralized estimation of ocean current field using underwater acoustic sensor networks. *J. Acoustical Soc. America* 149, 3106–3121. doi: 10.1121/10.0004795

Chen, M., Kaneko, A., Lin, J., and Zhang, C. (2017). Mapping of a typhoon-driven coastal upwelling by assimilating coastal acoustic tomography data. *J. Geophys. Res.: Oceans* 122, 7822–7837. doi: 10.1002/2017JC012812

Chen, Y., and Oliver, D. S. (2017). Localization and regularization for iterative ensemble smoothers. *Comput. Geosci.* 21, 13–30. doi: 10.1007/s10596-016-9599-7

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

Cornuelle, B., Munk, W., and Worcester, P. (1989). Ocean acoustic tomography from ships. *J. Geophys. Res.* 94 (C5), 6232–6250. doi: 10.1029/JC094iC05p06232

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: 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

Evensen, G. (2009). *Data assimilation: The ensemble kalman filter. earth and environmental science* (Berlin Heidelberg: Springer).

Evensen, G., Vossepoel, F., and Leeuwen, P. J. (2022). *Data Assimilation Fundamentals*. Cham: Springer International Publishing. p. 245. doi: 10.1007/978-3-030-96709-3

Gaspari, G., and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* 125, 723–757. doi: 10.1002/qj.49712555417

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Rev.* 129, 2776–2790. doi: 10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Howe, B. M., Worcester, P. F., and Spindel, R. C. (1987). Ocean acoustic tomography: Mesoscale velocity. *J. Geophys. Res.: 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

Kawanisi, K., Zhu, X.-H., Fan, X., and Nistor, I. (2017). Monitoring tidal bores using acoustic tomography system. *J. Coast. Res.* 33, 96–104. doi: 10.2112/JCOASTRES-D-15-00172.1

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. *Geophys. Res. Lett.* 32, L14607. doi: 10.1029/2005GL022914

Munk, W., Worcester, P., and Wunsch, C. (1995). *Ocean acoustic tomography. Cambridge monographs on mechanics* (Cambridge: 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. Oceanogr. Res. Papers* 26, 123–161. doi: 10.1016/0198-0149(79)90073-6

Ogasawara, H., and Mori, K. (2016). Acoustical environment measurement at a very shallow port: Trial case in hashirimizu port. *Japanese J. Appl. Phys.* 55, 07KE17. doi: 10.7567/jjap.55.07ke17

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

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

Trump, C. L., and Marmorino, G. O. (1997). Calibrating a gyrocompass using ADCP and DGPS data. *J. Atmospheric Oceanic Technol.* 14, 211–214. doi: 10.1175/1520-0426(1997)014<0211:CAGUAA>2.0.CO;2

van Leeuwen, P. J. (2019). Non-local observations and information transfer in data assimilation. *Front. Appl. Mathematics Stat* 5. doi: 10.3389/fams.2019.00048

Wang, T., Zhang, Y., Yang, T. C., Chen, H., and Xu, W. (2018). Physics-based coastal current tomographic tracking using a kalman filter. *J. Acoustical Soc. Am.* 143, 2938–2953. doi: 10.1121/1.5036755

Williams, P. D. (2009). A proposed modification to the Robertâ€"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

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. 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

Zhu, Z.-N., Zhu, X.-H., Guan, W., Zhang, C., Chen, M., Liu, Z.-J., et al (2022). Synchronous Assimilation of Tidal Current-Related Data Obtained Using Coastal Acoustic Tomography and High-Frequency Radar in the Xiangshan Bay, China. *Remote Sens.* 14, 3235. doi: 10.3390/rs14133235

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fmars.2023.1113224

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Xin Qing,
Harbin Engineering University, China
Haocai Huang,
Zhejiang University, China

*CORRESPONDENCE
Rongbin Lin
✉ rblin@xmu.edu.cn

SPECIALTY SECTION
This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 01 December 2022
ACCEPTED 03 January 2023
PUBLISHED 07 February 2023

CITATION
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

Data augmentation and deep neural network classification based on ship radiated noise

Zhuofan Xie[1,2,3], Rongbin Lin[1,2,3,4]*, Lingzhe Wang[2],
Anmin Zhang[5], Jiaqing Lin[1] and Xiaoda Tang[6]

[1]Shenzhen Research Institute of Xiamen University, Shenzhen, China, [2]School of Informatics, Xiamen University, Xiamen, China, [3]School of Electronic Science and Engineering (National Model Microelectronics College), Xiamen University, Xiamen, China, [4]Key Laboratory of Southeast Coast Marine Information Intelligent perception and Application, Ministry of Natural Resources, Xiamen, China, [5]School Marine Science and Technology, Tianjin University, Tianjin, China, [6]Department of R&D, Xiamen Weilai Marine Science and Technology Ltd., Xiamen, Fujian, China

Introduction: Various types of ships sail at sea, and identifying maritime ship types through shipradiated noise is one of the tasks of ocean observation. The ocean environment is complex and changeable, such rapid environmental changes underline the difficulties of obtaining a huge amount of samples. Meanwhile, the length of each sample has a decisive influence on the classification results, but there is no universal sampling length selection standard.

Methods: This study proposes an effective framework for ship-radiated noise classification. The framework includes: i) A comprehensive judgment method based on multiple features for sample length selecting. ii) One-dimensional deep convolution generative adversarial network (1-DDCGAN) model to augment the training datasets for small sample problem. iii) One-dimensional convolution neural network (CNN) trained by generated data and real data for ship-radiated noise classification. On this basis, a onedimensional residual network (ResNet) is designed to improve classification accuracy.

Results: Experiments are performed to verify the proposed framework using public datasets. After data augmentation, statistical parameters are used to measure the similarity between the original samples and the generated samples. Then, the generated samples are integrated into the training set. The convergence speed of the network is clearly accelerated, and the classification accuracy is significantly improved in the one-dimensional CNN and ResNet.

Discussion: In this study, we propose an effective framework for the lack of scientific sample length selection and lack of sample number in the classification of ship-radiated noise, but there aret still some problems: high complexity, structural redundancy, poor adaptability, and so on. They are also long-standing problems in this field that needs to be solved urgently.

KEYWORDS

ship-radiated noise, data augmentation, generative adversarial network, deep learning, target classification

1 Introduction

Ship-radiated noise is the unavoidable noise emitted by ships during movement. Traditional classification for ship-radiated noise mainly relies on the artificial extraction of features, such as time, frequency and time-frequency domain characteristics Zhang et al. (2020a).

Ship-radiated noise is generally modelled as a composition of mechanical, hydrodynamic and propeller noise components Yan et al. (2021); Zhang and Yang (2022). Mechanical noise is generated by diesel engines, the generators and air condition units, and mainly comprises line spectra Zhang et al. (2020b). Hydrodynamic noise is a time-stationary signal with a continuous spectrum. Considered environmental noise, as well as multi-path and Doppler effects in the propagation process, which make ship-radiation noise an unstable process with nonlinear, non-Gaussian and non-stationary characteristics Zhang et al. (2022b).

With the development of noise reduction technology, the distinguishing features have been significantly weakened, which increases the difficulty of collecting data. Meanwhile, the sample size in existing publicly available dataset is small. Most researchers commonly use feature extraction algorithms to solve this problem Qi et al. (2020); Zhang et al. (2021). In this paper, on the basis of traditional feature extraction and data augmentation Ian et al. (2014), a new idea of noise data augmentation is proposed to deal with the problem of small sample sizes in ship-radiated noise classification.

Traditional approaches in ship-radiated noise classification with small sample size are based on an analysis of machine learning Li et al. (2019); Xie et al. (2020). The core steps are to extract the features of the signal, increase the differences between different samples, and use these features to identify or classify signals Pethiyagoda et al. (2018). In earlier works, researches classify the signal using amplitude and frequency content Yang et al. (2021). Based on the difficulties of effective features extracting, some researchers start to use artificial feature extraction techniques, such as principal component analysis Wei (2016).

In Li et al. (2019), Support vector machine (SVM) is used for ship classifying because its ability to perform non-linear classification. In Xie et al. (2020), permutation entropy is combined with the normalized correlation coefficient and enhanced variational mode decomposition, and an SVM is used to classify three types of ships. In Jiao et al. (2021), fluctuation-based dispersion entropy is combined with intrinsic time-scale decomposition, and SVM is used for classification. The common classification methods are compared in Table 1. GAN is proposed by Goodfellow et al. Ian et al. (2014) for solving the imbalanced data problem. It is used in many fields Arniriparian et al. (2018). In Atanackovic et al. (2020), authors use GAN to augment ship-radiated noise data. In this study, we augment ship-radiated noise data using GAN models to obtain sufficiently similar and diverse samples.

With the development of neural networks in recent years, faster convergence times and higher accuracy have been achieved. This requires large amounts of balanced data to fully train the deep network structures. Some researchers have focused on manually extracted features combined with SVMs for ship-radiated noise classification Li and Yang (2021). If the number of samples is small or there is a big gap in the proportion of different samples, then there will be reduced classification accuracy Buda et al. (2018). When considering a one-dimensional speech signal, the length of the sample needs to be considered Zhang et al. (2022c). Under normal circumstances, collected ship-radiated noise is a piece of audio, which needs to be split, and thus the appropriate length of each sample needs to be determined. In this paper, we propose a comprehensive judgment method based on multiple features to select an appropriate sample length. The model used for data augmentation and the deep neural networks are designed with the selected sample length.

In Pan et al. (2019), two ways are proposed for the optimization GAN models: architecture optimization and objective function optimization. Convolutional neural networks (CNNs) are a kind of network that can adaptive extract the characteristics of the signal Zhang et al. (2022a). Moreover, there are also deep features extracted by convolution that cannot be easily extracted manually. For the above reasons, we choose convolution-based GANs for structural optimization. Convolution-based GANs are a development of CNNs, and DCGAN is one of the main models used Arniriparian et al. (2018). However, these GAN models are mostly used in the two-dimensional image processing domain. Based on DCGAN, in this paper we propose a one-dimensional DCGAN (1-DDCGAN) to generate ship-radiated noise samples.

In summary, in this paper, by using the public dataset ShipsEar Santos-Domínguez et al. (2016), we propose a data augmentation method with a comprehensive length select algorithm. Then, construct a deep neural framework for ship classification. The main contributions can be summarized as follows:

1. A new classification method is proposed, which is based on deep neural network classification combined with data augmentation. The proposed method is capable of performing accurate classification in the presence of small numbers of training samples.
2. We propose a decision-level fusion method to select sample length and combine 1-DDCGAN and ResNet for data augmentation and feature classification without manual feature extraction. The convergence time is reduced Xie et al. (2020) and the classification accuracy is improved Li et al. (2017); Li et al. (2019).
3. The performance of the proposed method is verified using experimental data Santos-Domínguez et al. (2016). First, data augmentation is carried out and the generated sample is mixed with the original in different proportions to increase the number of sample in training sets. From using only the original sample to a ratio of the original sample to the generated sample being 1:2, the classification rate in CNN improved from 79% to 90%, while the classification rate in ResNet improved from 87.5% to 99.17%.

The overall approach is shown in Figure 1. The rest of this paper is organized as follows. In Section 2, we discuss cepstrum coefficients and entropy characteristics. Then, an SVM is used for classification in order to determine the appropriate sample length. A 1-DDCGAN is then designed and the generated augmentation data are analyzed in Section 3. Section 4 describes merging the real and generated data in different proportions, and the classification accuracy under these training sets is compared. Conclusions are provided in Section 5.

TABLE 1 Comparison of common ship-radiated noise classification methods.

Classification Method	Feature Engineering	Drawbacks
Statistical analyze	Time-frequency features	High complexity, low adaptability
Machine learning	Principle component analysis	High feature extraction difficulties
Deep learning	Neural network	Large demand for data volume

2 Sample length selection of ship-radiated noise

The cepstrum coefficients and entropy features reflect different aspects of the signal characteristics. In order to extract different aspects of the information contained in the signal, we select the appropriate sampling length according to the contribution of these characteristics in the classification process. The details are shown in Figure 2.

Mel-frequency cepstrum coefficient (MFCC)-based features reflect the timbre of the signal Noda et al. (2019), while Gammatone cepstrum coefficients (GFCC)-based features reflect the robustness, the degree of influence of data disturbance, noise, and outliers in the model Zhang et al. (2018). On the other hand, entropy-based features describe the extent of signal regularity, which can be useful for the classification of nonlinear non-stationary ship-radiated noise. These characteristics can be used to measure the amount of information contained in the sample from many aspects. Sample entropy measures the complexity of a time series by measuring the probability of encountering new patterns in the signal Richman et al. (2004). Specifically, dispersion entropy is more suitable for long signals and represents the amplitude characteristics of the signals. Multi-scale weighted permutation entropy can retain the amplitude information of the signal when calculating the sequential mode of the time series Shi et al. (2021). Since these features reflect different characteristics of signals, we selected them for comprehensive consideration and selection of an appropriate sample length of ship-radiated noise. Meanwhile, when the GAN is used for sample

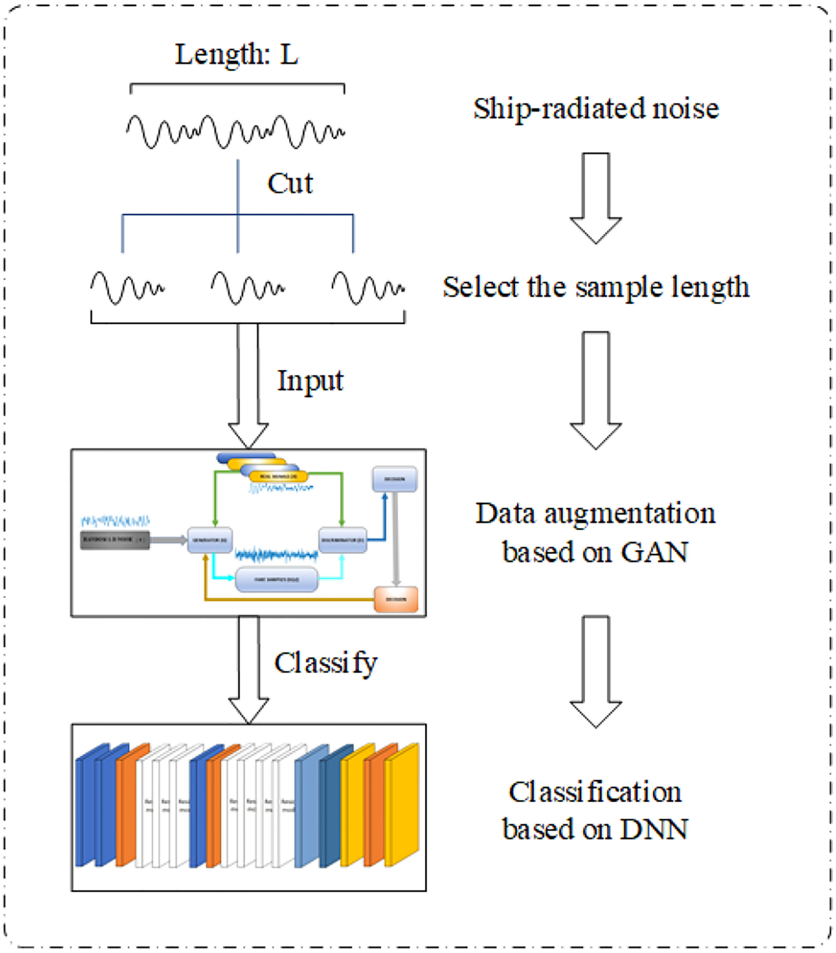


FIGURE 1
General algorithm framework.

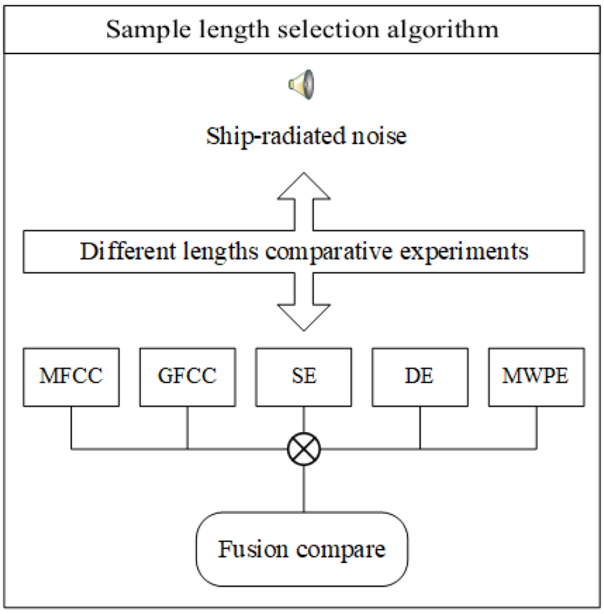


FIGURE 2
Sample length selection algorithm framework.

generation, GAN uses random noise as input to generate, longer inputs require longer iteration times, so longer sample lengths lead to longer computation times. The purpose of sample length selection is to determine the optimal trade-off between having sufficient features while maintaining the computational load to a minimum.

2.1 Cepstrum coefficient characteristics

MFCC is a characteristic parameter based on the characteristics of simulated acoustic signals passing through the cochlea Lin et al. (2021). The Mel scale describes the nonlinear characteristics of human ear frequencies, and MFCC is the cepstrum coefficient extracted under the nonlinear characteristics of this frequency. The relationship between frequency and the Mel scale can be approximated using the following equation:

$$M(f) = 2595log(1 + f/700) \tag{1}$$

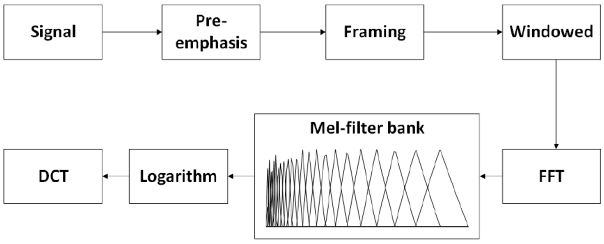


FIGURE 3
MFCC algorithm flow chart.

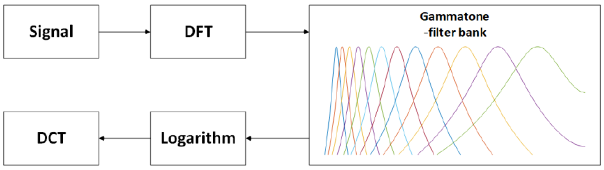


FIGURE 4
GFCC algorithm flow chart.

where f is the frequency of the signal, while the other fixed parameters are derived from the cochlea. The specific steps of the algorithm are shown in Figure 3:

In Figure 3, the signal is pre-weighted through a high-pass filter and divided into frames with a length of 256, while a Hamming window is applied to each frame to enhance the continuity at both ends. Next, a fast Fourier transform is applied on these frames, which then pass through the triangular filter banks based on the Mel scale. After that, the logarithmic energy of each filter output is calculated and the discrete cosine transform (DCT) is applied to finally obtain MFCC feature.

The Gammatone filter simulates the spectrum analysis and frequency selection characteristic of human ears to achieve strong noise resistance and maintain good classification performance in environments with strong interference. It is a bandpass filter, defined in time domain as:

$$g(t, f_c) = At^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \phi) \tag{2}$$

where n is the filter order, A is the filter gain, b is the filter attenuation factor, f_c is the center frequency, and φ is the filter phase.

A Gammatone filter bank is formed using a series of Gammatone filter combinations with different center frequencies [282018Zhang et al.Zhang, Wu, Wang, Wang, Wang, and Zhang]. The center frequencies are first obtained by dividing the equivalent bandwidth scale in equal parts, and are then mapped to a linear scale to determine the center frequency of each Gammatone filter. After that, the characteristic GFCC parameters can be obtained using a logarithmic operation and the DCT. The detailed process is shown in Figure 4.

2.2 Entropy feature

In this section, we describe sample entropy, dispersion entropy and multi-scale weighted permutation entropy.

2.2.1 Sample entropy

Sample entropy is a nonlinear entropy metric commonly used to describe complexity Richman et al. (2004). Sample entropy measures the likelihood of the occurrence of a new pattern in a time series, that is, it is a prediction of the amplitude distribution of future signals based on the current amplitude distribution. In essence, it is the quantification of the complexity and regularity of a sequence. The algorithm is described in Algorithm 1.

```
Input: Time signal $X(i)$, $i$ = 1, 2, ...,$N$, embedding dimension $m$, threshold value $r$
Output: $SampleEn\ (m,r)$
1 Form a set of vector sequences of dimension m according to the sequence number, $X_m(1),\ldots,X_m(N-m+1)$, where $X_m(i) = x(i), x(i+1),\ldots,x(i+1),\ldots,x(i+m-1), 1 \le i \le N-m+1$, and these vectors represent the values of m consecutive x's starting from the i-th point.;
2 Definethe distance between $X_m(i)$ and $X_m(j)$ as $d_{ij}$, which is the absolute value of the largest difference between them: $d_{ij} = d[X_m(i), X_m(j)] = max_{k-0.1,\ldots,m-1}[|x(i+k)-x(j+k)|]$;
3 for $m$ or $m+1$ do
4     Given $X_m(i)$, count the number of $j$ whose distance between $X_m(i)$ and $X_m(j)$ is less than or equal to $r$ and denote it as $B_i$, define $B_i^m(r) = \frac{1}{N-m-1}B_i$ when $1<i<N-m$;
5     Calculate $B^{(m)}(r) = \frac{1}{N-m}B_i\sum_{i=1}^{N-m}B_i^m(r)$
6 end
7 Obtain sample entropy $SampleEn(m,r) = \lim_{N\to\infty}[-\mathrm{In}\frac{B^{(m+1)}(r)}{B^{(m)}(r)}]$
```

ALGORITHM 1
Sample Entropy.

2.2.2 Dispersion entropy

Dispersion entropy is a metric used to measure the complexity and irregularity of time series, which is sensitive to the variation of the frequency, amplitude, and time series' bandwidth, and it does not require the sorting of the amplitude values of each embedded vector Li et al. (2022). Because of the above characteristics, the calculation efficiency of the dispersion entropy is high. The calculation steps of dispersion entropy are given in Algorithm 2.

```
Input: Time signal $x(i)$,$i$=1,2,…,$N$, embedding dimension $m$, class number $c$, time delay $d$
Output: Dispersion Entropy $DE(x,m,x,d)$
1 Use the normal cumulative distribution function to map $x$ to $y$=$y_j$, $j$ = 1,2,…,$N$, $y_j \in (0,1)$, using $y_j = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_j} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t$ ;
2 Use a linear transformation to assign $y$ to the range [1,2, …, $c$]: $z_j^c = round(c \times y_j + 0.5)$;
3 Calculate the embedded vector $z_i^{m,c} = z_i^c, z_{i+d}^c, \ldots, z_{i+(m-1)d}^c, i = 1,2,\ldots,N-(m-1)d$;
4 Calculate the dispersion model $\pi_{v_0}, v_1, \ldots, v_{m-1}$ ($v$=1,2,…,$c$), if $z_i^c = v_0$, $z_{i+(m-1)d}^c = v_{m-1}$, the dispersion model corresponding to $z_i^{m,c}$ is $\pi_{v_0}, v_1, \ldots, v_{m-1}$;
5 Calculate the probability of each dispersion model and $N_b(\pi_{v_0}, v_1,\ldots,v_{m-1})$ is the number of maps from $z_i^{m,c}$ to $\pi_{v_0}v_1,\ldots,v_{m-1}$;
```

ALGORITHM 2
Dispersion Entropy.

2.2.3 Multi-scale weighted permutation entropy

Permutation entropy (PE) is suitable for time series, because it measures their randomness and detects their dynamic changes. It can be obtained at fast speeds through a comparison of neighboring values. Weighted PE (WPE) is an improved algorithm based on PE, which fully considers that the amplitude of adjacent vectors of the same order may be different. PE and WPE are already used in different fields, such as underwater acoustic signal denoising Li et al. (2018). The detailed calculation steps are described in Algorithm 3.

```
Input : Time signal $x(i)$,$i$=1,2,…,$N$, embedding dimension $m$, time delay $\tau$
Output: Weighted Permutation Entropy $WPE(x, m, \tau)$
1 Reconstruct time signal $X_i = x(i), x(i+\tau),\ldots, x(i+(m-1)\tau), i = 1,2,\ldots,N-(m-1)\tau$;
2 Rearrange $Xi$ elements in increasing order:
    $X(i) = x(i+(j_1-1)\tau) \le x(i+(j_2-1)\tau) \le \ldots \le x(i+(j_m-1)\tau)$. In case of two if the rearranged elements are equal, then obtain new order: $x(i+(j_1-1)\tau) \le x(i+(j_2-1)\tau)\ (j_1 \le j_2)$ and obtain one of the $m!$ symbol sequences in phase space: $S(g) = (j_1, j_2, \ldots, j_m)$, $j$ = 1,2,…, $k$.$k \le m!$
3 If the probability distribution of the symbol sequence is $P_1$, $P_2$,…, $P_k$, obtain the permutation entropy: $H_p(m) = -(\mathrm{In}(m!)^{-1}\sum_{g=1}^{k}P_g\ \ \mathrm{In}\ \ P_g$;
4 Calculate the weights wi of all adjacent vectors $X_i : w_j = \sum_{k=1}^{m}[x_j + (k-1)_T - \bar{X}_j^{m,\ T}]^2)$, $\bar{X}_j^{m,T} = \frac{1}{m}\sum_{k=1}^{m}x_{k+(k+1)_T}$;
5 Calculate the weighted relative frequency: $p_w(\pi_i^{m\cdot T} = \frac{\sum j \le N^{1}{}_{u:type(u)=\pi_i(\bar{X}_j^{m,T})w_j}}{\sum j \le N^{1}{}_{u:type(u)\in\Pi}{}^{(\bar{X}_j^{m,T})w_j}})$;
6 Obtain WPE: $WPE(x,m,d) = -\sum_{i:\pi_i^{m,T}\in\Pi}p_w(\pi_i^{m,T})\ \mathrm{In}\ p_w(\pi_i^{m,T})$.
```

ALGORITHM 3
Weighted Permutation Entropy.

The multi-scale analysis algorithm is developed by Costa, and can be used to estimate the complexity of the original data at different scales Chen et al. (2019). Based on the concept of multi-scale analysis, multi-scale permutation entropy (MPE) is proposed by Aziz and Arif (2005). The MPE algorithm is divided into two steps. The first step is the application of a coarse-grained procedure to obtain multi-scale time series from the original time series. The second step is the calculation of the WPE at each coarse-grained time series. The details of the algorithm are shown in Algorithm 4.

```
Input: Time signal $x(i)$, $i$ = 1,2,…,$N$, embedding dimension $m$, time delay $\tau$, scale factor $s$
Output: Dispersion Entropy $MWPE(x,s,m,\tau)$
```

```
1 Construct coarse-grained time series $y_j^s$ at a scale factor of s: $s: y_j^s = \frac{1}{s}\sum_{i=(j-1)s+1}^{js} x_i, 1 \le j \le \frac{N}{s}$;;
2 Calculate the WPE of each coarse-grained time series through Algorithm 3.;
3 Obtain MWPE: $MWPE(x, s, m, \tau) = [WPE(1),\ldots, WPE(s)$.
```

ALGORITHM 4
Multi-scale Weighted Permutation Entropy.

2.3 Experimental results

In the ship-radiated noise generated by the four types of ships, we randomly selected 64 samples for each type, used 200 of the 256 samples as the training set and 56 as the test set for classification, and analyzed the classification results under different sample lengths. Support vector machine can achieve good classification results in a short time in the case of fewer samples and manual feature extraction. We used MFCC, GFCC, MWPE, SE, and DE to express different characteristics of signals for feature extraction, and the corresponding features are independently input into the SVM classifier, starting from a sample length of 100 sampling points and up to 2,500 sampling points at an interval of 100 sampling points.

Figure 5 shows the classification results for the above-mentioned features. In classification using MFCC, 256 or 512 sampling points are usually used as a frame when framing. We selected 256 sampling points as a basic frame. The algorithm cannot work when the sampling length is less than two basic frames, the classification rate is 0 when the sampling length is less than 500 sampling points. When the sample length increases, the classification accuracy of each feature also improves. At this time, the sample length has a greater impact on classification. The classification rate tends stabilize when the sample length is about 1,600 sampling points. We concluded that the ship-radiated noise sample at this time already contained enough information for classification. In a CNN, the time complexity of a single convolutional layer is $TimeO(M^2\ K^2\ C_{in}C_{out})$, where M is the side length of the convolution kernel; K is the side length of the convolution kernel; C_{in} is the number of input channels; C_{out} is the number of output channels; and the size of the output feature map M itself is determined by four parameters, the input matrix size X, the size of the convolution kernel K, *Padding*, and *Stride*, which are expressed as $M=(X-K+2*Padding)/Stride+1$. The space complexity of the CNN is determined by the total parameter amount and the feature map output by each layer as follows: $Space \sim O(\sum_{l=1}^{D} K^{2_l} \cdot \quad C_{l-1} \cdot \quad C_l + \sum_{l=1}^{D} M^2 \cdot \quad C_l)$ .The total parameter amount is only related to the size of the convolution kernel K, the number of channels C, and the number of layers D; it has nothing to do with the size of the input data. The space occupation of the output feature map is related to the space size M and the number of channels C. In other words, the performances of the network are closely related to the input size. The sampling length we selected after a comprehensive analysis of multiple features is 1,600 sampling points. A sample of this size can not only reduce the time and space complexity of the network but also obtain results efficiently and prevent overfitting.

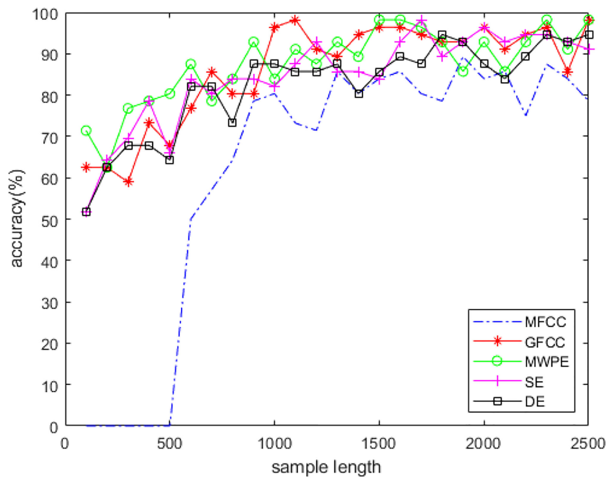


FIGURE 5
Classification results of different features.

3 1-DDCGAN for ship-radiated noise data augmentation

When training deep neural networks, too few training samples results in limited features learned by the network, making it difficult to perform more complex classification tasks. We used 1-DDCGAN to enhance the data and analyze the generated samples.

3.1 1-DDCGAN structure

GANs provide a way to learn deep representations without extensive annotated training data by using backpropagation through a competitive process involving a pair of networks. GANs are used in a variety of applications, including image synthesis, semantic image editing, style transfer, image super-resolution, and classification Arniriparian et al. (2018); Pan et al. (2019). GANs are inspired by the game theory, where the generator and discriminator

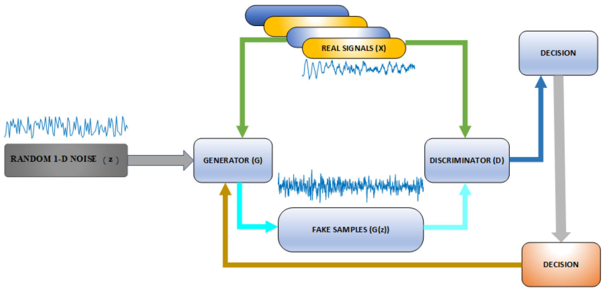


FIGURE 6
GAN architecture in general.

will compete with each other to achieve a Nash equilibrium during training. The architecture of GAN is shown in Figure 6.

The function of generator *G* is to generate fake data to fool the discriminator as much as possible, while the discriminator *D* is trained to distinguish real data samples from synthesized samples. The input of the generator is a random noise vector *z*. The noise is mapped to a new data space through generator *G* to obtain a fake sample *G*(*z*), which is a multi-dimensional vector. Then, the discriminator *D*, which is a binary classifier, accepts either real data or fake data from generator *G* as an input and outputs the probability of the input being true and false. Then, the training process continues until the discriminator *D* cannot determine whether the data comes from the real dataset or from *G*. Finally, we obtained a model *G*, which can generate data that are similar to the real data. In prior work Ian et al. (2014), the discriminator *D* is defined as a binary classifier, whose loss function is represented by the cross entropy as follows:

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{x\sim pdata}\log \quad D(x) - \frac{1}{2}\mathbb{E}_z \log(1 - D(G(z))), \tag{3}$$

where *x* is the real sample, *z* is the random noise vector, *G*(*z*) are the data generated by the generator *G*, and *E* is the expectation. *D*(*G*(*z*)) indicates the probability that discriminator *D* determines the data are generated by *G*. The goal of *D* is for *D*(*G*(*z*)) to approach 0, while *G* aims to bring it closer to 1. Therefore, the loss of the generator can be derived *via* the discriminator *D*: $f^{(G)}$=-$f^{(D)}$

For the above reasons, the optimization problem of GANs is transformed into the minimax game as shown below:

$$\begin{aligned} &\min_G \max_D \quad V \quad (D,G) \\ &= \mathbb{E}_{x\sim pdata}\log \quad D(x) + \mathbb{E}_{z\sim p(z)}\log(1 - D(G(z))) \end{aligned} \tag{4}$$

The whole process of the algorithm is roughly as follows:
1. update the discriminator by ascending its stochastic gradient:

$$\nabla\ \theta_d \sum_{i=1}^{m} [\log D(x^{(i)}) + log 1 - D(G(z^{(i)}))] \tag{5}$$

2. Sample mini-batch of *m* noise samples $z^{(1)}$, $z^{(2)}$, …,$z^{(m)}$;
3. Update the generator by descending its stochastic gradient:

$$\nabla\ \theta_d \frac{1}{m}\sum_{i=1}^{m} + \log 1 - D(G(z^{(i)})) \tag{6}$$

These training sessions are then alternated until equilibrium is achieved.

The performance of the GAN is basically determined by its structure. Due to the deficiencies of the original GANs, various derived GANs models have been proposed. DCGAN Radford et al. (2015) is based on the use of a CNN; it is regarded as an effective network model of supervised learning, and is the most common generator and discriminator structure. The general two-dimensional convolution kernel is mainly used to process images, and it is difficult to directly process the one-dimensional ship-radiated noise signal. We use a one-dimensional convolution filter to fit ship-radiated noise. The structure is illustrated in Figure 7.

3.2 Generated sample results analysis

Due to the non-linear, non-Gaussian and non-stationary characteristics of ship-radiated noise, it is difficult to evaluate the quality of the generated data with traditional amplitude and frequency characteristics. We used the statistical distribution characteristics of the generated data, such as Mutual information (MI) and correlation coefficient (CC), to conduct a preliminary similarity analysis of the generated samples. MI is a measure of the degree of interdependence between variables:

$$MI(X;Y) \sum_{y\in Y} \sum_{x\in Y} p(x,y) \log(\frac{p(x,y)}{p(x)p(y)}) \tag{7}$$

We obtained MI between the generated samples and the original samples to measure the correlation between the samples. CC is the coefficient of correlation between different statistics:

$$CC(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y} \tag{8}$$

where Cov(X,Y) is the covariance, and σ is the standard deviation. Fifty pairs of samples are randomly selected from the original samples and the generated samples, and the average values of the MI and CC are calculated. The specific results are shown in Tables 2, 3.

MI and CC are indices used to evaluate the similarity between two variables and for both indices, higher values correspond to a greater similarity between the variables. We see from the tables that MI and

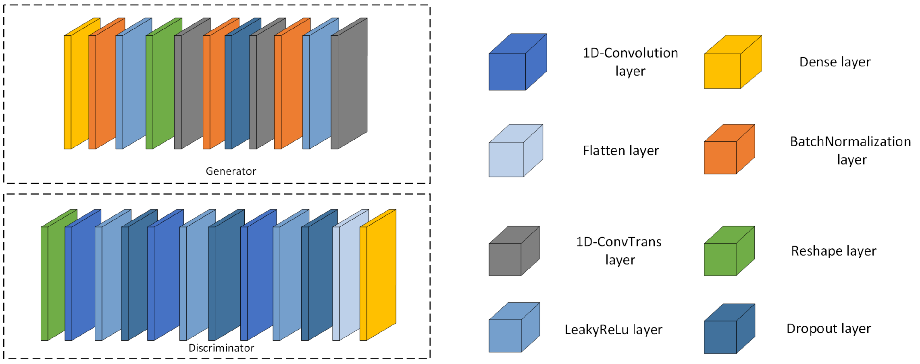


FIGURE 7
One-dimensional DCGAN architecture.

TABLE 2 Mutual information results.

MI	GSample 1	GSample 2	GSample 3	Gsample 4
Sample 1	2.8937	2.3979	2.4599	2.2281
Sample 2	2.2112	2.2814	1.8438	1.8728
Sample 3	2.1499	2.4278	3.1285	2.2975
Sample 4	2.2590	2.2396	1.9204	2.6094

TABLE 3 Correlation coefficient results.

CC	GSample 1	GSample 2	GSample 3	Gsample 4
Sample 1	0.1151	0.0085	0.0775	0.0446
Sample 2	0.0045	0.0825	0.0026	0.0162
Sample 3	0.0095	0.0095	0.1438	0.0126
Sample 4	0.0130	0.0046	0.0142	0.0402

CC between samples of the same type, such as Sample 1 and Gsample 1, are higher than those between other samples. However, the gap between classes is not very obvious. To further verify the reliability of the generated samples, we used the DE, which is more suitable for nonlinear signal analysis, to compare the generated samples with the original samples. Similarly, we selected 50 samples of each kind and calculated the average value. The results are shown in Table 4.

As shown in Table 4, the DE values of the generated samples of types 2, 3, and 4 are very close to the original samples, which indicates that the samples generated by GAN are consistent with the original samples, and there is diversity among samples. However, the DE values of type 2 and 4 are also relatively close, which is also one of the difficulties in the classification of ship-radiated noise. The DE values of type 1 had some gaps, but CC and MI performed better, and this also verifies the diversity of the generated samples. The generated samples could be therefore be used to form a new training dataset for training a ship-radiated noise classifier.

4 Classification based on deep learning

Given the connection between different network layers, deep neural networks extract signal features for classification inherently. Convolutional neural networks are a traditional type of deep neural network. Compared with SVMs, they require more data to ensure that the network will not be overfitted, but they perform better on unprocessed data. Deep networks with better performance have emerged with the continuous development of deep learning. ResNet He et al. (2015) uses direct mapping to connect different layers, solves the problem of gradient disappearance or gradient explosion problems, and greatly improves the model fitting capability than networks without this module. In this study, we used the sample length of 1,600 sampling points selected in Section 2 as the input size of the network for model construction, and the samples generated in Section 3 are used to mix with the original samples in different proportions to construct a new dataset as the training dataset of the network.

To observe the impact of generated samples on the final classification, we set up five training sets with different proportions of the original samples and the generated samples as follows: only 400 original samples, 400 original samples and 100 generated samples, 400 original samples and 200 generated samples, 400 original samples and 400 generated samples, and 400 original samples and 800 generated samples. The number of original samples in the test set is set according to the 10% ratio of the training set. The structure of the CNN is similar to the GAN discriminator, with the last dense classification layer changed to a layer using a four-way Softmax activation function, and the structure is shown in Figure 8. The parameters of three 1D-convolution layers are (64,3,2), (128,3,2), and (256,3,1). The first parameter is the number of convolution kernels; the second parameter is the kernel size; the last parameter is stride; and the value of padding is set to *same*, which means that it complements 0 uniformly during the sliding process of the convolution kernel. The parameter in the dense layer is set to (4, *softmax*), which means that the CNN used Softmax as the activation function for the four classifications. The structure of ResNet is shown in Figure 9. The parameters of the 1D-convolution layers before the residual modules are (64,3,1) and (64,3,1); the parameters of the 1D-

TABLE 4 Dispersion entropy results.

Sample name	GSample 1	GSample 2	GSample 3	GSample 4
DE	3.2164	4.5132	2.5607	4.5823
Sample name	Sample 1	Sample 2	Sample 3	Sample 4
DE	2.6951	4.3608	2.5028	4.4172

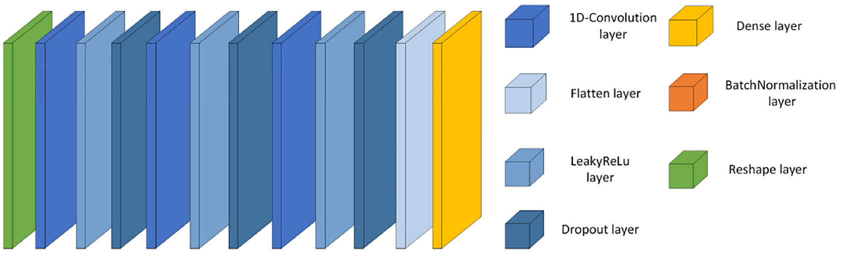


FIGURE 8
CNN structure.

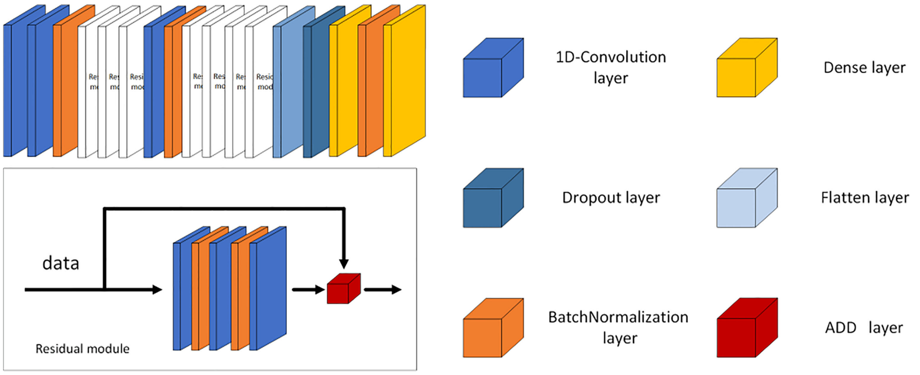


FIGURE 9
ResNet structure.

convolution layer in the first batch of residual modules are (16,1,1), (16,3,1), and (64,1,1); and the parameters of the next 1D-convolution layer are (128,3,1). The parameters of the 1D-convolution layer in the first batch of residual modules are (32,1,1), (32,3,1), and (128,1,1). Then, the parameter of the first dense layer is (2048, *relu*), which means that the dimension of the output is 2,048, and the dense layer used the ReLU function as the activation function to fully connect the data. The parameter of the second dense layer is (4,*softmax*), and the function is same to the dense layer in CNN. The results are shown in Figures 10, 11.

The images on the left are the training accuracy results, while the images on the right are the training loss results. Figures 10A and 11A show 400 original samples of 4 categories, including 100 samples in each category. They are then input into the traditional CNN and one-dimensional ResNet for classification.

The rest of the figures are the classification results obtained from the datasets formed when the ratio of original to generated samples is 4:1 in Figure 10B, 2:1 in Figure 10C, 1:1 in Figure 10D, and 1:2 in Figure 10E, respectively. In Figures 10, 11, the lines “train-acc” and “train-loss” are the results obtained in the training set, and “val-acc” and “val-loss” are the results obtained in the testing set.

The classification accuracy of the CNN using the dataset with only original samples is 79% in Figure 10A. When 100 generated samples are added to the dataset, the classification rate significantly improved, reaching 85.6% in Figure 10B. Adding another 100 generated samples to the dataset caused the classification accuracy to reach 86.68% in Figure 10C, with a 2:1 ratio of the original to generated samples. When the ratio is 1:1, the accuracy rose to 89.5% in Figure 10D. However, when the ratio is 1:2, the accuracy reached 90% in Figure 10E and did not increase significantly.

At this point, the sample size already met the requirements of the network structure, and the accuracy reached convergence, which also verifies the validity of the generated samples. ResNet has more network layers than a traditional CNN, and thus more training epochs are required. Therefore, the epoch number is set to 150. As shown in Figure 11, when we used the original dataset, the accuracy and loss converged slowly, and the accuracy only reached an eventual value of 87.5% in Figure 11A. As the number of samples increased, the accuracy gradually improved. After adding 100 samples into the dataset, the accuracy reached 92.5% in Figure 11B, but it is not stable and there are still some problems with overfitting.

Then, another 100 generated samples are added to the dataset, and the accuracy reached 98.33% in Figure 11C. The accuracy reached 97.5% in Figure 11D when the ratio of the original sample to the number of generated samples is 1:1. Finally, the accuracy reached 99.17% in Figure 11E when the ratio is 1:2. In the process of network model training and testing, each epoch took about 16 seconds. The number of epochs needed for convergence is reduced from 150 to 50 on testing set, and the convergence time is reduced from 40 minutes to 13.3 minutes. After network training is completed, when the new ship-radiated noise data are obtained, we only needed to perform sample segmentation and input them into the network directly without complex signal decomposition and manual feature extraction for classification. A classification result with an accuracy of 99% could be obtained within a few minutes.

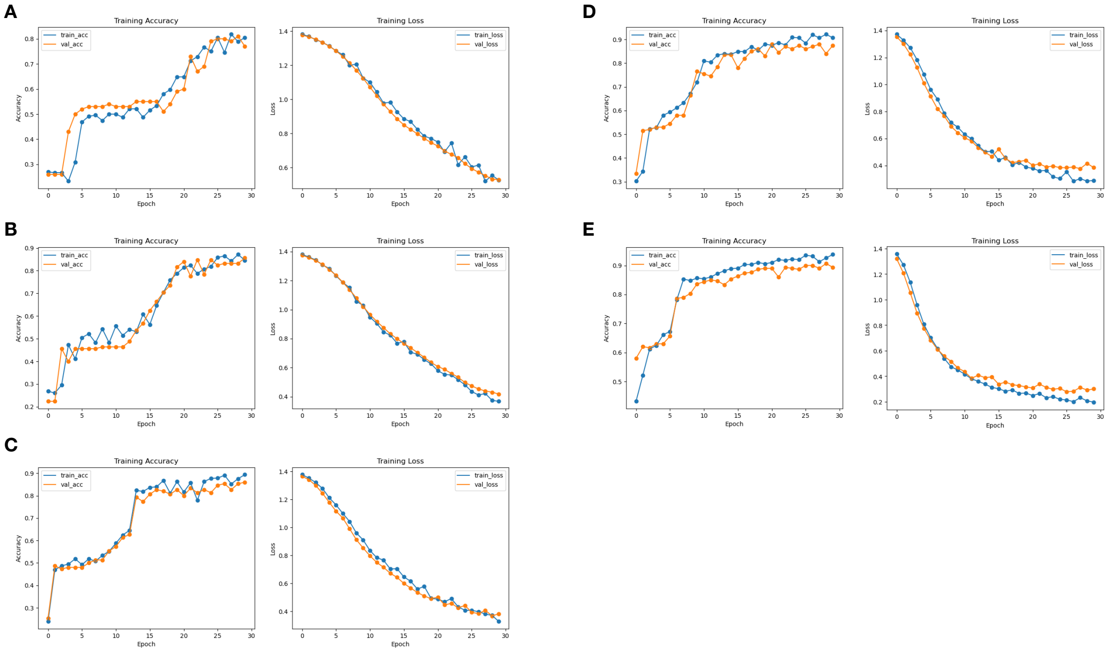


FIGURE 10
CNN classification results. **(A)** Results of the CNN on the original dataset. **(B)** Results of the CNN at 4:1 dataset. **(C)** Results of the CNN at 2:1 dataset. **(D)** Results of the CNN at 1:1 dataset. **(E)** Results of the CNN at 1:2 dataset.

5 Conclusion

In this paper, a 1-DDCGAN based on GAN is applied to extend ship-radiated noise under small sample size conditions. Deep neural network can extract sufficiently deep and essential features through the network structure when the sample size is large enough, and these features are used as a basis for classification. One-dimensional CNN and ResNet are designed for data analysis. And the accuracy is improved from 79% to 90% in CNN, from 87.5% to 99.17% in ResNet as the sample size increases.

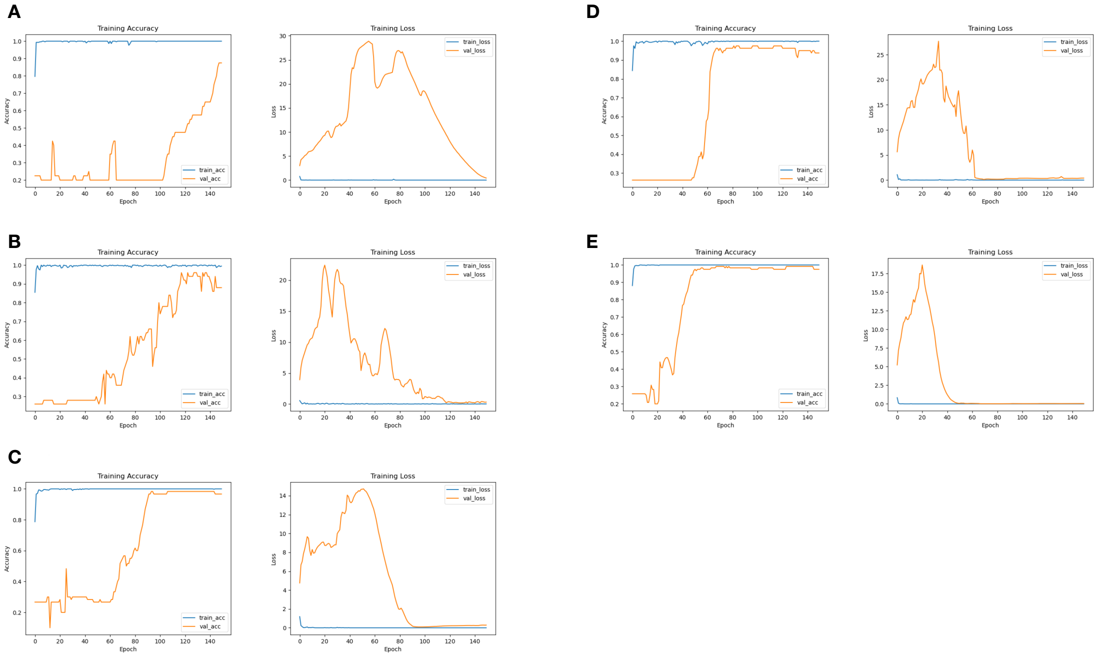


FIGURE 11
ResNet classification results. **(A)** Results of ResNet on the original dataset. **(B)** Results of ResNet at 4:1 dataset. **(C)** Results of ResNet at 2:1 dataset. **(D)** Results of ResNet at 1:1 dataset. **(E)** Results of ResNet at 1:2 dataset.

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

ZX: Conceptualization, visualization, writing, project administration, original draft. RL: supervision, writing review. LW: supervision, writing review, editing. AZ: Supervision, conceptualization. JL: Writing review, editing. XT: Writing review, editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Program of Marine Economy Development Special Foundation of Department of Natural Resources of Guangdong Province (GDNRC [2022]19), and Natural Resources Science and Technology Innovation Project of Fujian Province (KY-080000-04-2021-030).

Conflict of interest

Author XT was employed by Xiamen Weilai Marine Science and Technology 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

Arniriparian, S., Freitag, M., Cummins, N., Gerczuk, M., Pugachevskiy, S., and Schuller, B. (2018). “A fusion of deep convolutional generative adversarial networks and sequence to sequence autoencoders for acoustic scene classification,” in *2018 26th European signal processing conference (EUSIPCO)* (Rome, Italy: IEEE), 977–981.

Atanackovic, L., Vakilian, V., Wiebe, D., Lampe, L., and Diamant, R. (2020). “Stochastic ship-radiated noise modelling via generative adversarial networks,” in *Global oceans 2020: Singapore–US gulf coast* (Biloxi, MS, USA: IEEE), 1–8.

Aziz, W., and Arif, M. (2005). “Multiscale permutation entropy of physiological time series,” in *2005 Pakistan section multitopic conference* (Karachi, Pakistan: IEEE), 1–6.

Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106, 249–259. doi: 10.1016/j.neunet.2018.07.011

Chen, Z., Li, Y., Cao, R., Ali, W., Yu, J., and Liang, H. (2019). A new feature extraction method for ship-radiated noise based on improved ceemdan, normalized mutual information and multiscale improved permutation entropy. *Entropy* 21, 624. doi: 10.3390/e21060624

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv*. doi: 10.48550/arXiv.1512.03385

Ian, G., Pouget-Abadie, J., Mirza, M., Xu, B., and Warde-Farley, D. (2014). “Generative adversarial nets,” in *Advances in neural information processing systems. arXiv*. doi: 10.48550/arXiv.1406.2661

Jiao, S., Geng, B., Li, Y., Zhang, Q., and Wang, Q. (2021). Fluctuation-based reverse dispersion entropy and its applications to signal classification. *Appl. Acoust.* 175, 107857. doi: 10.1016/j.apacoust.2020.107857

Li, Y., Geng, B., and Jiao, S. (2022). 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, Y., Li, Y., Chen, X., and Yu, J. (2017). Denoising and feature extraction algorithms using npe combined with vmd and their applications in ship-radiated noise. *Symmetry* 9, 256. doi: 10.3390/sym9110256

Li, Y., Li, Y., Chen, X., Yu, J., Yang, H., and Wang, L. (2018). A new underwater acoustic signal denoising technique based on ceemdan, mutual information, permutation entropy, and wavelet threshold denoising. *Entropy* 20, 563. doi: 10.3390/e21010011

Li, Z., Li, Y., and Zhang, K. (2019). A feature extraction method of ship-radiated noise based on fluctuation-based dispersion entropy and intrinsic time-scale decomposition. *Entropy* 21, 693. doi: 10.3390/e21070693

Lin, Z., Di, C., Chen, X., and Hou, Y. (2021). Acoustic recognition method in low snr based on human ear bionics. *Appl. Acoust.* 182, 108213. doi: 10.1016/j.apacoust.2021.108213

Li, J., and Yang, H. (2021). The underwater acoustic target timbre perception and recognition based on the auditory inspired deep convolutional neural network. *Appl. Acoust.* 182, 108210. doi: 10.1016/j.apacoust.2021.108210

Noda, J. J., Travieso-González, C. M., Sánchez-Rodríguez, D., and Alonso-Hernández, J. B. (2019). Acoustic classification of singing insects based on mfcc/lfcc fusion. *Appl. Sci.* 9, 4097. doi: 10.3390/app9194097

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., and Zheng, Y. (2019). “Recent progress on generative adversarial networks (gans): A survey,” in IEEE Access, Vol. 7. 36322–36333. doi: 10.1109/ACCESS.2019.2905015

Pethiyagoda, R., Moroney, T. J., Macfarlane, G. J., Binns, J. R., and McCue, S. W. (2018). Time-frequency analysis of ship wave patterns in shallow water: modelling and experiments. *Ocean Eng.* 158, 123–131. doi: 10.1016/j.oceaneng.2018.01.108

Qi, J., Yang, H., and Sun, H. (2020). “Mod-rrt*: A sampling-based algorithm for robot path planning in dynamic environment,” in *IEEE Transactions on Industrial Electronics*, Vol. 68. 7244–7251. doi: 10.1109/TIE.2020.2998740

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*. doi: 10.48550/arXiv.1511.06434

Richman, J. S., Lake, D. E., and Moorman, J. R. (2004). “). sample entropy,” in *Methods in enzymology*, vol. 384. (Cambridge, MA, USA: Academic Press), 172–184. doi: 10.1016/S0076-6879(04)84011-4

Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A., and Pena-Gimenez, A. (2016). Shipsear: An underwater vessel noise database. *Appl. Acoust.* 113, 64–69. doi: 10.1016/j.apacoust.2016.06.008

Shi, Y., Li, B., Du, G., and Dai, W. (2021). Clustering framework based on multi-scale analysis of intraday financial time series. *Physica A: Stat. Mechan. its Appl.* 567, 125728. doi: 10.1016/j.physa.2020.125728

Wei, X. (2016). “On feature extraction of ship radiated noise using 11/2 d spectrum and principal components analysis,” in *2016 IEEE international conference on signal processing, communications and computing (ICSPCC)* (Hong Kong, China: IEEE), 1–4.

Xie, D., Esmaiel, H., Sun, H., Qi, J., and Qasem, Z. A. H. (2020). Feature extraction of ship-radiated noise based on enhanced variational mode decomposition, normalized correlation coefficient and permutation entropy. *Entropy* 22, 468. doi: 10.3390/e22040468

Yang, H., Cheng, Y., and Li, G. (2021). A denoising method for ship radiated noise based on spearman variational mode decomposition, spatial-dependence recurrence sample entropy, improved wavelet threshold denoising, and savitzky-golay filter. *Alexandria Eng. J.* 60, 3379–3400. doi: 10.1016/j.aej.2021.01.055

Yan, X., Song, H., Peng, Z., Kong, H., Cheng, Y., and Han, L. (2021). Review of research results concerning the modelling of shipping noise. *Polish Maritime Res.* doi: 10.2478/pomr-2021-0027

Zhang, H., Junejo, N. U. R., Sun, W., Chen, H., and Yan, J. (2020a). “Adaptive variational mode time-frequency analysis of ship radiated noise,” in *2020 7th international conference on information science and control engineering (ICISCE)* (Changsha, China: IEEE), 1652–1656.

Zhang, X., Wu, H., Sun, H., and Ying, W. (2021). Multireceiver sas imagery based on monostatic conversion. *IEEE J. Selected Topics Appl. Earth Observat. Remote Sens.* 14, 10835–10853.

Zhang, W., Wu, Y., Wang, D., Wang, Y., Wang, Y., and Zhang, L. (2018). “Underwater target feature extraction and classification based on gammatone filter and machine learning,” in *2018 international conference on wavelet analysis and pattern recognition (ICWAPR)* (Chengdu, China: IEEE), 42–47.

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., Feng, X., and Sun, H. (2022a). Efficient imaging method for multireceiver sas. *IET Radar Sonar Navigat.*

Zhang, X., Yang, P., and Sun, H. (2022b). Frequency-domain multireceiver synthetic aperture sonar imagery with chebyshev polynomials. *Electron. Lett.*

Zhang, X., Yang, P., and Sun, M. (2022c). Experiment results of a novel sub-bottom profiler using synthetic aperture technique. *Curr. Sci. (00113891)* 122, 461-464.

Zhang, X., Ying, W., Yang, P., and Sun, M. (2020b). Parameter estimation of underwater impulsive noise with the class b model. *IET Radar Sonar Navigat.* 14, 1055–1060.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 28 April 2023
DOI 10.3389/fmars.2023.1130898

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Alec Duncan,
Curtin University, Australia
Naveed Ur Rehman Junejo,
University of Lahore, Pakistan

*CORRESPONDENCE
Robin André Rørstadbotnen
✉ robin.a.rorstadbotnen@ntnu.no

RECEIVED 23 December 2022
ACCEPTED 28 March 2023
PUBLISHED 28 April 2023

CITATION
Rørstadbotnen RA, Eidsvik J, Bouffaut L, Landrø M, Potter J, Taweesintananon K, Johansen S, Storevik F, Jacobsen J, Schjelderup O, Wienecke S, Johansen TA, Ruud BO, Wuestefeld A and Oye V (2023) Simultaneous tracking of multiple whales using two fiber-optic cables in the Arctic. *Front. Mar. Sci.* 10:1130898.
doi: 10.3389/fmars.2023.1130898

Simultaneous tracking of multiple whales using two fiber-optic cables in the Arctic

Robin André Rørstadbotnen[1,2]*, Jo Eidsvik[2,3], Léa Bouffaut[4], Martin Landrø[1,2], John Potter[1,2], Kittinat Taweesintananon[1,2,5], Ståle Johansen[1,2,6], Frode Storevik[2,7], Joacim Jacobsen[2,8], Olaf Schjelderup[2,7], Susann Wienecke[2,8], Tor Arne Johansen[9], Bent Ole Ruud[9], Andreas Wuestefeld[2,10] and Volker Oye[2,10]

[1]Acoustics Group, Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [2]Centre for Geophysical Forecasting, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [3]Department of Mathematical Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [4]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States, [5]PTT Exploration and Production Public Company Limited, Bangkok, Thailand, [6]Department of Geoscience and Petroleum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [7]Sikt, Trondheim, Norway, [8]Alcatel Submarine Networks Norway AS, Tiller, Norway, [9]Department of Earth Science, University of Bergen, Bergen, Norway, [10]NORSAR, Kjeller, Norway

Climate change is impacting the Arctic faster than anywhere else in the world. As a response, ecosystems are rapidly changing. As a result, we can expect rapid shifts in whale migration and habitat use concurrent with changes in human patterns. In this context, responsible management and conservation requires improved monitoring of whale presence and movement over large ranges, at fine scales and in near-real-time compared to legacy tools. We demonstrate that this could be enabled by Distributed Acoustic Sensing (DAS). DAS converts an existing fiber optic telecommunication cable into a widespread, densely sampled acoustic sensing array capable of recording low-frequency whale vocalizations. This work proposes and compares two independent methods to estimate whale positions and tracks; a brute-force grid search and a Bayesian filter. The methods are applied to data from two 260 km long, nearly parallel telecommunication cables offshore Svalbard, Norway. First, our two methods are validated using a dedicated active air gun experiment, from which we deduce that the localization errors of both methods are 100 m. Then, using fin whale songs, we demonstrate the methods' capability to estimate the positions and tracks of eight fin whales over a period of five hours along a cable section between 40 and 95 km from the interrogator unit, constrained by increasing noise with range, variability in the coupling of the cable to the sea floor and water depths. The methods produce similar and consistent tracks, where the main difference arises from the Bayesian filter incorporating knowledge of previously estimated locations, inferring information on speed, and heading. This work demonstrates the simultaneous localization of several whales over a 800 km area, with a relatively low infrastructural investment. This approach could promptly inform management and stakeholders of whale presence and movement and be used to mitigate negative human-whale interaction.

KEYWORDS

Distributed Acoustic Sensing, Baleen whales, fiber-optic, bioacoustics, tracking, Passive acoustic monitoring, localization and fin whales

1 Introduction

Baleen whales play crucial ecosystemic roles in the oceans, from predators to prey, climate regulators, nutrient reservoirs, niche constructors enhancing biodiversity, and connectors between ecosystems (Lavery et al., 2014; Roman et al., 2014). After being brought to the brink of extinction, many species are recovering following the cessation of large-scale commercial whaling (Thomas et al., 2016). Nevertheless, their recovery is hampered by anthropogenic stressors associated with modern and industrialized ocean exploitation where pollution (acoustic, chemical, and thermal) coupled with ocean acidification, add to the primary threats of ship strikes and entanglement in fishing gear (Greene and Pershing, 2004; Thomas et al., 2016).

In the Arctic, the climate is changing faster than anywhere else in the world (IPCC, 2022), and the Svalbard archipelago is one of the fastest warming regions (Maturilli et al., 2013; Nordli et al., 2014). This can be observed in many ways; e.g., through the rapid retreat of glaciers in the Svalbard area (Hagen et al., 1993; Schuler et al., 2020) and by the decrease in Arctic sea-ice (Stroeve et al., 2007; Comiso et al., 2008). The "Atlantification" of the Arctic alters sea-surface temperatures and circulation patterns, forcing some cetaceans to change their seasonal habits; e.g., the timing of their migration, the migration route itself, or even forcing them to seek alternative habitats (van Weelden et al., 2021). This impacts both endemic Arctic species and boreal visitor species (Hamilton et al., 2021); e.g., fin whales have recently been observed to change their observed presence in Arctic regions from late spring/early summer to the fall to year-round (Klinck et al., 2012). At the same time, human activities, their intensity, and impacts are also evolving with sea-ice loss; e.g., with the impending development of cross-Arctic shipping routes and openings for natural resource exploitation (Jaskolski, 2021; Townhill et al., 2022). Hence, we can expect that human activities will intensify in species-rich areas (Hamilton et al., 2021).

Considering the region's dynamism, it is urgent to develop robust and scalable methods to draw the baseline of species' geographic range and habitat use to understand their ecology, which is challenging for highly mobile and pelagic baleen whales (Ahonen et al., 2021). The methods should also enable close to real-time monitoring to identify rapid changes and mitigate anthropogenic impacts. A key element is to be able to evaluate whales' positions both at large- and fine-scales. Current and common methods for tracking baleen whales include visual surveys (Cummings and Thompson, 1971), satellite tracking (Lydersen et al., 2020; Hoschle et al., 2021) or deploying widespread arrays of hydrophones to determine a whale's position from time difference of arrival and hyperbolic intercepts of its calls (McDonald et al., 1995). Nowacek et al. (2016) presents a review of recent technologies used for conservation-oriented behavioral studies of cetaceans while Harcourt et al. (2019) focuses specifically on methods applied to the conservation of right whales. Under the scope of localization and tracking, passive acoustic monitoring (PAM), using long-term hydrophone installations or towing arrays during short field trial periods, has proven reliable and cost-effective and has been used extensively since the late 1980's to quantify the seasonal presence of whales (Sirovic et al., 2013; Ahonen et al., 2021) and to estimate locations (see, e.g., McDonald and Fox (1999); Thode et al. (2000); Bouffaut et al. (2021)). However, individual hydrophones are often unevenly spaced, making fine-scale movement analysis difficult, and they are often sparse, resulting in undersampling of the vast ocean habitat of baleen whales (Ahonen et al., 2017; Ahonen et al., 2021). The sparse instrumentation can be largely attributed to the cost of purchasing, installing and maintaining these systems and the finite life span of their batteries. Therefore, we are in need of continuous-monitoring PAM systems that are cost-effective, spread over large areas and with sufficient spatial density to fill the gaps in existing capabilities. In addition, it would be highly desirable to be able to get measurements in near-real-time, rather than having to wait until recording instruments are recovered.

Over the last two decades, a new technology, Distributed Acoustic Sensing (DAS), has emerged as a game-changer in remote acoustic sensing, with the potential to fill many of the monitoring gaps in the ocean. Connecting an 'interrogator' to the end of a 'dark' (unused) fiber in a fiber-optic cable allows acoustic data to be acquired continuously. Virtual acoustic sensors can be generated at spatial intervals along the cable as frequently as every meter, for up to 171 km along the fiber (Waagaard et al., 2021). Over 1.3 million km of offshore telecommunication cables are installed worldwide, creating an opportunity to increase remote acoustic sensing coverage both onshore and offshore. Currently, these virtual sensors do not have the sensitivity to rival dedicated hydrophones. However, with thousands lying along an extended cable, the beamforming gain available through signal processing, together with the ability to 'focus' on sources using the very long array, offers unique potential. DAS technology has already been applied to many different fields ranging from earthquake seismology (Lindsey et al., 2017) to geophysical exploration (Mestayer et al., 2011; Taweesintananon et al., 2021), near-surface monitoring (Dou et al., 2017), oceanography (Taweesintananon et al., 2023), water-born sound sources (Matsumoto et al., 2021), and passive acoustic monitoring of ships (Rivet et al., 2021) and baleen whales (Bouffaut et al., 2022; Landrø et al., 2022). Until recently, DAS has been collected from single fibers. However, there is growing interest in combining two different fibers, either in the same or in separate telecommunication cables, when configuration and access allow it (Wilcock et al., 2023).

Fiber-optic cables are often trenched for protection against anchoring and fish trawling. Depending on the sea bed type, the cables are usually laid in a line that covers the shortest possible distance to minimize cost and reduce latency. When interrogating straight-line cable segments, we generally encounter the well-known left-right ambiguity (also known as 180° ambiguity) associated with linear arrays, in which it is impossible to determine which side of the array the source is located. Bouffaut et al. (2022) experienced this problem when tracking whales using the inner cable between Longyearbyen and Ny-Ålesund from a similar experiment in 2020 and Landrø et al. (2022) had the same issue when tracking a cargo ship. While it may sometimes be possible to resolve the left-right ambiguity if the environment breaks the symmetry, the only sure solution is to make the array two-dimensional. This can be achieved by curving the array (as it negotiates a 'corner', for example) or using a second array separated by a sufficient distance to provide a resolvable time of arrival difference, yet close enough to receive the same source signals. We exploit the availability of two such cables in this work.

We investigate and compare two different methods to track baleen whales using DAS: a Grid Search (GS) method (Havskov and Ottemoller, 2010) and a Bayesian Filter (BF; see, e.g., Sarkka (2013)). These two methods were applied to data recorded on two fiber-optic cables connecting Longyearbyen and Ny-Ålesund in the Svalbard archipelago (Figure 1). We first exploited the ground truth generated by a single geo-referenced air gun source for calibration and evaluation of the accuracy of these two methods. A small air gun was used, in accordance with the permissions given by the authorities in Svalbard, to minimize disturbance and potential harm to marine life. Then, over a 5.1 h period recorded on a 60 km portion of the cable, up to eight fin whales (*Balaenoptera physalus*) were successfully tracked.

The paper is organized into five main parts. First, the distributed acoustic sensing data is presented, where the experimental set-up, the air gun, and fin whale data are introduced. The second part contains information on the methods used to obtain the results. The data conditioning, time of arrival picking and computation of empirical detection range are first presented, followed by the localization methods, the GS and the BF. Then, the results from the air gun and fin whale tracking are given, as well as individual fin whale song characteristics. The results are discussed in part four. Firstly, the localization errors are discussed, followed by comparing the two localization methods and DAS-based localization for whale conservation. Finally, we present some conclusions.

2 Distributed acoustic sensing data

2.1 Experimental set up

Currently, it is common practice to include more optical fibers than initially required within one fiber optic telecommunication cable, to enable network growth and increase redundancy at a minimal incremental cost. Using Alcatel Submarine Network OptoDAS interrogators, we tapped into two out of 24 standard single mode G.652D fiber bundles within each of two existing submarine telecommunication cables connecting Longyearbyen and Ny-Ålesund in Svalbard, Norway (Figure 1). The OptoDAS interrogators send linear optical frequency-modulated swept pulses into the fiber and interrogate the Rayleigh backscattering caused by inherent inhomogeneities in the fibers (Waagaard et al., 2021). Such inhomogeneities are displaced by, e.g., acoustic waves from whales impinging on the fiber, which can be detected as phase changes in the Rayleigh backscatter. The time-differentiated phase is obtained by continuously comparing the backscattered response from one pulse to the next. This is typically done in two steps: first, the phase is

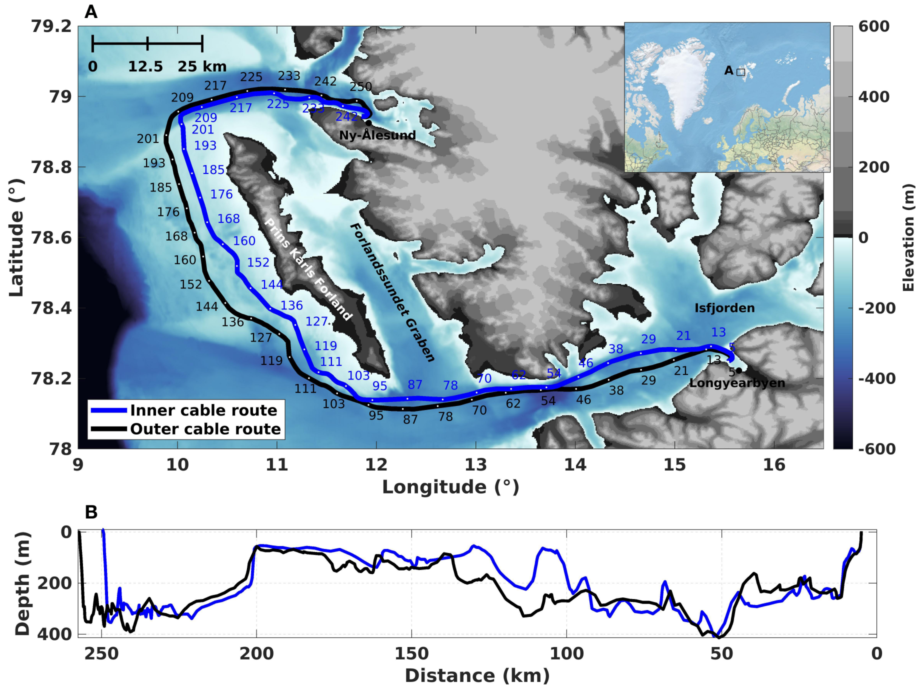


FIGURE 1
An overview map of the study area. **(A)** The fiber-optic telecommunication cables are located between Longyearbyen and Ny-Ålesund. Starting in Longyearbyen, the cables enter the Adventsfjorden 5 km after the interrogator unit; they then cross Isfjorden into the open ocean, bypassing Prins Karls Forland before entering the Kongsfjorden north of the settlement at Ny-Ålesund. The entire length of both telecommunication cables was interrogated using four interrogator units. However, only data from the two units located in Longyearbyen have been used in this work. **(B)** The ocean depth profiles along the fiber paths, the cables are trenched an additional 0-2 m into the seafloor.

spatially differentiated between regularly spaced channels along the optical fiber, then compared to the next pulse's response. The time-differentiated phase is stored by the interrogator and later converted to fiber strain during the data processing (see, e.g., Hartog (2017), for more information on this conversion). Normal signal strength decay along the fiber is about 0.2 dB/km, depending on the quality of the optical fibers and connectors. The returned signal strength from 100 km is, therefore, ≃-40 dB with respect to 1 km. To date, the maximum cable length that has been interrogated with a signal above the noise floor is 171 km, applying low-loss cables. On commercial telecommunication cables, the range is typically 140–150 km (Waagaard et al., 2021). Therefore, to collect data on the entire 260 km length of each separate telecommunication cable, four interrogator units were deployed; two in Ny-Ålesund and two in Longyearbyen, one for each direction on each telecommunication cable. Both optical fibers were dark, i.e., not used for data transfer. We denote the two telecommunication cables as the 'inner cable' (the cable closest to Prins Karls Forland) and the 'outer cable' (the cable furthest from Prins Karls Forland; see Figure 1). The first 5 km portions of the telecommunication cables at the Longyearbyen and Ny-Ålesund ends are trenched on land. The subsequent 248 km portion of the inner cable and the 252 km for the outer cable are sub-sea cables buried 0–2 m below the seafloor. Only data recorded on the sub-sea part of the telecommunication cables have been analyzed in this work.

The four interrogators were installed over two periods in the summer of 2022. The first interrogator started recording in Ny-Ålesund 2022.06.02 and is still recording as of 2023.03.20. The latter three were installed roughly two months later, the first two in Longyearbyen on 2022.08.17 and the last unit on 2022.08.19 in Ny-Ålesund. The OptoDAS interrogators were each connected to a different optical fiber within the inner and outer telecommunication cables. The same optical fiber was not used with more than one interrogator for two reasons: (1) The laser pulses emitted by an interrogator at one end of a fiber could damage a unit connected to the same fiber at the other end. (2) There could be unknown interference effects arising from two overlapping laser pulses traveling in opposite directions in the same fiber. All four interrogators used a gauge length of 8.16 m with virtual channels sampled every 4.08 m over a 136 km distance. The recorded data were transferred near-real-time to NTNU in Trondheim, Norway, using the network infrastructure described in Landrø et al. (2022). The data were also saved locally on Network Attached Storage (NAS) discs on the interrogator units as a backup. We used light pulses with a free-space wavelength of 1550 nm with a sampling period of 1×10^{-8} s at the optical receiver. The data were recorded at a sampling frequency of 625 Hz. The interrogated distance ensured that the entire length of both telecommunication cables was covered, with a ≃10 km overlap east of Prins Karls Forland (roughly ±5 km of the 130 km mark in Figure 1A). The three interrogators installed in mid-August recorded data continuously for roughly two months (with some breaks due to, e.g., power outages) before being shipped back to mainland Norway. Only data recorded by the interrogators in Longyearbyen have been analyzed in this work.

2.2 Air gun data

A dedicated air gun survey was performed on 2022.09.06 to allow us to estimate the errors related to the two localization methods used in this work. Permission was given by the authorities in Svalbard to minimizing disturbance and potential harm to marine life. No whale vocalizations were observed in the DAS data from the survey period. The air gun was towed 10 m behind the ship, which recorded its position every second from an onboard Global Positioning System (GPS).

The chamber of the air gun was 0.1311 m^3 (20 in^3), and the average chamber pressure used was 12000 kPa (120 bar), producing a source of 100–200 kPa-m (1–2 bar-m; the pressure one meter from the source, corresponding to 220 dB re. 1 μPa at 1 m). Note that fin whale source levels observed along the Norwegian coast is reported to be 190.5 ± 7.4 dB (Garcia et al., 2019). The air gun shots produced a valuable data set that we have used to calibrate the analysis in this paper.

The air gun data were acquired for roughly 3.5 h on 2022.09.06 with an inter-shot interval of ≃60 s (with some longer breaks due to malfunctions in the acquisition system), resulting in a total of 183 shots. For this work, we selected a 41-shot sub-set covering source locations above each cable and in between (see Figure 2). When fired directly above one of the cables, an air gun signal could be observed over ≃6 km in each direction along the cable as well as on the adjacent fiber optic cable if the inter-cable distance was lower than 6 km (see Supplement Material Figure S1A for an example of a recorded air gun shot). It is possible to increase the detection range by applying a filter in the frequency-wavenumber domain (*f-k* filter), which provides processing gain.

2.3 Fin whale data

To test our methods for baleen whale tracking, we selected a 5.1 h portion of the DAS dataset acquired on 2022.08.22 that proved to contain 1808 fin whale 20 Hz-calls. Note that we performed laser sweep calibrations on both interrogator units between 11:13 and 11:23 on that day. The fin whale song is described in the literature as a series of 20-Hz-centered down sweeps of ≃ 1 s duration (Thompson et al., 1992; McDonald et al., 1995). These calls, thought to be produced only by males, have been recorded around Svalbard between July and September/October (Ahonen et al., 2021), in association with local ship-based visual surveys. A subset of 188 calls was used to demonstrate the feasibility of using two parallel fiber-optic cables to track several whales simultaneously.

Figure 3 shows an example of a fin whale 20 Hz-call series recorded over a 90 s time window on both the inner and outer cables between 50 and 100 km on 2022.08.22 at 12:27:57 UTC, where (A, B) are spatio-temporal representations underlining the presence of at least three vocalizing whales at ≃60 km, ≃80 km and ≃95 km, (C, D) the waveforms of the calls emitted by the whale at 60 km and (E, F) the associated spectrograms. Note that the distances given (in this case, 50 to 100 km) are the length of fiber from the interrogator to the channel recording the vocalization.

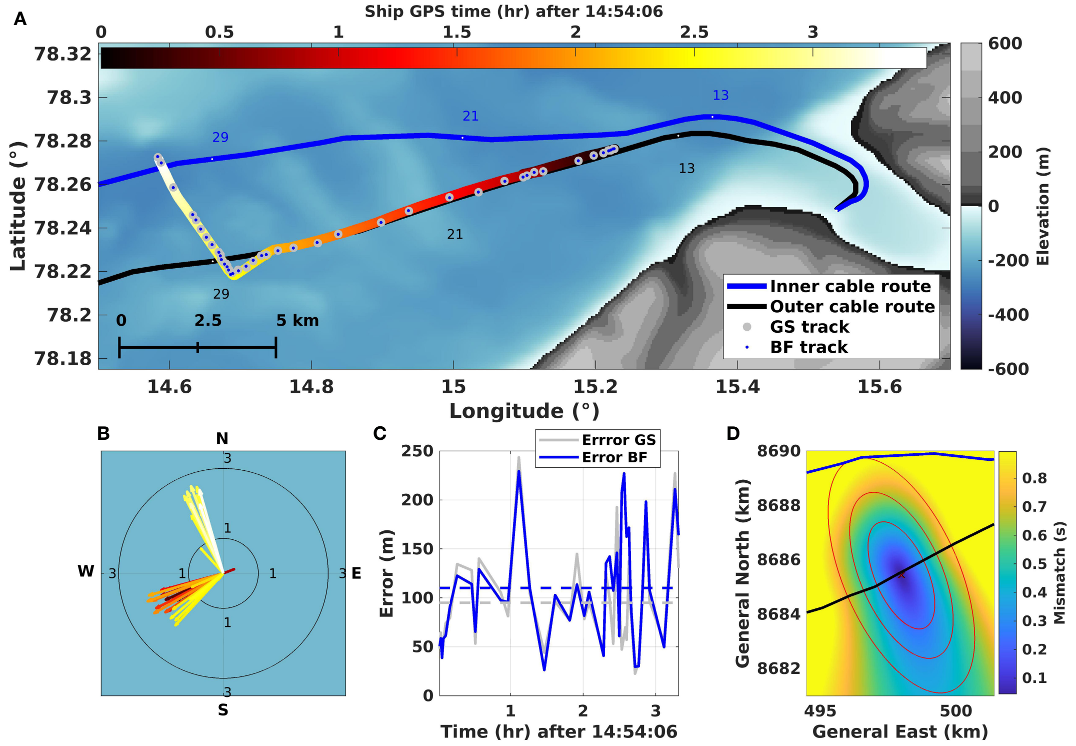


FIGURE 2
Tracking the air gun with known shot locations. **(A)** Overview of the first 35 km section of cables, showing ship track, color-coded from dark to light (by origin time of shots) overlaid with the estimates from the grid search (GS; big gray circles) and the Bayesian filter (BF; small blue circles). **(B)** The associated velocities of the ship are estimated by the BF with the same color coding as **(A)**. **(C)** The error between the estimated location and the known position of the ship at the shot origin time. **(D)** RMS mismatch from GS overlaid by the error ellipse (1x, 2x, 3x standard deviation) from the BF based on a representative shot with an error (100 m) close to the mean error of all shots.

From the repetitive 20 Hz down-sweep signals, their frequency span (between [18-25] Hz) with a duration less than 2 s and average inter-call intervals of 13 s (for the whale at 60 km; the whales at 80 km and 95 km have respective average intervals of 10 s and 9 s), we can clearly identify these as characteristic fin whale 20 Hz song, as well as an example of a back-beat at roughly 90 s (Watkins et al., 1987; Thompson et al., 1992).

3 Methods

3.1 Data conditioning

The following pre-processing steps were carried out to prepare the data for analysis and improve the signal-to-noise ratio (SNR). Initially, the time-differentiated phase was converted to longitudinal strain (Hartog, 2017). Then the data were resampled in time and space to cover the regions of interest. Recorded air gun signals contained frequencies between 5 and 45 Hz, and the sample rate was reduced by a factor of 5, from 625 Hz to 125 Hz. Similarly, the fin whale vocalization carried frequencies between 18 and 25 Hz and was resampled by a factor of 8 to 78 Hz. The data were tapered and band-pass filtered to focus on the frequency band of interest. A Tukey window was applied to the data and subsequently bandpass-filtered using a fourth-order Butterworth filter to focus on each signal of interests' dominant frequencies. The cut-off frequencies were chosen to focus on the frequency range listed above. Next, a 2D median filter over 3×3 data-points was applied to suppress common-mode noise. Finally, a frequency-wavenumber (*f*-*k*) fan filter was applied to preserve only acoustic waves propagating around sound speed in sea-water and sediments, keeping everything with a propagation speed between 1000 and 3000 m/s.

3.2 Time of arrival picking

The data were visually inspected using spatio-temporal representations to identify signals of interest (Bouffaut et al., 2022). This visualization gave an overview of the recorded signals over several channels and was used by one observer to pick the first times of arrivals manually. The first time of arrivals were chosen at a sudden amplitude and frequency changes in the onset of the acoustic signal. To constrain the time of arrival picking burden, a maximum of 12 arrival times were selected from both cables, with an average inter-channel distance between the picks of 600 m (depending on the range from source to receiver). An exact inter-channel distance was not used due to the variation of SNR along the cable. We avoided using arrivals on the tail and focused on arrivals near the apex to avoid picking in the portions of the data dominated by normal modes. However, due to the known directivity of DAS

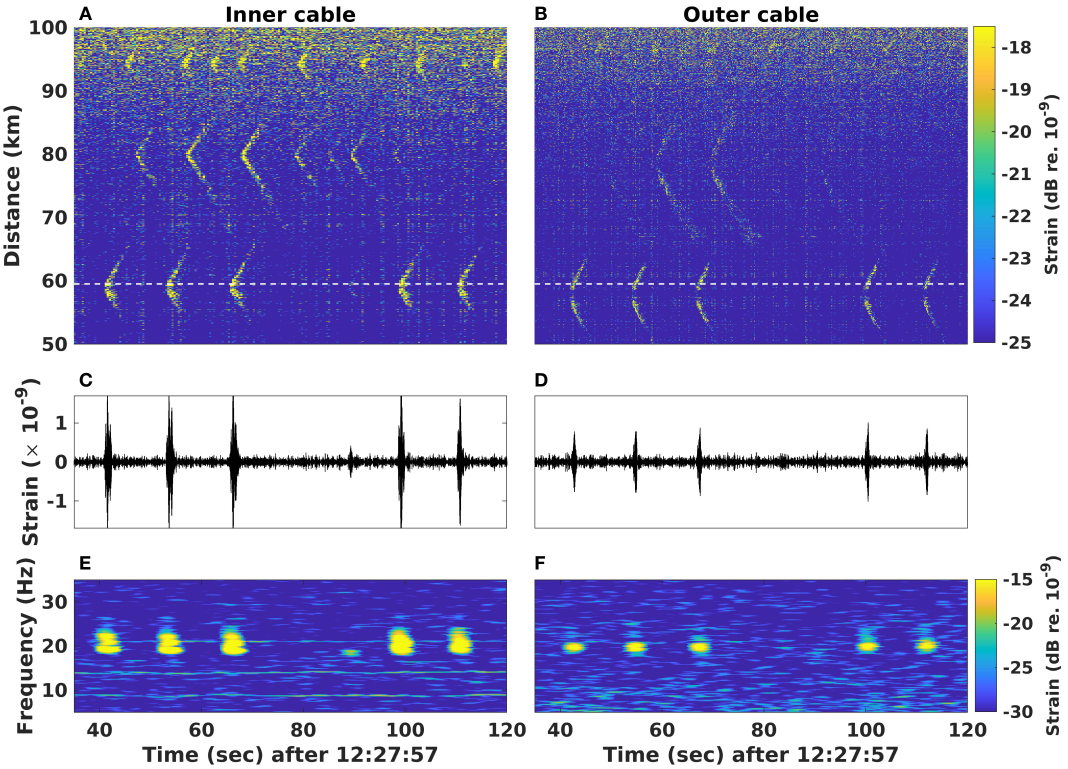


FIGURE 3
Distributed acoustic sensing recording of fin whale vocalizations on two fiber-optic cables on 2022.08.22 at 12:27:57 UTC. **(A, B)** Spatio-temporal ($t - x$) representation of three simultaneously vocalizing whales on the inner cable **(A)** and the outer cable **(B)** between 50 and 100 km. Waveforms **(C, D)** and corresponding spectrograms ($t - f$; **E, F)** at 59.52 km, represented by the white dashed line in **(A, B),** displaying portions of two series of fin whale 20 Hz calls and a back-beat at ≃95 is recorded on both cables with an average inter-call interval of 13 s. The spectrograms were computed using a Hann window of 512 samples with 98% overlap.

arrays (see, e.g., Papp et al. (2017)), we did not pick arrivals in close vicinity to the apex. When the signal quality from the two cables was significantly different, which was often the case when a whale was vocalizing near one cable, and at a greater distance to the other, fewer picks were selected from the cable with the poorer quality. In the worst cases, we used only one pick from the cable with lower signal strength only to resolve the left-right ambiguity.

3.3 Track estimation

Two different localization methods were developed to estimate the source positions from travel time information detected at the fiber-optic cables: a Grid Search (GS) method and a Bayesian Filter (BF). We denote the observed time of arrivals at the fiber channel locations by $T_{i,j}^{(obs)}$, $i = 1, \ldots, n_j$, $j = 1, 2$. Here, index i refers to the channel used to pick the time of arrival, and j denotes the telecommunication cable used (e.g., $j = 1$ corresponds to the inner cable and $j = 2$ to the outer cable). Alternatively, one could list these arrival times data in a length $n_1 + n_2$ vector, but the i and j index clarify that data are acquired and processed on two different cables.

In the forward model, we assume that a signal is emitted at time η from location (x, y, z), where x and y are the geographical coordinates and z is the depth. We use a reference depth of $z = z_{ref}$ set to 20 m for whales (known to be a typical call depth for fin whales (Stimpert et al, 2015) and 3.5 m for the air gun (set prior to the acquisition). The theoretically modelled arrival time, $T_{i,j}(x, y)$, is computed from the origin time of the event studied, η, and the travel time for the Euclidean distance from the event to the receiver, $t_{i,j}(x, y)$:

$$T_{i,j}(x, y) = \eta + t_{i,j}(x, y), \tag{1}$$

$$t_{i,j}(x, y) = \frac{\sqrt{(x - x_{i,j})^2 + (y - y_{i,j})^2 + (z_{ref} - z_{i,j})^2}}{c}.$$

where $(x_{i,j}, y_{i,j}, z_{i,j})$ is the coordinate of channel i at fiber j and $c = 1440$ m/s is the speed of sound in water. The sound speed is based on previous study of Arctic water (Gavrilov and Mikhalevsky, 2006) and the minimum error obtained in the air gun study.

3.4 Grid search position estimation

The GS method uses equally spaced grid points around a prior guess of the source location. Such a systematic search computes the travel time for all possible locations in a survey area to find the best matching source location. This kind of GS procedure has been adapted from earthquake seismology (see, e.g., Havskov and Ottemoller (2010)) and it has been previously applied to earthquakes recorded on DAS arrays by Rørstadbotnen et al. (2022). GS methods are also commonly used for whale localization (see, e.g., Wilcock (2012); Abadi et al. (2017)).

In this work, the grid points are defined from the channels receiving the first signal and laid out on a horizontal grid of (x, y) coordinates. The grid points were set at 25 m intervals, covering a 20 km by 20 km area centered around the receiving channel. For each element of the grid, the reference time η is calculated by shifting the computed origin time to the 'correct' timing, given by:

$$\eta = \frac{1}{n_1 + n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left[T_{i,j}^{(obs)} - t_{i,j}(x,y)\right]. \tag{2}$$

When the theoretical travel time and reference time in equation (2) are known for each grid point, the root mean square (RMS) misfit between observed and computed travel times is given by:

$$M(x,y) = \left(\frac{\mathrm{O(x,y)}}{\mathrm{N}}\right)^{1/2}, \mathrm{O(x,y)} = \sum_{j=1}^{2}\sum_{i=1}^{n_j}\left(\mathrm{T}_{i,j}^{(obs)} - \mathrm{T}_{i,j}(x,y)\right)^2, \tag{3}$$

where the normalization in the RMS computation is typically performed based on the total number of time of arrival picks used, $N = n_1 + n_2$. The estimated location of the source is the grid point with the lowest misfit, i.e., the global minimum RMS across the GS volume. The misfit function depends on the geometry of the source, the time of arrival picks, and cable pick locations.

Rather than evaluating the mismatch on a dense grid, Newton's method can be used to solve for the position (x, y). Using boldface for vectors and matrices, we start with an initial guess for the position $\boldsymbol{p}^0 = (x^0, y^0)$ and iterate by:

$$\boldsymbol{p}^r = \boldsymbol{p}^{r-1} - \left[\frac{d^2 l(\boldsymbol{p}^{r-1})}{d\boldsymbol{p}^2}\right]^{-1}\frac{dl(\boldsymbol{p}^{r-1})}{d\boldsymbol{p}}, r = 1, \ldots . \tag{4}$$

Here, $l(x,y) = -O(x,y)/2\tau^2$ is a statistical likelihood version of equation (3), where we assume that the travel time observations are made with additive independent and Gaussian distributed noise with variance τ^2. The guided search provided by equation (4) converges within a few iterations r. We denote the resulting optimum value by $\hat{\boldsymbol{p}}$. This kind of optimization involves re-setting the reference time η (see equation (2)) at each step of the iterative scheme. By investigating residuals from modeled and known air gun shot positions, the noise standard deviation was found to be $\tau = 5.4$ ms.

A probabilistic view enables uncertainty quantification. In particular, the curvature of $l(\boldsymbol{p})$ at the optimum value can be used to assess the covariance matrix of the position. Together with the optimum, this leads to a Gaussian approximation for the position:

$$\boldsymbol{p} \sim N(\hat{\boldsymbol{p}}, \hat{\Sigma}), \hat{\Sigma} = -\left[\frac{d^2 l(\hat{\boldsymbol{p}})}{d\boldsymbol{p}^2}\right]^{-1}, \tag{5}$$

where $N(\hat{\boldsymbol{p}}, \hat{\Sigma})$ denotes a normal distribution with mean $\hat{\boldsymbol{p}}$ and $\hat{\Sigma}$ covariance matrix holding the variance and covariance of the position estimate. The ellipse-shaped contours of the Gaussian probability density function (PDF) provide a comparison with that achieved using the GS, as shown in Figure 2D. If there is prior information about the position $\boldsymbol{p}$, a Bayesian approach with a Gaussian approximation for the posterior PDF works similarly.

3.5 State space model and Bayesian filtering

With the association of a state space model and a BF, we can connect the predictions at different locations and also consistently estimate the source speed and direction. Similar uncertainty assessments to travel time data have been applied to, e.g., satellite positioning accuracy (Yigit et al., 2014) and location estimation with borehole seismic data (Eidsvik and Hokstad, 2006).

Let s_k, $k = 1, \ldots$ denote time steps, and represent positions and swim velocities at time s_k by $\boldsymbol{m}_k = (x_k, y_k, u_k, v_k)$. Position $\boldsymbol{p}_k = (x_k, y_k)$ and east and north velocities u_k and v_k are next coupled over time in a state space model. Travel time data $T_{i,j,k}^{(obs)}$ at time s_k are modelled *via* the likelihood model $l(\boldsymbol{m}_k)$ as described below equation (4). Note that measurements carry direct information only about the position $\boldsymbol{p}_k$. However, now that a model connects positions and velocities over time, data are implicitly informative about all process variables.

For the dynamical model part, taking state variables from one time to the next, we assume constant velocity and additive noise $\boldsymbol{\varepsilon}_k$ such that:

$$\boldsymbol{m}_k = \boldsymbol{A}_{k-1,k}\boldsymbol{m}_{k-1} + \boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_k \sim N(\boldsymbol{0}, \boldsymbol{S}_{k-1,k}), \tag{6}$$

$$\boldsymbol{A}_{k-1,k} = \begin{bmatrix} 1 & 0 & \Delta_{k-1,k} & 0 \\ 0 & 1 & 0 & \Delta_{k-1,k} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where the time interval $\Delta_{k-1,k} = s_k - s_{k-1}$ and the matrix $\boldsymbol{S}_{k-1,k}$ is a 4×4 positive definite covariance matrix for the noise in this process model, accounting for potential acceleration. Equation (6) defines a Markovian structure where the state at time s_k depends on the previous states only through the one at time s_{k-1}. An associated formulation of this model in equation (6) is provided by a conditional PDF $p(\boldsymbol{m}_k|\boldsymbol{m}_{k-1})$. Initially, at the time of the first detection, we assume a vaguely informative prior model $\boldsymbol{m}_1 \sim N(\boldsymbol{\mu}, \Sigma)$, where the mean $\boldsymbol{\mu}$ is composed of the average positions of the cable locations where the first picks are detected and with 0 velocity. The covariance matrix Σ is set to have significant standard deviations (1000 m for positions and 5 m/s for velocities).

We denote all data available up to time s_k by:

$$\boldsymbol{D}_k = \left\{T_{i,j,k'}^{(obs)}; i = 1, \ldots, n_{j,k'}; j = 1, 2; k' \leq k\right\}.$$

The data are assumed to be conditionally independent over time. With the Markovian assumptions and that of conditionally independent measurements, methods from Bayesian filtering allow us to estimate the location and velocity $\boldsymbol{m}_k$ at a time s_k, given all data up to that time k (see, e.g., Sarkka (2013)). For efficient calculations that can be run online, we fit a Gaussian PDF to the filtering distribution at each observation time. This represents extensions of equations (4) and (5), where the mean of the predictive distribution given earlier data is used to initiate the Newton search. In terms of probability distributions, this means we are approximating the filtering PDF $p(\boldsymbol{m}_k|\boldsymbol{D}_k)$ by a Gaussian model. This filtering procedure is run recursively over time steps $k = 1, 2, \ldots$, and it offers a highly applicable tool for conducting real-time analysis of the travel time data.

By having positions and velocities in a state space model, we can obtain probabilistic state estimates and associated uncertainties at any time s_k, not only at the measurement sites and times.

3.6 Computation of empirical DAS detection range

Using the estimated whale position $(x_w, y_w, z_{ref} = 20 \text{ m})$ from either the GS or the BF, we estimated an empirical detection range from this position to the position of the DAS channel that detected the last whale signal $(x_{i,j}, y_{i,j}, z_{i,j})$ before it is lost in the background noise level:

$$r_{max} = \sqrt{(x_w - x_{i,j})^2 + (y_w - y_{i,j})^2 + (z_{ref} - z_{i,j})^2} \tag{7}$$

where r_{max} denotes the maximum empirical Detection Range and is taken as a straight line distance.

4 Results

4.1 Air gun tracking

The geo-referenced air gun survey detailed in Section 2.2 is used to calibrate and evaluate the error of both localization methods (Wuestefeld et al., 2018). Figure 2A shows the track obtained from the ship's GPS overlaid with the locations of the 41 shots as obtained using the grid search (GS; Section 3.4) and the Bayesian filter (BF; Section 3.5). Figure 2B displays the BF estimate of the direction and speed of the ship for the whole track. The warm color scale indicates time on the GPS trajectory on both sub-figures. Figure 2C displays the localization error of each method evaluated for air gun locations 10 m behind the ship's GPS positions. Figure 2D shows the mismatch between observed and calculated arrival times from the GS overlaid with error ellipses from the BF.

Both methods follow the ship's track for the selected 3.5 h of recordings. The estimated speed is roughly 1.8 m/s, which agrees with the average speed obtained from the GPS position, 1.9 m/s (see Supplementary Material Figure S1C). There are three areas where the BF underestimates the ship speeds, each associated with the ship making a sharp turn or a loop. In these cases, the BF predictably estimates the ship to travel a shorter distance than it actually does. The estimated mean error standard deviation is 94.9 ± 51.1 m and 108.8 ± 53.7 m for the GS and the BF, respectively. Errors are highest when the ship makes turns and when the signal quality on one cable is poor.

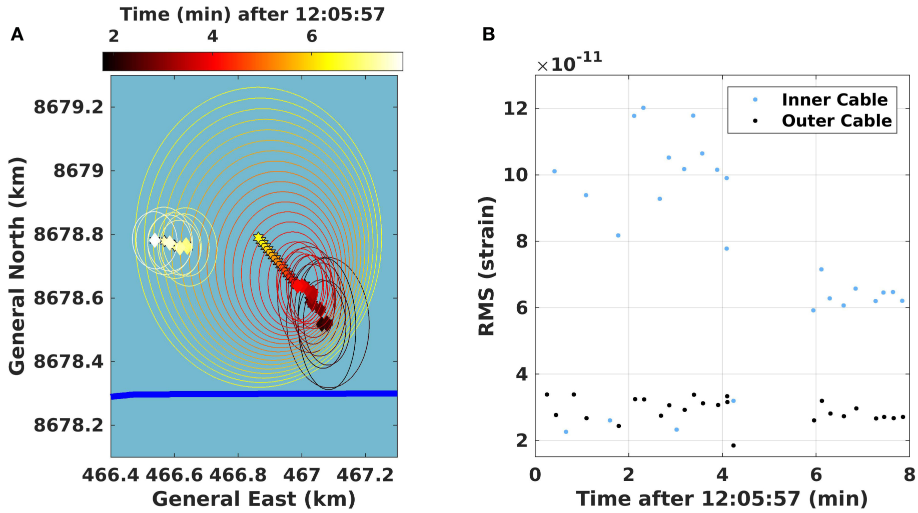


FIGURE 4
Prediction of whale location during an inter-series interval. **(A)** A subsection of a whale track near the inner cable (bold blue line; 12:06 – 12:14) with a 1 min 51 s inter-series interval. BF tracks are computed roughly every 10 s, and in inter-pulse intervals, we predict the position without updating (indicated by hexagram). Diamonds indicate the predictions with observed data, while the ellipses show 90% coverage regions constructed by the BF. **(B)** The associated RMS amplitude of the calls used in the localization.

4.2 Fin whale tracking

Figure 4A shows the application of the BF to evaluate a fin whale track over an interrupted series of calls (data at 12:07 to 12:14, later referred to as an 8 min portion of whale track B) and Figure 4B the corresponding RMS levels of the calls. The RMS levels are computed in the spatio-temporal domain by a rectangular window enclosing the recorded whale call. For example, for the whale call at 60 km in Figures 3A, B, a rectangular window covering 5 km on each side of the apex and from the start of the apex to 6 s after the apex is used.

Each diamond in Figure 4A indicates predictions based on observed data, while hexagrams represent intermediate predictions without observed data. The ellipses correspond to the error expressed in the BF version of equation (5), with a 90% coverage region. The plot shows a portion of whale B's track lasting roughly 8 min and an Inter-Series interval (ISI) of 117 s. From the BF error ellipses, it is clear that the computed variances decrease as the whale track builds up over many time steps. Locations are computed roughly every 10 s for the track. When the Inter-Call interval (ICI) is longer than 10 s, the BF predicts the location based on the previous track locations and velocities without updated uncertainty. Here, we have an ISI of 2 min, and we compute 12 predictions with a straight-line prediction and increasing uncertainty. When the whale vocalizes after the 2 min pause, the BF finds this new position based on the last prediction, and the new location is within the uncertainty of the prediction. This means that we are, most likely, still tracking the same whale. However, the whale did change course, heading more West during the ISI and it increased its swim speed somewhat, traveling slightly further than the BF prediction. We are thus able to show that interruptions in vocalization can be associated with changes in swim speed and direction.

Figure 5A shows the overview plot of all whale tracks color-coded by the time of vocalizations, where Figures 5B, D, F, H are zoomed-in representations of the tracks, and Figures 5C, E, G, I, the directional swim speeds estimated by the BF. Following the protocol described in the previous paragraph, we found eight distinct whale tracks within the five analyzed hours, denoted by the sequential letters (A) to (F) according to the time of the first vocalization in each track.

Whale track (A) started at a distance of 50 km from the interrogator unit at 09:28 (all times are given in UTC) and headed East for 72 min before contact loss. Just before losing contact with whale (A), whale track (B) started at 54 km, heading North-West for 160 min before the signal ceased. During this period, we also show the start of a vocalization series from three

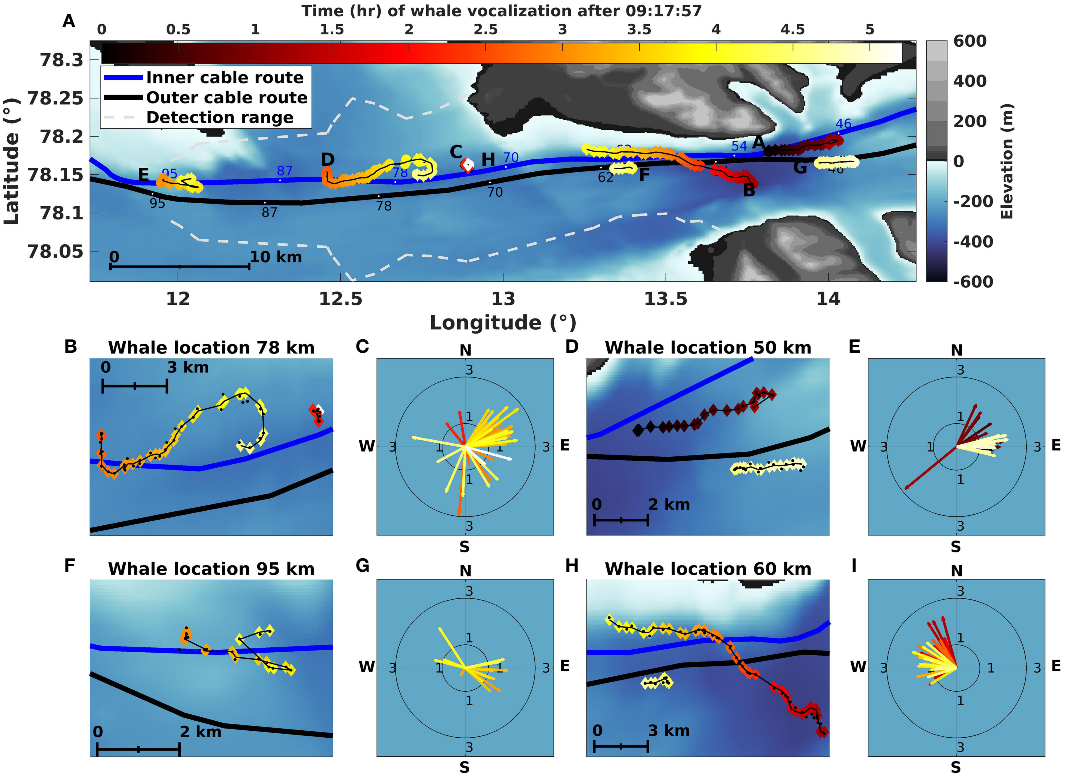


FIGURE 5
Simultaneously tracking multiple whales using fiber-optic cables in the Arctic. **(A)** Overview of a 60 km long section of the cables, showing the positions and tracks of up to eight acoustically-detected whales, color-coded from dark to light over a 5.1-hour period. Note that the GS and the BF have been plotted in the same color and shape to better illustrate the tracks. The two dashed lines show an attempt to estimate an empirical whale detection range directly from the DAS data using vocalizations every 10 min for each whale track. **(B, D, F, H)** Show zoomed-in detailed positions and tracks of the four areas with whale detections. We use the same color coding as in **(A)** for the BF, while GS has been plotted as black dots to better see the differences in the estimated positions. **(C, E, G, I)** show the corresponding swim velocities as polar plots for these tracks.

other positions along the cable, around 60 km, 80 km, and 95 km (see Figure 3), associated with tracks (F), (D), and (E), respectively.

Whale track (C) started around 72 km and headed North. The contact lasted roughly 10 min from 11:23. The interrogator calibration mentioned in Section 2.3 happened between 11:13 and 11:23, and vocalizations from both whale (B) and (C) were lost. No calls associated with whale track (C) are observed in the data prior to 11:13.

Whale track (D) started just after whale track (C), at 11:42, at roughly 80 km from the interrogator unit, lasting 113 min. This track went South before heading North-East over roughly 7 km before looping back West before contact loss. A few vocalizations are localized near track (C) but three hours later (14:20). They are denoted by track (H).

Whale track (E) started the furthest from the interrogator. Calls started at 12:20, South of Prins Karls Forland at 95 km and were detected for 71 min. The background noise level is significant at these long distances, making arrival time picking challenging, especially for the outer cable due to elevated noise and the high inter-cable distance. This increased the uncertainty of the estimated track. Nevertheless, the track headed East for roughly 2.7 km, where it turned and shifted to the North-East. Considering the uncertainty mentioned above and localization error of ≃100 m, two whales could have vocalized simultaneously and parted ways after 1.5 km, one continuing East and the other North-East.

Whale track (F) started at 62 km at 13:30, 18 min after the end of track (B), and headed West. The 4.8 km distance between the two tracks would have required the whale of track (B) to have more than doubled its maximum swim speed (from 2 m/s to 4.4 m/s). While this is not entirely unrealistic (fin whales have been reported to be able to increase swim speed between singing bouts (Clark et al., 2019)) we chose the conservative approach of considering the two tracks separately.

The first column of Figure 6 shows the RMS of received fin whale vocalization amplitudes for whales tracks (B), (C), (D), (F), and (H) over the 5.1 h studied (see Supplementary Material Figure S2 for whales (A), (E), and (G)), calculated as in Figure 4B. We assume that a given fin whale produces 20 Hz songs with an almost stable source level over a short period (Watkins et al., 1987; Garcia et al., 2019). The peaks traced out by the RMS values of detected strain reflect the fact that whale B crosses the outer fiber at a more perpendicular angle than the inner cable, the second peak being less sharp than the first. The minimum strain at which we can detect the vocalizations is observed to be $\simeq 2 \times 10^{-11}$, which constrains the maximum detection range. It is also interesting to note that the RMS amplitude peak of the two whale B crossings are similar, indicating that the two cables have similar sensitivities. The second column in Figure 6 shows a zoomed-in version of the calls of whale tracks (B) and (D) to illustrate the ICI and the ISI (for the same plot for whales (A) and (E) see Figure S2).

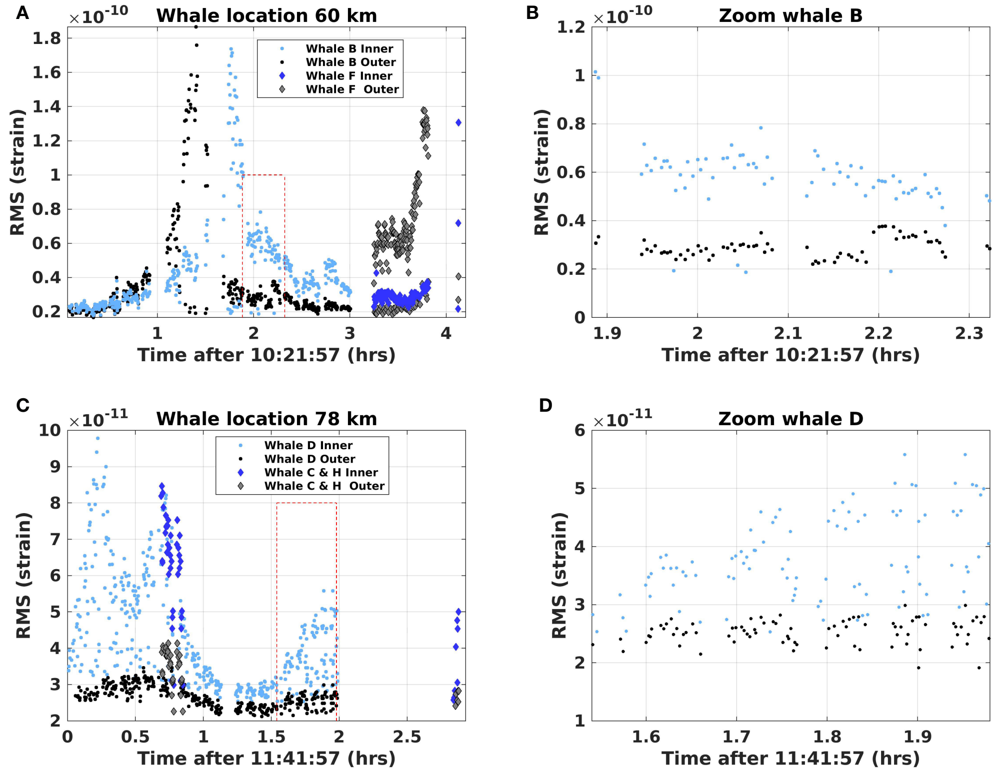


FIGURE 6
Observed RMS amplitudes of fin whale vocalizations in whale locations 60 km and 78 km (see Figure 5) computed by a rectangular window around the calls. **(A, C)** All RMS levels from the vocalizations in the respective locations. **(B, D)** Zoomed-in versions to illustrate periods with whale calls and the inter-series intervals.

We evaluated an empirical DAS detection range following Section 3.6, every 10 min, for all whale tracks. The detection range is represented as white dashed lines in Figure 5A. The obtained range does not monotonically decrease with fiber distance as whale track (D) at 80 km shows a longer range, 9.4 km, compared to, e.g., whale track (H) at 60 km with a range of 6.0 km. This suggests that other effects, such as coupling or the sediment properties, play a role or that the whales we observed at 80 km produced higher intensity calls than the whales at 60 km. This is not uncommon, e.g., Garcia et al. (2019) observed a significant variation in call levels and estimated fin whale sound intensity levels along the Norwegian coast to range from 186.3 ± 7.1 to 195.8 ± 4.4 dB. Moreover, whale (E) displays the lowest range (2.4 km). The large difference for whale (E) is believed to mainly be due to the increased background noise level at the location.

4.3 Fin whale individual characteristics

Over the 5.1 h of analyzed data, five fin whale tracks were analyzed over a minimum of 45 min, providing insight on a few critical measurements for the species at the individual scale, based on a male 20 Hz song. We analyzed the sounds associated with each fin whale trajectory, reporting the total number of calls, average ICI, average ISI, and average swim speed in Table 1. Averages are associated with their standard deviation denoted by the ± symbol. Note that we excluded the first and last call series in each track as it is not clear if they contain a complete set of vocalizations.

Blue whale song is geographically distinct with differences in call time-frequency characteristics and repetition [ICI and ISI; McDonald et al. (2006)]. Our description of fin whale song follows Watkins et al. (1987), where a call series is defined as a group of at least five fin whale vocalizations, ICI corresponds to the time stretch between two consecutive 20 Hz calls that are < 60 s, and ISI when the call interval is longer. Intervals are measured on a spatio-temporal representation of the signals, where the reference timings are picked along the same channel. Note also that the interrogator calibration break interval (11:13 to 11:23) was neither included in the ICI nor the ISI estimations.

While peak frequencies are fairly stable for fin whales, the most regionally-distinctive parameter is the ICI, e.g., Delarue et al. (2009) used ICIs in the North East Atlantic for stock differentiation. ICIs were found between 9.3 and 13.8 s, aligning with reports of systematically shorter ICIs in the Barents sea, measured post-2017 close to ≃10 s (Romagosa et al., 2022). Focusing only on the tracks with more than five observed ISI, we denote that the smallest interval was computed for whale (D; 96.9 s) and the longest for whale (B; 167.9 s), consistent with ISI reported by Watkins et al. (1987). ISIs varied substantially along track (B) with a standard deviation of 124.5 s, whereas whale D had the smallest ISI deviation of 22.2 s.

The BF uses 3–4 vocalizations to estimate swim direction. We found swim speeds between 1 and 3 m/s, with “instant” horizontal speeds between 1.24 m/s and 1.86 m/s (4.5–6.7 km/h). It is consistent with previously-reported fin whale swim speeds of 1.2–3.9 m/s during vocal activity (McDonald et al., 1995; Clark et al., 2019), generally lower than cruising speeds (Lydersen et al., 2020).

5 Discussion

5.1 Localization error

We used the known ship’s GPS associated with the air gun data to evaluate the GS and BF localization errors. The distance between the ship’s GPS position and the source was compensated by moving the logged source position from the GPS 10 m behind the ship following the known track. However, this is only a partial solution as the source will drift with abrupt changes in ship’s heading and with variations in local conditions, e.g., ocean currents and weather, during the acquisition. These variations could introduce some errors into the airgun position estimate.

Comparing the errors in airgun position estimation to the errors in the telecommunication cable coordinates and the time of arrival picks, we consider the error in the airgun shot location as negligible. The fiber track locations were logged while the cables were trenched into the substrata. Moreover, the DAS system saves data into channels related to the distance the laser has traveled

TABLE 1 Information on the different whale tracks, displaying the call period, the inter-call intervals, the inter-series interval, the mean swim speed and the number of calls during the different intervals.

Whale	Call period	Inter-Call interval	N_{call}	Inter-Series interval	N_{rest}	Mean swim speed
		(s)		(s)		(m/s)
(A)	09:20 - 10:32	9.3 ± 5.7	243	126.3 ± 58.8	12	1.53
(B)	10:23 - 13:19	13.8 ± 4.3	452	167.9 ± 124.5	17	1.40
(C)	11:23 - 11:33	13.6 ± 8.7	39	80.0	1	1.27
(D)	11:42 - 13:33	11.7 ± 6.0	447	96.9 ± 22.2	19	1.86
(E)	12:20 - 13:31	10.0 ± 7.4	324	116.7 ± 71.1	8	1.34
(F)	13:37 - 14:25	11.4 ± 3.1	154	83.4 ± 1.8	2	1.24
(G)	13:58 - 14:29	11.2 ± 2.8	137	–	0	1.70
(H)	14:28 - 14:31	12.2 ± 5.6	12	–	0	1.56

along the cable, which must be calibrated to the known telecommunication cable coordinates. As we know the coordinates along the fiber track, we can compute distances along the cable that roughly correspond to the distances logged by the interrogator equipment. However, the cable is not necessarily fully stretched; hence the distance calculated from the known fiber track position could deviate from the distances logged by the interrogator unit. Using the geo-referenced shot position, a future study could focus on developing an inversion procedure to match the distances that the interrogator logs with the known cable positions and quantify the potential errors.

First arrival times were manually picked on conditioned spatio-temporal representations of the data. We identify manual time-picking as a source of error directly related to the apparent SNR. For a specific DAS configuration and stable source characteristics as in the air gun example, the SNR will change with the distance between the source and the fiber (both on the horizontal plane and with changes in the water column depth), the ambient noise but also, changes in DAS recording capabilities all along the fiber.

Considering the configuration of (1) a pelagic sound source in the [5-45] Hz (air gun) and [15-25] Hz (fin whale) frequency band, (2) bottom-mounted receivers and, (3) a constrained detection range of 9.4 km we assumed a constant sound speed (and, therefore, straight ray travel time; see equation 1) to limit the computational costs of the localization methods. When estimating the air gun track, we tested various sound speeds and found the value with the lowest error to be 1440 m/s. The obtained sound speed agrees with average values given for sea-water at 4-6°C, which are typical for Svalbard (Timmermans and Labe, 2022), also used in previous studies of Arctic waters (see, e.g., Gavrilov and Mikhalevsky (2006)). We carried out a simple test to see the change in average errors between the known shot position and the estimated track. Increasing the sound speed by 25 and 50 m/s produced a 16 and 31 m increase in average errors. In the event longer detection ranges were available or to reduce the obtained 100 m localization error, 3D propagation modeling might be necessary to account for the complex bathymetry in fjord environments (e.g., using BELLHOP (Ocean-Acoustic-Library, 2022)), as highlighted in neighboring fjords (Richard et al., 2023).

The typical depth range for fin whale vocalizations is between 15 and 30 m (Oleson et al., 2014). We investigated a similar depth range using the air gun data to quantify the errors related to using a fixed source depth by changing the source depth of the air gun in the localization algorithms to 20 m while keeping the sound speed constant at 1440 m/s. The BF localization errors deviate less than 1 m, whereas the GS deviates with just over 1 m. The very low deviations in errors are likely because we are investigating depth variation well below the location accuracy.

5.2 Comparison of the localization methods

The two localization algorithms provide similar results for both the air gun and the whale vocalization and show similar localization errors of 100 m. For the air gun track, the GS showed a smaller error than the BF (Figure 2), whereas, for the whale tracks, the GS generated positions that are more scattered than the BF (Figure 5). This is attributed to the BF incorporating knowledge of previous locations, whereas the GS location estimates are independent for each vocalization.

Because of the manual first time of arrival picking, the methods from raw data to a position, swim speed and direction, and ultimately the tracks still need to be fully automated. The manual picks were limited by the SNR of the observed signal, which translated into a minimum received RMS strain of 0.2×10^{-10} on Figure 6 which in turn limited the detection and localization range to a maximum offset of 9.4 km from the fiber, for both methods. For fin whales' 20 Hz song, the arrival times selection could be automated, e.g., by a time-frequency matched filter (also known as spectrogram correlation; Mellinger and Clark (2000)) or by adapting recent advances in machine learning and convolutional neural networks to analyze the large amount of DAS-collected data and detect whale calls with high accuracy (Shiu et al., 2020). Automating the time of arrival picking could help extend the localization's spatial reach and potentially reduce the localization error.

The implementations of the algorithms used to find the tracks in this paper are computationally different. The GS is a brute force algorithm that finds arrival times for grid points over a pre-defined area and uses, for instance, 1.9 s to locate one air gun shot. Conversely, the BF is more computationally efficient and uses 0.15 s to locate one seismic shot. These computational times denote the applicability and potential for near-real-time whale tracking. The computational efficiency of the methods strongly depends on the computer used, the processing unit (CPU vs. GPU), the programming language, and the algorithms' efficiency, and it can be optimized for better performance.

Under the considerations of near-real-time tracking, the BF stands out as the most complete method as it estimates the heading and speed of the whales in addition to its geographical position. In contrast, the GS only estimates the geographical position of the whale call. However, it is a global optimization method with few assumptions in the inversion procedure. It is limited by the grid spacing, the size of the area covered by grid points, and the complexity of travel times calculations. The latter limitation is also valid for the BF. Additionally, it computes the misfit function for the whole grid area, which might show features in the objective function not captured by the BF. This can serve as a comparison to the BF (as done here) to find local minima in the grid area or provide an accurate starting model for the BF (or a similar local optimization method dependent on an accurate initial model). For example, when one cable is used, it will highlight the left-right ambiguity problem in the misfit function. The contours from the GS position, as depicted in Figure 2D, align well with the various error ellipses from the BF, which indicates that the Gaussian approximation used in the BF is a good approximation for situations like those studied here.

5.3 Implication of DAS-based localization for whale conservation

Telecommunication cables are available worldwide, and most new cables have more optical fibers than required to create redundancy. These could be repurposed to create distributed sensors from their onshore termination point to the first repeater. However, there are a limited number of places where two parallel fibers are laid a few km apart, as in this study. Even with one fiber, we can still detect and estimate the direction (with a left/right ambiguity) of vocalizing whales, with range inferred by, e.g., the received levels and the directional ambiguity perhaps broken by an asymmetry in the bathymetry (see, e.g., Bouffaut et al. (2022)). However, additional work is needed to quantify the instrument response of DAS, its directivity, and sensitivity at higher frequencies to understand which species can be monitored with this technology and under what circumstances.

The localization accuracy of the two methods presented in this work has been shown to be ≃100 m while offering the possibility to localize whales over a ≃1800 km^2 area (considering whales being detected at a distance of up to 95 km from the interrogators in Longyearbyen, with a detection range of ≃9.4 km). This combination of spatial coverage, relatively low infrastructural investment, and potential for real-time monitoring could bring great value to a range of coastal conservation applications.

For example, driven by the North Atlantic right whale case, several acoustic-based methods are being developed or applied to mitigate the risk of ship strikes (e.g., Gervaise et al. (2021); Baumgartner et al. (2019)). For this application, real-time methods are essential, and the obtained 100 m accuracy compares well with current practices (Hendricks et al., 2019). Such a collision management system is becoming increasingly important in the Arctic, where the decrease in sea-ice coverage opens up new shipping routes that might impact previously untouched whale habitats. At the same time, climate change is likely causing whales to change their habitat use and migration behavior. By automating the tracking procedure presented here, it would be possible to produce near-real-time tracks of whales with species identification, swim speed, and direction. At the very least, this information could be used to inform local management and stakeholders of whale presence and movement in a timely manner and at a very fine spatial resolution, supporting them to take appropriate mitigation measures. Furthermore, detailed information on whale locations and behavior would support science-based decisions and the management of conservation areas where anthropogenic activities should be kept at a minimum. Svalbard is a known summer feeding ground for baleen whales (Storrie et al., 2018), and mapping the common or active feeding areas is important. A final possibility is to find information on how whales respond to the changing marine ecosystem in the Arctic, induced by climate change, which has already been shown to alter baleen whale behavior (Moore et al., 2019). A global effort is needed to obtain such information, in which DAS systems can provide detailed information through high spatial and temporal sampling.

6 Conclusion

Examining five hours of data from two fiber-optic telecommunication cables, we have detected 1808 fin whale vocalizations. A subset of these data has been used to track eight whales during this period, with up to four being tracked simultaneously at a detection range from the cable of up to 9.4 km. This capability opens up new possibilities for detailed mapping of the presence and location of whales over large areas (at least 60 km long and ≃20 km wide) over long periods, in near real-time. The simultaneous tracking of multiple whales also has the potential to provide new insights into the behavior and interaction between whales along a corridor up to ≃9.4 km on either side of the cable. Using shots from a single air gun towed behind a ship, we were able to calibrate our two localization estimators (a grid search and a Bayesian filter) with deterministic signals of similar source level to fin whale calls and found both methods to be computationally efficient and accurate to ≃100 m. Using two fiber cables, separated by a few km, breaks the symmetry of a single straight-line array, resolving the usual left-right ambiguity. The capabilities demonstrated here establish the potential for a near-real-time whale tracking capability that could be applied anywhere in the world where there are whales and fiber-optic cables. Coupled with ship detection, using a similar approach and/or with fused data from other sources such as AIS, a real-time collision avoidance system could be developed to reduce ship strikes.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.18710/Q8OSON.

Ethics statement

Ethical review and approval were not required for the animal study because it was conducted by remote sensing using passive acoustics without any kind of contact, interactions or disturbances of the animals.

Author contributions

RR wrote the manuscript with support from JE, LB and JP, which was read and approved by all co-authors. RR carried out the acoustic analysis; RR and JE developed and conducted the whale tracking; RR produced the results shown in the manuscript, interpreted together with JE, LB, ML, JP, and KT. ML, RR, SJ, FS, JJ, OS, SW, VO, and AW, planned the experiment, and RR, JJ, and FS collected the DAS data. ML, VO, and AW provided the OptoDAS interrogator used. BR and TJ conducted the air gun experiment. All authors contributed to the article and approved the submitted version.

Funding

RR and KT are funded by the Geophysics and Applied Mathematics in Exploration and Safe production Project

(GAMES) at NTNU (Research Council of Norway; Grant No. 294404). We acknowledge support from the Centre for Geophysical Forecasting - CGF (grant no. 309960). The acquisition of the data was financed by CGF.

Conflict of interest

Author KT was employed by PTT Exploration and Production Public Company Limited.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1130898/full#supplementary-material

References

Abadi, S. H., Tolstoy, M., and Wilcock, W. S. (2017). Estimating the location of baleen whale calls using dual streamers to support mitigation procedures in seismic reflection surveys. *PloS One* 12, e0171115. 564. doi: 10.1371/journal.pone.0171115

Ahonen, H., Stafford, K. M., de Steur, L., Lydersen, C., Wiig, Ø., and Kovacs, K. M. (2017). The underwater soundscape in western fram strait: Breeding ground of spitsbergen's endangered bowhead whales. *Mar. Pollut. Bull.* 123, 97–112. doi: 10.1016/j.marpolbul.2017.09.019

Ahonen, H., Stafford, K. M., Lydersen, C., Berchok, C. L., Moore, S. E., and Kovacs, K. M. (2021). Interannual variability in acoustic detection of blue and fin whale calls in the northeast Atlantic high Arctic between 2008 and 2018. *Endangered Species Res.* 45, 209–224. doi: 10.3354/esr01132

Baumgartner, M. F., Bonnell, J., Van Parijs, S. M., Corkeron, P. J., Hotchkin, C., Ball, K., et al. (2019). Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: System description and evaluation. *Methods Ecol. Evol.* 10, 1476–1489. doi: 10.1111/2041-210X

Bouffaut, L., Landrø, M., and Potter, J. R. (2021). Source level and vocalizing depth estimation of two blue whale subspecies in the western Indian ocean from single sensor observations. *J. Acoustical Soc. Am.* 149 (6), 4422–4436. doi: 10.1121/10.0005281

Bouffaut, L., Taweesintananon, K., Kriesell, H. J., Rørstadbotnen, R., Potter, J. R., Landrø, M., et al. (2022). Eavesdropping at the speed of light: Distributed acoustic sensing of baleen whales in the Arctic. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.901348

Clark, C. W., Gagnon, G. J., and Frankel, A. S. (2019). Fin whale singing decreases with increased swimming speed. *R. Soc. Open Sci.* 6, 180525. doi: 10.1098/rsos.180525

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L. (2008). Accelerated decline in the Arctic sea ice cover. *Geophysical Res. Lett.* 35. doi: 10.1029/2007GL031972

Cummings, W. C., and Thompson, P. O. (1971). Underwater sounds from the blue whale, balaenoptera musculus. *J. Acoustical Soc. Am.* 50, 1193–1198. doi: 10.1121/1.1912752

Delarue, J., Todd, S. K., Van Parijs, S. M., and Di Iorio, L. (2009). Geographic variation in northwest atlantic fin whale (balaenoptera physalus) song: Implications for stock structure assessment. *J. Acoustical Soc. Am.* 125, 1774–1782. doi: 10.1121/1.3068454

Dou, S., Lindsey, N., Wagner, A. M., Daley, T. M., Freifeld, B., Robertson, M., et al. (2017). Distributed acoustic sensing for seismic monitoring of the near surface: a traffic-noise interferometry case study. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-11986-4

Eidsvik, J., and Hokstad, K. (2006). Positioning drill-bit and look-ahead events using seismic traveltime data. *Geophysics* 71, F79–F90. doi: 10.1190/1.2212268

Garcia, H. A., Zhu, C., Schinault, M. E., Kaplan, A. I., Handegard, N. O., Godø, O. R., et al. (2019). Temporal–spatial, spectral, and source level distributions of fin whale vocalizations in the norwegian sea observed with a coherent hydrophone array. *ICES J. Mar. Sci.* 76, 268–283. doi: 10.1093/598icesjms/fsy127

Gavrilov, A. N., and Mikhalevsky, P. N. (2006). Low-frequency acoustic propagation loss in the arctic ocean: Results of the arctic climate observations using underwater sound experiment. *J. Acoustical Soc. Am.* 119, 3694–3706. doi: 10.1121/1.2195255

Gervaise, C., Simard, Y., Aulanier, F., and Roy, N. (2021). Optimizing passive acoustic systems for marine mammal detection and localization: Application to real-time monitoring north atlantic right whales in gulf of st. lawrence. *Appl. Acoustics* 178, 107949. doi: 10.1016/j.apacoust.2021.107949

Greene, C. H., and Pershing, A. J. (2004). Climate and the conservation biology of north Atlantic 606 right whales: the right whale at the wrong time? *Front. Ecol. Environ.* 2, 29–34. doi: 10.1890/1540-9295(2004)002[0029:CATCBO]2.0.CO;2

Hagen, J. O., Liestøl, O., Roland, E., and Jørgensen, T. (1993). *Glacier atlas of Svalbard and Jan mayen* 129, 1–141. (Norsk Polarinstitutt Oslo).

Hamilton, C. D., Lydersen, C., Aars, J., Biuw, M., Boltunov, A. N., Born, E. W., et al. (2021). Marine mammal hotspots in the Greenland and barents seas. *Mar. Ecol. Prog. Ser.* 659, 3–28. 612. doi: 10.3354/meps13584

Harcourt, R., van der Hoop, J., Kraus, S., and Carroll, E. L. (2019). Future directions in eubalaena spp.: comparative research to inform conservation. *Front. Mar. Sci.* 5, 530. doi: 10.3389/fmars.2018.00530

Hartog, A. H. (2017). *An introduction to distributed optical fibre sensors. 1 edn* (Boca Raton, FL: CRC Press). doi: 10.6171201/9781315119014

Havskov, J., and Ottemoller, L. (2010). *Routine data processing in earthquake seismology: With sample data, exercises and software* (Springer Science & Business Media). 347

Hendricks, B., Wray, J. L., Keen, E. M., Alidina, H. M., Gulliver, T. A., and Picard, C. R. (2019). Automated localization of whales in coastal fjords. *J. Acoustical Soc. America* 146, 4672–4686. doi: 10.1121/1.5138125

Hoschle, C., Cubaynes, H. C., Clarke, P. J., Humphries, G., and Borowicz, A. (2021). The potential of satellite imagery for surveying whales. *Sensors* 21, 963. doi: 10.3390/s21030963

IPCC (2022). "Summary for policymakers," in *Climate change 2022: Mitigation of climate change* (Cambridge, UK and New York, NY, USA: Cambridge University Press). Intergovernmental Panel on Climate Contribution: Contribution from working Group III to the Sixth Assessment Report.

Jaskolski, M. W. (2021). For human activity in Arctic coastal environments–a review of selected interactions and problems. *Miscellanea Geographica* 25, 127–143. doi: 10.2478/mgrsd-2020-0036

Klinck, H., Nieukirk, S. L., Mellinger, D. K., Klinck, K., Matsumoto, H., and Dziak, R. P. (2012). Seasonal presence of cetaceans and ambient noise levels in polar waters of the north Atlantic. *J. Acoustical Soc. America* 132, EL176–EL181. doi: 10.1121/1.4740226

Landrø, M., Bouffaut, L., Kriesell, H. J., Potter, J. R., Rørstadbotnen, R. A., Taweesintananon, K., et al. (2022). Sensing whales, storms, ships and earthquakes using an Arctic fibre optic cable. *Sci. 635 Rep.* 12, 1–10. doi: 10.1038/s41598-022-23606-x

Lavery, T. J., Roudnew, B., Seymour, J., Mitchell, J. G., Smetacek, V., and Nicol, S. (2014). Whales sustain fisheries: blue whales stimulate primary production in the southern ocean. *Mar. Mammal Sci.* 30, 638 888–904. doi: 10.1111/mms.12108

Lindsey, N. J., Martin, E. R., Dreger, D. S., Freifeld, B., Cole, S., James, S. R., et al. (2017). Fiber optic network observations of earthquake wavefields. *Geophysical Res. Lett.* 44, 11–792. 641. doi: 10.1002/2017GL075722

Lydersen, C., Vacquie-Garcia, J., Heide-Jørgensen, M. P., Øien, N., Guinet, C., and Kovacs, K. M. (2020). Autumn movements of fin whales (Balaenoptera physalus) from Svalbard, Norway, revealed by satellite 644 tracking. *Sci. Rep.* 10, 1–13. doi: 10.1038/s41598-020-73996-z

Matsumoto, H., Araki, E., Kimura, T., Fujie, G., Shiraishi, K., Tonegawa, T., et al. (2021). Detection of hydroacoustic signals on a fiber-optic submarine cable. *Sci. Rep.* 11, 1–12. doi: 10.1038/647s41598-021-82093-8

Maturilli, M., Herber, A., and Konig-Langlo, G. (2013). Climatology and time series of surface meteorology in ny-alesund, Svalbard. *Earth System Sci. Data* 5, 155–163. doi: 10.5194/essd-5-155-2013

McDonald, M. A., and Fox, C. G. (1999). Passive acoustic methods applied to fin whale population density estimation. *J. Acoustical Soc. Am.* 105, 2643–2651. doi: 10.1121/1.426880

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (1995). Blue and fin whales observed on a seafloor array in the northeast pacific. *J. Acoust. Soc Am.* 98, 712–721. doi: 10.1121/1.413565

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (2006). Biogeographic characterization of 655 blue whale song worldwide: using song to identify populations. *J. Cetacean Res. Manage.* 8, 55–65. doi: 10.47536/jcrm.v8i1.702

Mellinger, D. K., and Clark, C. W. (2000). Recognizing transient low-frequency whale sounds 658 by spectrogram correlation. *J. Acoustical Soc. Am.* 107, 3518–3529. doi: 10.1121/1.429434

Mestayer, J., Cox, B., Wills, P., Kiyashchenko, D., Lopez, J., Costello, M., et al. (2011). "Field trials of distributed acoustic sensing for geophysical monitoring," in *Seg technical program expanded abstracts 2011* (Society of Exploration Geophysicists), 4253–4257. doi: 10.1190/1.3628095

Moore, S. E., Haug, T., V´ıkingsson, G. A., and Stenson, G. B. (2019). Baleen whale ecology in arctic and subarctic seas in an era of rapid habitat alteration. *Prog. Oceanography* 176, 102118. doi: 10.1016/j.pocean.2019.05.010

Nordli, Ø., Przybylak, R., Ogilvie, A. E., and Isaksen, K. (2014). Long-term temperature trends and variability on spitsbergen: the extended Svalbard airport temperature series 1898–2012. *Polar Res.* 33, 21349. doi: 10.3402/polar.v33.21349

Nowacek, D. P., Christiansen, F., Bejder, L., Goldbogen, J. A., and Friedlaender, A. S. (2016). Studying cetacean behaviour: new technological approaches and conservation applications. *Anim. Behav.* 120, 235–244. doi: 10.1016/j.anbehav.2016.07.019

Ocean-Acoustic-Library (2022) *Acoustic toolbox* (Accessed 2023-02-12). Dataset.

Oleson, E. M., Sirovic, A., Bayless, A. R., and Hildebrand, J. A. (2014). Synchronous seasonal change in fin whale song in the north pacific. *PloS One* 9, e115678. doi: 10.1371/journal.pone.0115678

Papp, B., Donno, D., Martin, J. E., and Hartog, A. H. (2017). A study of the geophysical response of distributed fibre optic acoustic sensors through laboratory-scale experiments. *Geophysical Prospecting* 65, 1186–1204. doi: 10.1111/1365-2478.12471

Rørstadbotnen, R. A., Landrø, M., Taweesintananon, K., Bouffaut, L., Potter, J. R., Johansen, S. E., et al. (2022). "Analysis of a local earthquake in the Arctic using a 120 KM long fibre-optic cable," in *83rd EAGE annual conference & exhibition (European association of geoscientists & engineers), vol. 2022 of conference proceedings*, 1–5. doi: 10.3997/2214-4609.202210404

Richard, G., Mathias, D., Collin, J., Chauvaud, L., and Bonnel, J. (2023). Three-dimensional anthropogenic underwater noise modeling in an arctic fjord for acoustic risk assessment. *Mar. pollut. Bull.* 187, 680 114487. doi: 10.1016/j.marpolbul.2022.114487

Rivet, D., de Cacqueray, B., Sladen, A., Roques, A., and Calbris, G. (2021). Preliminary assessment of ship detection and trajectory evaluation using distributed acoustic sensing on an optical fiber telecom cable. *J. Acoustical Soc. America* 149, 2615–2627. doi: 10.1121/10.0004129

Romagosa, M., Nieukirk, S., Cascao, I., Marques, T. A., Dziak, R., Royer, J.-Y., et al. (2022). Fin whale singalong: evidence of song conformity. *bioRxiv*, 2022–2010. MiriamRomagosa doi: 10.1101/2022.10.05.510968

Roman, J., Estes, J. A., Morissette, L., Smith, C., Costa, D., McCarthy, J., et al. (2014). Whales as marine ecosystem engineers. *Front. Ecol. Environ.* 12, 377–385. doi: 10.1890/130220

Sarkka, S. (2013). *Bayesian Filtering and smoothing* (Cambridge university press).

Schuler, T. V., Kohler, J., Elagina, N., Hagen, J. O. M., Hodson, A. J., Jania, J. A., et al. (2020). Reconciling Svalbard glacier mass balance. *Front. Earth Sci.* 8. doi: 10.3389/feart.2020.00156

Shiu, Y., Palmer, K., Roch, M. A., Fleishman, E., Liu, X., Nosal, E.-M., et al. (2020). Deep neural networks for automated detection of marine mammal species. *Sci. Rep.* 10, 607. doi: 10.1038/697s41598-020-57549-y

Sirovic, A., Williams, L. N., Kerosky, S. M., Wiggins, S. M., and Hildebrand, J. A. (2013). Temporal ´ separation of two fin whale call types across the eastern north pacific. *Mar. Biol.* 160, 47–57. 700. doi: 10.1007/s00227-012-2061-z

Stimpert, A. K., DeRuiter, S. L., Falcone, E. A., Joseph, J., Douglas, A. B., Moretti, D. J., et al. (2015). Sound production and associated behavior of tagged fin whales (Balaenoptera physalus) in the southern California bight. *Anim. Biotelemetry* 3, 1–12. doi: 10.1186/s40317-015-0058-3

Storrie, L., Lydersen, C., Andersen, M., Wynn, R. B., and Kovacs, K. M. (2018). *Determining the species assemblage and habitat use of cetaceans in the svalbard archipelago, based on observations from 2002 to 706 2014* (Norwegian Polar institute) 37 (1). doi: 10.1080/17518369.2018.1463065

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. (2007). *Arctic Sea ice decline: Faster than forecast. geophysical research letters*, (American geophysical union) 34 (9). doi: 10.1029/2007GL029703

Taweesintananon, K., Landrø, M., Brenne, J. K., and Haukanes, A. (2021). Distributed acoustic sensing for near-surface imaging using submarine telecommunication cable: A case study in the trondheimsfjord, Norway. *Geophysics* 86, B303–B320. doi: 10.1190/geo2020-0834.1

Taweesintananon, K., Landrø, M., Potter, J. R., Johansen, S. E., Rørstadbotnen, R. A., Bouffaut, L., et al. (2023). Distributed acoustic sensing of ocean-bottom seismo-acoustics and distant storms: A case study from svalbard, norway. *Geophysics* 88, 1–65. doi: 10.1190/geo2022-0435.1

Thode, A. M., D'Spain, G., and Kuperman, W. (2000). Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations. *J. Acoustical Soc. America* 107, 1286–1300. doi: 10.1121/1.428417. 5.

Thomas, P. O., Reeves, R. R., and Brownell, J. R.L. (2016). Status of the world's baleen whales. *Mar. Mammal Sci.* 32, 682–734. doi: 10.1111/mms.12281

Thompson, P. O., Findley, L. T., and Vidal, O. (1992). 20-Hz pulses and other vocalizations of fin whales, b alaenopteraphysalus, in the gulf of California, Mexico. *J. Acoustical Soc. America* 92, 3051–3057. doi: 10.1121/1.404201

Timmermans, M. L., and Labe, Z. (2022) *Sea Surface temperature* (Accessed Accessed: 2022-01-13). Dataset.

Townhill, B. L., Reppas-Chrysovitsinos, E., Suhring, R., Halsall, C. J., Mengo, E., Sanders, T., et al. (2022). Pollution in the Arctic ocean: An overview of multiple pressures and implications for ecosystem services. *Ambio*, 51 (2), 471–483. doi: 10.1007/s13280-021-01657-0

van Weelden, C., Towers, J. R., and Bosker, T. (2021). Impacts of climate change on cetacean distribution, habitat and migration. *Climate Change Ecol.* 1, 1009. doi: 10.1016/j.ecochg.2021.100009

Waagaard, O. H., Rønnekleiv, E., Haukanes, A., Stabo-Eeg, F., Thingbø, D., Forbord, S., et al. (2021). Real-time low noise distributed acoustic sensing in 171 km low loss fiber. *OSA Continuum* 4, 688–701. doi: 10.1364/OSAC.408761. 731.

Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (1987). The 20-Hz signals of finback whales (Balaenoptera physalus). *J. Acoustical Soc. America* 82, 1901–1912. doi: 10.1121/7341.395685

Wilcock, W. S. (2012). Tracking fin whales in the northeast pacific ocean with a seafloor seismic network. *J. Acoustical Soc. America* 132, 2408–2419. doi: 10.1121/1.4747017

Wilcock, W. S., Abadi, S., and Lipovsky, B. P. (2023). Distributed acoustic sensing recordings of low-frequency whale calls and ship noise offshore central oregon. *JASA Express Lett.* 3, 026002. doi: 10.1121/10.0017104

Wuestefeld, A., Greve, S. M., Nasholm, S. P., and Oye, V. (2018). Benchmarking earthquake location algorithms: A synthetic comparison. *Geophysics* 83, KS35–KS47. doi: 10.1190/geo2017-0317.1

Yigit, C. O., Gikas, V., Alcay, S., and Ceylan, A. (2014). Performance evaluation of short to long term GPS, GLONASS and GPS/GLONASS post-processed PPP. *Survey Rev.* 46, 155–166. doi: 10.1179/7431752270613Y.0000000068

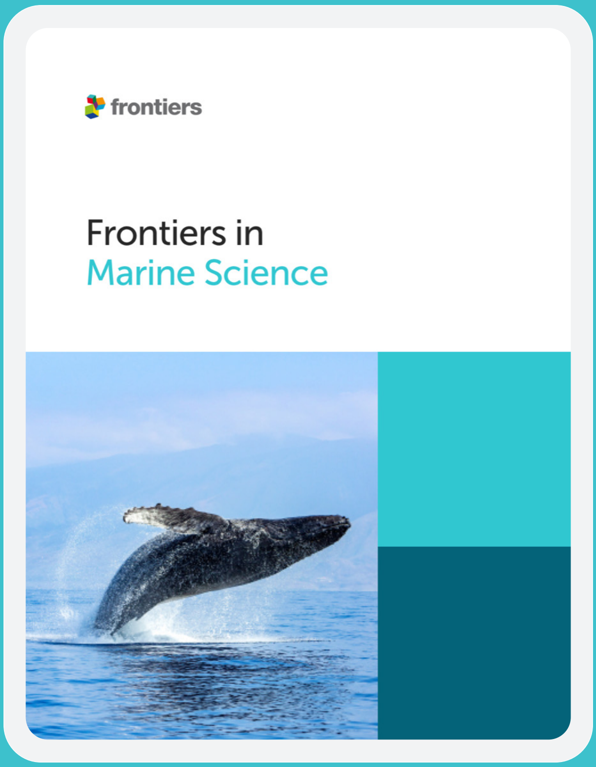
Frontiers in
Marine Science

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