frontiers | Research Topics

Remote sensing for environmental monitoring

Edited by
Samy Ismail Elmahdy, Maged Marghany and Jie Luo

Published in
Frontiers in Environmental Science
Frontiers in Ecology and Evolution
Frontiers in Earth Science

ISSN 1664-8714
ISBN 978-2-8325-4432-7
DOI 10.3389/978-2-8325-4432-7

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

Remote sensing for environmental monitoring

Topic editors

Samy Ismail Elmahdy — United Arab Emirates University, United Arab Emirates

Maged Marghany — Syiah Kuala University, Indonesia

Jie Luo — Zhejiang Lab, China

Citation

Elmahdy, S. I., Marghany, M., Luo, J., eds. (2024). *Remote sensing for environmental monitoring*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-4432-7

Table of contents

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fenvs.2023.1171660

OPEN ACCESS

EDITED BY
Prof-Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Kai Zhu,
Hubei University, China
Subodh Chandra Pal,
University of Burdwan, India

*CORRESPONDENCE
Huijun An,
deanimau@163.com

RECEIVED 22 February 2023
ACCEPTED 15 May 2023
PUBLISHED 25 May 2023

CITATION
Ri A and An H (2023), Health assessment of natural larch forest in arxan guided by forestry remote sensing integrated with canopy feature analysis.
Front. Environ. Sci. 11:1171660.
doi: 10.3389/fenvs.2023.1171660

Health assessment of natural larch forest in arxan guided by forestry remote sensing integrated with canopy feature analysis

Ana Ri and Huijun An*

College of Forestry, Inner Mongolia Agricultural University, Hohhot, China

This work aims to propose a more accurate assessment method for forest health in natural larch pine forests of the Arxan by integrating remote sensing technology with tree crown feature analysis. Currently, forest health assessment of natural Larch pine forests relies mainly on ground surveys, and there is a gap in the application of remote sensing technology in this field. This work introduces deep learning technology and proposes a spectral-Gabor space discrimination and classification model to analyze multi-spectral remote sensing image features. Additionally, quantitative indicators, such as tree crown features, are incorporated into the forest health assessment system. The health status of natural Larch pine forests is evaluated using forest resource survey data. The results show that the health levels of natural Larch pine forests in different areas vary and are closely related to factors such as canopy density, community structure, age group, and slope. Both quantitative and qualitative indicators are used in the analysis. The introduction of this innovative method enhances the accuracy and efficiency of forest health assessment, providing significant support for forest protection and management. In addition, the classification accuracy of the health assessment model suggested that the maximum statistical values of average classification accuracy, average classification effectiveness, overall classification accuracy, and Kappa were 74.19%, 61.91%, 63.18%, and 57.63%, respectively. This demonstrates that the model can accurately identify the health status of natural larch forests. This work can effectively assess the health status of the natural larch forest in the Arxan and provide relevant suggestions based on the assessment results to offer a reference for the sustainable development of the forest system.

KEYWORDS

forestry remote sensing, health assessment, natural larch forests, spectrum-gabor, sustainable development

Introduction

As one of the important components of the Earth's ecosystem, forests contain substantial species essential for soil and water conservation, climate improvement, and air purification (Torres et al., 2021; Ecke et al., 2022). Statistics show that the Three-North Shelter Forest Program extends from northeast China to northern China and northwest China, with a total area of nearly 1.5 million square kilometers. Its afforestation area in desertified areas has exceeded 300,000 square kilometers, and the total proportion of green area has reached more than 80% of the desertification land area in China (Cherubini et al., 2021). The resources and services provided by this protective forest are highly correlated with the health of the forest ecosystem. For

example, species rich and healthy forests produce sufficient timber and ecological services while maintaining the food chain and biodiversity in the forest well; forests with simple structures and in a critical state of health are only able to sustain basic tree survival (Iglhaut et al., 2019; Guimarães et al., 2020; Carnegie et al., 2022). The natural Larch pine forests in the Greater Khingan Mountains are currently facing a state of system destabilization, decreased vitality, and forest area reduction due to various factors, including climate change, human activities, and pests. These changes have resulted in more frequent occurrences of severe weather phenomena, such as dust storms, which have had adverse effects on both the ecological environment and socio-economic development in the area. Therefore, the understanding of tree characteristics and health assessment in forest ecosystems has become the focus of research scholars in related fields.

Of course, due to the complexity of the forest ecosystem, the understanding of the health status of forest trees in the forest isn't the same, and there is no unified forest health assessment standard. Saha et al. (2021a) employed advanced geospatial techniques to evaluate the forest health status of the Buxa Natural Reserve in the Himalayan Mountains, revealing varying degrees of forest pests and human interference in the region. In a similar vein, Malik et al. (2020) utilized geospatial analysis to investigate seasonal changes in the vegetation status of the subtropical deciduous forest, indicating significant fluctuations throughout the seasons. Moreover, Saha et al. (2021b) assessed and analyzed forest cover dynamics using a forest canopy density model, which demonstrated a close correlation between forest cover spatial distribution and changes in topography and land use types. Forestry remote sensing technology has been extensively studied by scholars as a means of evaluating forest species, structure, growth, and health status using airborne or satellite remote sensing data. Pal et al. (2018) combined the forest canopy density model with satellite data to achieve remote sensing monitoring and mapping of forest cover in the Sali Basin of West Bengal. Ahmadi et al. (2023) used high-resolution remote sensing technology to predict the distribution of major tree species in forests with insufficient climate data. Their findings suggested that integrating high-resolution remote sensing data could improve the accuracy of species distribution models. Despite these advancements, a unified standard for forest health assessment has yet to be established. Additionally, remote sensing data quality is affected by factors such as clouds, fog, and occlusion, resulting in limited accuracy in tree crown feature extraction. Therefore, the effective extraction of spectral features in remote sensing images is extremely important. The deep learning algorithm can mine the relationship between data through independent learning of remote sensing images, which can greatly improve the accuracy and stability of image classification and recognition while improving efficiency, like reference (Tian et al., 2021).

In conclusion, it is of great significance to explore the health status and assessment methods of natural larch forests under the guidance of forestry remote sensing integrated with the analysis of canopy characteristics for the sustainable development of subsequent forest systems and the formation of biodiversity. The innovations of this work are as follows. First, the health status of natural larch forests is analyzed by collecting 2A sentinel standard reflectance images with a spatial resolution of 10 m and forest class II survey data using Gurban Forest Farm of the Arxan Forestry Bureau as the study area. Secondly, the health of natural larch forests is assessed by quantitative and qualitative indicators. Third, a feature extraction and classification model based on spectral-Gabor spatial discriminant analysis is proposed to analyze the features of the acquired multi-spectral remote sensing images. Finally, the performance of the health assessment method and model reported here is verified through experimental analysis to provide a reference for the follow-up health status analysis and ecologically sustainable development of the forest system.

Recent related work

Advances in forest health evaluation and analysis

Many scholars have researched the state of forest health. Cortés et al. (2020) discussed predictive genomic approaches. They found that these approaches promise to increase the accuracy of adaptive selection and reduce generation intervals. Besides, exploring genome-wide prediction methods can help detect novel allelic variants in tree germplasm and reveal the genomic potential for adaptation to different environments. Jakovljević et al. (2021) conducted passive ozone measurements and monitoring of forest health indicators in holly oak (*Quercus ilex* L.), brash oak (*Quercus stellata*), Larch pine (*Pinus koraiensis* Sieb. et Zucc.), and black pine (*Pinus thunbergii* Parlatore) forests. Results showed that the ozone levels of all species were close to reaching a reasonable upper limit (100 ppb) for passive monitoring of air quality in forest sites, with the highest values for the uptake-based indicator on black pepper. At the same time, the relationship between environmental variables and forest health response indicators could be found to be significantly related to the soil moisture content at different depths of canopy defoliation; besides, tree growth is related to different forest health response indicators. Kayet et al. (2022) assessed and predicted the forest health risk in forest areas affected by mining based on the Analytic Hierarchy Process model of multi-criteria analysis. In total, they considered parameters including climate, natural or landform, forestry, topography, environmental, and anthropogenic factors. Very high-risk grades were found in mines surrounding forest subdivisions, based on FHR assessment and prediction results, with FHR negatively correlated with distance from mine and leaf flour dust concentration. Their research provided fundamental guidance for effective planning and management of forestry research in mining-affected areas. Tan et al. (2021) assessed the impact of nature-based solutions in urban design on forest ecosystem service performance based on a spatially explicit modeling approach by using a recent nature-centered town in Singapore as a case study. The authors found that designing towns with ecosystem services in mind and incorporating nature-based solutions into the urban design can help improve performance in delivering ecosystem services, with significant benefits for tree health assessments in forests.

Current situation of AI applied to remote sensing image recognition

Since each sample in hyperspectral remote sensing images consists of high-dimensional features and contains rich remote

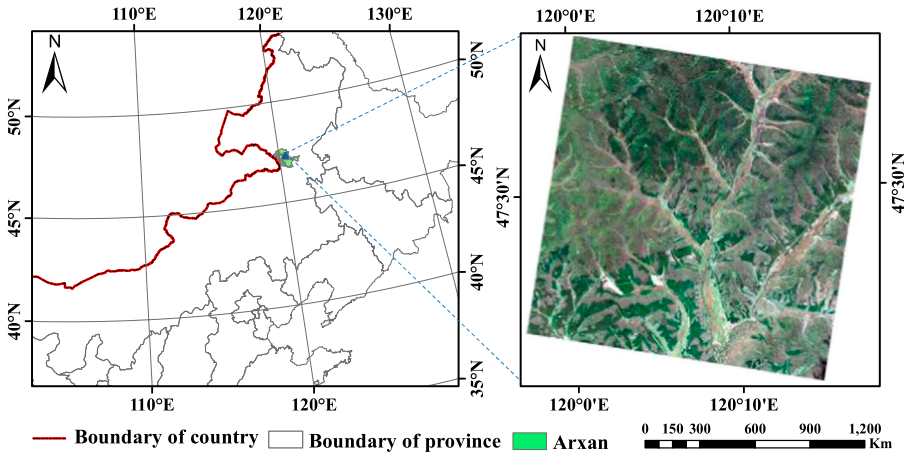


FIGURE 1
Geographical map of the study area.

sensing features, feature selection, and mining become even more difficult. Algorithms such as deep learning in artificial intelligence technology can independently learn and mine features in images, effectively improving the ability to identify features in remote sensing images. Scholars in many related fields have conducted research. Ge et al. (2020) compared the classification performance of k-nearest neighbor, random forest (RF), support vector machine (SVM), and artificial neural network (ANN) in land use and cover change (LUCC) in Chinese oases. They found that ANN, SVM, and RF achieved statistically similar accuracy. The RF algorithm performed well in several aspects, such as stability, ease of use, and processing time during parameter adjustment. It was capable of analyzing spectral indices (e.g., NDVI, MSAVI2, and MNDWI) and providing a reference for the extraction of LUCC information in arid regions with oasis-desert mosaic landscapes. Alam et al. (2021) applied the convolutional neural network to the semantic segmentation of remote sensing images. The authors also improved the encoder-decoder CNN structure SegNet with index pool-ing and U-net, making it suitable for remote sensing Multi-object Semantic Segmentation of Images. The results showed that the comprehensive algorithm could segment multi-object remote sensing images. Cai et al. (2021) proposed a multi-attention residual integrated network algorithm. The analysis of multiple multi-class public data sets revealed that the algorithm adds feature fusion while reducing redundant features, which makes the recognition capability of hyperspectral images effectively improved. Han et al. (2022) presented a building extraction method for remote sensing images combining traditional digital image processing methods and convolutional neural networks. Experiments showed that the method improves detection accuracy and reduces computation time compared with the Region-CNN algorithm.

To sum up, the research and analysis of the above-mentioned scholars indicate that forests, as one of the important components of the Earth's ecosystem, are closely related to the sustainable development of the economy and the improvement of social benefits to evaluate the health of forests.

Research area and methodology

Study area

Arxan Forestry Bureau is located in the northeast of the Inner Mongolia Autonomous Region and the middle and low mountainous area in the south of the main line of the Greater Khingan Mountains. It belongs to the transitional zone type of forest and forest grassland and has a typical volcanic landform type (Quesada-Román and Mata-Cambronero, 2021). Figure 1 displays the geographical location of the study area.

As shown in Figure 1, the research area is located in the Gurban Forest Farm of Arxan Forestry Bureau. The dominant tree species are white birch and larch, and the natural larch forest is taken as the research object. The elevation of Gurban Forestry Station of the Arxan Forestry Bureau gradually decreases from east to west, with an average elevation of 1023m; the annual average temperature is −3.2°C; the annual extreme maximum temperature is 34.1°C; the annual ex-treme minimum temperature is −45.7°C; the frost-free period is 77 days; the rainfall is 451.2 mm. Overall, it is characterized by cold and humid, long winter, and large temperature differences between day and night. At the same time, due to the influence of climate factors, the snow cover is thinner in sunny slopes and river valleys where there are no trees and vegetation, and the snow cover is unevenly distributed due to the effect of wind blowing (Dainelli et al., 2021). Ruidas et al. (2022b) and Ruidas et al. (2021) have conducted a vulnerability assessment of water resources based on hydro-geochemical characteristics. The study revealed that water resources in lakes were impacted by both human activities and natural processes, emphasizing the importance of effective management and planning of water resources.

The soil in the study area was gray forest soil and brown coniferous forest soil in order with the elevation increasing. Moreover, the regional distribution law is mainly reflected in the formation of different recessed soils due to the redistribution of water and heat caused by changes in topography and landform. For example, meadow soil is distributed on both sides of rivers

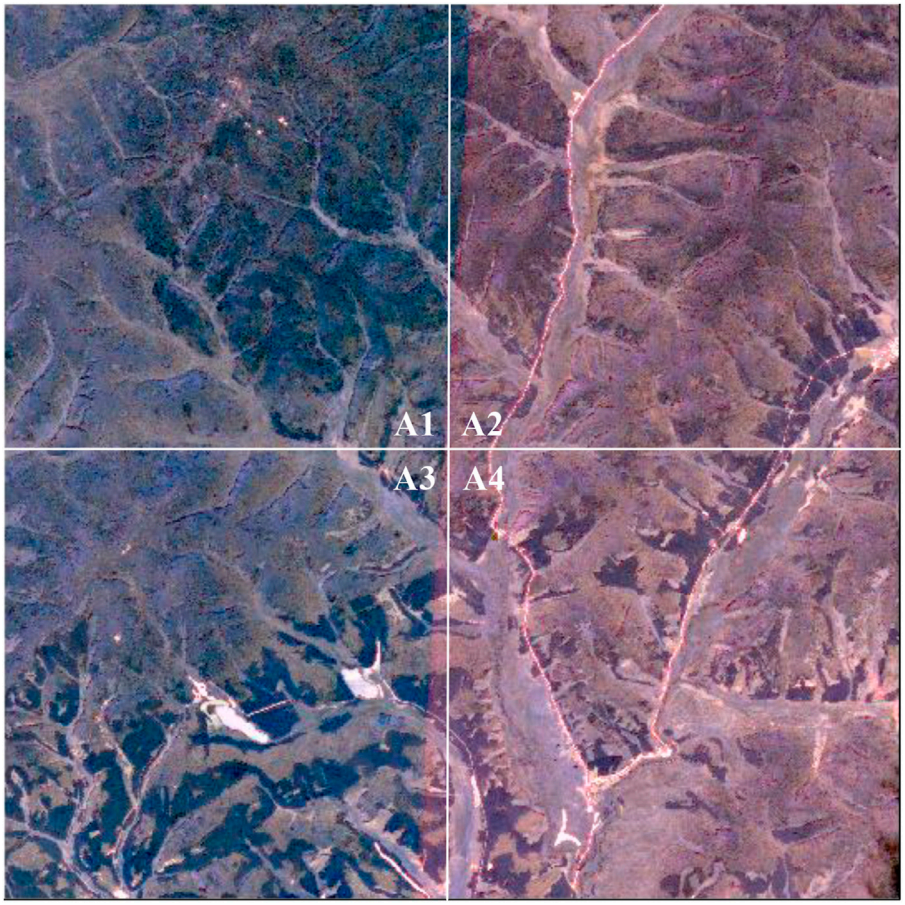


FIGURE 2
Sentries in the study area: 2A preprocessing result map (four sample plots, namely, **(A1–A4)**).

and valleys and in the low and flat areas of marshes; swamp soil is distributed lower than meadow soil, and the soil is humid, with seasonal or perennial water accumulation on the surface; stony soil is only distributed on the gravel ejected by modern volcanoes, and there are large pores. The Gurban Forest Farm of the Arxan Forestry Bureau is located on the gentle slope of the middle section of Greater Khingan Mountains, a forest-grass transition area from forest land to meadow, and the soil type is brown coniferous soil (Chen et al., 2022). Ruidas et al. (2022a) proposed a metaheuristic optimization process integrating rainfall and geological data to develop 15 flood-susceptibility factors. The study found that this approach could accurately assess flood susceptibility.

The vegetation types of the Arxan Forestry Bureau include forest, shrub, meadow, swamp, and aquatic plants. Its forest vegetation is clearly distributed vertically and roughly divided into four vegetation zones: the forest zone domi-nated by the zonal plant Xing'an larch (*Larix gmelinii* (Rupr.) Kuzen) (above 1200 m in elevation), the mixed conifer-broad forest formed by Xing'an larch and white birch (*Betula platyphylla Suk.*) (1000–1200 m in elevation), the forest-steppe zone (below 1000 m), and the stone pond forest (1100–1250 m) (Wang et al., 2021; Yang, 2022). The larch forest in Gurban Forest Farm of Arxan Forestry Bureau has neat margins and is a middle-aged forest with a forest age of 29 years. The average tree height is 16 m; the average diameter is 15 cm; the canopy density is 0.85; the soil type is brown coniferous soil. The understory shrub is rose thorn with a coverage of 5%, and the herb is sedge with a coverage of 60%. There are logging residues piled up under the forest. The birch forest is a natural forest with an age of about 21 years. The average tree height is 7 m, the average diameter is 8 cm, and the canopy density is 0.7.

Data acquisition and processing

This work downloads the Sentinel-2A remote sensing image data of the research area of the study area from https://scihub.copernicus.eu/. The Sentinel-2A data has a spatial resolution of 10 m. Among all the optical images, the Sentinel-2A data is the only one that contains three bands in the red edge range. This feature is useful for monitoring the vegetation health of the natural larch forest in the study area. The information is very effective. The image was produced at 12 September 2019 and radiation calibration and atmospheric correction are

TABLE 1 List of data sources.

Type of data	Resolution	Acquisition method
Sentinel-2A remote sensing image map data	10 m	Public data (from https://scihub.copernicus.eu/)
Forest resources type ii survey data	-	Survey data

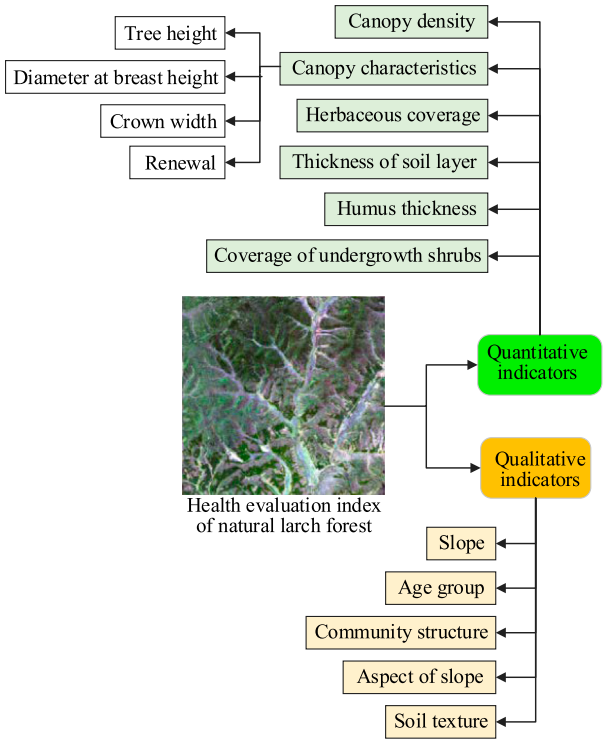


FIGURE 3
Health evaluation index system of natural larch forest.

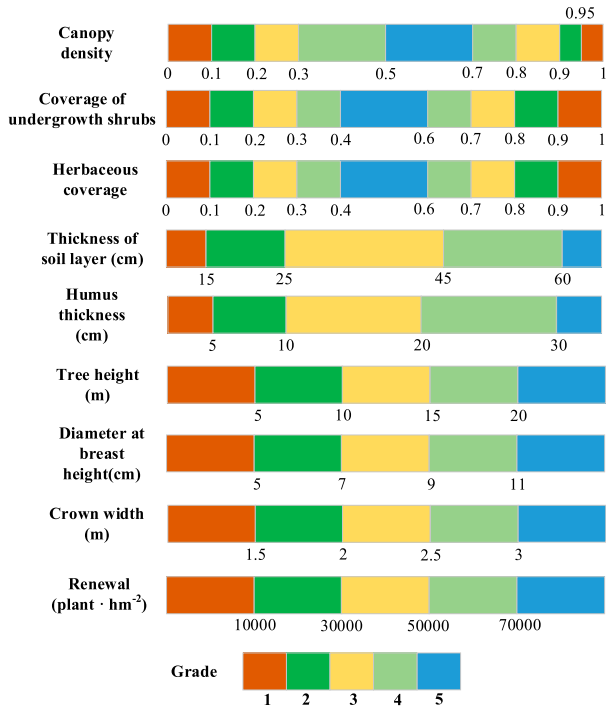


FIGURE 4
Classification of quantitative indicators for forest health assessment.

further performed on the Sentinel-2A image. Figure 2 reveals the preprocessed Sentinel-2A image in the study area.

As shown in Figure 2, the acquired Sentinel-2A image is preprocessed to obtain a Sentinel-2A image pixel with a size of 2000 * 2000. It is evenly divided into four sample plots for research, namely, A1, A2, A3, and A4. The natural larch forest remote sensing images in the study area were extracted from the sentinel spectral image. The classification accuracy was evaluated for the health status of the larch forest in the study area by using the following spectral-Gabor spatial discriminant method (Diez et al., 2021).

Further, the forest resource type II survey data were collected in the study area. The type II data includes community conditions such as tree species structure canopy density, understory shrub coverage, herbaceous coverage, and tree crown characteristics, as well as environmental conditions such as slope, aspect, soil type, and soil thickness. Arc GIS 10.1 software was utilized to confirm the sample data, reasonably move slightly to reduce the edge effect, and divide the obtained samples into training samples and verification samples. There are 337 training samples and 186 validation samples. Table 1 lists the specific statistics.

Analysis of health assessment methods of natural larix pine

According to the characteristics of the basic data used here, this work uses different indicators to comprehensively evaluate the health level of each evaluation object (small class) in the study area. The state of each indicator is divided into several grades. Then, a score is assigned to each grade. Finally, the health evaluation index system of the natural larch forest is constructed, as shown in Figure 3.

The health evaluation system illustrated in Figure 3 presents a comprehensive approach for assessing the health status of natural larch forests in the Gurban Forest Farm, utilizing both quantitative and qualitative indicators. In addition to the quantitative indicators such as depression, shrub cover, herbaceous cover, soil thickness, and humus thickness, this study incorporates canopy characteristics including tree height, diameter at breast height, crown width, and regeneration as qualitative indicators. The quantitative indices are categorized into five classes based on index values measured in the forest resources second-class survey data. The qualitative indices, such as community structure, age group, slope, slope direction, and soil texture, are classified into 3 to 8 classes according to the forest resources second-class survey protocols

TABLE 2 Forest health evaluation qualitative indicator ranking.

Indicator factors	Affiliation level							
	1	2	3	4	5	6	7	8
Community structure	Complete structure (100 points)	Relatively complete structure (80 points)	Simple structure (60 points)					
Age group	Young forest (60 points)	Middle-aged forest (80 points)	Near mature forest (90 points)	Mature forest (100 points)	Overripe forest (60 points)			
Slope	Flat slope (100 points)	Gentle slope (90 points)	Slopes (75 points)	Steep slopes (60 points)	Rapid slope (50 points)	Dangerous slope (35 points)		
Slope direction	Southwest slope (100 points)	Southeast slope (90 points)	Western slope (80 points)	Eastern slope (70 points)	Northwest slope (60 points)	Northeast slope (50 points)	South slope (40 points)	North slope (30 points)
Soil texture	Loam (100 points)	Sandy loam (90 points)	Light loamy soil (80 points)	Heavy loam (70 points)	Sandy soil (60 points)	Silt (50 points)	Clay (40 points)	gravel soil (30 points)

in China. The inclusion of both quantitative and qualitative indicators provides a more comprehensive evaluation of forest health status.

Among the quantitative indicators, the indicator of depression reflects the proportion of the total projected area of the tree canopy on the ground in direct sunlight to the total area of the forest stand and the density of the stand. Both un-derstory shrub cover and herbaceous cover indicators influence forest ecosystem health from the perspective of forest community hierarchy and species diversity. Soil and humus provide support for various life activities of trees, shrubs, and grasses in the forest through nutrient supply. Therefore, soil layer thickness and humus thickness ultimately affect the health of the forest ecosystem. In addition, the tree height, DBH, crown width, and regeneration of pine trees in natural deciduous forests reflect the health status of forest resources from the perspective of system vitality (Falťan et al., 2021; Pan et al., 2022). Figure 4 illustrates the classification of each indicator.

As shown in Figure 4, among the quantitative indicators, the scores of each level are calculated as follows. Level 1 means 100 points; level 2 means 80 points; level 3 means 65 points; level 4 means 50 points; level 5 means 35 points.

The qualitative indicators used in this work include community structure, age group, slope, slope direction, and soil texture, as shown in Table 2. Community structure reflects the species diversity and stability of the forest community, age group is used to assess the age structure in the forest, while slope and slope direction are important factors affecting the moisture in the terrain. Soil texture is another important physical property that influences soil aeration, water retention, and fertility status.

The weight value of the health evaluation index of the larch forest was determined by AHP to evaluate the health degree of the natural larch forest in the forest. Finally, the health level is divided into four levels according to the health value of each index and the specific conditions of the research area: healthy (>4 points), sub-health (3-4 points), general health (2-3 points), and unhealthy (<2 points).

Spectral feature extraction and health classification model design and analysis

This work preprocesses the acquired spectral data information to obtain the Sentinel-2A multi-spectral map and extracts its features to reduce the information redundancy caused by the high-dimensionality and band nature of the data set.

First, the principal component analysis (PCA) method is used to obtain the principal components of remote sensing images. Secondly, Gabor filters of different scales and directions are performed on the extracted principal component images to obtain Gabor spatial features of many different scales and directions. Third, the original spectral features and the obtained Gabor space features are fused in a certain way to form a fusion feature. Besides, the fusion feature is further modeled. The scatter matrix between classes and the scatter matrix within a class in the fusion feature space is constructed to describe the compactness of different types of objects and the same type of objects. The optimal transformation matrix is obtained by transforming the feature space. Finally, the transformation matrix is used to transform the original high-dimensional data into the fusion space to enhance the nonlinear feature learning ability and generalization performance of subsequent classifiers, thereby improving the overall classification accuracy.

When using the spectral-Gabor space discriminant method to extract information from the Sentinel-2A multi-spectral image, the training sample set is defined as, which can be written as Eq. 1.

$$\chi = \left\{x_i : x_i \in R^d\right\}_{i=1}^{(m^*n)} \tag{1}$$

In Eq. 1, d stands for the number of band features of the original image data, that is, the number of original spectral features; m^*n signifies the size of each spectral remote sensing image. Assume that the data set has c categories, and each category contains n_i pixels. m^*n is expressed as Eq. 2.

$$(m^*n) = \sum_{i=1}^{c} n_i \tag{2}$$

The data set needs to be preprocessed before feature extraction due to the high dimensionality of the acquired spectral remote sensing image data set and the correlation and redundancy between bands. Firstly, PCA (Cruz-Ramos et al., 2021) is used to extract $p_i\,(1 \le i < d)$ principal components. Secondly, Gabor filters of different scales and directions are performed on the extracted p_i principal components to obtain $40^* p_i$ individual spatial features. Usually, a two-dimensional Gabor filter is a Gaussian kernel function adjusted by a complex sinusoidal plane wave, which is defined as Eq. 3.

$$\psi_{f,\theta} = \frac{f^2}{\pi\gamma\eta}\exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(j2\pi f x' + \phi\right) \tag{3}$$

In Eq. 3, f refers to the central angular frequency of the complex sinusoidal plane wave; ϕ represents the phase; σ denotes the standard deviation; γ refers to the space ratio used to specify the ellipticity supported by the Gabor function. x' and y' indicate different scales and different directions, which can be written as Eq. 4 and Eq. 5.

$$x' = x\cos\theta + y\sin\theta \tag{4}$$

$$y' = -x\sin\theta + y\cos\theta \tag{5}$$

In Eq. 4 and Eq. 5, θ refers to the normal parallel stripe direction of the Gabor function. The Gabor filters with different scales μ and different directions ν are defined to ensure that the frequency and direction of Gabor filtering can be changed within a certain range to cover the entire two-dimensional image area:

$$f_\mu = \frac{\pi}{2} \times \sqrt{2}^{\mu} \tag{6}$$

$$\theta_\nu = \frac{\pi}{8} \times \nu \tag{7}$$

The two-dimensional convolution result G_i of the Gabor features of each scale and direction of the principal component of the spectral remote sensing image is obtained through the convolution of the spectral remote sensing image A_i and the Gabor filter cluster, as shown in Eq. 8.

$$G_i = A_i \psi_{f_\mu,\theta_\nu} \tag{8}$$

Denote $Z_i\,(1 \le i < d)$ as the spatial filtering features of the image at different scales μ and different directions ν, as shown in Eq. 9.

$$Z_i = \left[Z_1, Z_2, \cdots, Z_{p_i}\right] \tag{9}$$

Moreover, the spectral feature matrix Y_i is fused with the obtained Gabo, space feature matrix Z_i, generating the Spectral-Gabor spatial fusion feature matrix F_i, as presented in Eq. 10.

$$F_i = [Y_i, Z_i] \tag{10}$$

It is necessary to minimize the distance within a class while maximizing the distance between classes to improve the classification performance of spectral features in spectral remote sensing images. Therefore, the fusion feature optimization model shown in Eq. 11 is established.

$$W = \underset{W}{\operatorname{argmax}}\left|\frac{W^T S_b^{SG} W}{W^T S_w^{SG} W}\right| \tag{11}$$

In Eq. 11, S_b^{SG} refers to the inter-class scatter matrix of the spectral-Gabor space, and S_w^{SG} represents the intra-class scatter matrix of the spectral-Gabor space. Assume that $(d + 40^* p_i)$ training samples of dimensions can be obtained from the spectral-Gabor space fusion feature matrix F_i. S_b^{SG} and S_w^{SG} are expressed as Eqs 12, 13.

$$S_b^{SG} = \sum_{i=1}^{c} n_i \left(f_i - f_0\right)\left(f_i - f_0\right)^T \tag{12}$$

$$S_w^{SG} = \sum_{i=1}^{c}\sum_{j=1}^{n_i}\left(x_i^j - f_i\right)\left(x_i^j - f\right)^T \tag{13}$$

In Eqs 12, 13, c refers to the number of classes; n_i represents the number of training samples in the ith class; f_i stands for the mean vector of the ith class; f_0 signifies the mean vector of all training samples; x_i^j denotes the jth fused feature vector of the ith class.

The projection feature matrix x a test sample f can be obtained through optimal transformation W_2, as shown in Eq. 14.

$$x = W_2^T f \tag{14}$$

Eqs 9, 10 indicate the Gabor space feature matrix Z_i and the Spectral-Gabor space fusion feature matrix F_i of the test set I in the spectral remote sensing image data set, respectively. In actual analysis, the fusion feature optimization model in Eq. 11 can be transformed into the eigenvalue problem in Eq. 15.

$$S_b^{SG} W = \lambda S_w^{SG} W \tag{15}$$

In solving Eq. 15, the first step is to maximize the inter-class scatter matrix in the spectral-Gabor space by Singular Value Decomposition (SVD). The second step is to solve the generalized eigenvalue problem. The key problem of using SVD to maximize the inter-class scatter matrix in the spectral-Gabor space is to deal with the optimization problem shown in Eq. 16. First, it is essential to deal with the following optimization problem A.

$$A = \underset{A^T A = I}{\operatorname{argmax}}\, tr\left(A^T S_b^{SG} A\right) \tag{16}$$

Based on the above discussion, the spectral remote sensing image dataset L acquired in this work is processed by PCA method. Then the spatial features are extracted by Gabor filter clusters, which can be written as Eq. 17.

$$F = \left\{f_i:\ f_i \in R^{\left(d+40^* p_i\right)}\right\}_{i=1}^{(m^* n)} \tag{17}$$

Then, the feature space transformation is used to obtain effective feature vectors to reduce the computational complexity while improving the classification accuracy. Eq. 18 describes the optimal transformation matrix $W_2^{'}$.

$$W_2^{'} = U_{b1}\sum_{b1}^{-\frac{1}{2}} U_w \sum_{w}^{-\frac{1}{2}} \in R^{\left(d+40^* p_i\right)\times r} \tag{18}$$

In Eq. 18, $U_{b1} \in R^{(d+40^* p_i)\times r}$ refers to a column-orthogonal matrix; $\sum_{b1} \in R^{r\times r}$ represents a diagonal matrix with non-increasing and positive diagonal elements; $U_w \in R^{r\times r}$ signifies an orthogonal matrix; $\sum_w \in R^{r\times r}$ denotes a diagonal matrix.

The Sentinel-2A multi-spectral remote sensing image obtained is used as a data set to evaluate the performance of the spectral feature extraction and health classification model

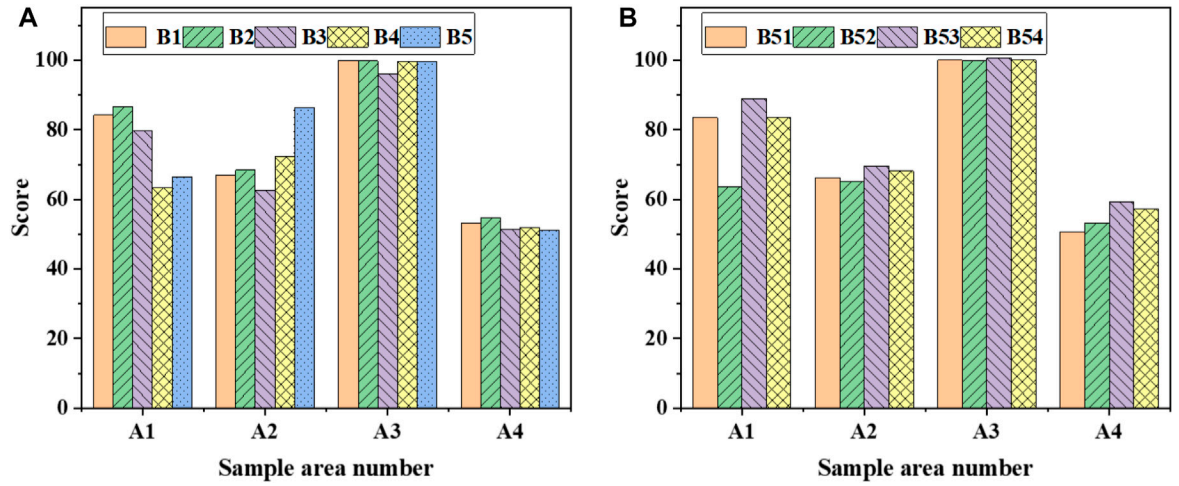


FIGURE 5
Results of quantitative indexes in various fields **(A)** quantitative indicators (B1: canopy density; B2: understory shrub coverage; B3: herb coverage; B4: soil layer thickness; B5: humus thickness); **(B)** quantitative indicators related to tree crown characteristics (B51: tree height; B52: DBH; B53: crown width; B54: renewal).

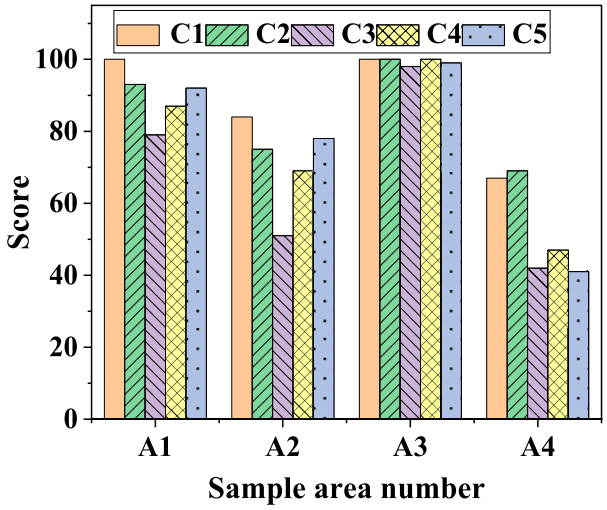


FIGURE 6
Score results of qualitative indicators of forest health assessment (C1: community structure; C2: age group; C3: slope; C4: aspect; C5: soil texture).

based on spectral-Gabor spatial discrimination reported here. It provides rich detailed features for the subsequent nonlinear feature learning of the classifier from different scales and different directions, including a total of 14 features. They are consistent with the quantitative and qualitative indicators mentioned above. At the same time, the model reported here is compared with the benchmark model PCA, linear discriminant analysis (LDA) (Dos Reis et al., 2020), Kernel PCA (KPCA) (He et al., 2021), Kernel Discriminant Analysis (KDA) (Shen et al., 2019), and the model proposed by Kayet et al. (2022) in terms of four evaluation metrics for performance verification. The four assessment criteria are the average classification accuracy (AA), the average classification effectiveness (AV), the overall classification accuracy (OA), and Kappa.

Results

Result analysis of factor indicators

Figures 5, 6 provide the analysis results of the quantitative and qualitative indicators to evaluate the health status of natural larch forests in different sites in the study area.

As shown in Figure 5, among the quantitative indicators, the scores of canopy density, understory shrub coverage, herbaceous coverage, soil layer thickness, and humus thickness in the A3 sample plot are all higher than 9.5 points, which are in a certain range. Specifically, the canopy density is 0.5–0.7; the understory shrub coverage is 0.4–0.6; the herb coverage is 0.4–0.6; the soil layer thickness is higher than 60 cm; the humus thickness is higher than 30 cm. In addition, the scores of plots A1 and A2 are all between 60 and 90, meaning that each quantitative index is at the second or third level. The score of plot A4 is about 50, indicating that each quantitative indicator performs at level four. In addition, the tree height, DBH, crown width, and regeneration grading effects reflecting the vitality of the forest system in each site are discussed. It is found that the overall performance is still the same as in A3 plots. The tree height, DBH, crown width, and regeneration in plots A1 and A2 are grade 2 or grade 3, while the scores of tree height, DBH, crown width, and re-generation in plot A4 are the lowest, which are grade 4. The health of natural larch forests in each sample plot is best in A3, followed by A1 and A2, and the worst in A4.

According to Figure 6, the scores of all qualitative indicators in the A3 sample plots are at level 1. In other words, the

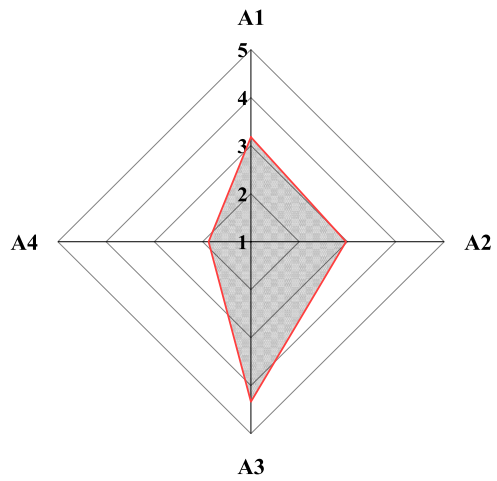


FIGURE 7
Health classification results of natural larch forests in various fields.

community structure is basically intact, the number of age groups is mostly mature forests, the slope is basically flat, the slope orientation is basically southwest, and the soil texture is loamy. However, plots A1 and A2 are classified as grade two or grade three; plot A4 has the worst performance in classification and is basically simple in terms of com-munity structure, age group, slope, aspect, and soil texture. In terms of structure, young forests or over-mature forests account for more, and the slope is steeper, and the slope also shows a northward direction, and the soil texture is mostly gravel soil type. From the analysis of the qualitative index factors, it can be found that the health of natural larch forests in each sample site is best in A3, followed by A1 and A2, and the worst in A4.

Figure 7 classifies the health classes of natural larch forests from healthy, sub-healthy, generally healthy, and un-healthy according to the specific conditions of various places in the study area.

According to the weights of each index in Figure 7, the health classification of natural larch forests in the four sample plots A1, A2, A3, and A4 are not the same. In plot A1, the health evaluation score of the natural larch forest is 3.1798 points, which can be divided into the sub-health level. In plot A2, the health evaluation score of the natural larch forest is 2.9687 points, which can be divided into the general health grade. In plot A3, the health evaluation score of the natural larch forest is 4.3354 points, which can be divided into the health grade. In plot A4, the health evaluation score of the natural larch forest is 1.8764 points, which can be classified as unhealthy.

Classification accuracy analysis

Furthermore, the spectral feature extraction based on the Spectrum-Gabor space discrimination and health classification model is employed to analyze the recognition accuracy of each index. The model is compared with benchmark models PCA, LDA, KPCA, KDA, and the model proposed by Kayet et al. (2022) in terms of four evaluation indicators: AA, AV, OA, and Kappa coefficient, as shown in Figure 8.

As shown in Figure 8, the statistical values of AA, AV, OA, and Kappa are analyzed with the Sentinel-2A multi-spectral remote sensing image as the data source. It can be found that with the increase in the number of features, the statistical values of AA, AV, OA, and Kappa all first increase rapidly and then reach a relatively stable state. The AA, AV, OA, and Kappa statistical values of the model reported here are the highest. This may be due to the fact that the method used here can demonstrate its superb feature extraction ability and the nonlinear feature learning ability and generalization performance of the subsequent classifier. The model proposed by Kayet et al. (2022) has the second-highest classification accuracy, and the KDA algorithm has the worst classification results. Moreover, the classification accuracy of the model reported here is optimal when the feature data is 7 or 8; the maximum statistical values of AA, AV, OA, and Kappa are 74.19%, 61.91%, 63.18%, and 57.63%, respectively. Therefore, the model can accurately identify the health status of natural larch forests.

Discussion and suggestion

Discussion

This work reveals that an accurate assessment of the health status of natural larch forests in the Arxan Forestry Bureau can be achieved through the analysis of remote sensing data and feature extraction methods. The results highlight significant differences in the health status of different plots. Plot A1 showed the lowest health status and was classified as sub-healthy, while plot A4 exhibited the worst health status and was classified as unhealthy. The most significant indicators for assessing the health status of natural forests were found to be quantitative indicators such as canopy density, understory shrub cover, humus thickness, tree height, and canopy regeneration. Among these, age group, soil texture, and community structure were identified as having the most significant impact on the health status of natural forests.

Further analysis of the spectral-Gabor spatial discriminant method proposed in this work for spectral feature extraction reveals higher precision and accuracy than those achieved by scholars in related fields (Reddy, 2021; Thakur et al., 2021). Compared with Fernandez-Carrillo et al. (2020), this study uses more refined remote sensing data and feature extraction methods, providing better reflection of the health status of natural Larch pine forests.

This work has some limitations that need to be acknowledged. Firstly, the sample size is relatively small, including only four plots, which may not comprehensively reflect the health status of natural larch forests in the Greater Khingan Mountains. Secondly, the impact of human activities and climate change has not been considered, indicating the need to include more factors in assessing the health status of natural larch forests. Therefore, future research should focus on increasing the sample size, integrating more data sources and feature extraction methods, and comprehensively considering the effects of various factors on natural larch forest health to improve the accuracy and reliability of health assessment.

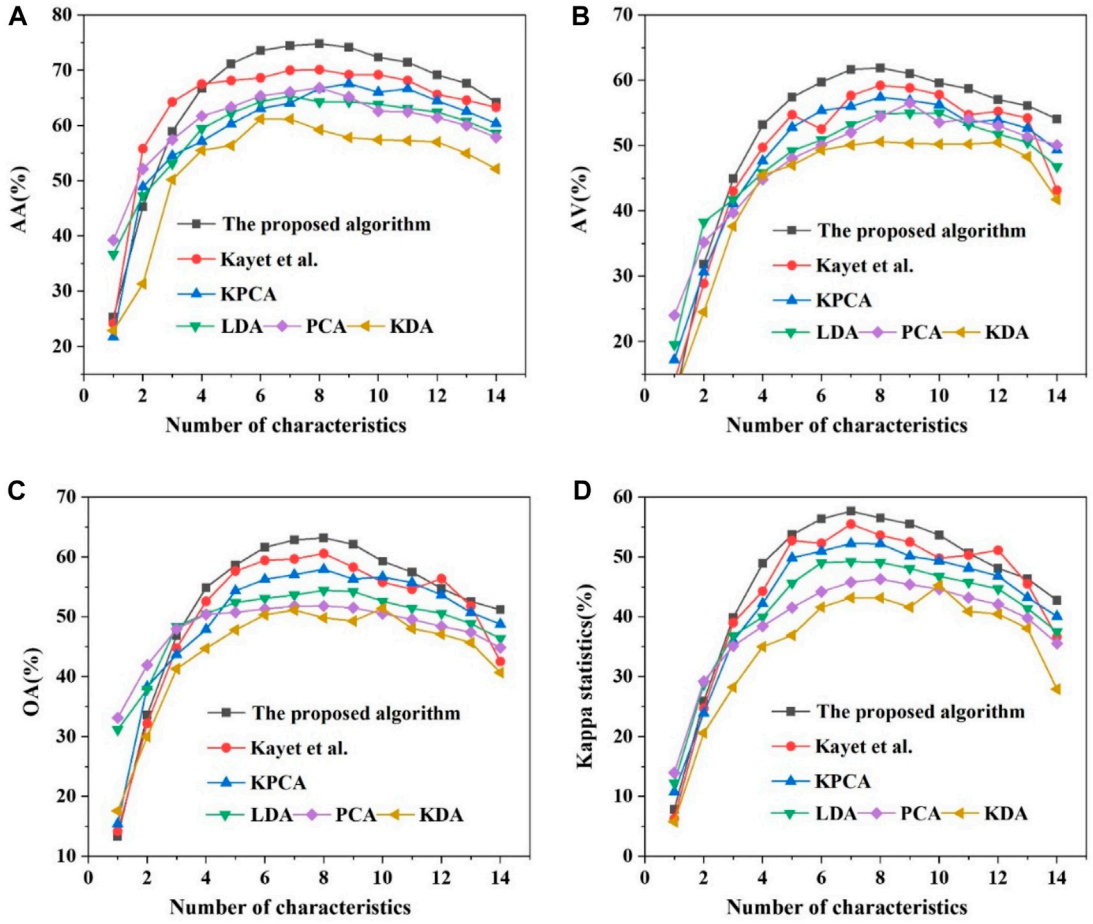


FIGURE 8
Classification results of each evaluation index under different algorithms **(A)** AA; **(B)** AV; **(C)** OA; **(D)** Kappa).

Suggestion

Based on the aforementioned discussion, this work proposes three recommendations:

Firstly, it is highly recommended to manage and protect the sub-healthy and unhealthy natural larch forests. The density of forest stands in natural larch forests can be optimized by adopting ecological thinning, replanting, understory mowing and irrigation, and pruning and shaping (Weller et al., 2021). In this way, the natural larch and understory irrigation can grow normally in the natural larch forest, increasing the health grade in the study area.

Secondarily, the community hierarchy in natural larch forests should be adjusted to enhance intra-forest permeability and ventilation by strengthening the management of canopy and depression and by thinning, single plant selection, or pruning (Schuldt et al., 2020; Lin et al., 2022). Ultimately, the natural larch forest will have a multi-level community structure. It is feasible to achieve biodiversity by adjusting the age structure, soil texture, and other measures. Therefore, the richer the biodiversity, the more stable the community structure, and the more pronounced the forest benefits.

Lastly, it is recommended to integrate multiple data sources for the health assessment of natural forests. Along with remote sensing data, other data sources such as topography and hydrological data should be incorporated to enhance the accuracy and reliability of the health assessment. Additionally, regular monitoring of the distribution and health status of natural Larch pine forests should be performed using remote sensing technology. The timely data acquisition capability of remote sensing should be combined with field survey data to promptly assess the health status of natural Larch pine forests.

Conclusion

In this work, a novel approach is proposed for assessing the health status of natural larch forests in the Arxan by incorporating tree crown features and utilizing a spectral-Gabor space discrimination analysis and classification model to extract multispectral remote sensing image features. The results demonstrate the effectiveness of this approach in accurately assessing the health status of natural larch forests. Furthermore, this work highlights the close relationship between forest health

status and environmental factors, human disturbances, and other factors, providing scientific evidence for developing targeted protection and management measures for sustainable development of natural larch forests. This work also presents a promising method for monitoring forest health on a global scale, which can rapidly and accurately evaluate forest health status and provide critical support for forest protection and management. The findings of this work have significant implications for the sustainable utilization of forest resources and ecological environment protection, indicating its substantial scientific research value and social significance. Still, there are some deficiencies here. For example, natural larch is the dominant species in natural larch forests in the study area, and a few other replanting species still present, such as artificial larch forests. Therefore, future research will further distinguish the tree species in the study area to enhance the precision and confidence of the results.

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

AR: methodology, data analysis and writing original draft preparation. HA: results validations, investigation, review and editing. All authors contributed to the article and approved the submitted version.

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

Ahmadi, K., Mahmoodi, S., Pal, S. C., Saha, A., Chowdhuri, I., Nguyen, T. T., et al. (2023). Improving species distribution models for dominant trees in climate data-poor forests using high-resolution remote sensing. *Ecol. Model.* 475, 110190. doi:10.1016/j.ecolmodel.2022.110190

Alam, M., Wang, J. F., Guangpei, C., Yunrong, L. V., and Chen, Y. (2021). Convolutional neural network for the semantic segmentation of remote sensing images. *Mob. Netw. Appl.* 26 (1), 200–215. doi:10.1007/s11036-020-01703-3

Cai, W., Wei, Z., Liu, R., Zhuang, Y., Wang, Y., and Ning, X. (2021). Remote sensing image recognition based on multi-attention residual fusion networks. *ASP Trans. Pattern Recognit. Intelligent Syst.* 1 (1), 1–8. doi:10.52810/TPRIS.2021.100005

Carnegie, A. J., Tovar, F., Collins, S., Lawson, S. A., and Nahrung, H. F. (2022). A coordinated, risk-based, national forest biosecurity surveillance Program for Australian forests. *Front. For. Glob. Change* 4, 218. doi:10.3389/ffgc.2021.756885

Chen, H., Cao, J., Zhu, H., and Wang, Y. (2022). Understanding household vulnerability and relative poverty in forestry transition: A study on forestry-worker families in China's greater khingan Mountains state-owned forest region. *Sustainability* 14 (9), 4936. doi:10.3390/su14094936

Cherubini, P., Battipaglia, G., and Innes, J. L. (2021). Tree vitality and forest health: Can tree-ring stable isotopes be used as indicators? *Curr. For. Rep.* 7 (2), 69–80. doi:10.1007/s40725-021-00137-8

Cortés, A. J., Restrepo -Montoya, M., and Bedoya- Canas, L. E. (2020). Modern strategies to assess and breed forest tree adaptation to changing climate. *Front. Plant Sci.* 11, 583323. doi:10.3389/fpls.2020.583323

Cruz-Ramos, C., Garcia-Salgado, B. P., Reyes-Reyes, R., Ponomaryov, V., and Sadovnychiy, S. (2021). Gabor features extraction and land-cover classification of urban hyperspectral images for remote sensing applications. *Remote Sens.* 13 (15), 2914. doi:10.3390/rs13152914

Dainelli, R., Toscano, P., Di Gennaro, S. F., and Matese, A. (2021). Recent advances in unmanned aerial vehicles forest remote sensing—a systematic review. Part II: Research applications. *Forests* 12 (4), 397. doi:10.3390/f12040397

Diez, Y., Kentsch, S., Fukuda, M., Caceres, M. L. L., Moritake, K., and Cabezas, M. (2021). Deep learning in forestry using uav -acquired rgb data: A practical review. *Remote Sens.* 13 (14), 2837. doi:10.3390/rs13142837

Dos Reis, A. A., Werner, J. P., Silva, B. C., Figueiredo, G. K., Antunes, J. F., Esquerdo, J. C., et al. (2020). Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery. *Remote Sens.* 12 (16), 2534. doi:10.3390/rs12162534

Ecke, S., Dempewolf, J., Frey, J., Schwaller, A., Endres, E., Klemmt, H. J., et al. (2022). UAV-based forest health monitoring: A systematic review. *Remote Sens.* 14 (13), 3205. doi:10.3390/rs14133205

Falťan, V., Petrovi č, F., Gábor, M., Šagát, V., and Hruška, M. (2021). Mountain landscape dynamics after large wind and bark beetle disasters and subsequent logging—case studies from the carpathians. *Remote Sens.* 13 (19), 3873. doi:10.3390/rs13193873

Fernandez-Carrillo, A., Patoč ka, Z., Dobrovolný, L., Franco-Nieto, A., and Revilla-Romero, B. (2020). Monitoring bark beetle forest damage in Central Europe. A remote sensing approach validated with field data. *Remote Sens.* 12 (21), 3634. doi:10.3390/rs12213634

Ge, G., Shi, Z., Zhu, Y., Yang, X., and Hao, Y. (2020). Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Glob. Ecol. Conservation* 22, e00971. doi:10.1016/j.gecco.2020.e00971

Guimarães, N., Pádua, L., Marques, P., Silva, N., Peres, E., and Sousa, J. J. (2020). Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sens.* 12 (6), 1046. doi:10.3390/rs12061046

Han, Q., Yin, Q., Zheng, X., and Chen, Z. (2022). Remote sensing image building detection method based on Mask R-CNN. *Complex & Intelligent Syst.* 8 (3), 1847–1855. doi:10.1007/s40747-021-00322-z

He, X., Yuan, X., Zhang, D., Zhang, R., Li, M., and Zhou, C. (2021). Delineation of urban agglomeration boundary based on multisource big data fusion—a case study of guangdong–Hong Kong–Macao greater bay area (GBA). *Remote Sens.* 13 (9), 1801. doi:10.3390/rs13091801

Iglhaut, J., Cabo, C., Puliti, S., Piermattei, L., O'Connor, J., and Rosette, J. (2019). Structure from motion photogrammetry in forestry: A review. *Curr. For. Rep.* 5 (3), 155–168. doi:10.1007/s40725-019-00094-3

Jakovljević, T., Lovreškov, L., Jelić, G., Anav, A., Popa, I., Fornasier, M. F., et al. (2021). Impact of ground-level ozone on Mediterranean forest ecosystems health. *Sci. Total Environ.* 783, 147063. doi:10.1016/j.scitotenv.2021.147063

Kayet, N., Pathak, K., Chakrabarty, A., Kumar, S., Chowdary, V. M., and Singh, C. P. (2022). Risk assessment and prediction of forest health for effective geo-environmental planning and monitoring of mining affected forest area in hilltop region. *Geocarto Int.* 37 (11), 3091–3115. doi:10.1080/10106049.2020.1849413

Lin, C. Y., Fu, C. Y., Liu, Y., Zhang, M. Q., Liu, Y., Wu, W. Y., et al. (2022). Assessing the changes of the monetary value of mangrove ecosystem services in China and its application. *Front. Environ. Sci.* 10, 1656. doi:10.3389/fenvs.2022.1018801

Malik, S., Pal, S. C., Das, B., and Chakrabortty, R. (2020). Intra-annual variations of vegetation status in a sub-tropical deciduous forest-dominated area using geospatial approach: A case study of Sali watershed, bankura, West Bengal, India. *Geol. Ecol. Landscapes* 4 (4), 257–268. doi:10.1080/24749508.2019.1633219

Pal, S. C., Chakrabortty, R., Malik, S., and Das, B. (2018). Application of forest canopy density model for forest cover mapping using LISS-IV satellite data: A case study of Sali watershed, West Bengal. *Model. Earth Syst. Environ.* 4, 853–865. doi:10.1007/s40808-018-0445-x

Pan, H., Liu, M., Ge, H., and Chen, S. (2022). Semi-supervised spatial–spectral classification for hyperspectral image based on three-dimensional Gabor and co-selection self-training. *J. Appl. Remote Sens.* 16 (2), 028501. doi:10.1117/1.JRS.16.028501

Quesada- Román, A., and Mata- Cambronero, E. (2021). The geomorphic landscape of the Barva volcano, Costa Rica. *Phys. Geogr.* 42 (3), 265–282. doi:10.1080/02723646.2020.1759762

Reddy, C. S. (2021). Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodivers. Conservation* 30 (10), 2617–2631. doi:10.1007/s10531-021-02216-5

Ruidas, D., Pal, S. C., Islam, A. R. M. T., and Saha, A. (2021). Characterization of groundwater potential zones in water-scarce hardrock regions using data driven model. *Environ. Earth Sci.* 80, 809–818. doi:10.1007/s12665-021-10116-8

Ruidas, D., Chakrabortty, R., Islam, A. R. M. T., Saha, A., and Pal, S. C. (2022a). A novel hybrid of meta-optimization approach for flash flood-susceptibility assessment in a monsoon-dominated watershed, Eastern India. *Environ. Earth Sci.* 81 (5), 145. doi:10.1007/s12665-022-10269-0

Ruidas, D., Pal, S. C., Saha, A., Chowdhuri, I., and Shit, M. (2022b). Hydrogeochemical characterization based water resources vulnerability assessment in India's first Ramsar site of Chilka lake. *Mar. Pollut. Bull.* 184, 114107. doi:10.1016/j.marpolbul.2022.114107

Saha, A., Ghosh, M., and Pal, S. C. (2021a). Forest health assessment using advanced geospatial technology in Buxa reserve forest, sub-Himalayan West Bengal, India. *For. Resour. Resil. conflicts*, 49–61. doi:10.1016/B978-0-12-822931-6.00004-6

Saha, A., Ghosh, M., Pal, S. C., Chowdhuri, I., Chakrabortty, R., Roy, P., et al. (2021b). "Assessment of forest cover dynamics using forest canopy density model in Sali river basin: A spill channel of damodar river," in *Spatial modeling in forest resources management: Rural livelihood and sustainable development*, 365–384. doi:10.1007/978-3-030-56542-8_15

Schuldt, B., Buras, A., Arend, M., Vitasse, Y., Beierkuhnlein, C., Damm, A., et al. (2020). A first assessment of the impact of the extreme 2018 summer drought on Central European forests. *Basic Appl. Ecol.* 45, 86–103. doi:10.1016/j.baae.2020.04.003

Shen, H., Zhou, M., Li, T., and Zeng, C. (2019). Integration of remote sensing and social sensing data in a deep learning framework for hourly urban PM2. 5 mapping. *Int. J. Environ. Res. public health* 16 (21), 4102. doi:10.3390/ijerph16214102

Tan, B. A., Gaw, L. Y. F., Masoudi, M., and Richards, D. R. (2021). Nature-based solutions for urban sustainability: An ecosystem services assessment of plans for Singapore's first "forest town". *Front. Environ. Sci.* 9, 610155. doi:10.3389/fenvs.2021.610155

Thakur, T. K., Patel, D. K., Thakur, A., Kumar, A., Bijalwan, A., Bhat, J. A., et al. (2021). Biomass production assessment in a protected area of dry tropical forest ecosystem of India: A field to satellite observation approach. *Front. Environ. Sci.* 9, 757976. doi:10.3389/fenvs.2021.757976

Tian, Y., Yang, C., Huang, W., Tang, J., Li, X., and Zhang, Q. (2021). Machine learning-based crop recognition from aerial remote sensing imagery. *Front. Earth Sci.* 15(1), 54–69. doi:10.1007/s11707-020-0861-x

Torres, P., Rodes -Blanco, M., Viana -Soto, A., Nieto, H., and García, M. (2021). The role of remote sensing for the assessment and monitoring of forest health: A systematic evidence synthesis. *Forests* 12 (8), 11341134. doi:10.3390/f12081134

Wang, S., Zhong, R., Liu, L., and Zhang, J. (2021). Ecological effect of ecological engineering projects on low-temperature forest cover in Great Khingan Mountain, China. *Int. J. Environ. Res. Public Health* 18 (20), 10625. doi:10.3390/ijerph182010625

Weller, D. L., Love, T. M., and Wiedmann, M. (2021). Comparison of resampling algorithms to address class imbalance when developing machine learning models to predict foodborne pathogen presence in agricultural water. *Front. Environ. Sci.* 9, 701288. doi:10.3389/fenvs.2021.701288

Yang, Y. (2022). Research hotspots and evolution trends of rural resilience in the perspective of biodiversity—based on citeSpace's visual analysis of bibliometrics. *Front. Environ. Sci.* 10, 2039. doi:10.3389/fenvs.2022.1029510

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 01 June 2023
DOI 10.3389/fevo.2023.1177849

OPEN ACCESS

EDITED BY
Jie Luo,
Aerospace Information Research Institute (CAS), China

REVIEWED BY
Peixiao Wang,
Wuhan University, China
Daniel Fiifi Tawia Hagan,
Nanjing University of Information Science and Technology, China

*CORRESPONDENCE
Yilin Lin
✉ 20200111@kust.edu.cn
Peng Zhou
✉ 212104020091@home.hpu.edu.cn

†These authors have contributed equally to this work

RECEIVED 02 March 2023
ACCEPTED 14 April 2023
PUBLISHED 01 June 2023

CITATION
Han Y, Lin Y, Zhou P, Duan J and Cao Z (2023) Dynamic change, driving mechanism and spatiotemporal prediction of the normalized vegetation index: a case study from Yunnan Province, China.
Front. Ecol. Evol. 11:1177849.
doi: 10.3389/fevo.2023.1177849

Dynamic change, driving mechanism and spatiotemporal prediction of the normalized vegetation index: a case study from Yunnan Province, China

Yang Han[1,2,3†], Yilin Lin[1,2,3*], Peng Zhou[4*†], Jinjiang Duan[5] and Zhaoxiang Cao[6]

[1]Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming, China, [2]Key Laboratory of Geospatial Information Integration Innovation for Smart Mines, Kunming, China, [3]Spatial Information Integration Technology of Natural Resources in Universities of Yunnan Province, Kunming, China, [4]School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China, [5]School of Geomatics and Spatial Information, Shandong University of Science and Technology, Qingdao, China, [6]College of Marine Sciences, Shanghai Ocean University, Shanghai, China

Vegetation indexes have been widely used to qualitatively and quantitatively evaluate vegetation cover and its growth vigor. To further extend the study of vegetation indexes, this paper proposes to study the spatial and temporal distribution characteristics and specific driving mechanisms of vegetation indexes based on the example of Yunnan Province, China, and also adds the study of spatial and temporal prediction methods of vegetation indexes. This paper used data on this region's normalized vegetation index (NDVI), three meteorological factors, and eight social factors from 1998 to 2019. The dynamic change in and driving mechanism of the NDVI were studied using mean value analysis, univariate linear trend regression analysis, and partial correlation analysis. In addition, the Fourier function model and the CA–Markov model were also used to predict the NDVI of Yunnan Province from 2020 to 2030 in time and space. The results show that: (1) The NDVI value in Yunnan Province is high, showing a significant growth trend. The increased vegetation coverage area has increased in the past 22years without substantial vegetation degradation. (2) The positive promotion of meteorological factors is greater than the negative inhibition. The partial correlation of relative humidity among meteorological factors is the highest, which is the main driving factor. (3) The NDVI value is significantly positively correlated with population and economy and negatively correlated with pasture land and agricultural area. (4) The NDVI values are predicted well in time (R=0.64) and space (Kappa=0.8086 and 0.806), satisfying the accuracy requirements. This paper aims to enrich the theoretical and technical system of ecological environment research by studying the dynamic change, driving mechanism, and spatiotemporal prediction of the normalized vegetation index. Its results can provide the necessary theoretical basis for the simulation and prediction of vegetation indexes.

KEYWORDS

normalized vegetation index, spatiotemporal distribution, driving mechanism, CA-Markov model, Fourier function model, prediction

1. Introduction

Vegetation is an essential component of terrestrial ecosystems and is the link between elements of natural geography such as the atmosphere, water, organisms, rocks, and soil (Hédl et al., 2017; Xiao et al., 2018; Li et al., 2021), and it is the most fundamental component of the ecosystem on which all organisms need to depend directly or indirectly. Any change in terrestrial ecosystems will inevitably affect vegetation type, quantity, or quality (Torres-García et al., 2022). Especially as a result of global warming, which has led to a series of changes in vegetation phenology events (Huang et al., 2019; Liu et al., 2020), the frequency of extreme climate events has increased significantly worldwide, causing severe negative impacts on agricultural production, water resources, and the socioeconomic development of ecosystems (Zhao et al., 2018). Therefore, long-term dynamic monitoring and the study of vegetation changes and their driving factors are of great value to understanding global changes' impact and achieving effective ecosystem management (Tong et al., 2019). Ecological problems such as global natural disasters, accelerated desert degradation, and soil erosion are frequent, especially in northern latitudes (Myneni et al., 1997; Tucker et al., 2001). In the past 30 years, ecosystems in more than half of the global regions and countries have experienced different degrees of degradation due to economic development and population growth. Issues such as land degradation, soil erosion, and desertification have occurred in China, especially in the northern regions (Wang et al., 2002). This has led researchers to conduct studies focusing on the northern part. In contrast, Yunnan Province, China, is a rare region with ultra-high vegetation cover and diverse vegetation, which needs to be studied due to the area's unique geomorphological and vegetation characteristics. Therefore, a timely, scientific, and accurate assessment of the dynamics of vegetation indices in Yunnan, understanding the driving mechanisms of vegetation indexes changes, and modeling and predicting the development of vegetation indexes are essential references for the ecological effects in other similar regions.

The NDVI is closely related to vegetation cover, leaf area index, biomass, and land use, which can reflect the greenness of vegetation from macroscopic aspects and is related to photosynthesis in the vegetation canopy (Zhu, 2016). Studies addressing the dynamics of the NDVI and its response characteristics to climate change and human activities have been better developed in recent years. For example, nationwide, Jin et al. (2020) used time series data of the NDVI and meteorological factors such as precipitation and temperature to establish a residual analysis model to achieve quantitative separation of climate change effects on the NDVI and arrived at the conclusion that there are significant spatial differences in the impact of climate change and human activities on the NDVI changes, but the positive contribution is generally dominant. Liu et al. (2014) analyzed the differences in the correlation between the NDVI and climate change for different vegetation types in China, and the differences in the relationship between NDVI change trends and temperature and precipitation in different eco-geographical zones were also studied. Piao et al. (2015) analyzed the driving factors of vegetation change in China based on the leaf area index and believed that the increase in atmospheric CO2 concentration and nitrogen deposition might be the main reasons for promoting vegetation recovery in China. Li et al. (2020) analyzed the dominant factors of vegetation productivity changes in China from 1992 to 2013, and the results showed that radiation made the largest contribution to vegetation productivity changes, followed by temperature and precipitation. Chen et al. (2020) explored the response of vegetation to precipitation anomalies under different climatic and eco-geographical conditions in China. With the refinement of the research scale, Wang et al. (2013) found that climate warming had caused the edge of the cold-temperate forest in the southern part of the Greater Hinggan Mountains to retreat 140 kilometers northward in the past century. The combined effects of moisture conditions and temperature promoted the growth of vegetation in the arid-humid transition zone in northern China (Sun R. et al., 2021). Decreased temperature led to reduced evaporation of soil moisture, alleviating the slight drought trend in southwestern China and resulting in a general increase in the vegetation index in the region (Sun et al., 2021a). The continuous warming of the climate has affected the senescence period and growing season length of grasslands in the temperate zone of China to varying degrees (He et al., 2022). Precipitation plays a decisive role in the changes in the NDVI in the Yarlung Zangbo River Basin on the Tibetan Plateau (Sun et al., 2019), while temperature is the dominant factor in the changes in the NDVI during the growing season in the permanently frozen area of Northeast China. Li et al. (2018) took the Inner Mongolia Plateau of China as their research object and concluded that the NDVI was positively correlated with extreme precipitation and extreme low temperature, negatively correlated with extreme high temperature, and that the sensitivity of different vegetation types to extreme climate was not the same. Many studies have analyzed the relationship between the NDVI and regional and meteorological factors (Tucker and Choudhury, 1987). Still, only a few studies have examined the driving mechanisms between the NDVI and relative humidity and other social factors, especially in Yunnan Province, China.

Therefore, this paper selects annual normalized difference vegetation index (NDVI) spatial distribution data (1998–2019) in China to study the dynamics of the vegetation index in Yunnan Province and analyzes the spatial and temporal changes in meteorological factors using air temperature, relative humidity data, and precipitation data and explores the partial correlation and significance with the NDVI. In addition, the influence of eight social factors on the NDVI is also investigated. These studies reveal the dynamics of the 22-year NDVI in Yunnan Province and its specific driving mechanisms in relation to meteorological and social aspects. Finally, the spatial distribution of the NDVI in Yunnan Province from 2020 to 2030 is predicted through simulations using mathematical and statistical models (Fourier function model) and physical statistical models (CA-Markov model). This paper is intended to provide theoretical support and a reference basis for ecological protection construction in the region.

2. Materials and methods

2.1. Study area

Yunnan Province lies between 21° 8′ and 29° 15′ north latitude and 97° 31′ and 106° 11′ east longitude, with a total area of 394,100 square kilometers, accounting for 4.1% of the total land area of the country and ranking eighth in size in China. In addition, the region is known as the Kingdom of Plants. It is the province with the most significant number of plant species, mainly including tropical, subtropical, temperate, and cold-temperate plants, among which ancient, derived, and exotic plants are abundant. Yunnan accounts for

more than 60% of China's 30,000 species of higher plants. More than 150 tree species are classified as national priorities for protection and development at the first, second, and third levels. The vegetation resources of Yunnan Province are remarkable, and the study of local vegetation indexes has essential reference values for ecological planning and conservation (Figure 1).

2.2. Data

2.2.1. Normalized difference vegetation index data

China's annual normalized vegetation index data comes from the registration and publication system of resources and environmental science data.[1]

2.2.2. Other data

The air temperature data were obtained from the National Centre for Atmospheric Sciences in the UK.[2] The relative humidity data were obtained from the National Earth System Science Data Center[3] in China. The TRMM precipitation data were obtained from NASA Release 7 data, and the monthly precipitation data from 1998 to 2019 were selected to calculate the annual average rainfall. In addition, the social factors associated with the NDVI changes were obtained by statistically and categorically classifying the statistical yearbook data.[4]

2.3. Methods

2.3.1. Average analyses of the normalized difference NDVI

In this paper, the annual NDVI and the three meteorological data were processed into an overall average of 22 years using the maximum value synthesis method. And the NDVI and the three meteorological factors are analyzed temporally and spatially. In addition, the annual data are analyzed temporally using the average of the NDVI and three meteorological data from raster images. The Maximum Value Composite (MVC) method is employed, with the specific calculation process as follows:

$$M_{NDVI_i} = \max\left(NDVI_{ij}\right)\left(i = 1,2\ldots, 12; j = 1,2\ldots, 12\right) \quad (1)$$

In the formula, M_{NDVIi} represents the maximized NDVI value for the i year; i is an integer from 1 to 22, representing the years from 1998 to 2019; $NDVI_{ij}$ represents the monthly NDVI values for the i year, and j is an integer from 1 to 12, representing January to December.

2.3.2. Spatial trend analysis methods

A one-dimensional linear regression analysis model was used to quantitatively analyze the change in vegetation cover trends in Yunnan Province from 1998 to 2019, calculated as follows:

$$S = \frac{t\sum_{i=1}^{t}\left(iN_i\right) - \sum_{i=1}^{t} i\sum_{i=1}^{t} N_i}{t\sum_{i=1}^{t} i^2 - \left(\sum_{i=1}^{t} i\right)^2} \quad (2)$$

Where S represents the slope of the trend line, N_i represents the NDVI value in the first year of the image, and t = 22 (1998–2019) represents the observation time series of the data; S > 0 represents an increase in the NDVI value and an increasing trend of vegetation cover; S < 0 illustrates a decrease in the NDVI value and a decreasing trend of vegetation cover. We use the Mann-Kendall significance test to evaluate the significance of the NDVI spatial trends. The M-K test is a non-parametric method suitable for detecting trends in time series data. By calculating the test statistic and the corresponding value of p, we can determine whether the NDVI spatial trend is significant.

2.3.3. Partial correlation analysis

Partial correlation analysis is a standard method for examining the relationship between changes in the NDVI (Normalized Difference Vegetation Index) and the climatic factors. In this study, we employed a second-order partial correlation analysis model to identify the primary driving factors of the NDVI. Among the four variables, the partial correlation coefficient between any two variables is calculated by excluding the influence of the other two variables, which is referred to as the second-order partial correlation coefficient. The significance of partial correlation coefficients is assessed using an F-test based on the partial correlation analysis. The calculation formula is as follows:

$$r_{ij\cdot hm} = \frac{r_{ij\cdot h} - r_{im\cdot h}r_{jm\cdot h}}{\sqrt{\left(1 - r_{im\cdot h}^2\right)\left(1 - r_{jm\cdot h}^2\right)}} \quad (3)$$

A positive correlation is indicated by $r > 0$, while a negative correlation is represented by $r \leq 0$. In this context, i, j, h, and m denote the combinations of the NDVI, the temperature, the precipitation, and the relative humidity, respectively. For instance, when investigating the partial correlation relationship between the NDVI and the temperature, the influence of the precipitation and the relative humidity is eliminated. This study also analyzes the spatiotemporal variations of the NDVI with the temperature, the precipitation, and the relative humidity, and explores the relationships between the NDVI and these factors. In addition, we integrate and select eight social factors to analyze the impact of anthropogenic factors on the NDVI based on local statistical bureau information. IBM SPSS statistical software is used for the analysis of the NDVI. By utilizing 22 years of average the NDVI data, this study examines the correlation between the NDVI and the eight social factors.

2.3.4. Prediction model

2.3.4.1. Fourier function model

This paper predicts the future NDVI over the study area based on the historical data from 2000 to 2019. Three functions (i.e., the polynomial function, trigonometric function, and Fourier function) are selected for fitting the NDVI variation during 2000–2019. Then, the mathematical and statistical parameters (i.e., R^2, SSE, and RMSE)

1 http://www.resdc.cn
2 https://crudata.uea.ac.uk/
3 http://www.geodata.cn
4 https://www.stats.gov.cn

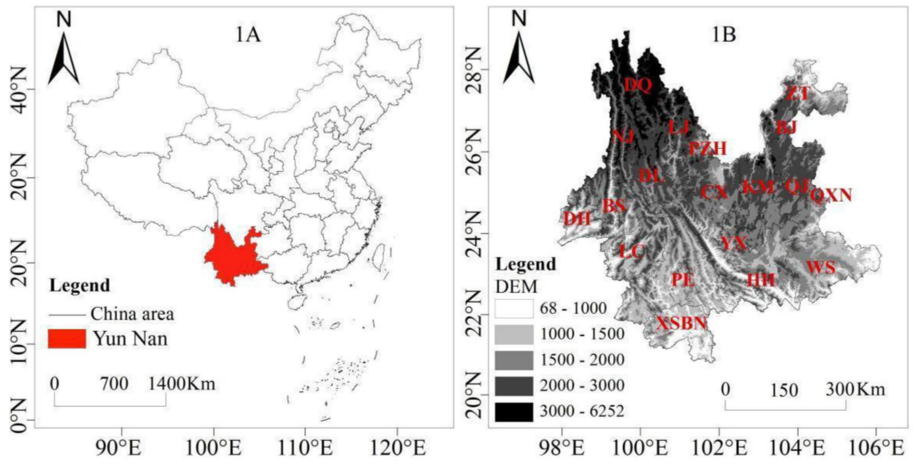


FIGURE 1
Location of the study area and its basic geomorphology.

were calculated to compare the capacities of these fitting functions. Finally, the best-fit process was selected to predict the future NDVI over the study area for 2020–2030.

2.3.4.2. CA-Markov model

The Markov chain is a "non-sequential" state of affairs process, where the state of the change process Sn at time n + 1 is only related to the state at time n and is not related to the state before Sn (Mokarram and Pham, 2022; Sun et al., 2022; Zhou et al., 2022). Thus, P_{ij} represents the probability of shifting from a vegetation index type E_i to another type of E_j and is calculated as:

$$P_{ij} = P\left(S_{n+1} = j / S_n = i\right) \tag{4}$$

In this paper, the number of NDVI change-type areas in 2012 and the NDVI change-type area in 2015 are constructed as a Markov transfer matrix:

$$Pij(N) = \begin{bmatrix} P_{11} & \cdots & P_{1N} \\ \cdots & & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & & \cdots \\ P_{N1} & \cdots & p_{NN} \end{bmatrix} \tag{5}$$

The NDVI type transfer state, where X(0) is the initial state vector, is calculated in this paper to obtain the probability of change of each the NDVI type in 2012 to each the NDVI type change in 2015 and the area size of each the NDVI type in 2017 and 2018 is inferred with this probability. The Markov model focuses on simulating quantitative changes. The CA model and Markov model were coupled to compensate for the deficiency of the Markov model in spatial analysis. Adaptive inertia mechanism-based meta-cellular automata simulations were applied to predict the distribution of each the NDVI type for 2020–2030.

2.4. Data preprocessing steps

To ensure the accuracy and consistency of the data, we adopted the following preprocessing steps:

1. The NDVI data preprocessing: Firstly, we used ArcGIS software and the maximum value composite method to eliminate the influence of clouds, atmosphere, and solar zenith angle, generating monthly the NDVI data. Next, we obtained the annual NDVI data by calculating the average values of the monthly data.
2. Data consistency processing: We projected all data to the same coordinate system and used ArcGIS to resample to achieve the same spatial resolution (1 km x 1 km). Furthermore, for the lower spatial resolution of the TRMM precipitation data, we applied the Kriging spatial interpolation method to achieve the same resolution.
3. Data synthesis: We integrated temperature, precipitation, and relative humidity data into yearly data for 1998–2019, ensuring they have the same temporal resolution. Then, we calculated the annual averages of the NDVI and meteorological data using the multivariate analysis - band collection statistics method.
4. Data clipping: By applying mask extraction and raster clipping, we retained the data required for the study area.
5. Data classification: Based on the spatial distribution of the meteorological data and the NDVI data, we divided them into five classes. At the same time, we classified the significance of the NDVI trends into five categories: significant degradation, moderately significant degradation, no significant change, moderately significant improvement, and significant improvement.
6. The NDVI spatial prediction preparation: Before conducting spatial predictions, we reclassified the NDVI, dividing it into five categories: Lower (<0.5), Low (0.5–0.6), Normal (0.6–0.7), High (0.7–0.8), and Higher (>0.8).

7. Correlated data selection: We selected eight statistical yearbook data closely related to the NDVI, including population density, GDP, industrial output, agricultural output, construction production output, forest area, grassland area, and cultivated land area, covering annual data from 1998 to 2019.

3. Results

3.1. Characteristics of NDVI time dimensional changes

In this study, we extracted the annual average NDVI values from raster images generated by the maximum value composite method and analyzed their temporal changes (Figure 2). We found that the NDVI values in Yunnan Province showed an overall upward trend, with an average NDVI of 0.768 and an annual increase of 0.00614. We also discovered that, over the 22-year period, the temporal changes in the NDVI values only experienced a continuous decline from 2009 to 2011, reaching the lowest point in 2011 and then steadily rebounding until 2014. These results indicate that the temporal changes in Yunnan Province's NDVI values not only reflect a significant increase in vegetation but also exhibit only one substantial fluctuation, further justifying our selection of this study area. Moreover, based on the unique change trends in the study area, research on the temporal changes in NDVI values in the study area is more targeted and facilitates the identification of the main driving factors behind vegetation changes.

In this section, we have discussed in detail the temporal change characteristics of the NDVI, analyzing its change trends and patterns on different time scales. To gain a deeper understanding of the dynamic change characteristics of the NDVI, in the following sections, we will focus on studying the spatial distribution and changes in the NDVI.

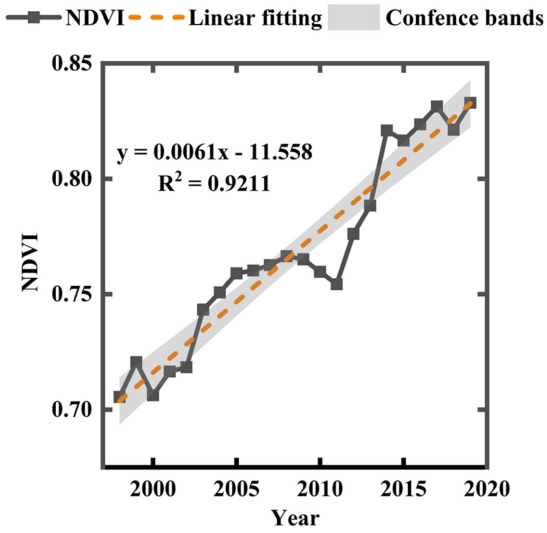


FIGURE 2
Temporal variation of NDVI in Yunnan from 1998 to 2019.

3.2. Characteristics of NDVI spatial distribution variation

In this paper, we analyzed the spatial variation in the NDVI in Yunnan Province over a 22-year period using the maximum value synthesis method (Figure 3A). We found that high vegetation cover areas and higher vegetation cover areas in Yunnan Province accounted for 51.45 and 34.17%, respectively, and low vegetation cover areas accounted for 0.95%. The NDVI values in Yunnan Province show an overall increase from northeast to southwest, with scarce vegetation areas mainly in the northwest and some central urban peripheral areas; high vegetation areas are distributed primarily in the southwest, some significant sites, and the northeast Zhaotong (ZT) area. We analyzed the spatial trends of the NDVI in Yunnan Province over a 22-year period using a one-dimensional linear regression equation model at the metascale (Figure 3B). We found that the direction of increasing NDVI values accounted for 95.02% and decreasing NDVI values accounted for 4.98%, with a vast difference between the two. The NDVI values were only reduced in urban and mountainous areas, such as Kunming (KM) and Yuxi (YX) in the central part of the study area, and urban areas in Lijiang (LJ) and Dali (DL) in the north-central part of the study area as well as in the spreading areas of the Hengduan Mountains in the northwest. We further tested the significance of the regression trend by using the F-trend test method (Figure 3C). We found that the significant increase and decrease in the NDVI accounted for 89.3 and 0.3%, respectively. The size of regions with an increasing trend in the NDVI was much larger than that of sites with a decreasing trend. Among them, the NDVI was significantly reduced in urban areas such as Yuxi (YX) and Qujing (QJ), and there was no significant change in many places in the northwest mountain range area and other mountain range areas. The distribution trend is more dispersed and patchier.

Therefore, the spatial variation and spatial movement of the NDVI and the spatial trend significance test results indicate that the areas with high NDVI values and the regions with increasing trends are much greater than the areas with low NDVI values and areas with decreasing trends in Yunnan Province.

In this section, we explored the spatial change characteristics of the NDVI and analyzed its distribution patterns in detail. To gain a deeper understanding of the various factors influencing the spatiotemporal changes in the NDVI, in the following sections, we focused on investigating the impact of climatic and socio-economic factors on the NDVI and attempted to identify the relationships and mechanisms involved.

3.3. Correlation analysis of the NDVI and climate factors

3.3.1. Impact of air temperature, precipitation, and relative humidity on the NDVI time dimension

This paper analyzes the temporal variation between meteorological factors and NDVI values through the annual mean values of air temperature, precipitation, and relative humidity extracted from raster images generated by the maximum value synthesis method (Figure 4). In terms of temporal variation, we found that the multi-year averages of NDVI values, air temperature, precipitation, and relative humidity

in Yunnan Province showed multi-peaked variations, with the annual average maximum value of NDVI occurring in 2019 at 0.833. The average yearly maximum values of air temperature, precipitation, and relative humidity occurred in 2019, 2001, and 2018, with maximum values of 15.7°C, 1377 mm, and 74.5%, respectively.

We further found that by comparing the temporal variation in NDVI values over the years the temporal variation in the NDVI values from 1998 to 2002 was not insignificant, while air temperature reached a minimum value of 14.5°C in 2000 and precipitation and relative humidity were much higher than other years during this period; the annual average NDVI values from 2014 to 2019 were much higher than other years, and vegetation grew better. However, the air temperature during this period differed significantly from the NDVI temporal variation, while precipitation and relative humidity were in solid agreement with the NDVI temporal variation. Therefore, air temperature showed a negative correlation with NDVI material changes, and precipitation and relative humidity correlated positively.

3.3.2. Effects of air temperature, precipitation, and relative humidity on the spatial dimension of the NDVI

This paper uses the maximum value synthesis method to process the three examined meteorological data into an overall average of 22 years to generate spatial variations in raster images (Figure 5). In terms of spatial variation, combined with (Figure 3), we found that the spatial distribution in the NDVI with air temperature, precipitation, and relative humidity all showed a decreasing trend from south to north. The relative humidity matched the spatial distribution of the NDVI the best.

We further found that the air temperature and relative humidity were numerically higher in the northeastern part of the study area, Zhaotong, at 15.5°C and 78.5%, respectively, while the precipitation was lower at 1050 mm, the NDVI was generally greater than 0.7, and the vegetation increased; in the central part of Kunming, the air temperature were higher at 16.9°C. The amount of precipitation and the relative humidity were lower at 67%, and 700 mm, respectively,

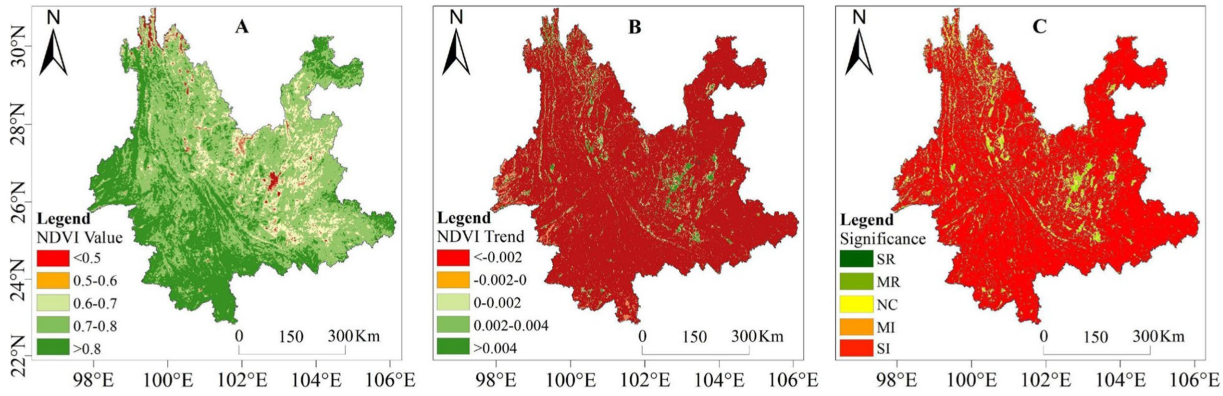


FIGURE 3
Spatial change and significance test distribution of NDVI in Yunnan from 1998 to 2019. **(A)** Spatial variation of NDVI values. **(B)** Trend changes in NDVI. **(C)** NDVI significance test. SR, significantly reduced; MR, moderately reduced; NC, no significant change; MI, moderately increased; SI, significantly increased.

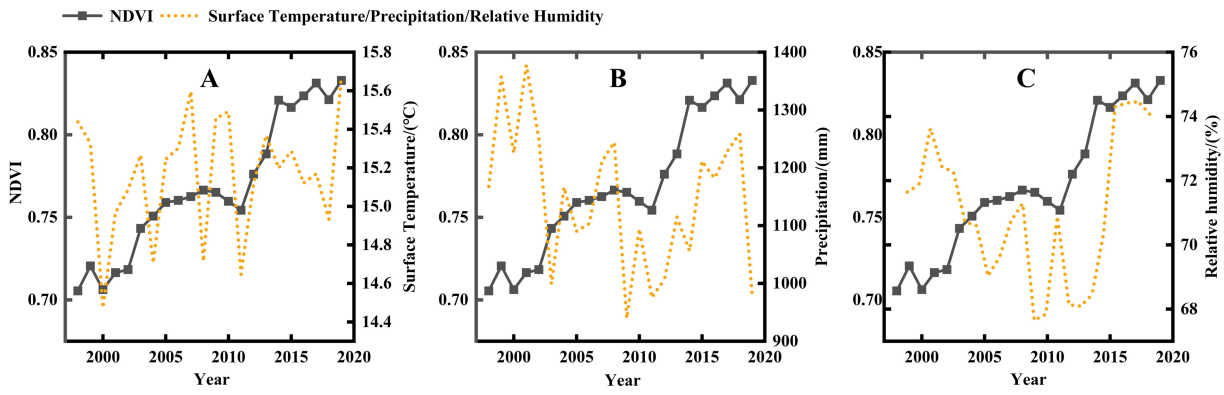


FIGURE 4
Evolution of annual mean air temperature, annual mean precipitation, and annual mean relative humidity in Yunnan from 1998 to 2019. **(A)** Time variation of NDVI and temperature. **(B)** Time variation of NDVI and precipitation. **(C)** Time variation of NDVI and relative humidity.

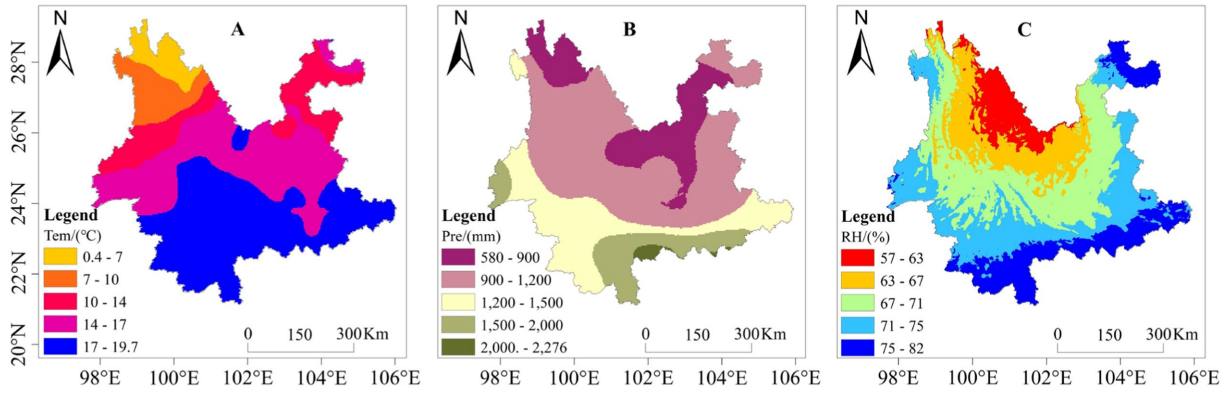


FIGURE 5
Spatial characteristics of air temperature, precipitation and relative humidity in Yunnan, 1998–2019. **(A)** Spatial distribution of temperature. **(B)** Spatial distribution of precipitation. **(C)** Spatial distribution of relative humidity.

and the vegetation did not increase as much as in the northern Zhaotong area. In the southern and southwestern parts of the study area, the air temperature, precipitation, and relative humidity were generally high in numerical values, 18.4°C, 2138 mm, and 79%, respectively, with the NDVI being 0.85, and the area observed the highest increase in vegetation. In the Northwest Transverse Range region, the air temperature, precipitation, and relative humidity were generally shallow, and vegetation showed an insignificant increase.

Therefore, there is some regional variability in the spatial variability and a spatial correlation between air temperature, precipitation, relative humidity, and the NDVI.

3.3.3. Partial correlation analysis between the NDVI and climatic factors

The related research results indicate that climate change is an important cause of increases in the NDVI, while air temperature, precipitation, and relative humidity are important indicators of climate change (Mao et al., 2022; Xu et al., 2022). Based on the findings of this paper on the temporal and spatial variation of the NDVI with air temperature, precipitation, and relative humidity, we further validated our results through a partial correlation model based on the image metric scale (Figures 6A–F).

We found an overall positive spatial correlation between the NDVI and air temperature, with this accounting for 61.6%. The positively correlated areas were mainly Pu'er (PE), Xishuangbanna (XSBN), Wenshan (WS), and Honghe (HH) in the south-central part of the study area. Negative correlations were dominant in areas such as Nujiang (NJ) in the northwestern part and Qujing (QJ) in the northeastern region. The positive and negative correlations were only 0.4% and were speckled in the study area. A total of 30.4% of the spatially negative correlations were found between NDVI and precipitation. The positive correlations were mainly in Zhaotong (ZT) and Wenshan in the northeastern part of the study area. The negative correlation was primarily in the central and western regions. Significant positive and negative correlation areas accounted for 1.70 and 11.20%, respectively. A significant positive correlation was found in southern Qujing (QJ) and Wenshan (WS) in the northeastern part of the study area. A significant negative correlation was found between Diqing and Nujiang (NJ) in the northwestern part of the study area. A total of 69.6% of the NDVI was spatially positively correlated with the relative humidity. The positive correlation areas were mainly Chuxiong (CX) and Yuxi (YX) in the central part of the study area and Pu'er (PE), Lincang (LC), and Baoshan (BS) in the west. The negative correlation was dominant in the Zhaotong (ZT) area in northeast China and Qujing (QJ) and Wenshan (WS) areas in south-central China. Significant positive and negative correlation areas accounted for 19.10 and 7.20%, respectively, with substantial positive correlation areas distributed in the central and western Dali (DL), Chuxiong (CX), and Lincang (LC) regions. The significant negative correlations were distributed in the local areas of Zhaotong (ZT), Qujing (QJ), and Wenshan (WS) in the north. They were especially significant in Zhaotong (ZT) and Qujing (QJ).

Figure 5 shows the spatial distribution of temperature, precipitation, and relative humidity within the entire study area, which helps to explain the differences in the spatial correlations between the NDVI and the climatic factors observed in Figures 6A–F. Notably, in Figure 5, we found that the spatial distribution of relative humidity was most closely aligned with the spatial distribution of the NDVI, which is consistent with the highest positive spatial correlation between the NDVI and the relative humidity shown in Figures 6A–F. It is the spatial distribution differences in the climatic factors within the study area, as shown in Figure 5, that cause significant disparities in the spatial correlations between the NDVI and climatic factors in various parts. These differences contribute to our in-depth understanding of the impact of climate change on vegetation growth in different regions, thereby providing a basis for developing appropriate vegetation conservation and management strategies.

Therefore, there is some regional variability in the spatial correlation between the NDVI and meteorological factors, which is consistent with previous findings (Cheng et al., 2022). There is an overall positive spatial correlation between the NDVI and air temperature and relative humidity, with the highest positive spatial correlation with relative humidity. The negative spatial correlation being with precipitation further suggests that a combination of meteorological factors influences the NDVI.

3.4. Characteristics of NDVI time dimensional changes

A social–ecological system (SES) is a complex adaptive system closely linked between humans and nature, with unpredictable,

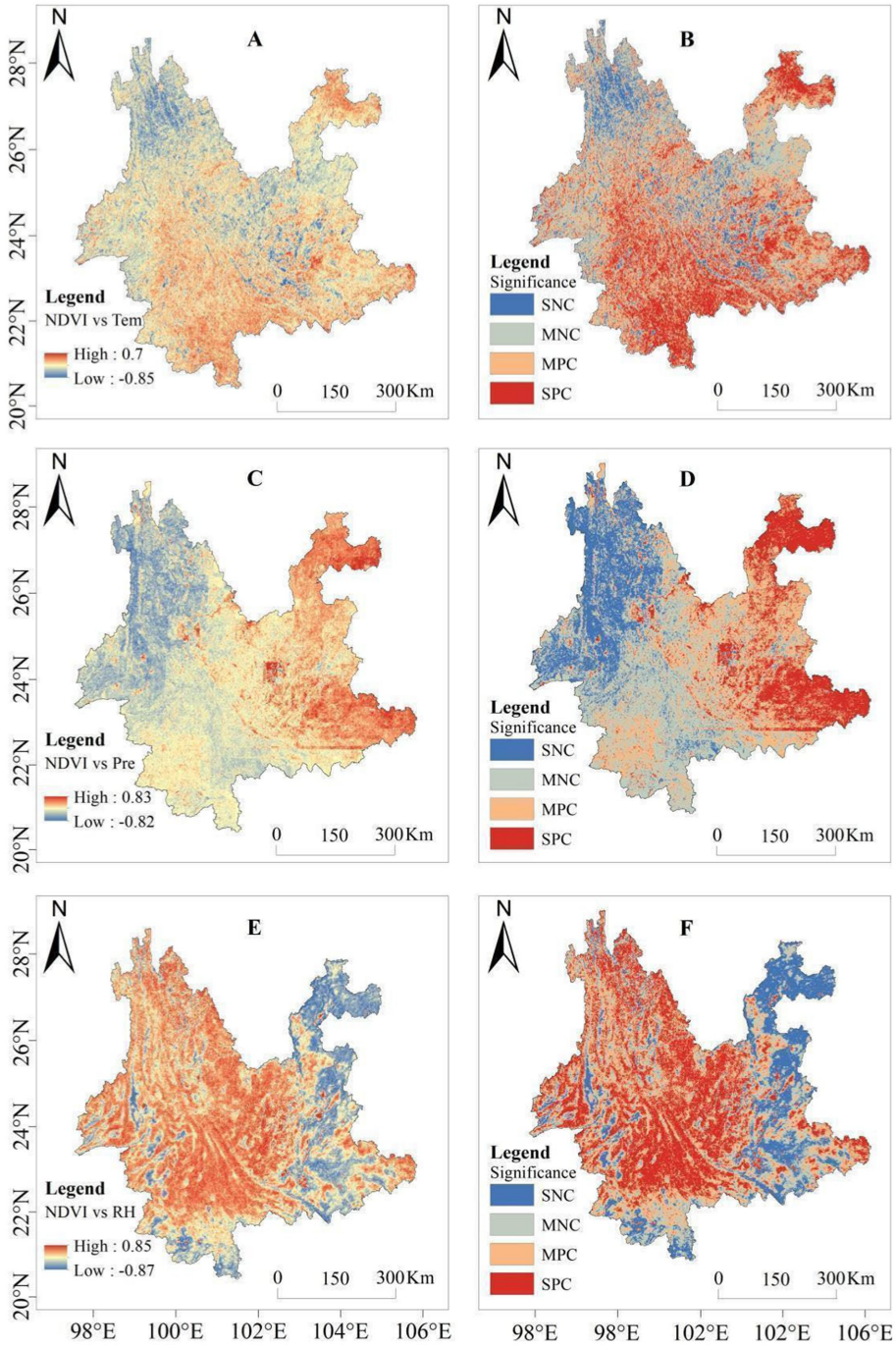


FIGURE 6
Partial correlation coefficients and significance of NDVI and meteorological factors in Yunnan Province from 1998 to 2019. **(A)** The partial correlation coefficient between NDVI and temperature; **(B)** the significance of the correlation between NDVI and temperature; **(C)** the partial correlation coefficient between NDVI and precipitation; **(D)** the significance of the correlation between NDVI and precipitation; **(E)** the partial correlation coefficient between NDVI and relative humidity, and **(F)** the significance of the correlation between NDVI and relative humidity. SNC, significantly negative correlation; MNC, moderately negative correlation; MPC, moderately positive correlation; SPC, significantly positive correlation.

self-organizing, and non-linear characteristics, and vegetation is an essential part of the ecosystem with it being more inextricably linked with society (Tong et al., 2016). In this paper, we analyzed the correlation between the NDVI of vegetation and eight social factors in Yunnan Province from 1998 to 2019 by using a bias correlation model (Table 1). As seen in Table 1, the NDVI had the strongest positive correlation with population density ($R^2 = 0.951$) and the weakest positive correlation with arable land. The NDVI was negatively correlated with agricultural and pasture area changes with R^2 of −0.460 and −0.724, respectively. There were also strong positive correlations between the NDVI and changes in gross product, industry, gross construction product, and forest area.

Following our previous research, we have gained a thorough understanding of the spatiotemporal distribution characteristics of the NDVI and analyzed the influence of climatic and socio-economic

factors on the NDVI. In the following sections, we employed advanced modeling methods to precisely simulate and predict the NDVI in both temporal and spatial dimensions.

3.5. Normalized vegetation index spatial and temporal simulation projections for 2020–2030

3.5.1. Impact of air temperature, precipitation, and relative humidity on the NDVI time dimension

A linear function and two periodic functions were constructed to fit the annual NDVI temporal variation for 1998–2019 (Table 2). The results show that the polynomial function includes the linear function better. In addition, the periodic functions (Fourier and trigonometric functions) performed much better than the linear function (primary function). The accuracy analysis (Table 2) showed that the Fourier function was the best-fitted function among the three functions constructed in this study. It effectively depicted the annual NDVI time-series variation in the study area.

The results of the prediction of the NDVI in Yunnan Province from 2020 to 2030 using the constructed Fourier function show (Figure 7) that the NDVI values in Yunnan Province show fluctuating changes. However, the NDVI values show a slight upward trend in general (0.0015/per year), reaching a maximum weight of 0.703 in September 2027 and a minimum of 0.498 in March 2030.

3.5.2. Modeled prediction of NDVI spatial distribution in 2020–2030

In this paper, based on the regional NDVI data of 2012 and 2015, the probability transfer matrix of each type of NDVI from 2012 to 2015 was obtained. The simulated data for 2017 and 2018 were obtained using the CA-Markov model. The decomposition results of the NDVI for 2017 and 2018 were compared with the simulation results (Figure 8). The simulation results were also verified with Kappa coefficients. The predicted Kappa coefficient values of the NDVI for 2015 and 2018 were 0.8086 and 0.806, respectively (generally, when the Kappa coefficient is greater than or equal to 0.75, the simulation prediction is considered to be more accurate) (Fu et al., 2018), so the

TABLE 1 Correlation between NDVI and social factors in Yunnan, 1998–2019.

	NDVI	Population density	Gross production	Industry	Agriculture	Gross building product	Forests	Pasture	Cropland
NAVI	1								
Population density	0.951**	1							
Gross production	0.922**	0.916**	1						
Industry	0.922**	0.947**	0.971**	1					
Agriculture	−0.460*	−0.552**	−0.363	−0.464*	1				
Gross building product	0.901**	0.886**	0.991**	0.946**	−0.295	1			
Forests	0.932**	0.880**	0.924**	0.890**	−0.410	0.922**	1		
Pasture	−0.724**	−0.671**	−0.838**	−0.705**	0.089	−0.886**	−0.803**	1	
Cropland	0.277	0.355	0.199	0.220	−0.097	0.184	0.177	−0.105	1

*The correlation is significant at the 0.05 level (one-tailed). **At the 0.01 level (two-tailed), the correlation is significant.

TABLE 2 The accuracy analysis of the results by different fitting functions.

Fitting function	Function expression	Number of terms	R^2	SSE	RMSE
Polynomial function	$y = -4.4984E11 + 1.35304E9*x - 1695626.94655*x^2 + 1133.25223*x^3 - 0.42601*x^4 + 8.5408E-5*x^5 - 7.13413E-9*x^6$	6	0.18**	1.40	0.006
Trigonometric function	$y = 0.59975 - 18.38915*\sin(pi*(x + 0.49874)/0.08951)$	–	0.61**	0.66	0.002
Fourier function	$y = 0.5996 - 0.006659*\cos(x*3.151) - 0.005518*\sin(x*3.151) - 0.08763*\cos(2*x*3.151) - 0.01923*\sin(2*x*3.151) + 0.002304*\cos(3*x*3.151) - 0.003563*\sin(3*x*3.151) + 0.001095*\cos(4*x*3.151) + 0.01228*\sin(4*x*3.151) + 0.001425*\cos(5*x*3.151) + 0.003636*\sin(5*x*3.151) - 0.007611*\cos(6*x*3.151) + 0.004719*\sin(6*x*3.151) - 0.001563*\cos(7*x*3.151) + 0.005697*\sin(7*x*3.151)$	7	0.64**	0.62	0.050

*The correlation is significant at the 0.05 level (one-tailed). **At the 0.01 level (two-tailed), the correlation is significant.

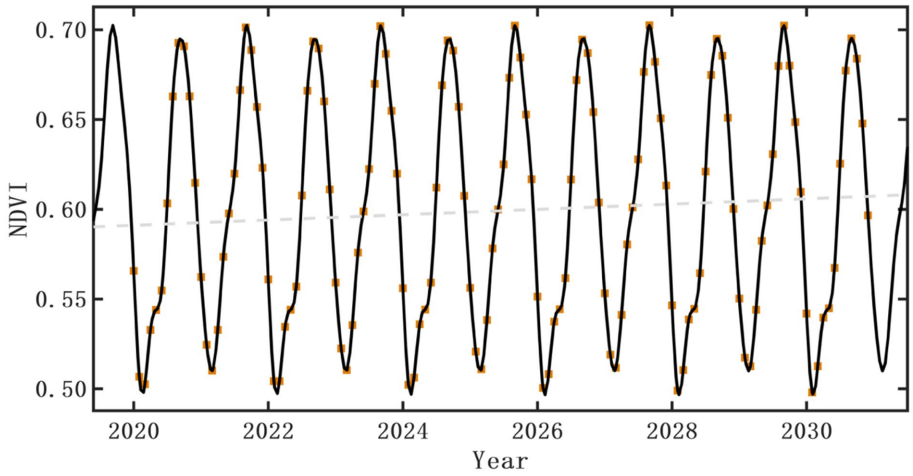


FIGURE 7
Fourier function model fitted prediction results of NDVI values for 2020–2030. The brown box represents the monthly NDVI values, the solid black line represents the finished prediction results, and the gray dashed line represents the linear fit. The linear fit function is y=0.0015*x-2.4.

simulation prediction of each type of NDVI change passed the accuracy test.

In this paper, the CA-Markov model further predicted the spatial variation in the NDVI values from 2020 to 2030 (Figure 9). Combined with Figure 10, we found that the NDVI Lower type will basically remain unchanged over the next 11 years from 2020 to 2030 and is still distributed in urban areas and the northwestern highlands. The NDVI Low type decreased by 0.4%. The NDVI Normal and High types decreased by 0.9 and 2.8%, respectively, while the NDVI Higher type increased by 4%. This indicates that the NDVI Normal and High types were transformed into Higher types, where by the decreasing and increasing areas were mainly distributed in the northwestern and eastern regions, respectively, with a scattered distribution. Therefore, the vegetation in Yunnan Province still has a clear increasing trend from 2020 to 2030.

Up to now, we have comprehensively understood the spatiotemporal distribution characteristics of the NDVI in the study area, the degree of influence of climatic factors on the NDVI, and successfully predicted the spatiotemporal change trends of the NDVI in the study area for 2020–2030. In the subsequent sections, we will continue to delve deeper into the relevant issues highlighted by these findings to ensure that our research conclusions are more rigorous and logical.

4. Discussion

4.1. Spatial and temporal distribution of the normalized difference vegetation index

Studying changes in vegetation dynamics in Yunnan Province is essential to improve vegetation ecological vulnerability assessment, especially as a result of climate change (Fernández et al., 2012; Min et al., 2015; Wang et al., 2016). In this study, we found a high increase in vegetation cover in Yunnan Province over the past 22 years (Figures 2, 3), with an annual growth rate of 0.00614, like previous results based on VIs (Li et al., 2021; Sun et al., 2021b). We also found that the temporal variation in NDVI values showed a significant decreasing trend in the mean NDVI values around 2011, with a faster rate of vegetation recovery in the subsequent years. On the one hand, this is attributed to the fact that meteorological factors (Figure 4) were at low values in around 2011, which made it unsuitable for vegetation growth. On the other hand, during this period, when the 11th and 12th Five-Year Plans converged, the development of industry and agriculture was promoted nationwide, and many natural areas in Yunnan Province were reclaimed. A total of 24.4% of agriculture and 56.7% of industrial enterprises above the scale were increased cumulatively in 4 years. The scale of arable land and buildings was fully expanded, and the vegetation growth environment was damaged, resulting in a sharp decrease in natural vegetation.

The spatial variation in NDVI values shows a trend of decreasing from south to north (Sun et al., 1998; Xie et al., 2021). The low NDVI values are mainly in the northwestern plateau, Gobi region, and urban areas. The plateau Gobi region indicates that the natural environment primarily influences vegetation, while the urban areas suggest that human activities affect vegetation changes. The higher NDVI values in the southern and southwestern regions are due to the favorable climate and more distribution of rivers in these regions, on the one hand, and a greater emphasis on vegetation ecosystem construction in these regions. Regarding spatial trend changes in NDVI values, vegetation in northwestern Yunnan Province is severely reduced, while vegetation in the southern and southwestern areas is better developed. This shows that the trend of plateau globalization in the northwest is due to the harsh natural environment. However, a series of policy measures implemented by the state to improve the ecological environment, such as ecological restoration and the planting of plantation forests, enhanced the environmental climate in the northwest (Du et al., 2019; Hu et al., 2022); more time is still needed for the poorer economy and small population in the northwest. In contrast, the environment is very suitable in the southern and southwestern regions, the population density is low, and the vegetation improvement shows more of a natural progression.

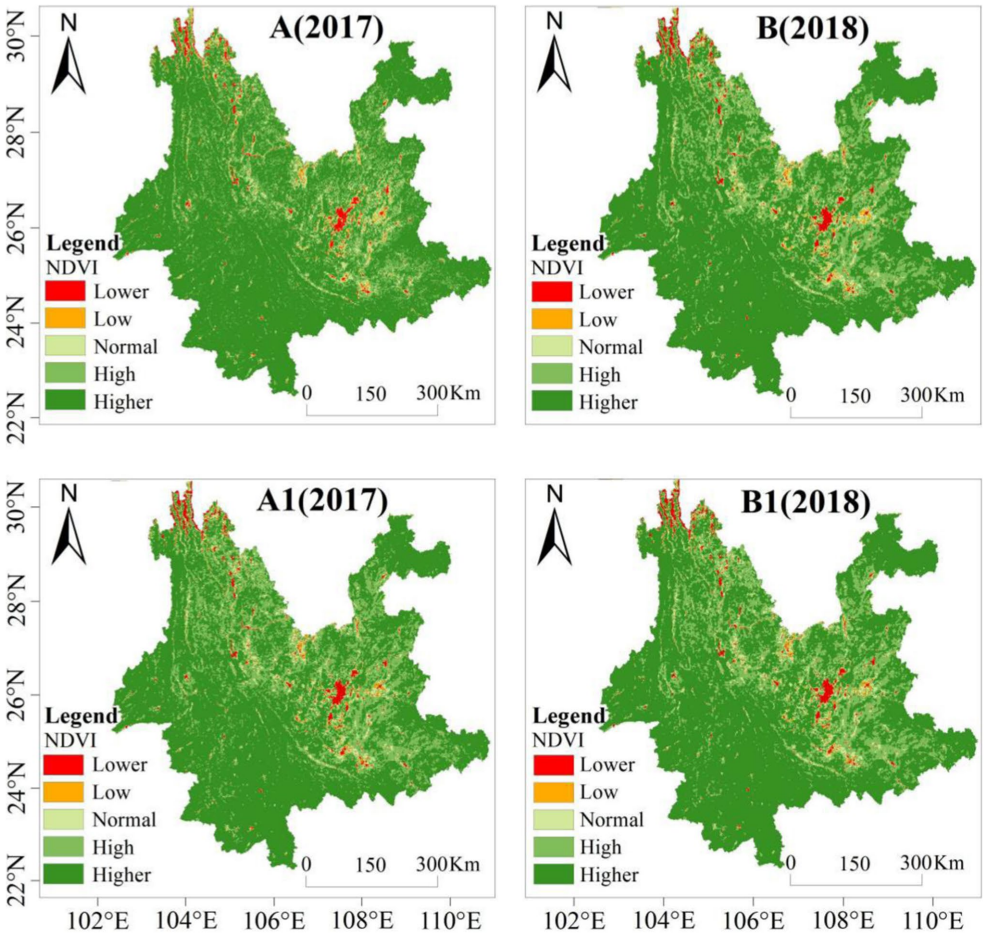


FIGURE 8
Comparative analysis of NDVI accuracy projections for 2017 and 2018. **(A,B)** Represent the original images for 2017 and 2018, respectively, while **(A1,B1)** represent the simulated prediction images. Note that lower indicates NDVI values <0.5; low represents NDVI values 0.5–0.6; normal denotes NDVI values 0.6–0.7; high corresponds to NDVI values 0.7–0.8; higher signifies NDVI values >0.8.

4.2. Effects of temperature, precipitation, relative humidity, and social factors on the normalized difference vegetation index

Meteorological factors influence vegetation growth, providing vegetation's necessary water and heat conditions (Aili et al., 1907). Air temperature, precipitation, and relative humidity are the main controls among meteorological factors that affect vegetation index changes (Jiang et al., 2019). We found that the vegetation index showed a significant decreasing trend in regions with lower air temperature, precipitation, and relative humidity, most likely due to the inconvenience caused by low temperature, low precipitation, and low humidity regions to the natural growth of vegetation and the artificial cultivation of vegetation and other behaviors. Therefore, beneficial human activities in warmer climatic regions are conducive to increasing vegetation growth and the vegetation index.

Our analysis of the biased relationship between vegetation and meteorological factors revealed an overall positive spatial correlation between the NDVI and air temperature and relative humidity, with the highest positive spatial correlation being with relative humidity. In addition, there is a negative spatial correlation between the NDVI and precipitation, further indicating that the NDVI is affected by multiple meteorological factors. It was found that air temperature and relative humidity had different effects on the growth and development of vegetation at various stages due to climate warming. In general, they promoted vegetation growth (Liu et al., 2018). We further found that the areas with a negative correlation between the NDVI and precipitation were mainly in the south-central region, attributed to the fact that this region is mostly in the tropics, where precipitation itself is high. If precipitation increases further, it will weaken vegetation photosynthesis and, thus, reduce vegetation.

In addition, the vegetation/land change caused by human activities is the main factor that affects the NDVI (Zhang et al., 2016; Bai and Li, 2022) and an essential driver of vegetation cover change (Zhang et al., 2011, 2019). We found the strongest positive correlation between the NDVI values and population density. Negative correlations were found with changes in agricultural and grazing land areas. The influence of population density on vegetation dominated densely populated areas, cities, and peri-urban areas. The balanced population growth in Yunnan Province has exceeded 20% in the last 22 years. The continuous rise of economic growth and urban population has led to the expansion of metropolitan construction land area, resulting in a decreasing trend in vegetation cover in some areas. Still, overall, it remained significantly and positively correlated with

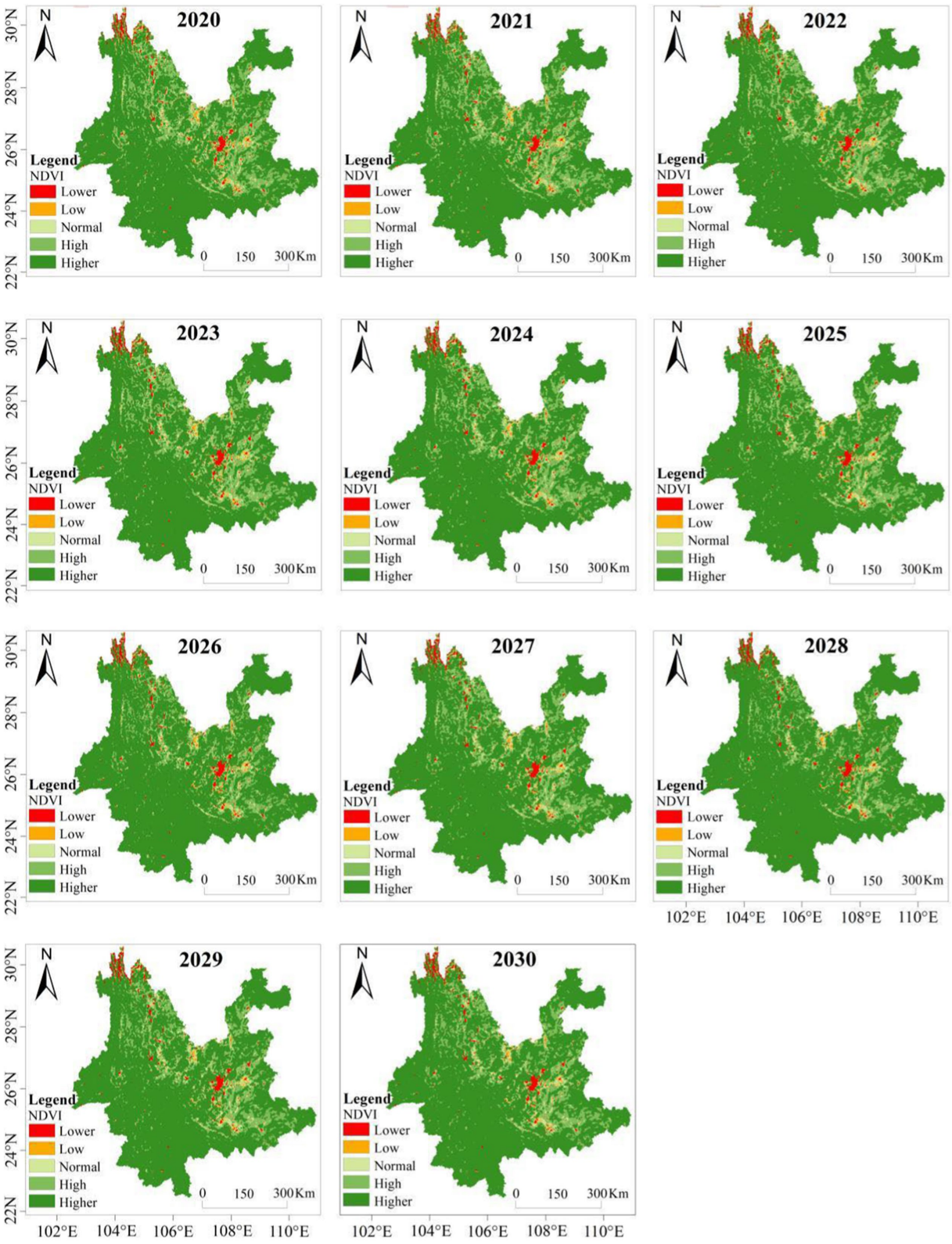


FIGURE 9
Projected spatial distribution of NDVI in Yunnan Province from 2020 to 2030.

the NDVI, stemming from the concentrated distribution of population growth in the study area.

We found that NDVI values were negatively correlated with changes in agricultural and grazing land areas. The increase in agriculture is attributed to the fact that, on the one hand, it represents an increase in agricultural land. On the other hand, excessive agricultural exploitation destroys the natural environment, causing problems such as land desertification, soil erosion, and soil pollution,

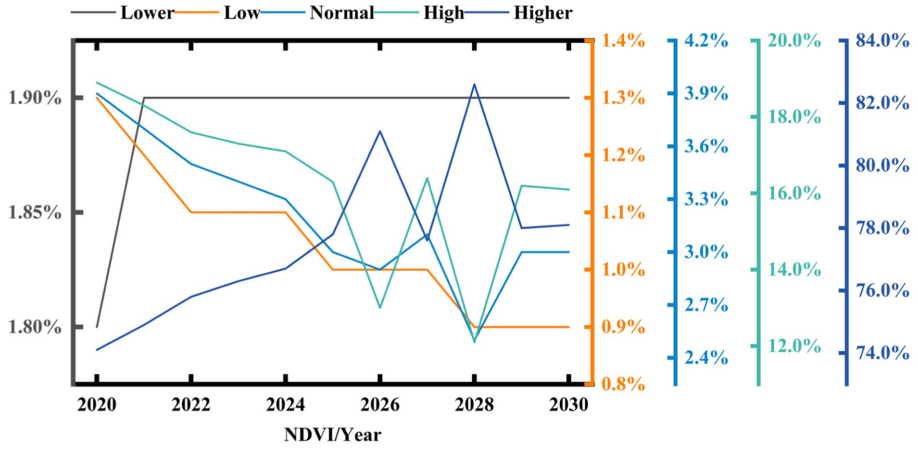


FIGURE 10
Percentage change of NDVI by type in Yunnan from 2020 to 2030.

which in turn leads to a significant decrease in vegetation cover. The livestock industry in the study area is more developed and has been further developed in recent years, leading to a further increase in grazing land. The increase in grazing land represents an increase in grazing, which mainly transforms the vegetation structure morphology and reduces the vegetation cover through behaviors such as animal foraging and trampling.

The NDVI is also significantly and positively correlated with GDP, industry, and construction GDP, indicating that improved socio-economics can increase natural vegetation cover. Since the 21st century, the state has attached great importance to preventing and controlling stone desertification and ecological restoration in Yunnan. It has implemented a series of environmental engineering measures that have played a vital role in the restoration of vegetation, not only improving the vegetation cover in the study area but also playing an essential role in improving the survival environment of vegetation communities, effectively reducing the degree of regional stone desertification, and improving the regional vegetation cover (Cheng et al., 2022).

Combining the results shown in Figure 6, we found that climatic and social factors synergistically affected vegetation index changes. The vegetation growth in the study area was positively and negatively inhibited by social factors, and the overall positive promotion of regional vegetation growth dominated. The development of ecological projects has led to an increase in vegetation cover. Still, the inhibitory effect of unreasonable human development in the development of the economy on vegetation growth should be addressed.

4.3. Normalized vegetation index spatial and temporal simulation prediction

In this paper, three functions were used to fit the time series of the historical NDVI in Yunnan Province, and it was found that the periodic function of the Fourier function performed the best. Roy et al. predicted future changes in the NDVI by using machine learning methods, but large data sets are needed for prediction (Ahmad et al., 2023). In contrast, this paper predicts the NDVI in Yunnan Province based on historical NDVI data in a simple and effective mathematical, statistical way (i.e., Fourier function). As confirmed by previous studies (Tchepel and Borrego, 2010; Güler and Özcan, 2019), these studies also used the Fourier function for analysis and prediction. Notably, the curve fit predicted using monthly data was better than that indicated by their use of interannual data by comparing it with Zhou et al. (2022). Thus, the Fourier function model was more suitable for relevant monthly data. In addition, the current use of CA-Markov models lies mainly in the simulation and prediction of land use patterns (Xu et al., 2022; Luan et al., 2023), and fewer research cases have been used for the simulation and prediction of the NDVI, which, as a type of cover, is a dynamic change in land cover. Simulating and predicting a single vegetation change is more straightforward than the interconversion between different land use types. Simulating and predicting a single vegetation change is shorter than the interconversion between different land use types.

Of course, there are uncertainties in the prediction set out in this paper. In time series prediction, the Fourier function is a mathematical, statistical method that uses historical data to fit and predict the NDVI, limited by the information provided by historical data; in addition, it can superimpose periodic information from historical data into the future predicted values. In spatial distribution prediction, since this study is the first attempt to directly simulate and predict the NDVI distribution now using the CA-Markov model, it focuses on affecting the spatial and temporal patterns under natural evolutionary conditions. At the same time, NDVI change is a complex process influenced by various uncertainties such as nature, human activities, and land use development policies. Therefore, how to adjust the model parameters based on a comprehensive analysis and integrated consideration of the effects of multiple factors also needs to be further explored in depth. In response to these situations, it is strongly recommended to use new and better methods in future investigations.

5. Summary and conclusion

To further expand upon research on the vegetation index, this paper plans to take Yunnan Province as an example; in addition to studying the spatiotemporal distribution characteristics and specific driving mechanism of its vegetation index, a new spatiotemporal

prediction method of the vegetation index is also added. In this paper, the normalized vegetation index (NDVI) data from 1998 to 2019, three meteorological factors, and eight social factors were used to analyze the topographic characteristics of the region. The dynamic change and driving mechanism of the NDVI were are studied using mean value analysis, univariate linear trend regression analysis, and partial correlation analysis. In addition, the Fourier function and CA-Markov models were used to predict the time and space of the NDVI in Yunnan Province from 2020 to 2030. The main conclusions are as follows:

1. From 1998 to 2019, the NDVI value of Yunnan Province showed a significant growth trend, and the annual growth rate was 0.00614. In terms of time, the NDVI value fluctuated but showed an upward trend. In space, the NDVI gradually increased from north to south.
2. The NDVI has a positive spatial correlation with air temperature and relative humidity and a spatial correlation with precipitation. The positive promotion of meteorological factors is more significant than negative inhibition. The partial correlation of relative humidity among the meteorological factors is the highest, which is the main driving factor.
3. The NDVI values had the strongest positive correlation with people, the weakest positive correlation with cropland, and a negative correlation with pasture and agricultural area.
4. In the time series prediction, The NDVI values in Yunnan Province fluctuated, but there was a slight upward trend in the NDVI values (0.0015/per year). In the spatial distribution projection, the vegetation in Yunnan Province still has a significant increasing trend from 2020 to 2030.

This study provides necessary theoretical support for NDVI simulation and forecasting. The predicted NDVI values offer valuable information for decision-makers and strategists in ecological environments.

Data availability statement

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

Author contributions

YH, YL, and PZ: conceptualization, methodology, visualization, and writing – first draft. YH and PZ: software and funding acquisition. YL, JD, and ZC: validation. YH: formal analysis and investigation. JD and ZC: supervision. PZ: writing – review and editing. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (grant no. 41761081), The Basic Research Program of Yunnan Province (grant no. 202201AU070112), The Kunming University of Science Technology introduced talent research start-up fund project (grant no. KKZ3202021055), and Yunnan Province Philosophy and Social Science Planning Project (grant no. PY202129).

Acknowledgments

The authors would like to thank the researchers who have provided the open-source algorithms, which have been extremely helpful to the research in this manuscript.

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

Ahmad, R., Yang, B., Ettlin, G., Berger, A., and Rodríguez-Bocca, P. (2023). A machine-learning based ConvLSTM architecture for NDVI forecasting. *Intl. Trans. in Op. Res.* 2020, 1–24. doi: 10.1111/itor.12887

Aili, A., Xu, H., Zhao, X., Zhang, P., and Yang, R. (1907). Dynamics of vegetation productivity in relation to surface meteorological factors in the Altay Mountains in Northwest China. *Forests* 2022:13.

Bai, Y., and Li, S. (2022). Growth peak of vegetation and its response to drought on the Mongolian plateau. *Ecol. Indic.* 141:109150. doi: 10.1016/j.ecolind.2022.109150

Chen, Z., Wang, W., and Fu, J. (2020). Vegetation response to precipitation anomalies under different climatic and biogeographical conditions in China. *Sci. Rep.* 10, 1–16. doi: 10.1038/s41598-020-57910-1

Cheng, Y., Zhang, L., Zhang, Z., Li, X., Wang, H., and Xi, X. (2022). Spatiotemporal variation and influence factors of vegetation cover in the Yellow River Basin (1982–2021) based on GIMMS NDVI and MOD13A1. *Water* 14:3274. doi: 10.3390/w14203274

Cheng, Z., Zhao, T., Zhu, Y., and Li, H. (2022). Evaluating the coupling coordinated development between regional ecological protection and high-quality development: a case study of Guizhou. *China. Land.* 11:e0228426 doi: 10.3390/land11101775

Du, Z., Zhao, J., Pan, H., Wu, Z., and Zhang, H. (2019). Responses of vegetation activity to the daytime and nighttime warming in Northwest China. *Environ. Monit. Assess.* 191:721. doi: 10.1007/s10661-019-7855-8

Fernández, G., María, E., Baival, B., and Batjav, B. (2012). Cross-boundary and cross-level dynamics increase vulnerability to severe winter disasters (dzud) in Mongolia. *Global Environ. Chang.* 22, 836–851. doi: 10.1016/j.gloenvcha.2012.07.001

Fu, X., Wang, X., and Jeffrey Yang, Y. (2018). Deriving suitability factors for CA-Markov land use simulation model based on local historical data. *J. Environ. Manage.* 206, 10–19. doi: 10.1016/j.jenvman.2017.10.012

Güler, E., and Özcan, B. (2019). PM2.5 concentration prediction based on winters' and Fourier analysis with least squares methods in Çerkezköy district of Tekirdağ. *Int. J. Environ. Pollut. Environ. Modell.* 4, 8–16.

He, Y. J., Kong, Z., Hu, X., Zhang, J., Wang, M., Peng, C. H., et al. (2022). Water and heat conditions separately controlled inter-annual variation and growth trend of NDVI in the temperate grasslands in China. *Acta Ecol. Sin.* 42, 766–777. doi: 10.5846/stxb202101130143 (in Chinese)

Hédl, R., Bernhardt-Römermann, M., Grytnes, J., Jurasinski, G., and Ewald, J. (2017). Resurvey of historical vegetation plots: a tool for understanding long-term dynamics of plant communities. *Science* 20, 161–163. doi: 10.1111/AVSC.12307

Hu, J., Zhou, Q., Cao, Q., and Hu, J. (2022). Effects of ecological restoration measures on vegetation and soil properties in semi-humid sandy land on the Southeast Qinghai-Tibetan plateau. *China. Glob. Ecol. Conserv.* 33:e02000. doi: 10.1016/j.gecco.2022.e02000

Huang, W., Ge, Q., Wang, H., and Dai, J. (2019). Effects of multiple climate change factors on the spring phenology of herbaceous plants in Inner Mongolia, China: evidence from ground observation and controlled experiments. *Int. J. Climatol.* 39, 5140–5153. doi: 10.1002/joc.6131

Jiang, H., Xu, X., Guan, M., Wang, L., Huang, Y., and Liu, Y. (2019). Simulation of spatiotemporal land use changes for integrated model of socioeconomic and ecological processes in China. *Sustainability* 11:3627. doi: 10.3390/su11133627

Jin, K., Wang, F., Han, J. Q., Shi, S. Y., and Ding, W. B. (2020). Effects of climate change and human activities on vegetation NDVI change in China from 1982 to 2015. *J. Geogr. Sci.* 75, 961–974. doi: 10.11821/dlxb202005006 (in Chinese).

Li, J., Chen, Q., Li, Q., Zhao, C., and Feng, Y. (2021). Influence of plants and environmental variables on the diversity of soil microbial communities in the Yellow River Delta wetland. *China. Chemosphere.* 274:129967. doi: 10.1016/j.chemosphere.2021.129967

Li, C., Wang, J., Hu, R., Yin, S., Bao, Y., and Ayal, D. Y. (2018). Relationship between vegetation change and extreme climate indices on the Inner Mongolia plateau, China, from 1982 to 2013. *Ecol. Indic.* 89, 101–109. doi: 10.1016/j.ecolind.2018.01.066

Li, M., Yin, L., Zhang, Y., Su, X., Liu, G., Wang, X., et al. (2021). Spatio-temporal dynamics of fractional vegetation coverage based on MODIS-EVI and its driving factors in Southwest China. *Acta Ecol. Sinica*. 41, 1138–1147. doi: 10.5846/STXB201907101451 (in Chinese)

Li, J., Yu, S. Y., and Liu, L. (2020). Determining the dominant factors of the variability of terrestrial ecosystem productivity in China during the last two decades. *Land Degrad. Dev.* 31, 2131–2145. doi: 10.1002/ldr.3580

Liu, Y., Wang, J., Dong, J., Wang, S., and Ye, H. (2020). Variations of vegetation phenology extracted from remote sensing data over the Tibetan plateau hinterland during 2000–2014. *J. Meteorol. Res.* 34, 786–797. doi: 10.1007/s13351-020-9211-x

Liu, S. H., Yan, D. H., Shi, X. L., and Yuan, Z. (2014). Interannual variation and correlation between vegetation NDVI and climatic factors in China. *Arid Land Geography.* 37, 480–489. doi: 10.13826/j.cnki.cn65-1103/x.2014.03.008 (in Chinese).

Liu, H., Zhang, M., Lin, Z., and Xu, X. (2018). Spatial heterogeneity of the relationship between vegetation dynamics and climate change and their driving forces at multiple time scales in Southwest China. *Agric. For. Meteorol.* 256–257, 10–21. doi: 10.1016/j.agrformet.2018.02.015

Luan, Y., Huang, G., and Zheng, G. (2023). Spatiotemporal evolution and prediction of habitat quality in Hohhot City of China based on the InVEST and CA-Markov models. *J. Arid. Land* 15, 20–33. doi: 10.1007/s40333-023-0090-8

Mao, X., Ren, H. L., and Liu, G. (2022). Primary interannual variability patterns of the growing-season NDVI over the Tibetan plateau and main climatic factors. *Remote Sens.* 14:5183. doi: 10.3390/rs14205183

Min, S. K., Son, S. W., and Seo, K. H. (2015). Changes in weather and climate extremes over Korea and possible causes: a review. *Asia-Pacific J. Atmos. Sci.* 51, 103–121. doi: 10.1007/s13143-015-0066-5

Mokarram, M., and Pham, T. M. (2022). CA-Markov model application to predict crop yield using remote sensing indices. *Ecol. Indic.* 139:108952. doi: 10.1016/j.ecolind.2022.108952

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* 386, 698–702. doi: 10.1038/386698a0

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., et al. (2015). Detection and attribution of vegetation greening trend in China over the last 30 years. *Global* 21, 1601–1609. doi: 10.1111/gcb.12795

Sun, C., Bao, Y., Vandansambuu, B., and Bao, Y. (2022). Simulation and prediction of land use/cover changes based on CLUE-S and CA-Markov models: a case study of a typical pastoral area in Mongolia. *Sustainability* 14:15707. doi: 10.3390/su142315707

Sun, R., Chen, S., and Su, H. (2021). Climate dynamics of the spatiotemporal changes of vegetation NDVI in northern China from 1982 to 2015. *Remote. Sens. Basel.* 13:187. doi: 10.3390/rs13020187

Sun, H., Wang, X., Fan, D., and Sun, O. J. (2021a). Contrasting vegetation response to climate change between two monsoon regions in Southwest China: the roles of climate condition and vegetation height. *Sci. Total Environ.* 802:149643 doi: 10.1016/j.scitotenv.2021.149643

Sun, W., Wang, Y., Fu, Y. H., Xue, B., Wang, G., Yu, J., et al. (2019). Spatial heterogeneity of changes in vegetation growth and their driving forces based on satellite observations of the Yarlung Zangbo River basin in the Tibetan plateau. *J. Hydrol.* 574, 324–332. doi: 10.1016/j.jhydrol.2019.04.043

Sun, H., Wang, C., Niu, Z., and Li, B. (1998). Analysis of the vegetation cover change and the relationship between NDVI and environmental factors by using NOAA time series data. *J. Remote Sens.* 3, 204–210.

Sun, H., Wang, J., Xiong, J., Bian, J., Jin, H., Cheng, W., et al. (2021b). Vegetation change and its response to climate change in Yunnan Province. *China. Adv. Meteorol.* 20:8857589. doi: 10.1155/2021/8857589

Tchepel, O., and Borrego, C. (2010). Frequency analysis of air quality time series for traffic related pollutants. *J. Environ. Monit.* 12, 544–550. doi: 10.1039/B913797A

Tong, L., Liu, Y., Wang, Q., Zhang, Z., Li, J., Sun, Z., et al. (2019). Relative effects of climate variation and human activities on grassland dynamics in Africa from 2000 to 2015. *Ecol. Inform.* 53:100979. doi: 10.1016/j.ecoinf.2019.100979

Tong, S., Zhang, J., Ha, S., Lai, Q., and Ma, Q. (2016). Dynamics of fractional vegetation coverage and its relationship with climate and human activities in Inner Mongolia. *China. Remote. Sens.* 8:776. doi: 10.3390/rs8090776

Torres-García, M. T., Oyonarte, C., Cabello, J., Guirado, E., Rodríguez-Lozano, B., and Salinas-Bonillo, M. J. (2022). The potential of groundwater-dependent ecosystems to enhance soil biological activity and soil fertility in drylands. *Sci. Total Environ.* 826:154111. doi: 10.1016/j.scitotenv.2022.154111

Tucker, C. J., and Choudhury, B. J. (1987). Satellite remote sensing of drought conditions. *Remote Sens. Environ.* 23, 243–251. doi: 10.1016/0034-4257(87)90040-X

Tucker, C. J., Slayback, D. A., Pinzon, J. E., Los, S. O., Myneni, R. B., and Taylor, M. G. (2001). Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *Int. J. Biometeorol.* 45, 184–190. doi: 10.1007/s00484-001-0109-8

Wang, R., Cherkauer, K. A., and Bowling, L. C. (2016). Corn response to climate stress detected with satellite-based NDVI time series. *Remote Sens.* 8:269. doi: 10.3390/rs8040269

Wang, X. Y., Zhao, C. Y., and Jia, Q. Y. (2013). Impacts of climate change on Forest ecosystems in Northeast China. *Adv. Clim. Chang. Res.* 4, 230–241. doi: 10.3724/SP.J.1248.2013.230

Wang, T., Zhu, Z., and Wu, W. (2002). Sandy desertification in the north of China. *Science* 45, 23–34. doi: 10.1007/BF02878385

Xiao, J., Wang, S., Bai, X., Zhou, D., Tian, Y., Li, Q., et al. (2018). Determinants and spatial-temporal evolution of vegetation coverage in the karst critical zone of South China. *Acta Ecol. Sinica.* 38, 8799–8812. doi: 10.5846/stxb201805061010 (in Chinese).

Xie, J., Lu, Z., Xiao, S., and Yan, C. (2021). Driving force and ecosystem service values estimation in the extreme arid region from 1975 to 2015: a case study of Alxa league. *China. Chinese Geogr. Sci.* 31, 1097–1107. doi: 10.1007/s11769-021-1244-2

Xu, X., Liu, J., Jiao, F., Zhang, K., Ye, X., Gong, H., et al. (2022). Ecological engineering induced carbon sinks shifting from decreasing to increasing during 1981-2019 in China. *Sci. Total Environ.* 864:161037 doi: 10.1016/j.scitotenv.2022.161037

Xu, D., Zhang, K., Cao, L., Guan, X., and Zhang, H. (2022). Driving forces and prediction of urban land use change based on the geodetector and CA-Markov model. *Int. J. Digit. Earth.* 15, 2246–2267. doi: 10.1080/17538947.2022.2147229

Zhang, C., Lu, D., Chen, X., Zhang, Y., Maisupova, B., and Tao, Y. (2016). The spatiotemporal patterns of vegetation coverage and biomass of the temperate deserts in Central Asia and their relationships with climate controls. *Remote Sens. Environ.* 175, 271–281. doi: 10.1016/j.rse.2016.01.002

Zhang, G., Xu, X., Zhou, C., Zhang, H., and Ouyang, H. (2011). Responses of grassland vegetation to climatic variations on different temporal scales in Hulun Buir grassland in the past 30 years. *J. Geogr. Sci.* 21, 634–650. doi: 10.1007/s11442-011-0869-y

Zhang, G., Yan, J., Zhu, X., Ling, H., and Xu, H. (2019). Spatio-temporal variation in grassland degradation and its main drivers, based on biomass: case study in the Altay prefecture. *China. Glob. Ecol. Conserv.* 20:e00723. doi: 10.1016/j.gecco.2019.e00723

Zhao, A., Zhang, A., and Liu, X. (2018). Spatiotemporal changes of normalized difference vegetation index (NDVI) and response to climate extremes and ecological restoration in the loess plateau. *China. Theor. Appl. Climatol.* 132, 555–567. doi: 10.1007/s00704-017-2107-8

Zhou, P., Zhao, D., Liu, X., Duo, L., and He, B.-J. (2022). Dynamic change of vegetation index and its influencing factors in Alxa league in the arid area. *Front. Ecol. Evol.* 10:922739. doi: 10.3389/fevo.2022.922739

Zhu, H. (2016). A biogeographical comparison between Yunnan, Southwest China, and Taiwan, Southeast China, with implications for the evolutionary history of the east Asian Flora. *Ann. Mo. Bot. Gard.* 101, 750–771. doi: 10.3417/2011037

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 14 June 2023
DOI 10.3389/fenvs.2023.1172977

OPEN ACCESS

EDITED BY
Jie Luo,
Zhejiang Lab, China

REVIEWED BY
Ming Chang,
Jinan University, China
Hirohiko Nagano,
Niigata University, Japan

*CORRESPONDENCE
Shailesh K. Kharol,
shailesh.kharol@atmoanalytics.com

RECEIVED 24 February 2023
ACCEPTED 02 June 2023
PUBLISHED 14 June 2023

CITATION
Kharol SK, Prapavessis C, Shephard MW, McLinden CA and Griffin D (2023), Cloud-based data mapper (CDM): application for monitoring dry deposition of reactive nitrogen.
Front. Environ. Sci. 11:1172977.
doi: 10.3389/fenvs.2023.1172977

Cloud-based data mapper (CDM): application for monitoring dry deposition of reactive nitrogen

Shailesh K. Kharol[1]*, Cameron Prapavessis[2], Mark W. Shephard[2], Chris A. McLinden[2] and Debora Griffin[2]

[1]AtmoAnalytics Inc., Brampton, ON, Canada, [2]Environment and Climate Change Canada, Toronto, ON, Canada

Spatiotemporal monitoring of reactive nitrogen atmospheric deposition is essential for understanding its impact on sensitive ecosystems and quantifying cumulative effects. However, the sparsity of direct surface flux measurements combined with barriers in dissemination are major limiting factors in providing this information to decision makers and non-experts in a timely manner. This work addresses both aspects of this information gap by, 1) utilizing satellite-derived reactive nitrogen dry deposition data products that can be used by decision-makers to supplement the sparse direct surface flux measurements and 2) fill in measurement gaps. Therefore, we have developed a Reactive Nitrogen Flux Mapper (RNFM) component of the interactive Cloud-based Data Mapper (CDM) for providing easy access of satellite-derived reactive nitrogen (defined here as nitrogen dioxide (NO_2) and ammonia (NH_3)) dry deposition flux spatial maps/data to decision-makers/stakeholders over North America. The RNFM component of CDM has a Graphical User Interface (GUI) that allows users to specify the geographical regions and time periods for computing the average fluxes on the fly using an integrated cloud-based computing platform. The CDM architecture is flexible and can be upgraded in the future to take advantage of upstream satellite data directly on cloud platforms to provide results in near real-time.

KEYWORDS

reactive nitrogen, deposition, satellite, RNFM, cloud-computing, CDM

Introduction

Atmospheric deposition is the process whereby gases and particles are removed from the atmosphere and transferred to the earth's surface. The main modes of transfer are wet (through precipitation) and dry (through a diffusive transfer process at the surface) deposition (Vet et al., 2014). The deposition of reactive nitrogen (defined here as nitrogen dioxide (NO_2) and ammonia (NH_3)) represents an essential source of nutrients to plants and a limiting element for growth in many ecosystems. However, when reactive nitrogen is in excess it has harmful effects on terrestrial and aquatic ecosystems, including soil acidification (Galloway et al., 2003), eutrophication (Bergstrom and Jansson, 2006) and loss of biodiversity (Fenn et al., 2010; Simkin et al., 2016). Human activities (i.e., burning of fossil fuels and production of nitrogen-based fertilizers) have doubled the reactive nitrogen inputs to the environment with a proportional increase of atmospheric deposition on the earth's surface since the start of the 20th century (Fowler et al., 2013).

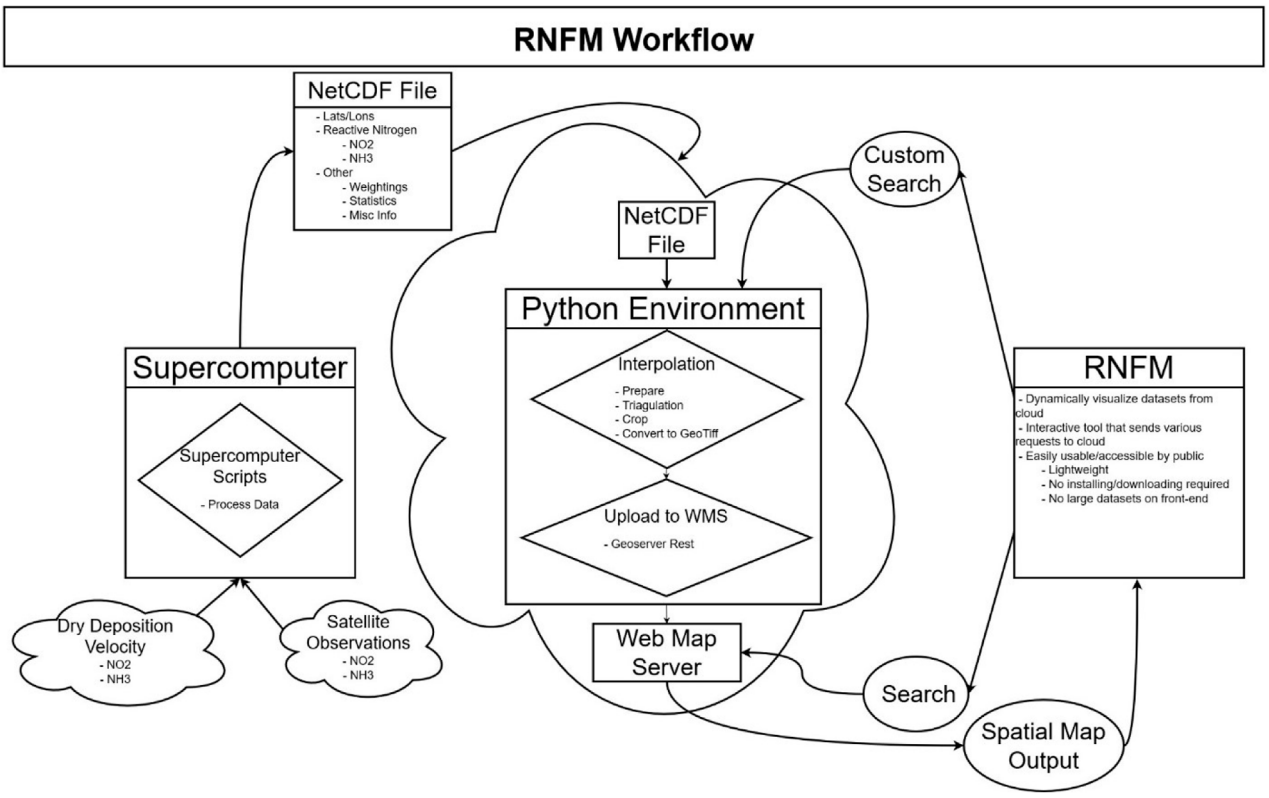


FIGURE 1
Schematic of the process flow of the Reactive Nitrogen Flux Mapper.

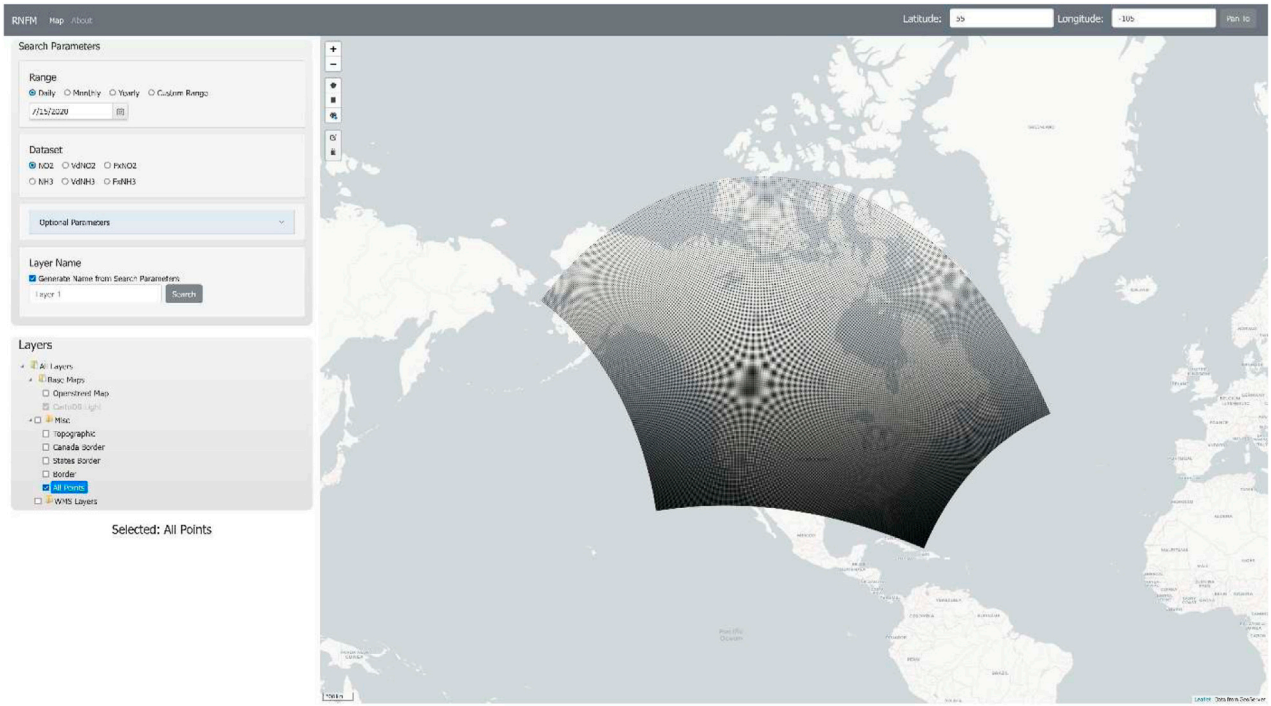


FIGURE 2
Domain boundaries of RNFM.

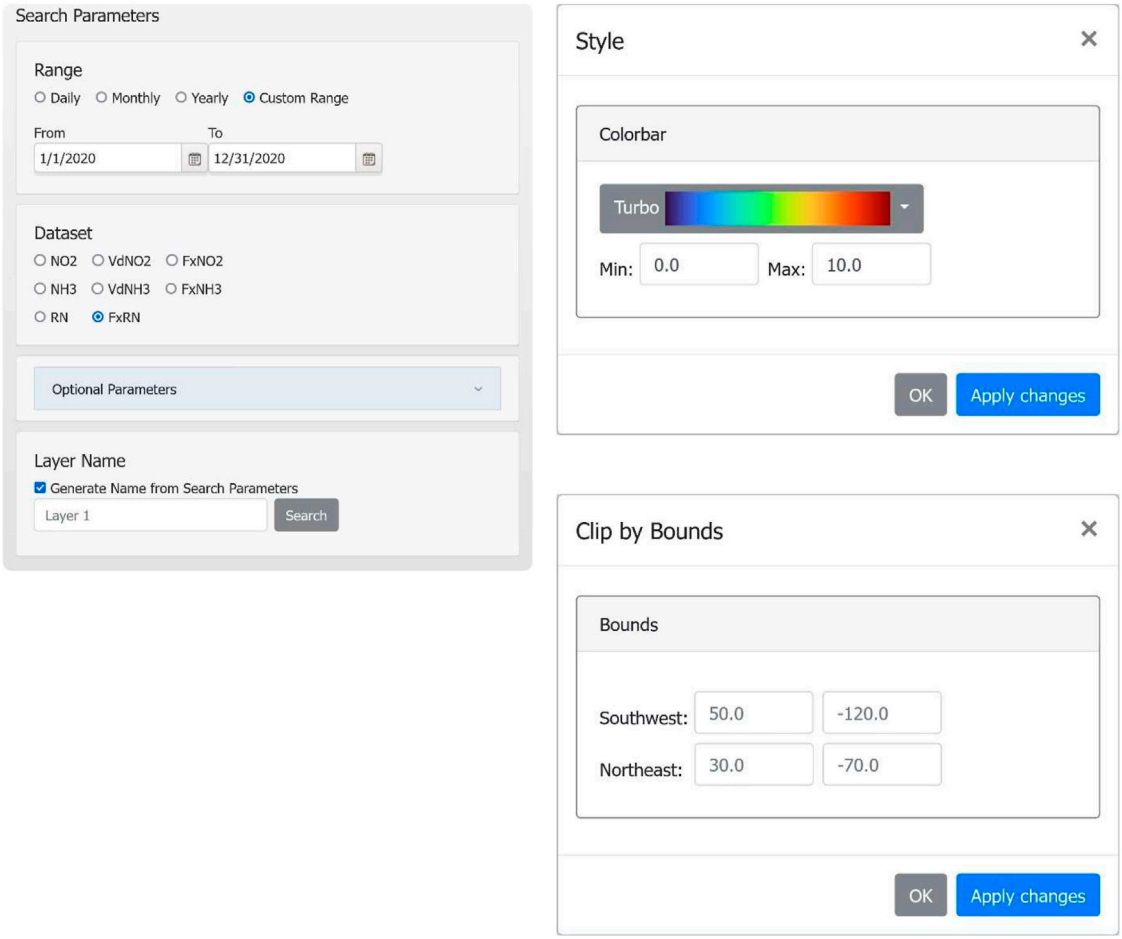


FIGURE 3
Popup that allows the user to define the search parameters and stylize the specified searched layer or clip the specified searched layer by Southwest/Northeast Latitudes/Longitudes bounds that were right-clicked from the layer tree.

In light of its importance, comprehensive monitoring of reactive nitrogen dry deposition flux is required to assess its ecological impacts. Yet, obtaining direct monitoring of dry deposition fluxes is limited as it is more challenging than wet deposition monitoring (Wolff et al., 2010), thus, at present, none of the measurement networks provide direct measurements of the former. The measurement networks for reactive nitrogen dry deposition are sparse in nature and lack the required spatial coverage. The existing measurement networks provide dry deposition flux estimates using the inferential method (which combines the concentration measurements with modelled dry deposition velocities; Wesely, 1989; Zhang et al., 2003), and can not be spatially interpolated like those of wet deposition due to the heterogeneity of dry deposition fields (Schwede and Lear, 2014).

On the contrary, satellite measurements of NO_2 and NH_3 with a daily global coverage offer a valuable data source to fill the measurement gaps and provide an opportunity to analyze the reactive nitrogen dry deposition fluxes spatially using the inferential method (Nowlan et al., 2014; Kharol et al., 2018). However, the processing of large datasets from satellites or models requires high-performance supercomputers that are not readily available to most users. Thus, providing easy access to this large data product to end-users (i.e., decision-makers and stakeholders) remains challenging. Presently, there is not any existing platform where decision-makers and stakeholders can easily access the spatiotemporal satellite-derived reactive nitrogen dry deposition flux information. The model or model-measurement fusion annual maps are available through regional/federal agencies (e.g., US EPA; https://www3.epa.gov/castnet/drydep.html), however, they do not provide the flexibility to users for custom selection (i.e., geographical region and time period) and require >2 years to be produced. In recent years, commercial cloud-computing platforms are becoming popular in the scientific community and have become a valuable alternative for large data processing and complex earth science model runs with its massive computing power and data storage capability. For example, recently, Amazon Web Services (AWS) cloud-computing platforms have successfully been used to run the Goddard Earth Observing System (GEOS)-Chem global 3-D chemical transport model (Zhuang et al., 2020) at 50-km horizontal resolution.

In an attempt to fill this gap we have utilized a cloud-computing platform for space-based earth observations. Here, we describe the

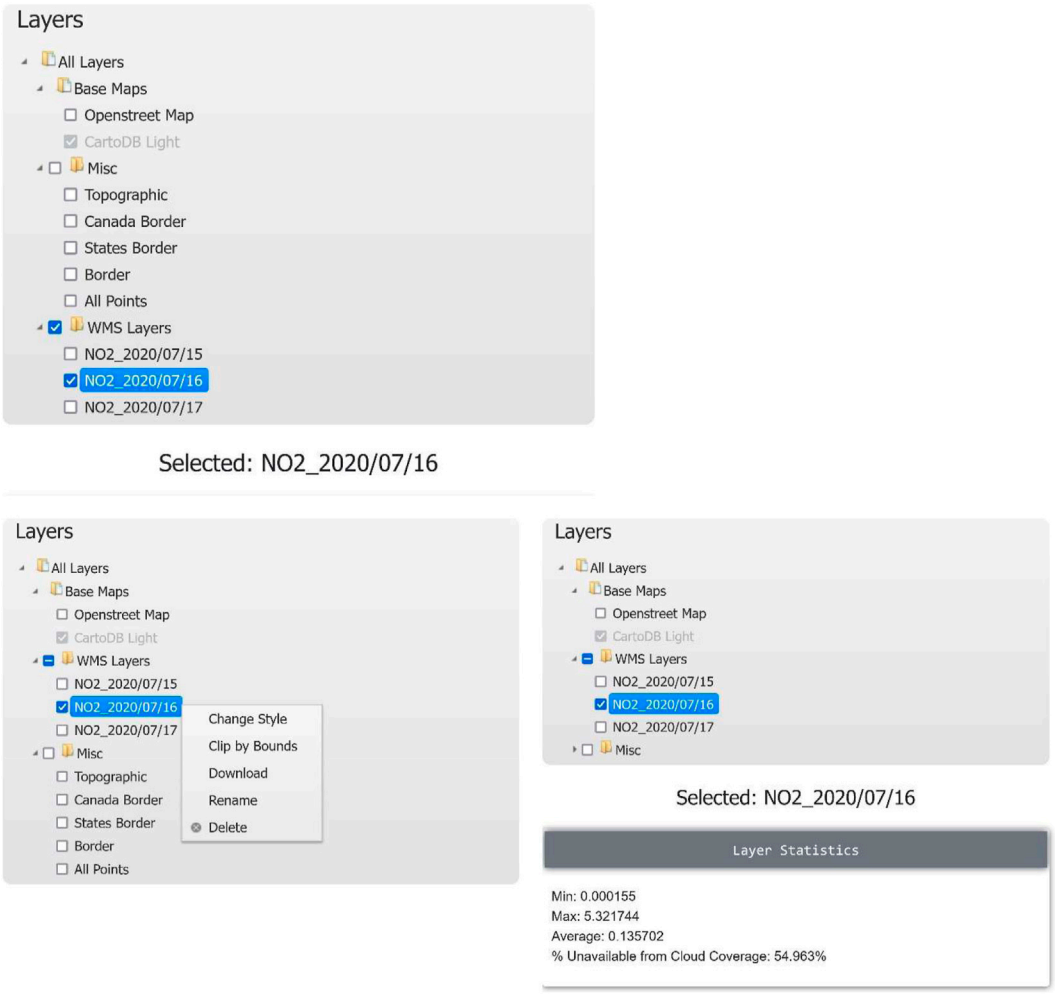


FIGURE 4
The layer tree of RNFM with different functions that can be performed on individual layers by right-clicking or selecting one.

newly developed Reactive Nitrogen Flux Mapper (RNFM) component of the interactive Cloud-based Data Mapper (CDM) with user-friendly features (such as custom selection of geographical region and date range through the Graphical User Interface (GUI)) that provides easy access of satellite-derived reactive nitrogen dry deposition fluxes to end-users.

Reactive nitrogen flux Mapper

The overall schematic of the processing flow describing the RNFM is provided in Figure 1. The upstream preprocessed daily reactive nitrogen dry deposition fluxes are currently computed offline (as described in Appendix A, B) and uploaded to a cloud computing platform together with NH_3 and NO_2 dry deposition velocities and concentrations information. Statistical information and pre-rendered files (i.e daily, monthly, yearly) are then processed, and hosted on the cloud virtual machine (VM) using a Web Map Service (WMS) server. Gridded averages of concentrations, dry deposition velocities and fluxes (i.e., monthly, annual) are calculated using the equations described in Appendix C on the cloud virtual machine (VM). Using the RNFM GUI, users can search for pre-rendered data and retrieve datasets hosted on the VM to be displayed using the WMS on the interactive map as described in Figure 1. The WMS server on the cloud in tandem with our GUI will dynamically load and display the pre-rendered datasets on an interactive map. This allows users to zoom/scroll along the interactive map while the dataset is dynamically loaded from the WMS. In addition to that, the RNFM GUI provides custom selection (i.e., date range and geographical region) flexibility to users where they can select any date range during 2018-to-2020 and geographical coordinates over North America to calculate the averages on the fly. This process is shown in Figure 4. Even though the RNFM GUI provides the flexibility to define a user-specified date range, we recommend using at least a month-long date range selection as this will significantly increase the flux signal compared to the noise in the measurements.

There are two processing streams used for generating datasets for the user through the WMS server on the cloud. The first is the aforementioned pre-rendering of data into standard time-series

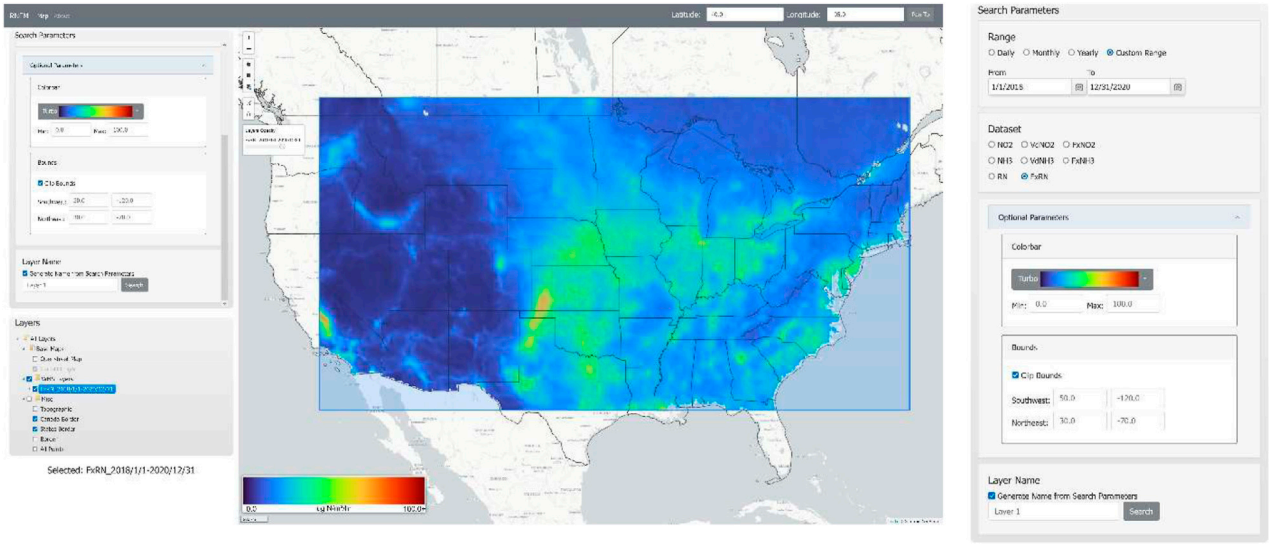


FIGURE 5
Example of clipped 3 year mean reactive nitrogen (Nr) dry deposition flux data for the period of 1/1/2018–12/31/2020 from Southwest to Northeast bounding box.

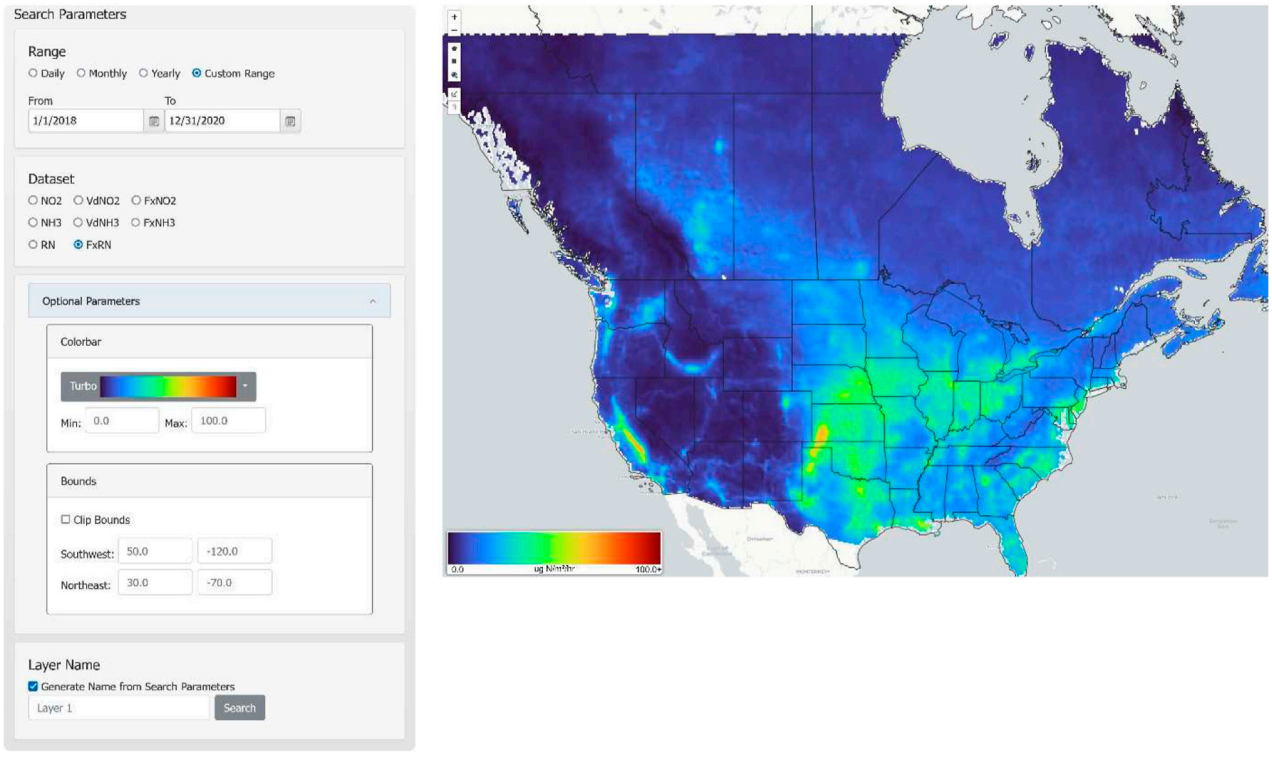


FIGURE 6
An example of RNFM with the inputted search parameters and the resultant map of 3-year average reactive nitrogen dry deposition flux map for 1/1/2018 to 12/31/2020 over North America.

(i.e., daily and monthly files) where the intensive workload is already done allowing for very quick viewing and loading of the datasets, and the second being more custom dynamic searches for subsets of data that have not already been rendered (pre-processed). An example of the custom dynamic process flow would be good if the client makes a request to the cloud using the RNFM custom selection that has not

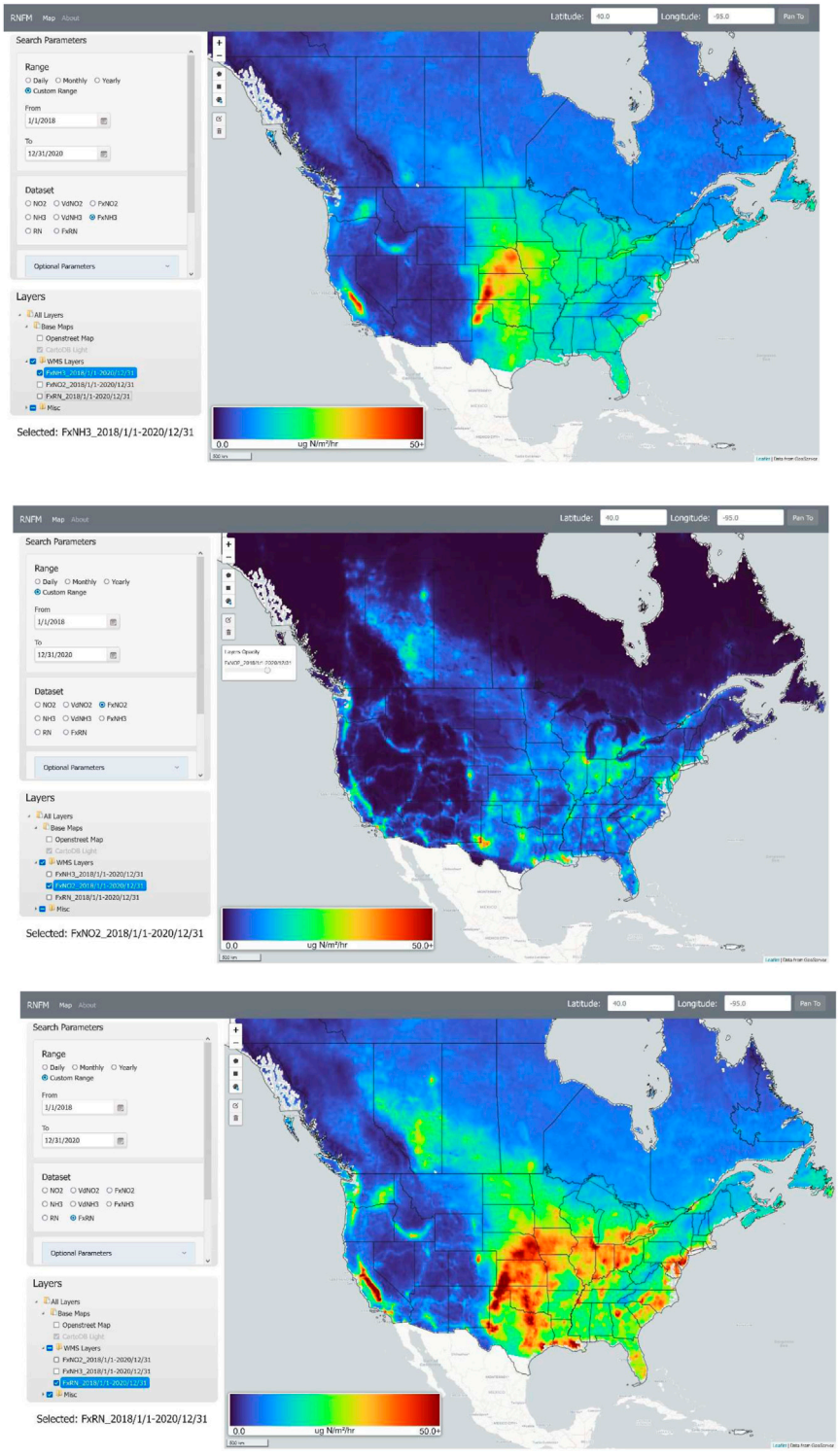


FIGURE 7
Three-year average dry deposition flux map of NH_3, NO_2, and Nr (NH_3+NO_2) for 1/1/2018 to 12/31/2020 over North America.

already been pre-processed (i.e., the weighted average of data from 15 March 2018 to 31 July 2018). In this case, the algorithms on the cloud are used to process and render the requested data in real time that will then be displayed on the user's local machine when completed. Since the rendering needs to be done, the trade-off for this is that it takes some time to complete before being displayed on the client's local machine. This means that for both workflows all intensive work is already pre-done on the cloud or will be done on the cloud. This provides a user-friendly way for users to quickly view and analyze various complex datasets with minimal

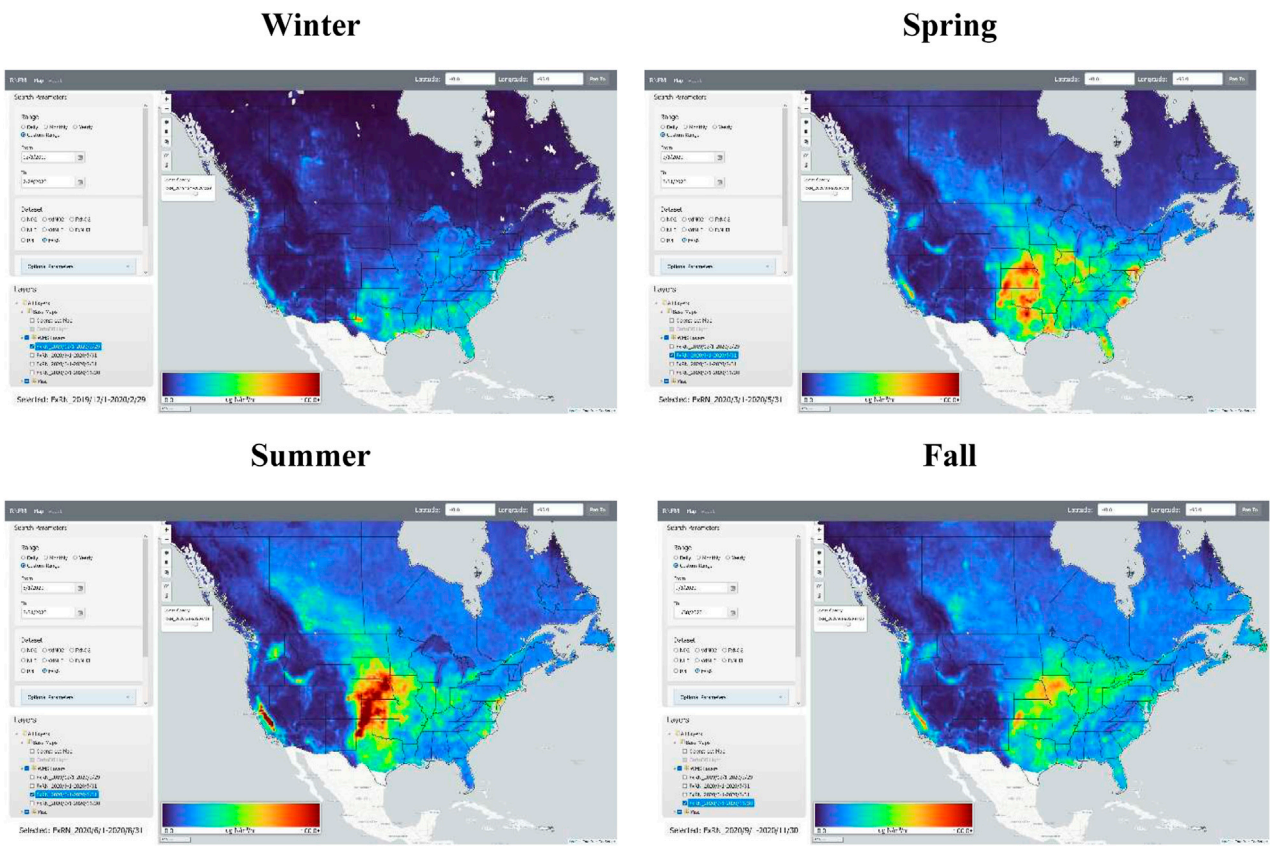


FIGURE 8
Seasonal maps of reactive nitrogen dry deposition fluxes for winter (December-January-February), spring (March-April-May), summer (June-July-August) and fall (September-October-November) over North America for the period of December 2019 to November 2020.

system requirements and performance. The RNFM application can be accessed through https://atmoanalytics.com/rnfm.html.

Figure 2 shows the study area domain map of RNFM where the points represent the latitude/longitude coordinates for which samples were attempted on each day. This region is defined by the GEM regional model output, but can be expanded depending on the availability of model outputs as the satellite observations themselves are global. Fill values (i.e., Nan) are assigned if good quality satellite observations for a specific region are not available (e.g., due to cloud cover). While the majority of data processing is currently performed offline on supercomputers, there are some extra steps needed to format this preprocessed data in such a way that it can be served *via* the WMS server hosted on the cloud to the end-users.

Since satellite observations are typically not on a regular grid, a Triangular Interpolation Network technique is used to format the satellite data into a gridded raster format that also allows datasets to be served to the end-users via the WMS from the cloud. Then these gridded raster files are clipped by the border file (as shown above in Figure 2), which is generated using a convex hull algorithm on all sampled point data. Once these steps are complete, the raster files are uploaded to the WMS server on the cloud and accessed from the RNFM. These algorithms are already performed for standard time-series, or are done on-the-fly from custom user searches on the cloud as stated earlier.

In order to visualize the datasets (which are residing on the cloud platform) on a user's local machine, the user will first need to specify the required inputs using the "Range" and "Dataset" tabs in the search parameters as shown in Figure 3. Here the daily, monthly, and yearly search parameters will display pre-rendered datasets from the specified day/month/year, whereas the custom search parameter will display datasets that will be dynamically processed on the cloud for the specified date range. After this, the user can specify the dataset (one selection at a time) to view the surface concentration, dry deposition velocity, and dry deposition flux for nitrogen dioxide and ammonia. Once these parameters are specified, the user can search for the specified dataset, with the option of providing a user-specified name for the layer that will be displayed on the map. If the layer name is not defined by the user, then a name will be generated automatically for the layer based on their search parameters. Optional parameters include displaying the layer with a custom color palette and searching for data within a specified geographical region. These optional parameters can also be changed later on for that layer as the user sees fit.

After searching for a specified dataset through search parameters (an example shown in Figure 3), the layer will be added to the layer tree with user-specified or automatically generated layer name as shown in Figure 4. The layer tree is a powerful tool that allows users

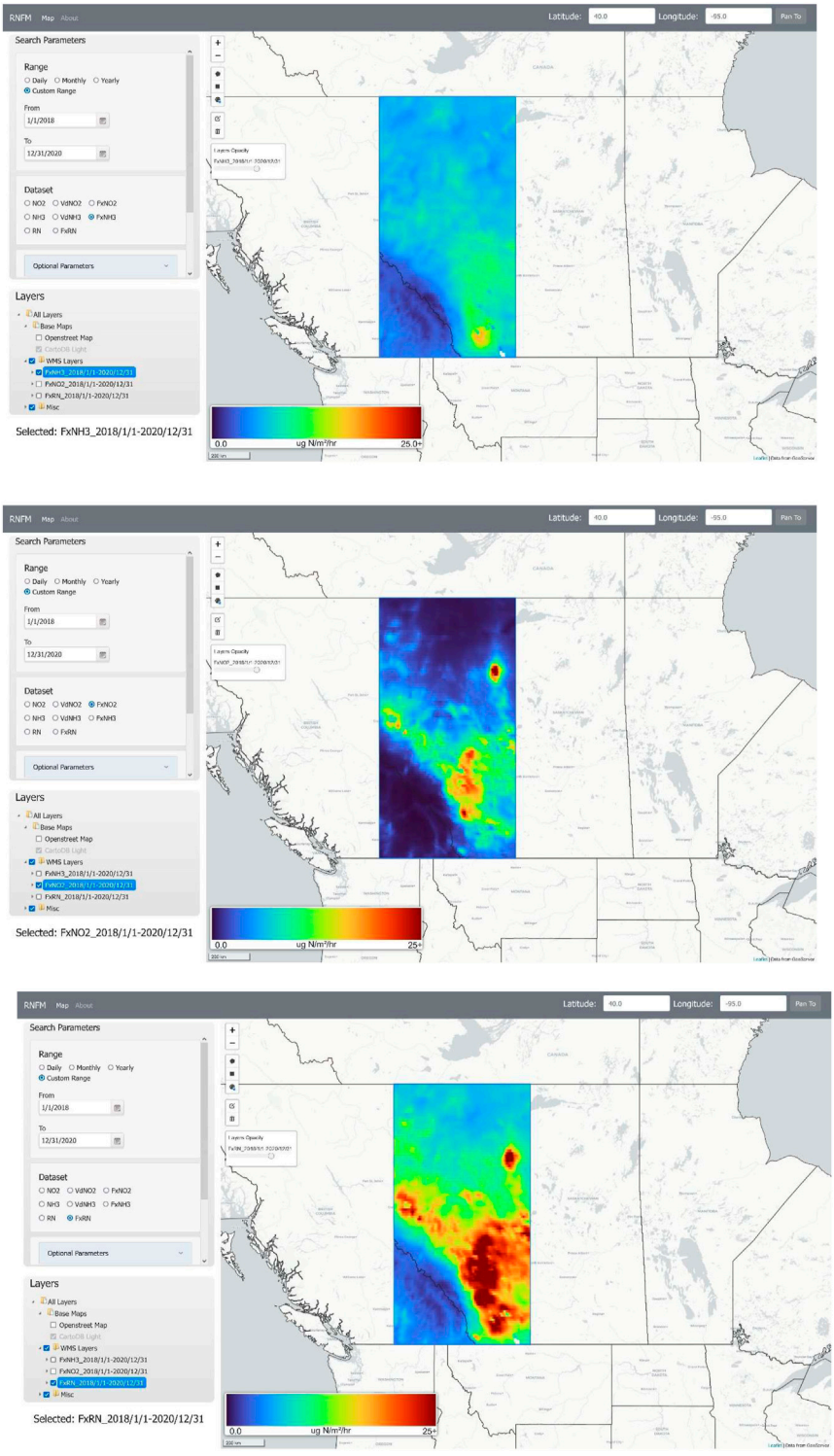


FIGURE 9
Average dry deposition flux map of NH_3, NO_2, and Nr (NH_3+NO_2) for 3-year period (1/1/2018 to 12/31/2020) over Alberta, Canada.

to work with multiple datasets on a single map. The layer tree also allows the users to select their choice of basemap, topographical layer, state/province boundaries to be displayed together with the selected datasets map. In addition to that, the interactive layer tree allows users to add/remove layers to the map by checking/unchecking their corresponding check box. Checked layers are then displayed in order from the bottom to top where the bottom layers of the layer tree overlay the layers above it on the map. The base maps at the top of the layer tree are rendered first on the map and are overlaid by the following checked layers. Users can

also drag and drop layers on the layer tree to new positions, which will affect the order they are displayed on the map.

Since layers have their own unique features, various functions can be performed on individual layers. The user has multiple options to customize the layer as they see fit including changing the layer's style, opacity and clippings by selecting (or "right-clicking") a layer in the layer tree. The general statistics for that given layer (i.e., min, max, average, percentage of points affected by cloud coverage) will be displayed by selecting a layer (as shown in a blue highlight when selected), and provide the flexibility to user's to modify the selected layer through multiple tools. These tools are selectable on the top left corner of the interactive map and allow the user to clip the layer by drawing rectangles or their own custom polygons as shown in Figure 5, and also provide an option to change the opacity of the layer using the opacity function.

Figure 6 shows the 3-year mean (2018–2020) reactive nitrogen dry deposition flux map generated by RNFM in real time using the custom range selection and preprocessed data files uploaded on the cloud. Similar to reactive nitrogen dry deposition flux map, users can generate the average map of any datasets (e.g., reactive nitrogen concentration, NH_3 and NO_2 concentration *etc.*) defined in search parameters under the datasets option for their choice of custom date range and geographical region. The selection of a custom date range is one of the key features of RNFM which provides more flexibility to users to select any time-period to generate the average maps instead of pre-defined averages within the RNFM domain region.

Application

To illustrate RNFM's ability to readily provide the reactive nitrogen flux information for interpretation we have applied it to example case studies. An overall example is provided in Figure 7 showing 3-year average maps of NH_3, NO_2 and reactive nitrogen (Nr (NH_3+NO_2)) dry deposition fluxes across North America that were generated using the RNFM's custom range on-the-fly options. The individual spatial distribution maps of NH_3 and NO_2 dry deposition fluxes in Figure 7 shows that the elevated NH_3 dry deposition generally coincides with the agricultural regions, whereas the NO_2 dry deposition hotspots are mainly located over the cities and industrial regions across North America (Kharol et al., 2018). The combined spatial map of reactive nitrogen dry deposition flux provides cumulative information of the reactive nitrogen deposition from both atmospheric species.

Another example application is the use of RNFM to investigate the changes in the reactive nitrogen dry deposition fluxes by season. Figure 8 shows the seasonal maps of reactive nitrogen dry deposition fluxes for winter (December-January-February), spring (March-April-May), summer (June-July-August) and fall (September-October-November) over North America for the period of December 2019 to November 2020. The NH_3 and NO_2 emissions and lifetime (McLinden et al., 2014; Shephard et al., 2020) as well as their dry deposition velocities (Zhang et al., 2003) vary by season and affect the ambient concentrations and its deposition. It is evident from Figure 8 that the reactive nitrogen deposition from NH_3 is greater during the growing season (i.e., spring and summer) over the agricultural regions (Shephard et al., 2020), whereas deposition from NO_2 is greater during fall and winter seasons in urban/industrial regions (Nowlan et al., 2014; Kharol et al., 2018).

As previously noted, the RNFM also provides an opportunity for users to zoom into a region of interest and visualize/analyze the reactive nitrogen dry deposition in greater detail for that area.

To demonstrate this, here we have clipped the 3-year average of NH_3, NO_2 and Nr dry deposition fluxes over the province of Alberta, Canada as shown in Figure 9. The province of Alberta, Canada is a good example as it has separated source regions of NH_3 and NO_2. As seen in Figure 9 the NO_2 dry deposition hotspots are mostly located near/over the Athabasca oil sands region (Latitude: 57.02 N, Longitude: 111.65 W), urban and industrial regions. Also shown in the figure are the large NH_3 hotspots in and around Lethbridge (Latitude: 49.69 N, Longitude: 112.84 W), which has many concentrated animal feeding operations (CAFOs)) and the main agricultural regions of the province. These example applications demonstrate how easily RNFM can make new research results available to make informed decisions on mitigation strategies for environmental protection in a timely manner.

Summary

The RNFM component of interactive CDM allows users to obtain easy access to the new satellite-derived dry deposition of reactive nitrogen from NO_2 and NH_3 using a user-friendly GUI. The RNFM component provides researchers, stakeholders, and other interested parties with access to new scientific research information on the cumulative effects of reactive nitrogen in land and water ecosystems that can lead to soil acidification, biodiversity loss, and eutrophication (e.g., algal blooms). The CDM is a powerful cloud-based platform application that generates and displays large satellite-derived datasets alleviating numerous hurdles that would otherwise make it much more difficult, time consuming, and resource demanding (storage and computational burden) for users to work with in a meaningful way. The RNFM component of CDM helps overcome the scarcity of ground-based reactive nitrogen dry deposition flux measurements by providing additional new satellite-based information that decision-makers can use to make more informed and timely decisions on mitigation strategies for environmental protection. This CDM architecture can also be enhanced in the future to take advantage of upstream near real time observations directly available on cloud computing platforms.

Data availability statement

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

Author contributions

SK developed the reactive nitrogen dry deposition results. CP developed an interactive cloud-based Reactive Nitrogen Flux Mapper (RNFM). MS developed the CrIS CFPR Level 3 gridded ammonia product. CM and DG developed the TROPOMI

NO2 Level 3 gridded product. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to acknowledge the Labs Canada and Environment and Climate Change Canada for the funding support.

Conflict of interest

Author SK was employed by the company AtmoAnalytics Inc.

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

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

References

Bergstrom, A.-K., and Jansson, M. (2006). Atmospheric nitrogen deposition has caused nitrogen enrichment and eutrophication of lakes in the northern hemisphere. *Glob. Change Biol.* 12, 635–643. doi:10.1111/j.1365-2486.2006.01129.x

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res.* 106 (D19), 23,073–23,095. doi:10.1029/2001jd000807

Fenn, M. E., Allen, E. B., Weiss, S. B., Jovan, S., Geiser, L. H., Tonnesen, G. S., et al. (2010). Nitrogen critical loads and management alternatives for N-impacted ecosystems in California. *J. Environ. Manag.* 91, 2404–2423. doi:10.1016/j.jenvman.2010.07.034

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, N., Reis, S., et al. (2013). The global nitrogen cycle in the twenty-first century. *Phil. Trans. R. Soc. B* 368, 20130164. doi:10.1098/rstb.2013.0164

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., et al. (2003). The nitrogen cascade. *Bioscience* 53, 341–356. doi:10.1641/0006-3568(2003)053[0341:tnc]2.0.co;2

Griffin, D., Zhao, X., McLinden, C. A., Boersma, F., Bourassa, A., Dammers, E., et al. (2019). High-resolution mapping of nitrogen dioxide with TROPOMI: First results and validation over the Canadian oil sands. *Geophys. Res. Lett.* 46, 1049–1060. doi:10.1029/2018GL081095

Kharol, S. K., Shephard, M. W., McLinden, C. A., Zhang, L., Sioris, C. E., O'Brien, J. M., et al. (2018). Dry deposition of reactive nitrogen from satellite observations of ammonia and nitrogen dioxide over North America. *Geophys. Res. Lett.* 45, 1157–1166. doi:10.1002/2017GL075832

McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., et al. (2014). Improved satellite retrievals of NO2 and SO2 over the Canadian oil sands and comparisons with surface measurements. *Atmos. Chem. Phys.* 14, 3637–3656. doi:10.5194/acp-14-3637-2014

Moran, M. D., Ménard, S., Talbot, D., Huang, P., Makar, P. A., Gong, W., et al. (2010). "Particulate-matter forecasting with GEM-MACH15, A new Canadian air-quality forecast model," in *Air pollution modelling and its application XX* (Dordrecht, Netherlands: Springer).

Nowlan, C. R., Martin, R. V., Philip, S., Lamsal, L. N., Krotkov, N. A., Marais, E. A., et al. (2014). Global dry deposition of nitrogen dioxide and sulfur dioxide inferred from space-based measurements. *Glob. Biogeochem. Cycles* 28, 1025–1043. doi:10.1002/2014GB004805

Pendlebury, D., Gravel, S., Moran, M. D., and Lupu, A. (2018). Impact of chemical lateral boundary conditions in a regional air quality forecast model on surface ozone predictions during stratospheric intrusions. *Atmos. Environ.* 174, 148–170. doi:10.1016/j.atmosenv.2017.10.052

Schwede, D. B., and Lear, G. G. (2014). A novel hybrid approach for estimating total deposition in the United States. *Atmos. Environ.* 92, 207–220. doi:10.1016/j.atmosenv.2014.04.008

Shephard, M. W., and Cady-Pereira, K. E. (2015). Cross-track infrared sounder (CrIS) satellite observations of tropospheric ammonia. *Atmos. Meas. Tech.* 8, 1323–1336. doi:10.5194/amt-8-1323-2015

Shephard, M. W., Dammers, E., Cady-Pereira, K. E., Kharol, S. K., Thompson, J., Gainariu-Matz, Y., et al. (2020). Ammonia measurements from space with the cross-track infrared sounder: Characteristics and applications. *Atmos. Chem. Phys.* 20, 2277–2302. doi:10.5194/acp-20-2277-2020

Simkin, S. M., Allen, E. B., Bowman, W. D., Clark, C. M., Belnap, J., Brooks, M. L., et al. (2016). Conditional vulnerability of plant diversity to atmospheric nitrogen deposition across the United States. *Proc. Natl. Acad. Sci.* 113 (15), 4086–4091. doi:10.1073/pnas.1515241113

Vet, R., Artz, R. S., Carou, S., Shaw, M., Ro, C. U., Aas, W., et al. (2014). A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. *Atmos. Environ.* 93, 1–2. doi:10.1016/j.atmosenv.2013.11.013

Wesely, M. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmos. Environ.* 23 (6), 1293–1304.

White, E., Shephard, M. W., Cady-Pereira, K. E., Kharol, S. K., Ford, S., Dammers, E., et al. (2023). Accounting for non-detects: Application to satellite ammonia observations. *Remote Sens.* 15 (10), 2610. doi:10.3390/rs15102610

Wolff, V., Trebs, I., Ammann, C., and Meixner, F. X. (2010). Aerodynamic gradient measurements of the NH3-HNO3-NH4NO3 triad using a wet chemical instrument: An analysis of precision requirements and flux errors. *Atmos. Meas. Tech.* 3, 187–208. doi:10.5194/amt-3-187-2010

Zhang, L., Brook, J. R., and Vet, R. (2003). A revised parameterization for gaseous dry deposition in air-quality models. *Atmos.Chem. Phys.* 3, 2067–2082. doi:10.5194/acp-3-2067-2003

Zhang, J., Moran, M. D., Zheng, Q., Makar, P. A., Baratzadeh, P., Marson, G., et al. (2018). Emissions preparation and analysis for multiscale air quality modeling over the Athabasca Oil Sands Region of Alberta, Canada. *Atmos. Chem. Phys.* 18 (14), 10,459–10,481. doi:10.5194/acp-18-10459-2018

Zhuang, J., Jacob, D. J., Lin, H., Lundgren, E. W., Yantosca, R. M., Gaya, J. F., et al. (2020). Enabling high-performance cloud computing for Earth science modeling on over a thousand cores: Application to the GEOS-Chem atmospheric chemistry model. *J. Adv. Model. Earth Syst.* 12, e2020MS002064. doi:10.1029/2020MS002064

Appendix A:

Datasets

Ammonia (NH_3)

We use the NASA/NOAA SNPP Cross-Track Infrared Sounder Satellite (CrIS, v1.6.3) retrieved level-3 gridded surface NH_3 concentrations obtained from the CrIS Fast Physical Retrieval algorithm (CFPR), which is described in detail by Shephard and Cady-Pereira (2015), and with updates in Shephard et al. (2020) and White et al. (2023) for the period of 2018–2020 over North America. CrIS is an infrared nadir pointing instrument in a sun-synchronous orbit (824 km) with a mean local daytime overpass time of 13:30, and a mean local nighttime overpass time of 1:30 in the descending node. Here, we only used daytime (i.e., 13:30 LST) satellite observations and filtered the data for clouds.

Nitrogen dioxide (NO_2)

Unlike NH3, the main retrieved parameter of NO_2 from the Sentinel-5P (S5P) TROPOspheric Monitoring Instrument (TROPOMI) version 2 (S5P-PAL) data is a total column values that then must be converted to a surface concentration values in a two-step process that utilizes output from the Environment and Climate Change Canada (ECCC) Global Environmental Multi-scale - Modelling Air quality and Chemistry (GEM-MACH) regional air quality model. First, the tropospheric vertical column densities (VCDs) are improved for North American monitoring using the approach described in Griffin et al. (2019), and then the VCDs are converted to surface concentrations following the approach described in McLinden et al. (2014) by scaling the satellite derived VCDs by the ratio of the model surface concentration to model VCD:

$$C = V_{t\_satellite} \times \left(\frac{C}{V_t}\right)_{model} \quad (1)$$

Where C is the surface concentration (ppb), and V_t is the tropospheric VCD. For the model profile and surface concentrations, the GEM-MACH operational model is utilized (Moran et al., 2010; Pendlebury et al., 2018) with a 10×10 km^2 grid cell size for the North American domain. The operational forecast makes use of 2013 emissions information (Zhang et al., 2018). For the conversion we select the daily model output with the closest coincidence to the observation. Since the operational GEM-MACH model is missing sources in the free troposphere, such as lightning and aircraft, a monthly mean GEOS-Chem free tropospheric VCDs (Bey et al., 2001) are added to the GEM-MACH VCDs. These correspond to adjustments of ~0.3-1e15 molec/cm^2.

TROPOMI is a nadir-viewing spectrometer on board the S5P satellite, launched on 13 October 2017. TROPOMI is in a sun-synchronous orbit with an overpass time of 13:30 LST and provides near-daily global coverage of NO_2 with a ground spatial resolution of 3.5×5.5 km^2. More details about TROPOMI are described in Griffin et al. (2019). Here, we only use the TROPOMI observations with "quality assurance value" (qa_value) ≥ 0.75 (the recommended pixel filter, https://sentinel.esa.int/documents/247904/3541451/Sentinel-5P-Nitrogen-Dioxide-Level-2-Product-Readme-File), which remove the less accurate observations (i.e., cloud-covered observations, snow/ice covered observations, errors, and problematic retrievals).

Appendix B:

DRY deposition flux calculation

The daily reactive nitrogen dry deposition fluxes are computed using an inferential method, which combines modeled dry deposition velocities according to the resistance analogy (Wesely, 1989; Zhang et al., 2003) and satellite-derived near-surface observations of NH_3 and NO_2 over North America between 2018–2020. The deposition velocity is a function of the surface type and properties, and meteorological parameters. Here, we calculated the dry deposition velocities of NO_2 and NH_3 according to Zhang et al. (2003) approach and used the meteorological inputs produced by the Environment and Climate Change Canada's Global Environmental Multiscale Model (GEM) together with MODIS land-use/land-cover and leaf area index (LAI) inputs. More details on this approach is described in Kharol et al. (2018).

The dry deposition fluxes are calculated on a 15 km × 15 km GEM grid. The satellite-derived surface concentration of NH_3 and NO_2 are first calculated on 0.1 × 0.1 (~10 × 10 km) grid and regridded to the GEM grid (i.e., 15 km × 15 km). The Gaussian distance weighting from the centre of the grid is used to place the averaged surface NH_3 concentrations on a 0.1 × 0.1 grid. The total sum of the weights also provides information on how well the area in the grid is sampled by the satellite observations. For example, low total weight in a grid indicates that a grid is not sampled well for a given day (e.g., due to cloud cover), where a high weight total indicates the grid was well sampled by the satellite observations (e.g., under clear-sky atmospheric conditions). Missing days are taken care of in the flux calculations by assigning a weight value of 0 for days with no observations. These weights are applied to the weighted average calculations in the scene as described in Appendix C. To easily manage missing observations in the gridding and averaging of TROPOMI measurements we also assign weights to the NO_2 surface concentrations of either 0 or 1, where 0 represents no NO_2 observations available, and 1 represents available NO_2 observations. Similar to the CrIS NH_3, these weights are applied to the weighted average calculation described in Appendix C.

Appendix C:

The averages of concentrations, dry deposition velocities and dry deposition fluxes are calculated on the cloud VM as follows:

$$Flux\ Average\left(F^{sat}{}_{wavg}\right) = \frac{\sum_{d=1}^{n} F^{sat}{}_{d} \times Weight_d}{\sum_{d=1}^{n} Weight_d} = \frac{\sum_{d=1}^{n} (C^{sat}{}_{d} \times V_d) \times Weight_d}{\sum_{d=1}^{n} Weight_d} \quad (2)$$

$$Concentration\ Average\left(C^{sat}{}_{wavg}\right) = \frac{\sum_{d=1}^{n} C^{sat}{}_{d} \times Weight_d}{\sum_{d=1}^{n} Weight_d} \quad (3)$$

$$Deposition\ Velocity\ Average\left(V_{wavg}\right) = \frac{\sum_{d=1}^{n} V_d \times Weight_d}{\sum_{d=1}^{n} Weight_d} \quad (4)$$

Where,

$F^{sat}{}_d$ = Satellite-derived daily dry deposition flux of reactive nitrogen from NH_3 or NO_2
$C^{sat}{}_d$ = Satellite-derived daily surface concentration of NH_3 or NO_2
V_d = Daily dry deposition velocity of NH_3 or NO_2
$Weight_d$ = Weight assigned to the satellite-derived surface concentration of NH_3 or NO_2
d = number of days (i.e., 1,2,3. . ..n)

The reactive nitrogen concentrations and dry deposition fluxes of NH3+NO2 is calculated as follows:

$$\left(C^{sat}{}_{wavg}\right)_{NH3+NO2} = \left(C^{sat}{}_{wavg}\right)_{NH3} + \left(C^{sat}{}_{wavg}\right)_{NO2} \quad (5)$$

$$\left(F^{sat}{}_{wavg}\right)_{NH3+NO2} = \left(F^{sat}{}_{wavg}\right)_{NH3} + \left(F^{sat}{}_{wavg}\right)_{NO2} \quad (6)$$

Where.

$(C^{sat}{}_{wavg})_{NH3}$ = Satellite-derived average concentration of NH_3
$(C^{sat}{}_{wavg})_{NO2}$ = Satellite-derived average concentration of NO_2
$(F^{sat}{}_{wavg})_{NH3}$ = Satellite-derived average dry deposition flux of reactive nitrogen from NH_3
$(F^{sat}{}_{wavg})_{NO2}$ = Satellite-derived average dry deposition flux of reactive nitrogen from NO_2

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 12 July 2023
DOI 10.3389/fevo.2023.1171358

OPEN ACCESS

EDITED BY
Jie Lou,
Zhejiang Lab, China

REVIEWED BY
Jinman Wang,
China University of Geosciences, China
Vishal Mishra,
Indian Institute of Technology Roorkee, India

*CORRESPONDENCE
Lei Han
hanshuanglei@chd.edu.cn

RECEIVED 22 February 2023
ACCEPTED 20 June 2023
PUBLISHED 12 July 2023

CITATION
Chen R, Han L, Zhao Y, Zhao Z, Liu Z, Li R, Xia L and Zhai Y (2023) Extraction and monitoring of vegetation coverage based on uncrewed aerial vehicle visible image in a post gold mining area.
Front. Ecol. Evol. 11:1171358.
doi: 10.3389/fevo.2023.1171358

Extraction and monitoring of vegetation coverage based on uncrewed aerial vehicle visible image in a post gold mining area

Rui Chen[1], Lei Han[2,3]*, Yonghua Zhao[2,3], Zilin Zhao[1], Zhao Liu[2,3], Risheng Li[4], Longfei Xia[4] and Yunmeng Zhai[2,3]

[1]School of Earth Science and Resources, Chang'an University, Xi'an, China, [2]School of Land Engineering, Shaanxi Key Laboratory of Land Consolidation, Chang'an University, Xi'an, China, [3]Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources of the People's Republic of China, Chang'an University, Xi'an, China, [4]Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group, Xi'an, China

Vegetation coverage reflects the degree of environmental degradation. Timely and effective monitoring of vegetation conditions is the basis for promoting vegetation protection and improving the ecological environment of mining areas. Exploring vegetation coverage extraction methods and selecting the optimal vegetation index in mining areas can provide scientific reference for estimating vegetation coverage based on vegetation index in mining areas. Uncrewed aerial vehicles (UAVs) are widely used because of their fast real-time performance, high spatial resolution, and easy accessibility. In this study, the performances of nine visible vegetation indices and two threshold segmentation methods for extracting vegetation coverage in a post-gold mining area in the Qinling Mountains were comprehensively compared using visible spectrum UAV images. Of the nine indices, the excess green index (EXG) and visible-band difference vegetation index (VDVI) were the most effective in discriminating between vegetation and non-vegetation by visual interpretation. In addition, the accuracy of the bimodal histogram threshold method in extracting vegetation coverage was higher than that of Otsu's threshold method. The bimodal histogram threshold method combined with EXG yielded optimal extraction results. Based on optimal methods, the total percentages of fractional vegetation coverage in 2019, 2020, and 2021 were 31.47%, 34.08%, and 42.77%, respectively, indicating that the vegetation in the mining area improved. These results provide valuable guidance for extracting vegetation information and evaluating vegetation restoration in mining areas.

KEYWORDS

remote sensing, uncrewed aerial vehicle, vegetation coverage, eco-monitoring, post-mining area

1 Introduction

Vegetation restoration and reconstruction are key components of ecological restoration in mining areas and are effective ways to improve the quality of the local ecological environment (Li et al., 2019). Therefore, obtaining vegetation information quickly and accurately in mining areas to evaluate the status of local ecological restoration is an urgent issue. As an important indicator of vegetation status, fractional vegetation coverage (FVC) is defined as the percentage ratio of the vertical projection area of vegetation (including leaves, stems, and branches) on the ground in a statistical area (Jia et al., 2015). FVC is not only the main indicator of regional environmental status and quality assessment but also an important part of terrestrial ecosystem research. Thus, accurate and rapid extraction of vegetation coverage requires timely monitoring of vegetation change, which is crucial for protecting biodiversity and promoting economic development.

Currently, remote sensing observations and land surface measurements are primarily used to monitor FVC (Lu et al., 2020). As a low-cost and highly efficient monitoring technology, remote sensing can provide objective and accurate environmental monitoring for large-scale mining areas. With the rapid development of satellite remote sensing technologies, many vegetation products, such as those derived from NOAA/AVHRR (Boyd et al., 2002), TM/Landsat (Voorde et al., 2008; Leng et al., 2019), and Terra & Aqua/MODIS (Song et al., 2017), have facilitated large-scale monitoring of vegetation coverage. However, for small-scale areas, such as mining areas with complex topography and heterogeneous habitats, monitoring FVC using satellite remote sensing technologies is challenging because of their relatively coarse spatial resolution and long revisit period. Furthermore, although *in situ* measurements have high accuracy, they are usually time consuming and labor intensive, rendering them unsuitable for real-time and long-term monitoring. Notably, uncrewed aerial vehicles (UAVs) have the advantages of strong real-time performance, high spatial resolution, and easy access; thus, they have attracted wide attention as a novel and improved method to extract vegetation coverage with high efficiency and precision on small spatial scales in agriculture, forestry, surveying, mapping, and other related fields (Watanabe et al., 2017; Schofield et al., 2019; Ana et al., 2021; Guo et al., 2021; Park et al., 2022; Mishra et al., 2023).

Compared with multispectral, hyperspectral, and other sensors, visible light sensors are better options for extracting vegetation coverage via UAV technology owing to their outstanding advantages, such as low cost and are less affected by weather and light (Coy et al., 2016; Jay et al., 2019; Ren et al., 2021). The vegetation index can effectively reflect vegetation vitality and information and is a commonly used method for extracting vegetation coverage (Woebbecke et al., 1995; Hague et al., 2006; Rasmussen et al., 2016; Kim et al., 2018; Geng et al., 2022). Various vegetation indices have been developed based on the spectral characteristics of green vegetation in the visible light band, such as the green leaf index (Shane et al., 2021), green-red vegetation index (Zhang et al., 2019), and difference-enhanced vegetation index (Zhou et al., 2021). An increasing number of studies have shown that vegetation coverage can be extracted using a vegetation index derived from UAV-visible images. Therefore, the limited wavelength information of UAV visible-light images must be used to construct a universally applicable and highly accurate vegetation index and effectively extract green vegetation information. Another key aspect of vegetation coverage extraction by vegetation indices is the determination of a suitable threshold, which can be based on threshold segmentation methods (Akash et al., 2019). However, few studies have used this method to determine the thresholds in mining areas, and the effectiveness of vegetation indices in mountainous mining areas has not yet been evaluated.

The Qinling Mountains are an important ecological security barrier in China and provide many ecosystem services, such as climate regulation, water yield, carbon sequestration, and biodiversity preservation (Fu et al., 2022). Rich gold mineral resources in the Qinling Mountains provide a good foundation for mining activities; however, long-term mining has resulted in serious vegetation destruction (Li et al., 2022), which has plagued sustainable local development (Huo et al., 2022). Therefore, a rapid and accurate method for acquisition of mine vegetation cover is required. Currently, research on vegetation coverage extraction based on visible vegetation index focuses mostly on cities, forests, grasslands, and farmlands with well-growing plants (Geng et al., 2022). However, an optimal vegetation index for extracting vegetation coverage suitable for Qinling gold mining areas with sparse vegetation and complex terrain has not yet been determined. Furthermore, previous studies focused on extraction methods for the current vegetation situation and lacked long-term monitoring. Therefore, an abandoned gold mining area in the Qinling Mountains was selected as the research area, and high spatial resolution visible spectrum images obtained by a UAV were used as the data source. The objectives of this study were to (1) compare the performances of nine visible light vegetation indices (RGRI, BGRI, EXG, EXGR, NGRDI, NGBDI, RGBVI, VDVI, and VEG) and two threshold segmentation methods (bimodal histogram method and Otsu's threshold method) in the extraction of vegetation coverage information; (2) select the optimal combination of the vegetation index and threshold segmentation method with high extraction accuracy and wide applicability; and (3) analyze the interannual variation of FVC in the study area using results obtained by the optimal combination. This study provides scientific guidance for rapidly and accurately extracting vegetation coverage and offers technical support for evaluating vegetation restoration in mining areas.

2 Materials and methods

2.1 Study area

The study area is located in the southeastern part of Shangluo City, Shaanxi Province, China (Figure 1). It is between 108°34′20″–111°1′25″ E and 33°2′30″–34°24′40″ N. The study area is located in the Qinling Mountains and has a warm, temperate climate. The mean annual temperature is 12.2°C, the mean annual precipitation is 804.8 mm, and the mean annual sunshine duration is 1947.4 h.

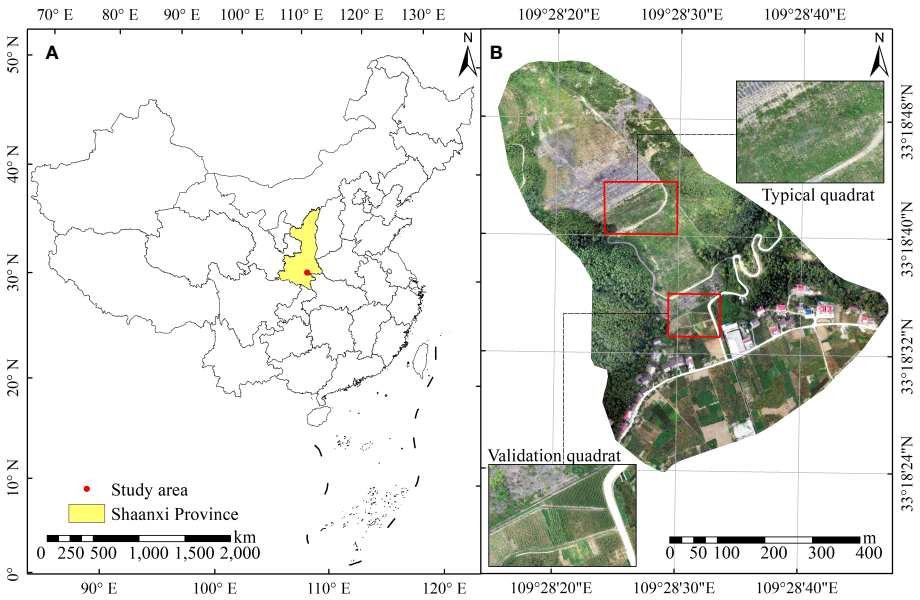


FIGURE 1
(A) Geographical location and **(B)** UAV image with a spatial resolution of 0.0436 m of the study area.

The soil type is yellow cinnamon. It is high in the northwest and low in the southeast. A gold production company in the research area began operations in 1999 and ceased production after a dam failure in 2006. Even after several years, bare slag still poses a serious threat to human health, and this research area has been listed as a key area for heavy metal prevention and control (Chen et al., 2022).

2.2 UAV image acquisition and processing

Field and UAV aerial surveys were conducted in August 2019, 2020, and 2021 to monitor the vegetation coverage at the research site in the post-mining area. The UAV flight test was conducted using a DJI Phantom 4 Pro on clear and cloudless days, and RGB-visible images were acquired. The flight parameters are listed in Table 1. The automatic cruise mode was used for route planning during the flight. The flight area and route were designed prior to conducting the experiment. The flight was 0.68 km^2. Orthoimages of the study area are shown in Figure 1B.

2.3 Calculation of visible light vegetation index

The basic principle behind the construction of a vegetation index is that vegetation absorbs and reflects light of different wavelengths. The corresponding vegetation index can be obtained by combining different bands of remote sensing images to enhance vegetation (Guilherme et al., 2018). The visible vegetation index is mainly constructed based on the red, green, and blue bands of the image because healthy green vegetation has a strong reflection in the green band and weak reflections in the red and blue bands. The nine commonly used visible light vegetation indices are listed in Table 2.

2.4 Vegetation information extraction based on threshold

The vegetation index threshold method is effective for discriminating between vegetation and non-vegetation

TABLE 1 Flight setting of the UAV and image parameters.

Flight setting	Parameter	Acquired image content	Parameter
Flight speed	14.1 m/s	Number of original images	300+
Photo interval	2 s	Picture resolution	72 dpi
Number of routes	13	Graphic form	JPEG
Number of waypoints	26	Shutter speed	1/1600
Course overlap rate	80%	ISO	800
Side overlap rate	60%		
Flight altitude	140 m		

TABLE 2 Nine commonly used visible light vegetation indices considered in this study and the calculation formulas based on the visible spectrum.

Visible vegetation index	Full name	Calculation formula	Reference
RGRI	Red–green ratio index	R/G	(Verrelst et al., 2008)
BGRI	Blue–green ratio index	B/G	(Romina et al., 2010)
EXG	Excess green index	2g−r−b	(Kim et al., 2018)
EXGR	Excess green minus red index	EXG−1.4r−g	(Sun et al., 2014)
NGRDI	Normalized green–red difference index	(G−R)/(G+R)	(Gitelson et al., 2002)
NGBDI	Normalized green–blue difference index	(G−B)/(G+B)	(Hunt et al., 2005)
RGBVI	Red–green–blue vegetation index	$(G^2-B\times R)/(G^2+B\times R)$	(Juliane et al., 2015)
VDVI	Visible-band difference vegetation index	(2G−R−B)/(2G+R+B)	(Wang et al., 2015)
VEG	Vegetative index	$g/r^{0.67}b^{0.33}$	(Geng et al., 2022)

information in an image. Three steps are required to extract the vegetation coverage using the visible light vegetation index. The first step is to calculate the vegetation index, the second is to set an appropriate threshold, and the final is to separate the vegetation and non-vegetation parts. The accuracy of vegetation coverage extraction largely depends on threshold selection (Wang et al., 2015). Two commonly used methods, the bimodal histogram threshold method and Otsu's threshold method, were applied to determine the threshold for each vegetation index.

2.4.1 Bimodal histogram method

A bimodal histogram is an image with two obvious peaks in a gray histogram (Zhou et al., 2021). These two wave peaks correspond to the internal and external target points. The wave trough between the two wave peaks corresponded to the target point near the edge of the object. Typically, the value at the wave trough is selected as the threshold. The calculation process of the bimodal histogram used in this study is as follows. (1) Calculate the average gray value (avg) and standard deviation of the pixels. (2) Considering the average pixel value as the dividing point, find the positions of the maximum values of the left (small peak) and right (large peak) parts. (3) If the two peak positions are close (within the standard deviation range), then one of the two peaks of the histogram is very low; hence, another low peak position must be found; otherwise, proceed to step (7). (4) Determine the position of the pixel gray median point (midpoint). (5) If midpoints>avg, then the small peak is on the left side of the large peak (lower gray level); otherwise, the small peak is on the right side of the large peak (higher gray level), and the position of the dividing point should be adjusted accordingly. (6) Re-find the positions of the large and small peaks. (7) The wave trough of the two peak positions is considered the required threshold (Liang, 2002).

2.4.2 Otsu's threshold method

Otsu's threshold method, also known as the maximum between-cluster variance method, is a global threshold selection method (Otsu, 2007). This method divides an image into background and target images based on a threshold. When the optimal threshold is considered, the variance between the background and target and the difference between the two parts of the image are the largest; that is, the optimal threshold is determined based on the maximum between-cluster variance. The calculation process of Otsu's method is as follows. (1) Identify the highest gray level in the image. (2) Take each gray level as a threshold. (3) Calculate the number of pixels and the average value of the two categories segmented by the threshold. (4) Calculate the variance between the two clusters. (5) Determine the threshold of the maximum variance (Xu et al., 2022).

2.5 Extraction accuracy evaluation

Accuracy, Precision, and Recall were calculated as follows to evaluate the classification accuracy (Shukla and Jain, 2020):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{TP + TN}{P + N}$$

$$\text{Presicion} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

where TP, which stands for "true positive," is the object that is correctly classified as vegetation among all the extracted objects; TN, which stands for "true negative," is the object that is correctly classified as non-vegetation among all the extracted objects; FP, which stands for "false positive," is the object that is misclassified as vegetation among all extracted objects; and FN, which stands for "false negative," is the object that is misclassified as non-vegetation among all the extracted objects.

2.6 Data analysis

The UAV images were converted into orthoimages using DJI Terra v.3.3 software developed by DJI (Shenzhen, China). Supervised classification, calculation of vegetation indices, threshold segmentation, and extraction of vegetation coverage were performed using ENVI 5.3 software.

3 Results

3.1 Calculation results of vegetation index

3.1.1 Visual interpretation and supervision classification

The performance of the visible vegetation indices in extracting vegetation was evaluated by comparison with the results of the maximum likelihood classification method (Figure 2). Fifty regions of interest (ROIs) with non-vegetation and fifty ROIs with vegetation were uniformly selected to verify the classification accuracy. The overall accuracies of the typical and validation quadrats were 99.99% and 99.39%, respectively.

3.1.2 Vegetation index calculation results

The vegetation indices derived from the gray image of a typical quadrat are shown in Figure 3. Most visible light vegetation indices can be used to effectively distinguish vegetation from non-vegetation information; however, the extraction effects are different. Some vegetation indices, such as EXG, EXGR, RGBVI, VEG, and VDVI, can clearly discriminate between vegetation and non-vegetation areas; however, BGRI, RGRI, NGBDI, and NGRDI cannot clearly distinguish between the two and resulted in some misclassifications, indicating poor extraction performance. Furthermore, to analyze the pixel value ranges of vegetation and non-vegetation in the gray image of each band and vegetation index, 75 representative ROIs were randomly selected to count the pixel eigenvalues of each visible band and vegetation index (Table 3). The results indicated that the reflectance in the green band of the vegetation was significantly higher than that of the non-vegetation. In the BGRI, RGRI, and NGRDI gray images, the pixel values of vegetation and non-vegetation overlapped over a large range; therefore, vegetation and non-vegetation areas overlapped. In addition, the calculation formulas for RGRI, NGBDI, and NGRDI only used blue + green or red + green bands, indicating that the red, green, and blue bands should be combined when calculating the visible light vegetation index.

3.1.3 Threshold segmentation and vegetation index selection

The bimodal histogram threshold and Otsu's threshold methods were employed to determine the threshold of each visible light vegetation index grey image. The vegetation and non-vegetation areas were discriminated based on the thresholds, and the extraction accuracy was verified by comparison with the supervised classification results. The threshold segmentation results are shown in Figures 4, 5. By jointly viewing the orthoimages and supervised classification results (Figure 2A1, A2), we found that the extraction results of the bimodal histogram threshold method had fewer misclassifications, and the

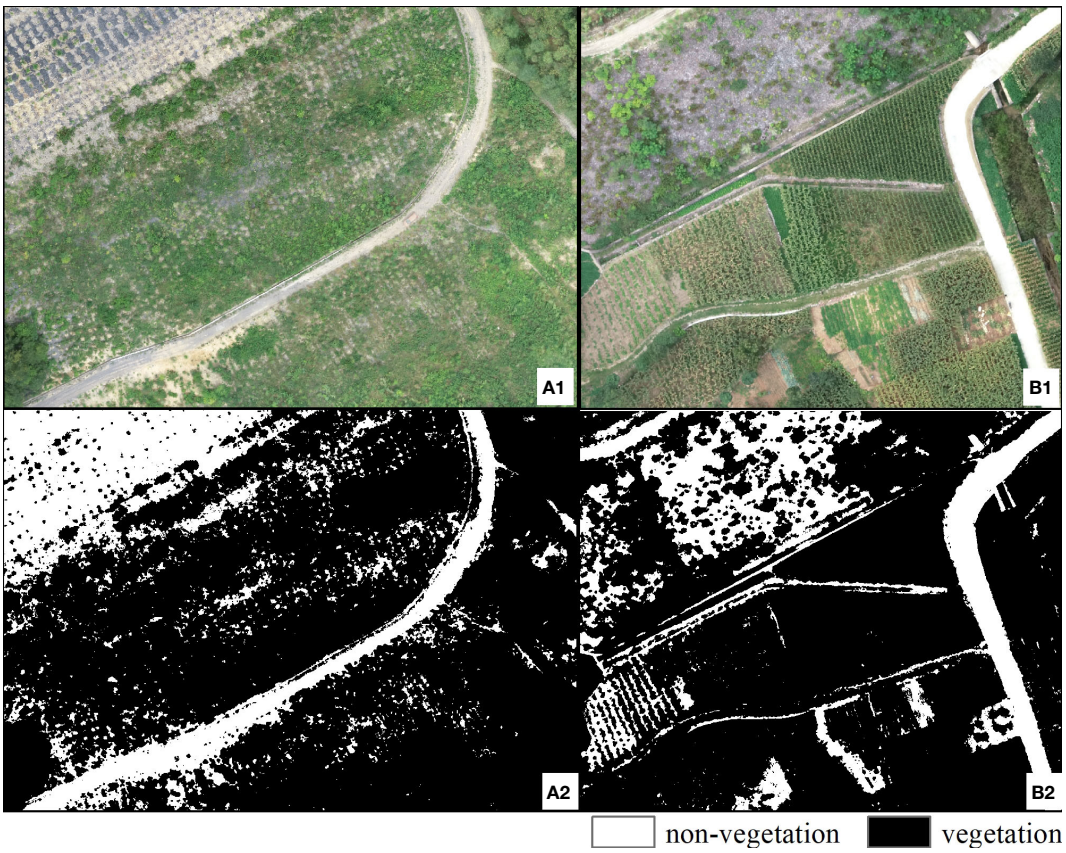


FIGURE 2
Original images of **(A1)** typical and **(B1)** validation quadrats and classification results of the **(A2)** typical and **(B2)** verification quadrats.

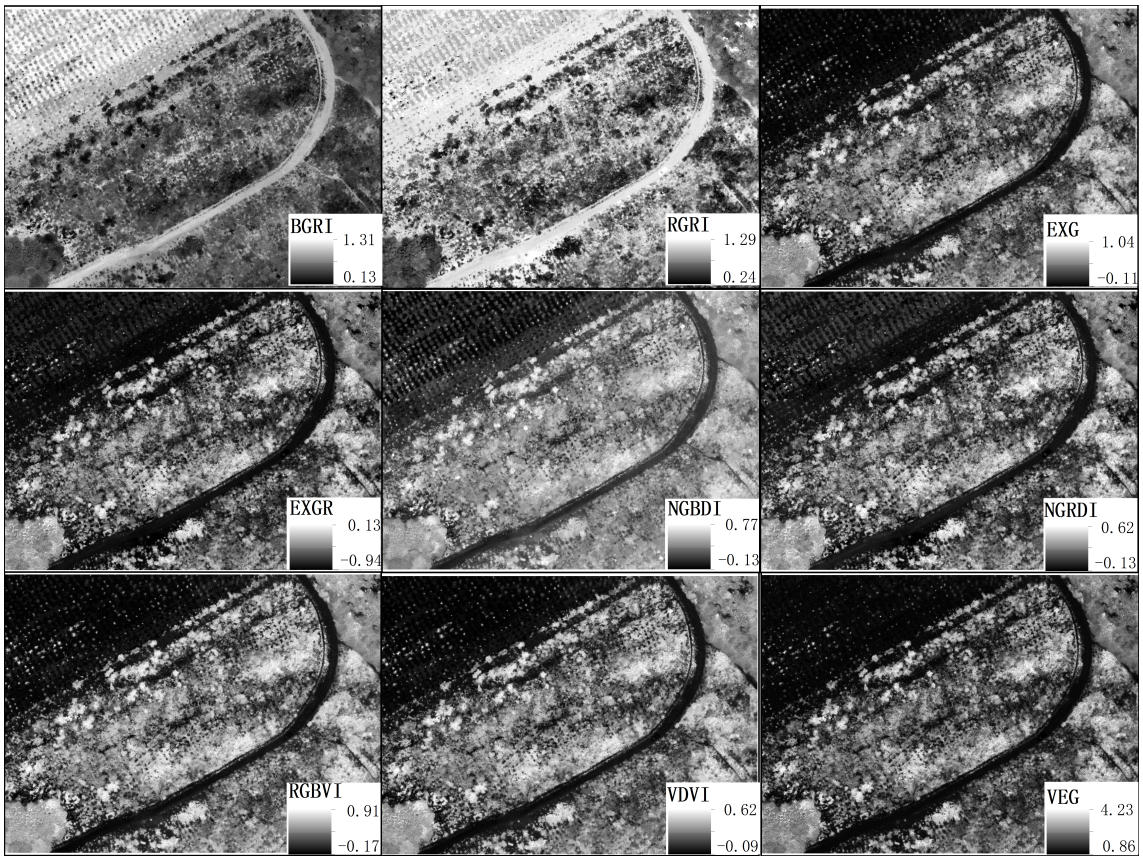


FIGURE 3
Calculation results of nine vegetation indices for the typical quadrat.

segmentation effect was obviously better than that of Otsu's threshold method. In the segmentation results of the bimodal histogram threshold method, RGRI, EXGR, NGRDI, and VEG misclassified vegetation as non-vegetation, whereas BGRI and EXG misclassified non-vegetation as non-vegetation, indicating relatively poor extraction accuracy. The quantitative accuracy must be evaluated to accurately evaluate the effects of the segmentation results. The Accuracy, Precision and Recall of the threshold segmentation results were calculated based on the maximum likelihood classification results (Table 4). Overall, the classification accuracy of the bimodal histogram method was higher than that of Otsu's threshold method. Among the visible light vegetation indices, EXG, based on the bimodal histogram method, had the highest classification accuracy, with the Accuracy was 98.264%, Precision was 99. 811% and 97.572% in vegetation and non-vegetation, and Recall was 99.913% and 94.847% in vegetation and non-vegetation.

3.2 Suitability performance test

The reliability and applicability of EXG, VDVI, and RGBVI for extracting vegetation coverage were verified based on the supervised classification results of the verification quadrat. The vegetation extracted based on EXG, VDVI, RGBVI, and the bimodal histogram threshold method (the thresholds were 0.047603, 0.041258, and 0.075669, respectively) are shown in Figure 6. The results of vegetation coverage extraction were compared with those of the maximum likelihood classification (Table 5). EXG combined with the bimodal histogram method still had the highest accuracy in extracting vegetation coverage, followed by VDVI and RGBVI, suggesting that EXG had the highest precision in extracting vegetation information and could be used to estimate vegetation coverage in mining areas.

3.3 Vegetation coverage assessment

According to the above results, EXG combined with the bimodal histogram threshold method was used to estimate vegetation coverage in 2019, 2020, and 2021 (the thresholds were 0.07848, 0.122353, and 0.125108, respectively). The extraction results were statistically classified as follows: vegetation coverage of 0–0.05 was considered a zero-coverage area, 0.05–0.2 was a low vegetation coverage area, 0.2–0.4 was a low–moderate vegetation coverage area, 0.4–0.6 was a moderate vegetation coverage area, 0.6–0.8 was a moderate–high vegetation coverage area, and 0.8–1 was a high vegetation coverage area (Zhao et al., 2022). Figure 7 shows that EXG can clearly discriminate between vegetation and non-vegetation areas. From 2019 to 2021, the non-vegetation area

TABLE 3 Differences in pixel values of visible bands and vegetation indices of the typical quadrat.

Indicators	Vegetation				Non-vegetation				ANOVA P value
	Min	Max	Mean	Standard deviation	Min	Max	Mean	Standard deviation	
Red band	30.00	194.00	96.57	23.89	68.00	254.00	177.23	43.07	0.054
Green band	66.00	211.00	132.26	20.03	78.00	249.00	176.73	38.20	0.033
Blue band	27.00	163.00	84.02	17.51	86.00	249.00	174.07	29.24	0.048
BGRI	0.32	0.85	0.63	0.08	0.85	1.20	1.00	0.07	0.007
RGRI	0.43	0.96	0.73	0.10	0.86	1.13	1.00	0.04	0.031
EXG	0.07	0.70	0.28	0.10	−0.03	0.05	0.00	0.01	0.001
EXGR	−0.76	−0.22	−0.57	0.09	−0.86	−0.73	−0.80	0.02	0.012
NGRDI	0.02	0.40	0.16	0.07	−0.06	0.08	0.00	0.02	0.151
NGBDI	0.08	0.51	0.23	0.06	−0.09	0.08	0.00	0.04	0.023
RGBVI	0.11	0.75	0.37	0.11	−0.05	0.07	0.00	0.02	0.017
VDVI	0.05	0.45	0.19	0.06	−0.03	0.03	0.00	0.01	0.001
VEG	1.10	2.51	1.47	0.20	0.95	1.06	1.01	0.01	0.002

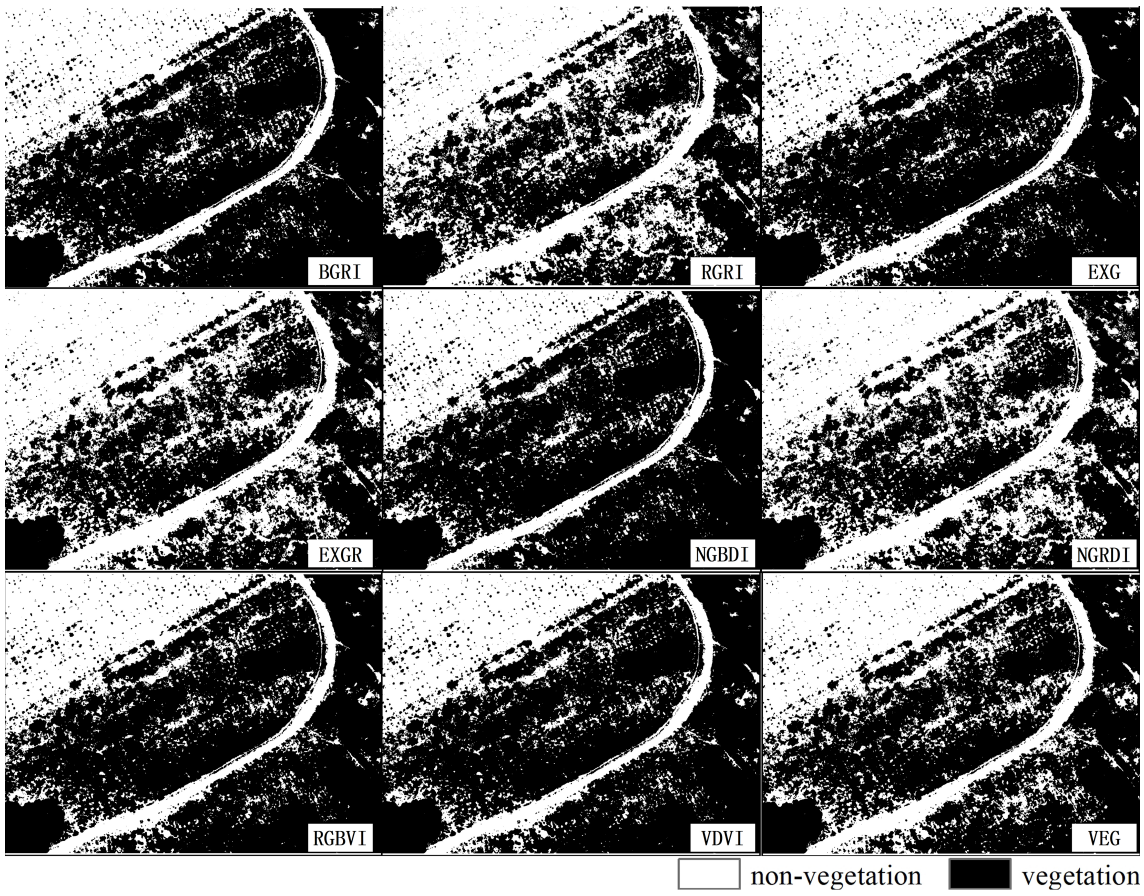


FIGURE 4
Segmentation results of the bimodal histogram threshold method.

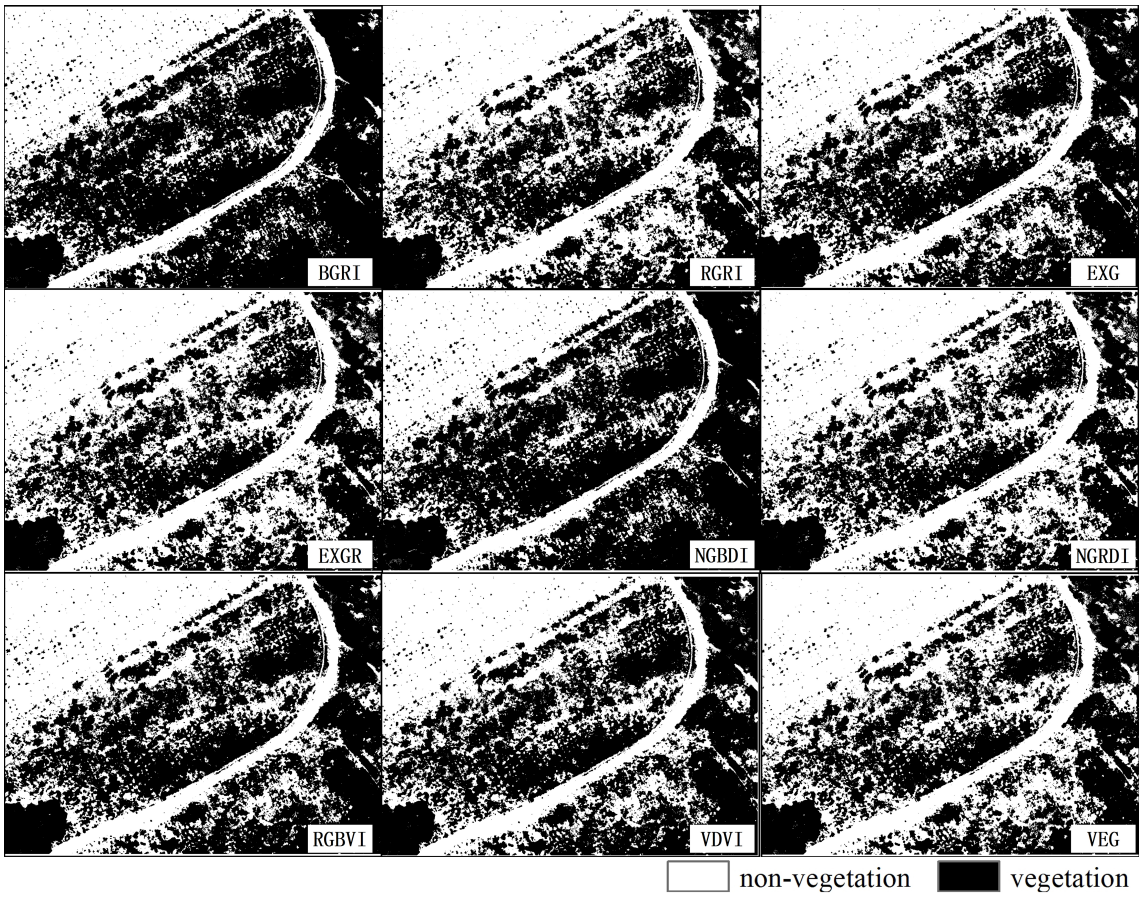


FIGURE 5
Segmentation results of Otsu's threshold method.

TABLE 4 Accuracy evaluation of the typical quadrat.

Segmentation		Accuracy (%)	Precision (%)		Recall (%)	
			Vegetation	Non-vegetation	Vegetation	Non-vegetation
BGRI	Otsu's	92.741	99.330	89.790	99.667	81.332
	Bimodal histogram	96.275	96.233	96.294	98.278	0.920
EXG	Otsu's	81.656	100.000	73.441	99.999	62.772
	Bimodal histogram	98.264	99.811	97.572	99.913	94.847
EXGR	Otsu's	75.787	99.998	64.956	99.998	56.086
	Bimodal histogram	82.903	99.999	75.315	99.998	66.471
MGRVI	Otsu's	75.024	99.996	64.967	99.997	56.106
	Bimodal histogram	82.524	99.811	74.389	99.886	63.573
NGBDI	Otsu's	91.216	99.627	87.449	99.809	78.045
	Bimodal histogram	96.421	95.192	96.979	97.828	93.384
NGRDI	Otsu's	75.184	99.997	64.006	99.998	55.439
	Bimodal histogram	80.634	99.924	71.996	99.953	61.507

(Continued)

TABLE 4 Continued

Segmentation		Accuracy (%)	Precision (%)		Recall (%)	
			Vegetation	Non-vegetation	Vegetation	Non-vegetation
RGBVI	Otsu's	84.722	99.999	77.880	99.999	66.936
	Bimodal histogram	97.824	99.795	96.941	99.905	93.595
RGRI	Otsu's	76.536	99.993	66.031	99.995	56.864
	Bimodal histogram	81.725	99.866	73.602	99.918	62.882
VDVI	Otsu's	83.005	99.999	75.395	99.999	64.539
	Bimodal histogram	97.993	99.895	97.141	99.952	93.994
VEG	Otsu's	77.195	99.997	66.982	99.996	57.560
	Bimodal histogram	91.142	99.999	87.175	99.999	77.737

FIGURE 6
Verification results of the verification quadrat.

TABLE 5 Accuracy evaluation of the verification quadrat.

Segmentation	Accuracy (%)	Precision (%)		Recall (%)	
		Vegetation	Non-vegetation	Vegetation	Vegetation
EXG	95.073	99.967	91.114	99.989	77.046
RGBVI	91.421	99.999	88.861	99.999	72.816
VDVI	93.107	99.928	91.072	99.976	76.957

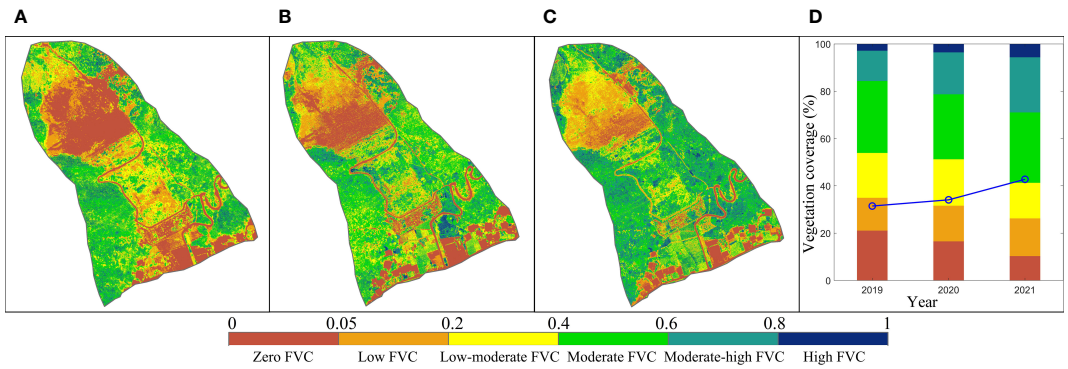


FIGURE 7
Estimated results of vegetation coverage based on EXG combined with the bimodal histogram threshold method in **(A)** 2019, **(B)** 2020, and **(C)** 2021. **(D)** Inter-annual variation of vegetation coverage from 2019 to 2021.

decreased, and the proportion of moderate–high and high vegetation coverage areas increased significantly. The average FVC values in 2019, 2020, and 2021 were 31.47%, 34.08%, and 42.77%, respectively, indicating that the FVC in the mining area increased. The results suggest that the effect of vegetation restoration was remarkable, and the quality of the ecological environment improved. However, most areas in the post-mining area had low, low–moderate, and moderate vegetation coverage, and vegetation restoration requires further strengthening.

4 Discussion

4.1 Extraction accuracy of vegetation coverage

The accuracy of vegetation coverage extraction was related to both the vegetation index and threshold segmentation method. In this study, EXG exhibited the highest extraction accuracy, followed by VDVI and RGBVI, which is consistent with the results of Wang et al. (2015) and Chen and Deng (2019). The calculation formulas for EXG, VDVI, and RGNVI show the reflectance characteristics of vegetation in the visible bands, which effectively increase the sensitivity of vegetation to green bands and make full use of the information in the red, green, and blue bands. Currently, the bimodal histogram threshold and Otsu's threshold methods are widely used for threshold segmentation. In this study, the results of the threshold methods for vegetation coverage extraction suggested that the accuracy of the bimodal histogram method was significantly better than that of Otsu's threshold method (Figure 5), reaffirming the results of Zhao et al. (2019). Using the bimodal histogram method, the accuracies of RGRI, EXGR, and NGRDI were relatively low, which may be related to the histogram characteristics. As shown in the histogram of each vegetation index (Figure 8), EXG, VDVI, and RGBVI showed similar changes and obvious bimodal characteristics, whereas the histograms of RGRI, EXGR, and NGRDI had no obvious bimodal characteristics. Therefore, the accuracy of vegetation coverage extraction varied greatly.

4.2 Characteristics of UAV visible vegetation indices

Satellite remote sensing images have advantages, such as large image areas and multiple bands (Xu et al., 2020; Guo and Guo, 2021). However, owing to the relatively coarse spatial resolution, the interpretation accuracy is relatively limited, and the temporal resolution often cannot meet the real-time requirements of vegetation monitoring on a small spatial scale, such as in mining areas. With the popularization of UAV technology, UAV images have compensated for the deficiencies in satellite remote sensing images in terms of spatial and temporal resolution. UAVs provide a new data source for the acquisition of vegetation coverage information in mining areas and offer new approaches for monitoring vegetation growth and recovery in mining areas (Sun et al., 2021). The results of this study indicate that vegetation coverage data can be accurately extracted from UAV images. As an unsupervised classification method, the visible light vegetation index can be used to extract vegetation coverage quickly and accurately without manual visual discrimination of vegetation areas or non-vegetation areas. Two major advantages are commonly associated with using visible spectrum images for extracting vegetation coverage. One is that RGB images are low

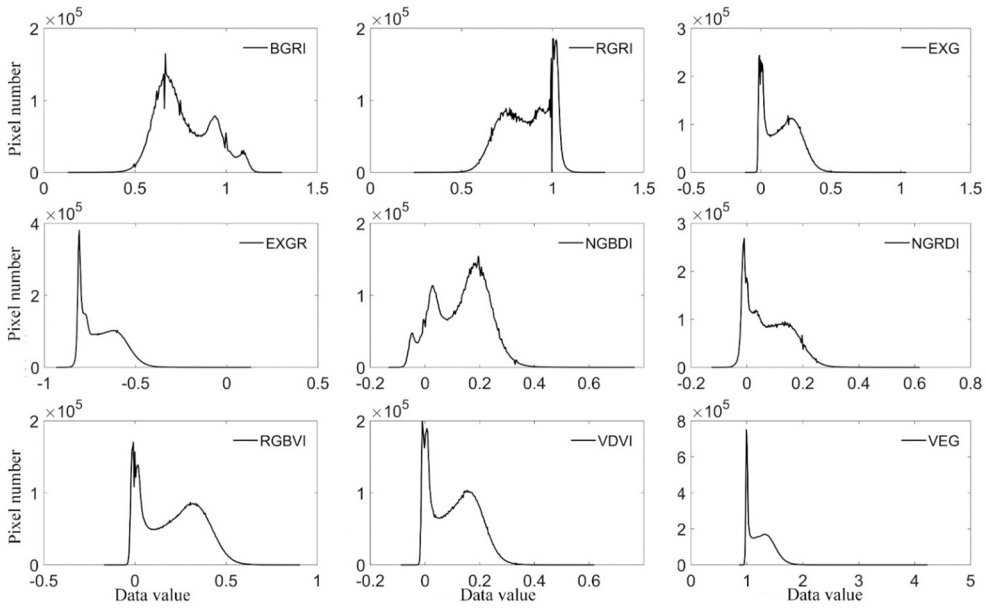


FIGURE 8
Statistical histogram of the nine vegetation indices considered in this study.

cost, convenient to process, and less affected by weather and light. The other is that RGB images have a relatively high spatial and temporal resolution, which is more suitable for local studies. For example, Marcial-Pablo et al. (2019) indicated that the accuracy of visible vegetation indices is higher than that of visible NIR vegetation indices for early crop cover. Furukawa et al. (2021) reported that RGB images provide reliable information for vegetation monitoring. For the mining areas, the land-use type was relatively single, and vegetation coverage could be quickly obtained via UAV images. Moreover, the UAV-visible images were acquired in summer, when vegetation growth was the best. For most vegetation, summer is the most vigorous period for plant growth, during which the vegetation exhibits the strongest reflected spectral features. Thus, vegetation coverage can be accurately estimated using the vegetation index.

4.3 Variation characteristics of vegetation coverage

According to previous investigation and research results, the soil arsenic contamination in the gold mining area is serious (the average soil arsenic content was 93.96 mg/kg) (Chen et al., 2022). Vegetation types are scarce (mostly herbaceous plants), and vegetation coverage is low. The results of the vegetation coverage change from 2019 to 2021 indicated that most natural vegetation restoration sites had low to low–moderate vegetation coverage. This was mainly because soil As contamination limited the normal growth and development of plants in the early stages of vegetation restoration (Yang et al., 2020), and community succession was relatively slow. Increased vegetation coverage improves the quality of regional ecological environments. With the progress of ecological restoration, the soil arsenic content has decreased, and plants have developed their own unique physiological and ecological characteristics after a period of adaptation. The number of pixels with zero vegetation and low and low-moderate vegetation coverage decreased, those with moderate–high and high vegetation coverage increased, and the overall vegetation coverage increased.

5 Conclusions

In this study, a disused gold mining area in the Qinling Mountains was selected as the research area, and UAVs were deployed to obtain image data with high spatial resolution in the visible light. The performance of different visible light vegetation indices combined with two threshold segmentation methods for extracting vegetation coverage was evaluated. The main conclusions are as follows. (1) Except for RGRI, NGRDI, and NGBDI, the other visible light vegetation indices effectively discriminated between vegetation and non-vegetation in the study area. (2) EXG, VDVI, and RGBVI combined with the bimodal histogram threshold method had higher extraction accuracy in distinguishing between vegetation and non-vegetation areas. (3) EXG and the bimodal histogram threshold method had the highest accuracy for vegetation identification, which was the closest to the results of the monitored and actual situations. (4) The spatiotemporal analysis of vegetation coverage in 2019, 2020, and 2021 showed that most mining areas had low, low–moderate, and moderate vegetation coverage, whereas the overall vegetation coverage was low. The average FVC for the three years were 31.47%, 34.08%, and 42.77%, respectively, indicating an increasing trend. Future studies should continue monitoring vegetation coverage changes to provide technical support for land reclamation and ecological restoration in mining areas.

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

RC: conceptualization, methodology, analysis, and writing – original manuscript. LH: conceptualization, reviewing and editing, and funding acquisition. YHZ: reviewing and editing, supervision. ZZ: methodology, statistical analysis, and reviewing and editing. ZL: reviewing and editing, supervision. RL: methodology, reviewing and editing. LX: reviewing and editing, providing revisions and comments, and supervision. YMZ: methodology, providing revisions and comments, and reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources of the People's Republic of China (Program No. SXDJ2017-9), and the Shaanxi Key Laboratory of Land Reclamation Engineering: (Program No. 2018-ZZ03).

Acknowledgments

We are grateful to the reviewers whose comments have helped to clarify and improve the manuscript.

Conflict of interest

RL and LX were employed by the Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group, Xi'an, China.

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

Publisher's note

References

Akash, A., Jinha, J., Anjin, C., Sungchan, O., Murilo, M., and Juan, L. (2019). A comparative study of RGB and multispectral sensor-based cotton canopy cover modelling using multi-temporal UAS data. *Remote Sens.* 11 (23), 2757. doi: 10.3390/rs11232757

Ana, D. C., Shi, Y. Y., Maja, J. M., and Peña, J. M. (2021). UAVs for vegetation monitoring: overview and recent scientific contributions. *Remote Sens-Basel* 13, 2139. doi: 10.3390/rs13112139

Boyd, D. S., Foody, G. M., and Ripple, W. J. (2002). Evaluation of approaches for forest cover estimation in the pacific Northwest, USA, using remote sensing. *Appl. Geogr.* 22 (4), 375–393. doi: 10.1016/S0143-6228(02)00048-6

Chen, X. D., and Deng, J. H. (2019). Study on extraction method of vegetation coverage of summer maize based on visible image. *Exp. Technol. Manage.* 36 (12), 131–136. doi: 10.6041/j.issn.1000-1298.2019.05.027

Chen, R., Han, L., Liu, Z., Zhao, Y. H., Li, R. S., Xia, L. F., et al. (2022). Assessment of soil-heavy metal pollution and the health risks in a mining area from southern shaanxi province, China. *Toxics* 10, 385. doi: 10.3390/toxics10070385

Coy, A., Rankine, D., Taylor, M., and Nielsen, D. C. (2016). Increasing the accuracy and automation of fractional vegetation cover estimation from digital photographs. *Remote Sens.* 8 (7), 474. doi: 10.3390/rs8070474

Fu, Y., Song, Y., Yang, C., Liu, X., Liu, Y., and Huang, Y. (2022). Relationship between brain size and digestive tract length support the expensive-tissue hypothesis in feirana quadranus. *Front. Ecol. Evol.* 10. doi: 10.3389/fevo.2022.982590

Furukawa, F., Laneng, L. A., Ando, H., Yoshimura, N., Kaneko, M., and Morimoto, J. (2021). Comparison of RGB and multispectral unmanned aerial vehicle for monitoring vegetation coverage changes on a landslide area. *Drones* 5, 97. doi: 10.3390/drones5030097

Geng, X., Wang, X. M., Fang, H. L., Ye, J. S., Han, L. K., Gong, Y., et al. (2022). Vegetation coverage of desert ecosystems in the qinghai-Tibet plateau is underestimated. *Ecol. Indic* 137, 108780. doi: 10.1016/j.ecolind.2022.108780

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80 (1), 76–87. doi: 10.1016/S0034-4257(01)00289-9

Guilherme, M. S., Daniel, G. D., Oriel, T. K., Ana, C. S. L., Sérgio, G., Fábio, M. O., et al. (2018). The potential for RGB images obtained using unmanned aerial vehicle to assess and predict yield in sugarcane fields. *Int. J. Remote Sens.* 39 (15-16), 5402–5414. doi: 10.1080/01431161.2018.1448484

Guo, X., and Guo, Q. (2021). And feng z detecting the vegetation change related to the creep of 2018 baige landslide in jinsha river, SE Tibet using SPOT data. *Front. Earth Sci.* 9. doi: 10.3389/feart.2021.706998

Guo, Z. C., Wang, T., Liu, S. L., Kang, W. P., Chen, X., Feng, K., et al. (2021). Biomass and vegetation coverage survey in the mu us sandy land - based on unmanned aerial vehicle RGB images. *Int. J. Appl. Earth Obs.* 94, 102239. doi: 10.1016/j.jag.2020.102239

Hague, T., Tillett, N. D., and Wheeler, H. (2006). Automated crop and weed monitoring in widely spaced cereals. *Precis. Agric.* 7 (1), 21–32. doi: 10.1007/s11119-005-6787-1

Hunt, E. R., Cavigelli, M., Daughtry, C., Mcmurtrey, J., and Walthall, C. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6 (4), 359–378. doi: 10.1007/s11119-005-2324-5

Huo, A. D., Wang, X., Zhao, Z. X., Yang, L. Y., Zhong, F. Q., Zheng, C. L., et al. (2022). Risk assessment of heavy metal pollution in farmland soils at the northern foot of the qinling mountains, China. *Int. J. Environ. Res. Public Health* 19 (22), 14962. doi: 10.3390/ijerph192214962

Jay, S., Baret, F., Dutartre, D., Malatesta, G., Héno, S., Comar, A., et al. (2019). Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sens. Environ.* 231, 110898. doi: 10.1016/j.rse.2018.09.011

Jia, K., Liang, S. L., Liu, S. H., Li, Y. W., Xiao, Z. Q., Yao, Y. J., et al. (2015). Global land surface fractional vegetation cover estimation using general regression neural networks from MODIS surface reflectance. *IEEE T Geosci Remote* 53 (9), 4787–4796. doi: 10.1109/TGRS.2015.2409563

Juliane, Be, Kang, Y., Helge, A., Andreas, B., Simon, B., Janis, B., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012

Kim, D. W., Yun, H. S., Jeong, S. J., Kwon, Y. S., Kim, S. G., Lee, W. S., et al. (2018). Modeling and testing of growth status for Chinese cabbage and white radish with UAV-based RGB imagery. *Remote Sens.* 10 (4), 563. doi: 10.3390/rs10040563

Leng, R. L., Zhang, Y. Y., Xie, J. Q., Li, F. N., Xu, G., and Cui, X. (2019). An analysis of fractional vegetation cover of the gannan grassland in the non-growing season based on multispectral data and small UVAs. *Pratacultural Sci.* 36 (11), 2742–2751. doi: 10.11829/j.issn.1001-0629.2019-0013

Li, X. L., Gao, J., Zhang, J., Wang, R., Jin, L. Q., and Zhou, H. K. (2019). Adaptive strategies to overcome challenges in vegetation restoration to coalmine wasteland in a frigid alpine setting. *Catena* 182, 104142. doi: 10.1016/j.catena.2019.104142

Li, T., Wu, M. H., Duan, C. G., Li, S. Y., and Liu, C. E. (2022). The effect of different restoration approaches on vegetation development in metal mines. *Sci. Total Environ.* 806 (2), 150626. doi: 10.1016/J.Scitotenv.2021.150626

Liang, H. W. (2002). Direct determination of threshold from bimodal histogram. *Pattern Recognition Artificial Intell.* 15 (2), 253–256.

Lu, M., Liao, X. H., Yue, H. Y., Huang, Y. H., Ye, H. P., Xu, C. C., et al. (2020). Optimizing distribution of drone ports for emergency monitoring of flood disasters in China. *J. Flood Risk Manag.* 13 (1), e12593. doi: 10.1111/jfr3.12593

Marcial-Pablo, M. D. J., Gonzalez-Sanchez, A., Jimenez-Jimenez, S. I., Ontiveros-Capurata, R. E., and Ojeda-Bustamante, W. (2019). Estimation of vegetation fraction using RGB and multispectral images from UAV. *Int. J. Remote Sens.* 40 (2), 420–438. doi: 10.1080/01431161.2018.1528017

Mishra, V., Avtar, R., Prathiba, A. P., Mishra, P. K., Tiwari, A., and Sharma, S. K. (2023). Uncrewed aerial systems in water resource management and monitoring: a review of sensors, applications, software, and issues. *Adv. Civ. Eng.* 28, 3544724. doi: 10.1155/2023/3544724

Otsu, N. (2007). A threshold selection method from gray-level histograms. *IEEE T Syst. Man Cy-S* 9 (1), 62–66. doi: 10.1109/TSMC.1979.4310076

Park, G., Park, K., Song, B., and Lee, H. (2022). Analyzing impact of types of UAV-derived images on the object-based classification of land cover in an urban area. *Drones* 6, 71. doi: 10.3390/drones6030071

Rasmussen, J., Ntakos, G., Nielsen, J., Svensgaard, J., Poulsen, R. N., and Christensen, S. (2016). Are vegetation indices derived from consumer-grade cameras mounted on UAVs sufficiently reliable for assessing experimental plots? *Eur. J. Agron.* 74, 75–92. doi: 10.1016/j.eja.2015.11.026

Ren, L., Liu, Y., Zhang, S., Cheng, L., Guo, Y., and Ding, A. (2021). Vegetation properties in human-impacted riparian zones based on unmanned aerial vehicle (UAV) imagery: an analysis of river reaches in the yongding river basin. *Forests* 12, 22. doi: 10.3390/f12010022

Romina, S., María, C., Santiago, A. T., Elizabeth, K., Ana, S. B., Constanza, R., et al. (2010). Cryptochrome as a sensor of the blue/green ratio of natural radiation in arabidopsis. *Plant Physiol.* 154 (1), 401–409. doi: 10.1104/pp.110.160820

Schofield, G., Esteban, N., Katselidis, K. A., and Graeme, C. H. (2019). Drones for research on sea turtles and other marine vertebrates – a review. *Biol. Conserv.* 238, 108214. doi: 10.1016/j.biocon.2019.108214

Shane, G., Kevin, L., John, J., and Kenny, S. (2021). Use of an unmanned aerial vehicle (UAV) to document vegetation coverage rate in managed grasslands following a historic river flood. *J. Anim. Sci.* 99 (2), 9–10. doi: 10.1093/jas/skab096.015

Shukla, A., and Jain, K. (2020). Automatic extraction of urban land information from unmanned aerial vehicle (UAV) data. *Earth Sci. Inform* 13, 1225–1236. doi: 10.1007/s12145-020-00498-x

Song, W. J., Mu, X. H., Ruan, G. Y., Gao, Z., Li, L. Y., and Yan, G. Y. (2017). Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method. *Int. J. Appl. Earth Obs.* 58, 168–176. doi: 10.1016/j.jag.2017.01.015

Sun, Z. Y., Wang, X. N., Wang, Z. H., Yang, L., Xie, Y. C., and Huang, Y. H. (2021). UAVs as remote sensing platforms in plant ecology: review of applications and challenges. *J. Plant Ecol.* 14 (6), 1003–1023. doi: 10.1093/jpe/rtab089

Sun, G. X., Wang, X. C., Yan, T. T., Li, X., Chen, M., Shi, Y. Y., et al. (2014). Inversion method of flora growth parameters based on machine vision. *Trans. Chin. Soc. Agric. Eng.* 30 (20), 187–195.

Verrelst, J., Schaepman, M. E., Koetz, B., and Kneubühlerb, M. (2008). Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sens. Environ.* 112 (5), 2341–2353. doi: 10.1016/j.rse.2007.11.001

Voorde, T. V., Vlaeminck, J., and Canters, F. (2008). Comparing different approaches for mapping urban vegetation cover from landsat ETM+ data: a case study on Brussels. *Sensors* 8, 3880–3902. doi: 10.3390/s8063880

Wang, X. Q., Wang, M. M., Wang, S. Q., and Wu, D. P. (2015). Extraction of vegetation information from visible unmanned aerial vehicle images. *Trans. Chin. Soc. Agric. Eng.* 31 (05), 152–159.

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., et al. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00421

Woebbecke, D., Meyer, G., Bargen, K., and Mortensen, D. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *T ASABE* 38, 259–269. doi: 10.13031/2013.27838

Xu, X., Liu, L., Han, P., Gong, X., and Zhang, Q. (2022). Accuracy of vegetation indices in assessing different grades of grassland desertification from UAV. *Int. J. Env. Res. Pub. He.* 19 (24), 16793. doi: 10.3390/ijerph192416793

Xu, K. X., Su, Y. J., Liu, J., Hu, T. Y., Jin, S. C., Ma, Q., et al. (2020). Estimation of degraded grassland aboveground biomass using machine learning methods from terrestrial laser scanning data. *Ecol. Indic.* 108 (C), 105747. doi: 10.1016/j.ecolind.2019.105747

Yang, G. Y., Zhong, H., Liu, X., Liu, C. E., Li, S. Y., Hou, L., et al. (2020). Arsenic distribution, accumulation and tolerance mechanisms of typha angustifolia in different phenological growth stages. *B Environ. Contam. Tox.* 104 (3), 358–365. doi: 10.1007/s00128-020-02796-y

Zhang, X. L., Zhang, F., Qi, Y. X., Deng, L. F., Wang, X. L., and Yang, S. T. (2019). New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV). *Int. J. Appl. Earth Obs.* 78, 215–226. doi: 10.1016/j.jag.2019.01.001

Zhao, D., Wang, Z. W., Zhang, G. Z., Xu, Y. M., and Sun, L. J. (2022). Exploration of influence factors on regional fractional vegetation cover based on a combination of factor regression and interaction–take the three-river headwaters region as an example. *China Environ. Sci.* 42 (8), 3903–3912. doi: 10.19674/j.cnki.issn1000-6923.20220329d011

Zhao, J., Yang, H. B., Lan, Y. B., Lu, L. Q., Jia, P., and Li, Z. M. (2019). Extraction method of summer corn vegetation coverage based on visible light image of unmanned aerial vehicle. *Trans. Chin. Soc. Agric. Machinery* 050 (005), 232–240.

Zhou, T., Hu, Z. Q., Han, J. Z., and Zhang, H. (2021). Green vegetation extraction based on visible light image of UAV. *China Environ. Sci.* 41 (05), 2380–2390.

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 13 July 2023
DOI 10.3389/feart.2023.1218823

OPEN ACCESS

EDITED BY
Jie Luo,
Zhejiang Lab, China

REVIEWED BY
Jiaye Li,
Dongguan University of Technology, China
Hasan Bilgehan Makineci,
Konya Technical University, Türkiye
Ahmed Sefelnasr,
United Arab Emirates University, United Arab Emirates

*CORRESPONDENCE
Vladimir Tabunshchik,
tabunshchyk@ibss-ras.ru

RECEIVED 08 May 2023
ACCEPTED 30 June 2023
PUBLISHED 13 July 2023

CITATION
Tabunshchik V, Gorbunov R, Gorbunova T, Pham CN and Klyuchkina A (2023), Identification of river basins within northwestern slope of Crimean Mountains using various digital elevation models (ASTER GDEM, ALOS World 3D, Copernicus DEM, and SRTM DEM). *Front. Earth Sci.* 11:1218823. doi: 10.3389/feart.2023.1218823

Identification of river basins within northwestern slope of Crimean Mountains using various digital elevation models (ASTER GDEM, ALOS World 3D, Copernicus DEM, and SRTM DEM)

Vladimir Tabunshchik[1]*, Roman Gorbunov[1], Tatiana Gorbunova[1,2], Cam Nhung Pham[1] and Aleksandra Klyuchkina[1]

[1]Geomatics Research Center, A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Sevastopol, Russia, [2]Department of Subtropical and Tropical Ecology, Institute of Environmental Engineering, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

Since the end of the 20th century, the use of geographic information systems and digital elevation models has reduced the time required for and improved the quality of morphometric analysis of the relief within river basins. However, researchers are constantly faced with the problem of choosing the most accurate and suitable digital terrain model for their task. Many global, regional, and local digital elevation models are available. In this study, we comparatively analyzed the accuracy of the ASTER GDEM, ALOS World 3D, Copernicus DEM, and SRTM DEM spatial datasets for the purpose of catchment basin modeling for the river basins of the northwestern slope of the Crimean Mountains (Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers) as an example. For each river basin, we calculated the systematic, root mean square, mean absolute, standard root mean square (Bessel's correction), and centered mean absolute errors by comparing ASTER GDEM, ALOS World 3D, Copernicus DEM, and SRTM DEM data with a 1:100,000 topographic map within the considered river basins. We found the smallest error values for the Copernicus DEM and ALOS World 3D datasets; furthermore, we used the Copernicus DEM dataset to model the river basins and sub-basins of the northwestern slope of the Crimean Mountains. As a result, we identified these river basins and sub-basins for the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers, which are represented by stream basins, valleys, gullies, and ravine systems.

KEYWORDS

GIS, digital elevation model, ASTER GDEM, ALOS World 3D, Copernicus DEM, SRTM, river, river basin

1 Introduction

Digital elevation model (DEM) is a generic term for digital topographic and/or bathymetric data in all their forms (Manune, 2007). A large number of studies have been devoted to assessing the accuracy of DEM, considering both the practical and theoretical aspects of this issue (del Rosario Gonzalez-Moradas and Viveen, 2020; Mesa-

Mingorance and Ariza-López, 2020; Uuemaa et al., 2020; Yamazaki et al., 2017). The number of these scientific studies has been constantly increasing due to the improvement in existing geodatasets and the market entry of new sets of geodata, which is of considerable interest to researchers. In recent years, DEMs created using unmanned aerial vehicles (UAVs) (Uysal et al., 2015; Hashemi-Beni et al., 2018; Escobar Villanueva et al., 2019; Annis et al., 2020) have been extensively used. UAVs have high accuracy but, in almost all cases, are inaccessible to a wide range of researchers. DEMs are actively used in the study of glaciers (Fischer et al., 2015; Bodin et al., 2018), forests (Balzter et al., 2015; Liu et al., 2018), and celestial bodies (Florinsky and Filippov, 2017; Fawdon et al., 2018), among others. One of the largest niche areas is occupied by studies of the morphometry of river basins (Pyankov and Shikhov, 2017; Fang et al., 2019; Sarkar et al., 2020; Yermolaev et al., 2021; Zhao et al., 2021), in which DEMs are actively used to identify and characterize river basins.

ASTER and SRTM DEMs have been thoroughly compared. For example, Rajasekhar et al. (2018) studied lineament extraction from ASTER DEM, SRTM, and Cartosat for the Jilledubanderu River basin, Anantapur district, India. Thomas and Prasannakumar, 2015 studied basin morphometry derived from topographic maps, ASTER, and SRTM DEMs, considering an example from Kerala, India. Nikolakopoulos et al. (2006) compared ASTER and SRTM DEMs in Greece using two regions of Crete Island. Zhao et al. (2021) compared the performance among typical open global DEM datasets for the Fenhe River Basin in China.

Due to the emergence of a large number of new open datasets, researchers have been increasingly using Copernicus DEM and comparing different datasets (Karlson et al., 2021; Garrote, 2022; Yuan et al., 2022). Mutar et al. (2021) considered the river basins flowing into the Mosul reservoir (Iraq), finding that the Copernicus DEM model was more accurate than SRTM DEM and ASTER GDEM. However, other authors (Kramm and Hoffmeister, 2021) reported that Copernicus DEM data can produce ambiguous results. Many studies (Karionov, 2010; Yeritsian, 2013; Trofimov and Filippova, 2014) have described the accuracy of the SRTM dataset. Nevertheless, other authors (Karionov, 2010; Yeritsian, 2013) have emphasized that the accuracy of the cartographic material obtained using the SRTM datasets is equivalent or close to that of topographic map. In recent years, ALOS World 3D DEM (Tadono et al., 2016) was also introduced and its accuracy, as well as advantages and disadvantages compared with other DEMs (Courty et al., 2019; Viel et al., 2020), have been analyzed. For the Crimean Peninsula, SRTM are mainly used by researchers, but comparisons with other DEMs are not given and measurement errors are not evaluated.

Several global datasets contain information on river basins on a global scale (Tang and Lettenmaier, 2012; Lehner and Grill G., 2013; Dallaire et al., 2019). For example, the HydroBASINS Version 1.0 dataset contains information on river basins and sub-basins worldwide. This dataset is not suitable for the Crimean Peninsula due to a large number of errors, in particular, incorrect allocation of the catchment basins' boundaries (for example, merged boundaries of the South Coast of the Crimean Peninsula river basins, unreasonable basin division of the largest rivers of the Crimean Peninsula into logically unreasonable parts, etc.). Most regional models in Europe do not include the Crimean Peninsula in the research area, which complicates further analysis and comparison of catchment basins (Vanham and Bidoglio, 2014). Additionally, the identification of small river basins or sub-basins for most large rivers of the Crimean Peninsula is limited only to the main tributaries, whereas the catchment basins of tributaries and their tributaries are practically not considered, with the exception of the most studied and largest river of the Crimean Peninsula, the Salgir River, as well as a small number of Crimean rivers.

Three groups can be distinguished among the studies on the catchment basins of the Crimean Peninsula. The first small group consists of studies (Dunaieva and Kovalenko, 2013; Narozhnyaya, 2021) that considered the river basins of the Crimean Peninsula as a whole. Almost always (with the exception of one study (Pozachenyuk, 2009), which is of historical value), river basins have been automatically identified using geographic information systems and DEMs. The second group consists of studies devoted to river basin groups in certain regions of the Crimean Peninsula: the river basins of the northwestern slope of the Crimean Mountains (Vermaat et al., 2012; Tabunshchik, 2018), the Kerch Peninsula (Krivoguz, 2016), the Sivash region (Timchenko et al., 2020), and the northern macroslope of the Crimean Mountains (Timchenko, 2000). The third and most numerous groups of studies has focused on the catchment basins of different separate rivers of the Crimean Peninsula (Vlasova, 2011; Pozachenyuk et al., 2014; Ergina and Timchenko, 2016; Kayukova, 2016; Amelichev et al., 2017). Moreover, in the third group, the most studied river basins are the basins of the largest rivers and their main tributaries.

The purpose of the study is to select the most accurate DEMs and, on its basis, to identify the basins and sub-basins of the rivers within northwestern slope of Crimean Mountains. Specifically, the main contents of this study are as follows. In the Section 2 "Materials and Methods," four DEMs are compared with a topographic map and a general scheme of research using geoinformation research methods are presented. In the Section 3 "Results," the calculation of measurement errors typical for various DEMs are shown. Also, the result of modeling the allocation of basins and sub-basins of the rivers of the northwestern slope of the Crimean Mountains are presented. In the Section 4, discussion of the obtained results and their comparison with other regions of the world are shown. Also, in the Section 4, the difficulties that the authors encountered while working on the article and ways to solve them are described. In the Section 5, conclusions and implications are given.

2 Materials and methods

2.1 Study area

The basins of the Zapadnyy Bulganak, Alma, Kacha, Belbek, Chernaya Rivers are located in the southwestern part of the Crimean Peninsula (Figure 1). The area of the studied territory comprises approximately 2,299 sq km. The rivers originate in the Crimean Mountains and flow into the Black Sea (Tabunshchik et al., 2022).

2.2 Materials and methods

We chose the DEMs for identifying river basins by selecting those most suitable for the study geodatasets, with the highest spatial resolution, and distributed under an open license. We thus selected

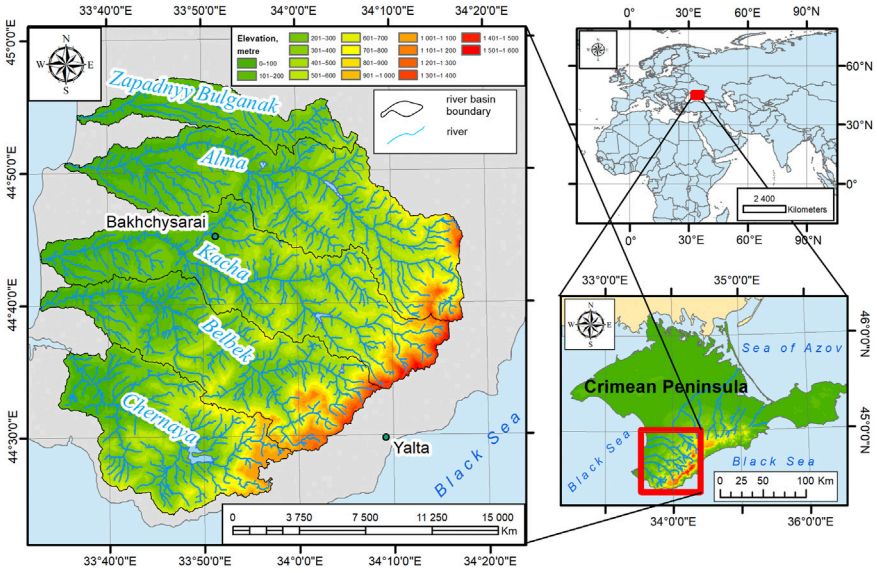


FIGURE 1
Geographical location of the study area (Tabunshchik et al., 2022).

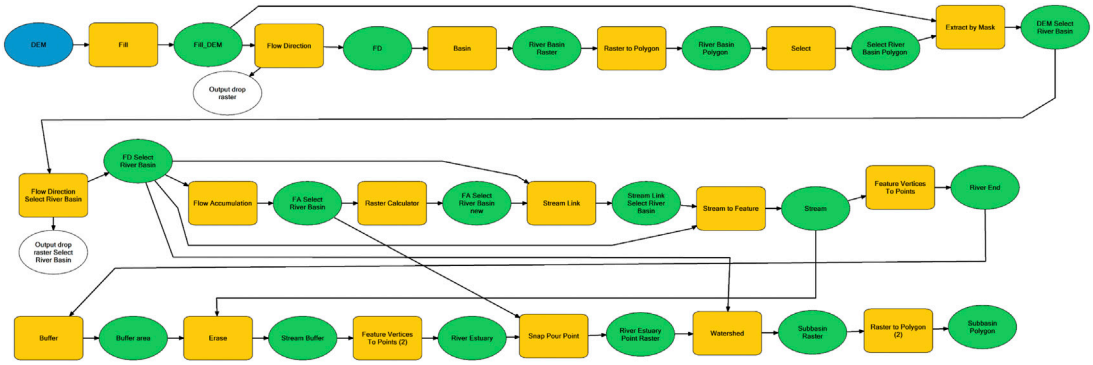


FIGURE 2
Model for identifying river basins on the northwestern slope of Crimean Mountains and their sub-basins within basins of Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers [compiled by us using (Samsonov, 2022)].

ASTER GDEM (Version 3, 2019), ALOS World 3D (Version 3.2, 2021), Copernicus DEM (Version 3, 2021), and SRTM DEM (Version 3, 2013) for this study among all DEMs available to us. The spatial resolution of these DEMs is 30 m/pixel.

As a DEM differs from the real terrain elevation, as for each pixel, an average value is given, we verified the selected DEMs by comparing the height marks and elevation values obtained from the topographic map. From a topographic map with a scale of 1: 100,000 [which was previously linked to the WGS 84 UTM zone 36 N (EPSG: 32,636) projection], we obtained sample values of several peaks (mountains, points) and isohypses, which we then compared with the elevation values of the same points on each of the considered DEMs through a simple spatial relationship. For these purposes, a point shapefile was created containing the elevation values of points from a topographic map. Then, using the tools "Spatial Join" and " Extract Values to Points," the elevation values from each DTM for each point were obtained. In total, we selected 100 points for each catchment basin, and then we compared the obtained data. Accuracy was calculated according to a previously reported method (Onkov, 2011). An additive error model was adopted during statistical data processing. According to this additive error model, we calculated the difference in heights of the DEM H_{DEM} and the topographic relief H_{TOPO} as

$$\Delta H = H_{DEM} - H_{TOPO} \quad (1)$$

Land was considered as the sum of systematic Δ_H and random Δh errors:

$$\Delta H = \Delta_H + \Delta h. \quad (2)$$

TABLE 1 Comparison of accuracy of absolute heights in study area according to a topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error, m	−7.7	−3.8	−3.4	−9.5
Root mean square error, m	14.0	12.8	8.7	14.9
Mean absolute error, m	10.3	7.4	6.0	10.9
Standard root mean square error, m	21.9	19.0	14.3	11.3
Centered mean absolute error, m	3.9	3.9	3.4	2.7

TABLE 2 Comparison of accuracy of absolute heights in Zapadnyy Bulganak River basin according to topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error, m	−3.8	−2.7	−1.5	−1.9
Root mean square error, m	6.7	4.4	3.6	3.7
Mean absolute error, m	4.7	2.8	2.2	2.4
Standard root mean square error, m	14.7	12.8	9.6	4.8
Centered mean absolute error, m	3.0	3.2	2.7	1.8

TABLE 3 Comparison of accuracy of absolute heights in Alma River basin according to topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error, m	−10.5	−7.0	−5.2	−11.4
Root mean square error, m	13.3	12.1	10.0	17.3
Mean absolute error, m	11.0	8.0	6.7	12.1
Standard root mean square error, m	23.3	20.8	16.5	13.6
Centered mean absolute error, m	4.2	4.3	3.8	2.9

TABLE 4 Comparison of accuracy of absolute heights in Kacha River basin according to topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error, m	−12.6	−9.1	−7.1	−16.1
Root mean square error, m	19.6	13.9	12.0	19.6
Mean absolute error, m	14.8	10.9	9.5	16.7
Standard root mean square error, m	28.2	22.9	18.7	14.2
Centered mean absolute error, m	4.5	4.5	4.0	3.3

10.3389/feart.2023.1218823

TABLE 5 Comparison of accuracy of absolute heights in Belbek River basin according to topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error. m	−6.2	−0.2	−3.4	−10.4
Root mean square error. m	15.3	20.5	8.8	16.3
Mean absolute error. m	11.7	10.2	6.8	13.4
Standard root mean square error. m	22.4	23.6	14.7	12.9
Centered mean absolute error. m	4.1	4.1	3.5	3.0

After eliminating the systematic error from the measurement results using

$$\Delta h_i = \Delta H_i - \bar{\Delta}_{Hi} \tag{3}$$

the parameters of the random component Δh were estimated.

The following types of errors were calculated in the study (where n is the number of measurements):

1. Average elevation difference (systematic error), m;

$$\bar{\Delta}_H = \frac{1}{n}\sum_{i=1}^{n}\Delta H_i \tag{4}$$

2. Root mean square error, m;

$$RMSE_{\Delta H} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta H_i^2} \tag{5}$$

3. Mean absolute error, м;

$$MAE_{\Delta H} = \frac{1}{n}\sum_{i=1}^{n}|\Delta H_i| \tag{6}$$

4. Standard root mean square error (Bessel's correction), m;

$$\sigma_{\Delta h} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\Delta h_i^2} \tag{7}$$

5. Centered mean absolute error, m.

$$\boldsymbol{\Theta}_{\Delta h} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}|\Delta h_i|} \tag{8}$$

The methodology for delineating river basins is based on the utilization of the ArcGIS software suite in conjunction with the DEM. It encompasses a systematic algorithm comprising a series of steps executed using the "Hydrology" toolbox within the "Spatial Analyst" tool.

1. The DEM is imported into the ArcGIS software suite.
2. The "Fill" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is employed to rectify erroneous depressions within the DEM.
3. The "Flow Direction" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is applied to ascertain the flow direction for each pixel of the DEM, which has been preprocessed (in step 2) using the "Fill" tool.
4. The "Flow Accumulation" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is utilized to compute the cumulative flow, representing the aggregated weight of all pixels that drain into each downslope pixel in the resulting raster. The flow direction raster created in step 3 serves as the input.
5. The "Raster Calculator" tool from the "Map Algebra" toolbox in the "Spatial Analyst" toolset is used to select pixels with a flow accumulation value exceeding 25. As a result, a new raster is generated with flow accumulation values above 25.
6. The "Stream Link" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is employed to create a raster linear network wherein each section of the network is assigned unique values, representing individual stream links. The input rasters consist of the flow direction raster generated in step 3 and the

TABLE 6 Comparison of accuracy of absolute heights in Chernaya River basin according to topographic map at scale of 1:100,000 for various DEM sets.

Error type	DEM			
	ASTER	ALOS	Copernicus	SRTM
Systematic error, m	−6.1	−1.6	−0.7	−9.2
Root mean square error, m	12.4	7.2	6.8	12.2
Mean absolute error, m	10.5	5.7	5.2	10.5
Standard root mean square error, m	20.6	14.7	11.8	8.2
Centered mean absolute error, m	3.9	3.7	3.3	2.6

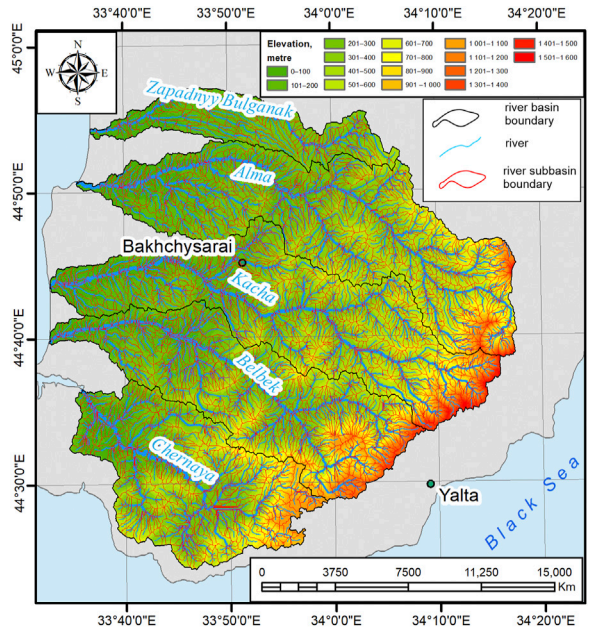


FIGURE 3
Stream sub-basins on the northwestern slope of the Crimean Mountains within the basins of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers.

flow accumulation raster with values exceeding 25, produced in step 5.

7. The "Stream Order" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is utilized to assign a stream order to each link within the stream network raster created in step 6.
8. The "Basin" tool from the "Hydrology" toolbox in the "Spatial Analyst" toolset is employed to generate a raster depicting river basins based on the constructed flow direction raster (step 3).
9. The "Raster to Polygon" tool from the "Conversion" toolbox within the "From Raster" toolset is applied to transform the stream network raster obtained in step 7 into a polygon shapefile.

We supplemented and implemented the method for identifying river basins and their sub-basins using ArcGIS 10.8 software and the builtin model editor "Model Builder," which allowed us to automate and speed up the delineation process (Figure 2). Our identification of river basins is based on a previously described method (Elkhrachy, 2018; Bajirao et al., 2019; Garrote, 2022; Samsonov, 2022). The theoretical and methodological foundations of delineating river basins and sub-basins were extensively discussed by Bai et al. (2015a) (Bai et al., 2015b). We automated it using the built-in ArcGIS Model Builder (Figure 2).

We note the sensitivity of this method to the incoming sets of spatial data: the type of DEM and its accuracy, as well as the accuracy of tying the points of the river mouths.

3 Results

We calculated the values of five different types of errors for the territory of the river basins of the northwestern slope of the Crimean Mountains. These values are presented in Table 1. Table 1 shows that the errors for ASTER GDEM and SRTM DEM were the largest; those of ALOS World 3D and Copernicus DEM were the smallest. Additionally, the values of errors within the basins of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers were analyzed (Table 2, Table 3, Table 4, Table 5, Table 6).

The Copernicus DEM is most suitable for the analysis of the morphometric characteristics of the river basins of the northwestern slope of the Crimean Mountains. This DEM has a spatial resolution of 30 m/pixel.

Based on the Copernicus DEM, the boundaries of the streams sub-basins within the basins of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers were identified as a result of modeling and partial manual correction of the obtained model results (we selectively checked for the presence of errors in the boundaries of the selected sub-basins). The obtained results are shown in Figure 3.

Figure 3 shows that we identified 3,293 sub-basins in the study area, which form the valley, gully, and ravine systems of the study area. The basin of the Western Bulganak River has 207 sub-basins, the Alma River has 860 sub-basins, the Kacha River has 855 sub-basins, the Belbek River has 747 sub-basins, and the Chernaya River has 624 sub-basins.

4 Discussion

After analyzing the data presented in Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, we found that the error values of the

TABLE 7 Comparison of basin areas of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers.

Basin	Area, km^2			Difference from literature data, %	
	SRTM DEM	Copernicus DEM	According to (Lisovsky et al., 2011)	SRTM DEM	Copernicus DEM
Zapadnyy Bulganak	177.1	174.6	180	2	3
Alma	641.8	631.8	635	−1	0
Kacha	570.9	573.3	573	0	0
Belbek	492.1	491.1	505	3	3
Chernaya	430.5	428.0	427	−1	0

Copernicus DEM, which has a spatial resolution of 30 m/pixel, were the lowest. Using the same technique with different data as the input to the model may have led to the obtained results slightly differing. This may concern both the boundaries of the study region and the software products on which the data were processed. For example, if we compare our earlier calculations of the area of catchment basins using SRTM DEM (Tabunshchik, 2021a) with those obtained using Copernicus DEM and data from the literature, insignificant differences are observed. Here, on the entire-basin scale, these changes are insignificant, and the differences are mainly related to the number of points along which the outer boundaries of the river basins are drawn.

We compared our data with those calculated by researchers for other regions. Karwel and Ewiak (2008) reported that the accuracy of SRTM within the flat part of the territory of Poland is 2.9 m, and 5.4 m for mountainous and foothill areas. Calculations (Orlyankin and Aleshina, 2019) showed that within the river basins of the northwestern slope of the Crimean Mountains, the systematic error of elevation calculated from the SRTM dataset, with a spatial resolution of 90 × 90 m, is +1 m.

Mutar et al. (2021) indicated that the RMSE of Copernicus DEM is 1.3 m in Iraq, which is 2.6 times more accurate than the SRTM DEM dataset and 5.2 times more accurate than the ASTER GDEM dataset. The accuracy of the Copernicus DEM dataset in China is 6.73 m (Li et al., 2022). Santillan and Makinano-Santillan (2016) found that when comparing datasets within the Philippines, the AW3D30 dataset most accurately represents true heights compared with the SRTM and ASTER GDEM datasets, because the AW3D30 dataset has the lowest mean error, RMSE, and standard deviation. Elkhrachy (2018), for the territory of Saudi Arabia, reported that when comparing DEM and a topographic map at a scale of 1: 10,000, which was chosen as a reference, the vertical accuracy of the SRTM and ASTER datasets is ±6.87 and ±7.97 m, respectively. Dong et al. (2015) conducted an accuracy assessment of ZY-3, SRTM, DLR-SRTM, and GDEM in Northeast China. GPS data was used as the accuracy evaluation criterion for ZY-3, and the RMSE for SRTM was found to be ±2.82 m. Zhang et al. (2019) compared ASTER, SRTM, ALOS, and TanDEM-X for flood risk mapping on the island of Hispaniola, using GPS and LiDAR measurements. They found that ASTER had the highest errors, while ALOS and TanDEM-X had the lowest errors. Karabörk et al. (2021) compared AlosPalsar, Sentinel-1A, AW3D30, SRTM, and ASTER GDEM with ground control points (GCP) obtained from digital aerial photographs, photogrammetric maps, or orthophotos. They found that the mean error values for ALOS were 1.1 m on flat terrain and 8.2 m in mountainous areas, while SRTM had mean errors of 1.8 m on flat terrain and 7.9 m in mountainous areas. ASTER had mean errors of 1.0 m on flat terrain and 8.4 m in mountainous areas. Purinton and Bookhagen (2021) compared the accuracy of SRTM, ASTER, ALOS, TanDEM-X, and Copernicus DEM in the Arid Central Andes. They found that the Copernicus DEM provided the most accurate representation of the landscape and should be the preferred DEM model for topographic analysis in areas where local high-quality DEM coverage is not available.

To demonstrate the changes in the areas and morphometric characteristics of river basins, Table 7 presents the results of a comparison of the basins area of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya rivers, calculated using the SRTM DEM and Copernicus DEM datasets, as well as a comparison with the data on river basin area given in the literature.

Using GIS, both the river basins (Ermolaev et al., 2014; Ali et al., 2023; Sharma et al., 2023) and sub-basins (Vanham and Bidoglio, 2014; Dallaire et al., 2019) of large rivers can be identified. However, the low accuracy of the DEM and new techniques can often lead to distortion of the output results. As an example, consider a previously described technique (Tabunshchik, 2021b; Samsonov, 2022), which is based on the PCRaster Python Library and automated by Van der Kwast as a PCRaster Tools plugin for QGIS. The application of this technique to the river basins of the northwestern slope of the Crimean Mountains showed a rather mixed picture that defies logical classification (Figure 4).

Figure 4 shows that the sub-basins of the rivers were not identified. The resulting processing result contained many errors, and the boundaries of sub-basins were identified without considering watersheds, which indicates the impossibility of applying this method in the study area.

The method proposed by Samsonov (2022) produced the best result of modeling the sub-basins of the five largest rivers of the northwestern slope of the Crimean Mountains (Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers). We propose using a buffer value that exceeds twice the DEM resolution's value to separate the points of the mouth of smaller streams flowing into the main stream. In our measurements, we found that this value should be equal to the pixel resolution, given the large error in constructing and visualizing data with a buffer size of two pixels (Figure 5).

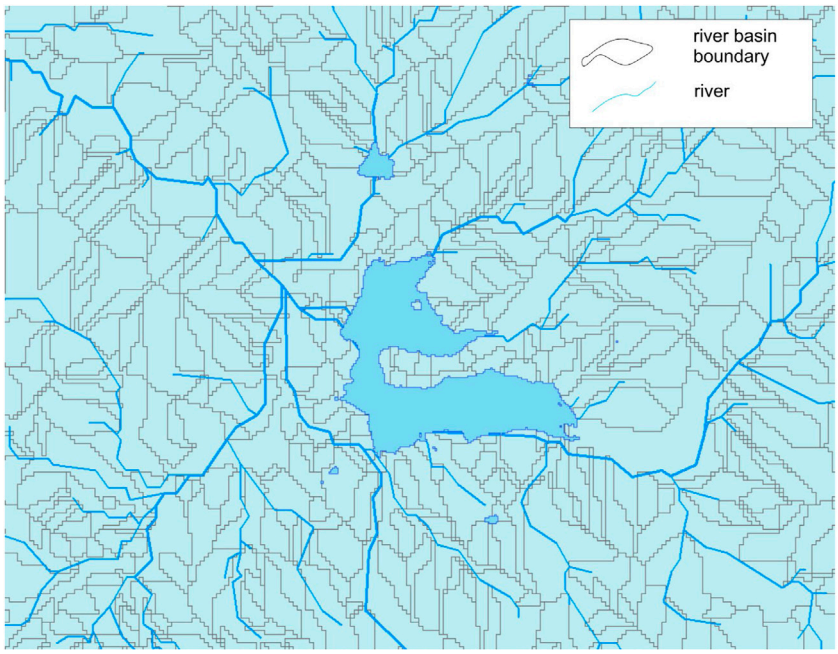


FIGURE 4
Example of the unsuccessful selection of sub-basins using PCRaster Tools plugin for QGIS on catchment basin fragment of Chernaya River.

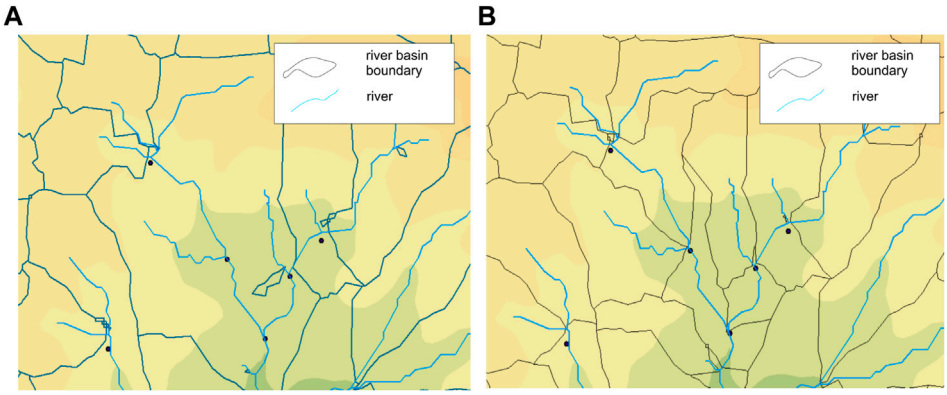


FIGURE 5
Fragment of a map of sub-basins compiled using a 60 m buffer **(A)**, containing construction errors and 30 m **(B)** with minimal construction errors.

In some areas, we identified single image pixels that did not belong to catchment basins, so we then manually identified them. Additionally, the resolution of various open DEM datasets impacts the accuracy of sub-basin identification. When comparing DEM data with a resolution of 90 m/pixel with those with a resolution of 30 m/pixel, the most accurate results were achieved by the latter.

Many studies have focused on the identification of river basins using SRTM DEM for the Crimean Peninsula (Vlasova, 2012; Pozachenyuk et al., 2014; Pozachenyuk et al., 2015; Tabunshchik, 2021a; Narozhnyaya, 2021; Drygval, 2022). Other DEMs have practically not been used, which is probably due to the popularity and widespread use of SRTM DEM.

Narozhnyaya (2021) provided a detailed description of the morphometric analysis of the river basins of the Crimean Peninsula; however, the description is based on the use of SRTM DEM. A detailed description of the individual basins of the large rivers of the Crimean Peninsula was not provided, including the five basins that we considered. Narozhnyaya (2021) did not distinguish the sub-basins of the rivers: only separate maps were presented that allowed judging the distribution of certain morphometric indicators of river basins.

We identified the sub-basins within the basins of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya Rivers for the first time, and we substantially refined the available information on the quantitative characteristics of the basins of these rivers. This, in turn, creates many opportunities for researchers of the nature of the Crimean Peninsula to solve many problems with more accurate initial data. Additionally, the use of DEMs is advisable when conducting complex engineering, geological, hydrological, and hydrogeological studies. Several directions for further research of the river basins and sub-basins of the northwestern slope of the Crimean Mountains emerge from our study, in particular, their anthropogenic transformation, searching for relationships between climate change and changes in vegetation and land cover types; assessing their geoecological state; preparing landscape planning maps; and developing and implementing recommendations for sustainable development.

In further research, it is necessary to continue comparing different DEM datasets and identifying the most accurate and suitable ones for specific tasks. While the study focused on analyzing specific DEM datasets for the studied region, there is potential in future research to examine the accuracy of various DEM datasets in different regions and compare them to each other. Additionally, advancements in remote sensing technologies and data processing methods present opportunities to improve the accuracy of DEM datasets. Further research can explore the integration of data from multiple sources, such as LiDAR, satellite imagery, and ground-based measurements, to enhance the accuracy and reliability of DEMs. The use of Unmanned Aerial Vehicles (UAVs) at the local level of investigation also holds great interest.

5 Conclusion

Despite more than two centuries of hydrological studies of the rivers and river basins of the Crimean Peninsula, many unsolved problems remain. In general, the Crimean Peninsula remains insufficiently hydrologically studied. This primarily applies to important tasks such as the hydrological characteristics of rivers and the morphometric characteristics of river basins, determining the types of water management use of rivers and developing schemes for optimizing river and river basin use, studying and predicting possible ecogeodynamic processes under the influence of anthropogenic factors within river basins, studying the degree of anthropogenic transformation of the river basin, etc. Only at the end of the 20th to the beginning of the 21st century did detailed work begin on the identification and description of the basins and sub-basins of the rivers of the Crimean Peninsula, which continues to this day. However, these studies are still extremely scarce, and the data are scattered. This study of the river basins and sub-basins identifying the northwestern slope of the Crimean Mountains provides a distinctive contribution to the unresolved history of hydrological research within the Crimean Peninsula.

The use of DEMs enables the study of the main morphometric characteristics of the river basins of the northwestern slope of the Crimean Mountains. However, when choosing the initial data, the least error-prone datasets should be used. The performed calculations showed that the smallest errors in the selection of DEM were obtained for Copernicus DEM, which has a resolution of 30 m/pixel. Copernicus DEM provides a sufficiently high level of accuracy and detail, which was shown in the calculation model of the Zapadnyy Bulganak, Alma, Kacha, Belbek, and Chernaya River basins, as well as their sub-basins.

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

VT conceived and designed the experiments; TG performed the experiments; VT and RG analyzed the data and contributed to the analysis tools; VT, TG, CP, AK, and RG wrote the paper, but all authors discussed the results and enhanced the final draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The research was conducted within the framework of the research topic “Studying the spatial and temporal organisation of aquatic and terrestrial ecosystems in order to develop an operational monitoring system based on remote sensing data and GIS technologies. Registration number: 121040100327-3.” The RUDN University Strategic Academic Leadership Program supported this research.

Acknowledgments

The authors are grateful to Makarova V.I. and Kolesnikova E.M. (Don State Public Library) for support of the study.

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

Ali, F., Srisuwan, C., Techato, K., and Bennui, A. (2023). Assessment of small hydropower in songkhla lake basin, Thailand using GIS-MCDM. *Sustain. Water Resour. Manag.* 9 (1), 25–22. doi:10.1007/s40899-022-00788-w

Amelichev, G. N., Oliferov, A. N., and Novikova, F. N. (2017). Hydrological features of the river abdalka (simferopol) in recharge area of the plain crimea artesian basin. *Sci. Notes V.I. Vernadsky Crime. Fed. Univ. Geogr. Geol.* 3 (1), 160–175.

Annis, A., Nardi, F., Petroselli, A., Apollonio, C., Arcangeletti, E., Tauro, F., et al. (2020). UAV-DEMs for small-scale flood hazard mapping. *Water* 12, 1717. doi:10.3390/w12061717

Bai, R., Li, T. J., Huang, Y. F., Li, J. Y., and Wang, G. Q. (2015a). An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree. *Geomorphology* 238, 56–67. doi:10.1016/j.geomorph.2015.02.028

Bai, R., Li, T. J., Huang, Y., Li, J. Y., Wang, G. Q., and Yin, D. Q. (2015b). A hierarchical pyramid method for managing large-scale high-resolution drainage networks extracted from DEM. *Comput. Geosciences* 85 (PA), 234–247. doi:10.1016/j.cageo.2015.06.019

Bajirao, T. S., Kumar, P. K., Kumar, P. K., Tarate, C., and Bajirao, S. (2019). Application of remote sensing and GIS for morphometric analysis of watershed: A review. *IJCS* 7.2, 709–713.

Balzter, H., Cole, B., Thiel, C., and Schmullius, C. (2015). Mapping CORINE land cover from sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens.* 7, 14876–14898. doi:10.3390/rs71114876

Bodin, X., Thibert, E., Sanchez, O., Rabatel, A., and Jaillet, S. (2018). Multi-annual kinematics of an active rock glacier quantified from very high-resolution DEMs: An application-case in the French alps. *Remote Sens.* 10, 547. doi:10.3390/rs10040547

Courty, L. G., Soriano-Monzalvo, J. C., and Pedrozo-Acuña, A. (2019). Evaluation of open-access global digital elevation models (AW3D30, SRTM, and ASTER) for flood modelling purposes. *J. Flood Risk Manag.* 12, e12550. doi:10.1111/jfr3.12550

Dallaire, C. O., Lehner, B., Sayre, R., and Thieme, M. (2019). A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environ. Res. Let* 14 (2), 024003. doi:10.1088/1748-9326/aad8e9

del Rosario Gonzalez-Moradas, M., and Viveen, W. (2020). Evaluation of ASTER GDEM2, SRTMv3. 0, ALOS AW3D30 and TanDEM-X DEMs for the Peruvian Andes against highly accurate GNSS ground control points and geomorphological-hydrological metrics. *Remote Sens. Environ.* 237, 111509. doi:10.1016/j.rse.2019.111509

Dong, Y., Chang, H. C., Chen, W., Zhang, K., and Feng, R. (2015). Accuracy assessment of GDEM, SRTM, and DLR-SRTM in northeastern China. *Geocarto Int.* 30 (7), 779–792. doi:10.1080/10106049.2014.985744

Drygval, A. V. (2022). "Morphometric analysis of the Shelen river basin (Southern coast of Crimea)," in *Geology and water resources of Crimea* (LEMA Publishing House).

Dunaieva, Y., and Kovalenko, P. (2013). River basins typification of crimea by agrolandscapes and ecological load. *Sci. J. Russ. Sci. Res. Inst. Land Improv. Problems* 4 (12), 157–167.

Elkhrachy, I. (2018). Vertical accuracy assessment for SRTM and ASTER digital elevation models: A case study of najran city, Saudi Arabia. *Ain Shams Eng. J.* 9 (4), 1807–1817. doi:10.1016/j.asej.2017.01.007

Ergina, E. I., and Timchenko, Z. V. (2016). The soil of the basin of the angara river. *Sci. Notes V.I. Vernadsky Crime. Fed. Univ. Geogr. Geol.* 2 (2), 88–96.

Ermolaev, O. P., Maltsev, K. A., and Ivanov, M. A. (2014). Automated construction of boundaries of basin geosystems for the Volga. *Fed. Dist. Geogr. Nat. Resour.* 3, 32–39.

Escobar Villanueva, J. R., Iglesias Martínez, L., and Pérez Montiel, J. I. (2019). DEM generation from fixed-wing UAV imaging and LiDAR-derived ground control points for flood estimations. *Sensors* 19, 3205. doi:10.3390/s19143205

Fang, G., Yuan, T., Zhang, Y., Wen, X., and Lin, R. (2019). Integrated study on soil erosion using RUSLE and GIS in yangtze River Basin of jiangsu province (China). *Arabian J. Geosciences* 12 (5), 173–213. doi:10.1007/s12517-019-4331-2

Fawdon, P., Gupta, S., Davis, J. M., Warner, N. H., Adler, J. B., Balme, M. R., et al. (2018). The Hypanis Valles delta: The last highstand of a sea on early Mars? *Earth Planet. Sci. Let* 500, 225–241. doi:10.1016/j.epsl.2018.07.040

Fischer, M., Huss, M., and Hoelzle, M. (2015). Surface elevation and mass changes of all Swiss glaciers 1980–2010. *Cryosphere* 9 (2), 525–540. doi:10.5194/tc-9-525-2015

Florinsky, I. V., and Filippov, S. V. (2017). A desktop system of virtual morphometric globes for Mars and the Moon. *Planet. Space Sci.* 137, 32–39. doi:10.1016/j.pss.2017.01.005

Garrote, J. (2022). Free global DEMs and flood modelling—a comparison analysis for the january 2015 flooding event in mocuba city (Mozambique). *Water* 14, 176. doi:10.3390/w14020176

Hashemi-Beni, L., Jones, J., Thompson, G., Johnson, C., and Gebrehiwot, A. (2018). Challenges and opportunities for UAV-based digital elevation model generation for flood-risk management: A case of princeville, North Carolina. *Sensors* 18 (11), 3843. doi:10.3390/s18113843

Karabörk, H., Makineci, H. B., Orhan, O., and Karakus, P. (2021). Accuracy assessment of DEMs derived from multiple SAR data using the InSAR technique. *Arabian J. Sci. Eng.* 46, 5755–5765. doi:10.1007/s13369-020-05128-8

Karionov, Yu.I. (2010). Assessment of the accuracy of the SRTM DEM. *Geoprofi* 1, 48–51.

Karlson, M., Bastviken, D., and Reese, H. (2021). Error characteristics of pan-arctic digital elevation models and elevation derivatives in northern Sweden. *Remote Sens.* 13, 4653. doi:10.3390/rs13224653

Karwel, A. K., and Ewiak, I. (2008). Estimation of the accuracy of the SRTM terrain model on the area of Poland. *Int Archives Photogrammetry Remote Sens Spatial Inf Sci* XXXVII (7), 169–172.

Kayukova, E. P., (2016). Using hydrochemical data to estimate the components of the water balance (the Bodrak river basin, southwestern Crimea). *Ser. 7. Geol. Geogr* 4, 25–36. Vestnik of Saint Petersburg University. doi:10.21638/11701/spbu07.2016.403

Kramm, T., and Hoffmeister, D. (2021)., 37. Germany, 7774–7799. Comprehensive vertical accuracy analysis of freely available DEMs for different landscape types of the Rur catchment, Germany*Geocarto Int.* doi:10.1080/10106049.2021.1984588

Krivoguz, D. O. (2016). *The role of hydrological processes in the processes of landslide formation of the Kerch Peninsula*. Yekaterinburg, Russia: Aeterna Limited Liability Company.

Lehner, B., and Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrol. Process.* 27 (15), 2171–2186. doi:10.1002/hyp.9740

Li, H., Zhao, J., Yan, B., Yue, L., and Wang, L. (2022). Global DEMs vary from one to another: An evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data. *Int. J. Digital Earth* 15 (5), 1149–1168. doi:10.1080/17538947.2022.2094002

Lisovsky, A. A., Novik, V. A., Timchenko, Z. V., and Gubskaya, U. A. (2011). *Surface water bodies of the Crimea. Management and use of water resources*. Krymuchpedgiz Publishing House.

Liu, Y., Gong, W., Hu, X., and Gong, J. (2018). Forest type identification with random forest using sentinel-1A, sentinel-2A, multi-temporal landsat-8 and DEM data. *Remote Sens.* 10, 946. doi:10.3390/rs10060946

Manune, D. F. (2007). *Digital elevation model. Technologies and applications: The DEM users manual.* Bethesd: American Society for Photogrammetry and Remote Sensing.

Mesa-Mingorance, J. L., and Ariza-López, F. J. (2020). Accuracy assessment of digital elevation models (DEMs): A critical review of practices of the past three decades. *Remote Sens.* 12 (16), 2630. doi:10.3390/rs12162630

Mutar, A. Q., Mustafa, M. T., and Hameed, M. A. (2021). The impact of (DEM) Accuracy on the Watersheds areas as a function of spatial data. *Periodicals Eng. Nat. Sci. (PEN)* 9 (4), 1118–1130. doi:10.21533/pen.v9i4.2585

Narozhnyaya, A. G. (2021). Morphometric analysis of the Crimean river basins relief by using GIS. *InterCarto. InterGIS.* 27 (2), 218–232. doi:10.35595/2414-9179-2021-2-27-218-232

Nikolakopoulos, K. G., Kamaratakis, E. K., and Chrysoulakis, N. (2006). SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. *Int. J. Remote Sens.* 27, 4819–4838. doi:10.1080/01431160600835853

Onkov, I. (2011). Evaluation of SRTM height accuracy for the orthotransformation of high resolution satellite images. *Geomatics* 3, 40–46.

Orlyankin, V. N., and Aleshina, A. R. (2019). The use of matrixes of heights srtm in the preliminary calculations and mapping of depths and potential flood inundation of fluvial plain. *Issled. Zemli iz Kosmosa* 5, 72–81. doi:10.31857/S0205-96142019572-81

Pozachenyuk, E. A., Ergina, E. I., Oliferov, A. N., Mikhailov, V. A., Vlasova, A. N., Kudrjan', E. A., et al. (2014). Analysis of factors of the Salgir River's water resources formation under the condition of climate changing. *Sci. Notes V.I* 2, 118–138.

Pozachenyuk, E. A., Lisetskii, F. N., Vlasova, A. N., Buryak, Z. A., and Marinina, O. A., (2015). Model of position-dynamic structure of river basins. *Res. J. Pharm. Biol. Chem. Sci.* 6, 1776–1780.

Pozachenyuk, E. A. (2009). "Modern landscapes of the crimea and adjacent water areas," in 19th International Multidisciplinary Scientific GeoConference SGEM 2019, At: Albena, Bulgaria.

Purinton, B., and Bookhagen, B. (2021). Beyond vertical point accuracy: Assessing inter-pixel consistency in 30 m global dems for the arid central andes. *Front. Earth Sci.* 9, 758606. doi:10.3389/feart.2021.758606

Pyankov, S. V., and Shikhov, A. N. (2017). *Geoinformation support for modeling hydrological processes and phenomena*. Perm State National Research University.

Rajasekhar, M., Raju, G. S., Raju, R. S., Ramachandra, M., and Kumar, B. P. (2018). Data on comparative studies of lineaments extraction from ASTER DEM, SRTM, and Cartosat for Jilledubanderu River basin, Anantapur district, AP, India by using remote sensing and GIS. *Data brief* 20, 1676–1682. doi:10.1016/j.dib.2018.09.023

Samsonov, T. E. (2022). *Fundamentals of geoinformatics: Practicum*. Moscow, Russia: Faculty of Geography of Moscow State University. doi:10.5281/zenodo.1167857

Santillan, J. R., and Makinano-Santillan, M. (2016). Vertical accuracy assessment of 30-m resolution ALOS, ASTER and SRTM global DEMS over Northeastern Mindanao, Philippines. XXIII ISPRS Congress. Comission IV. *Int. Archives Photogrammetry Remote Sens. Spatial Inf. Sci.* 41 (B4), 149–156. doi:10.5194/isprsarchivesXLI-B4-149-2016

Sarkar, D., Mondal, P., Sutradhar, S., and Sarkar, P. (2020). Morphometric analysis using SRTM-DEM and GIS of nagar River Basin, indo-Bangladesh barind tract. *J. Indian. Soc. Remote Sens.* 48, 597–614. doi:10.1007/s12524-020-01106-7

Sharma, P., Singh, M. M., Chaurasia, R. S., and Sabir, M. (2023). Remote sensing and GIS based approach in morphometric analysis of birma River Basin (central India). *Sustain. Agri, Food Environ. Res.* 11. doi:10.7770/safer-V11N1-art2387

Tabunshchik, V. A. (2021b). Estimation of stream power index (SPI) in the River Basins of the north-western slopes of the Crimean Mountains (River Basins Zapadnyy Bulganak, Alma, Kacha, Belbek, Chernaya). *Geopolit. i ecogeodinamika Reg.* 7 (2), 344–354.

Tabunshchik, V. A. (2021a). Morphometric characteristics of River Basins of the northwestern slope of the Crimean Mountains (Zapadnyy Bulganak, Alma, Kacha, Belbek, Chernaya River Basins). *Sci. Notes V.I. Vernadsky Crime. Fed. Univ. Geogr. Geol.* 7 (3), 267–278. doi:10.37279/2413-1717-2021-7-3-267-278

Tabunshchik, V. A. (2018). Relief of river basins of the north-western slope of the Crimean Mountains (river basins Zapadnyy Bulganak, Alma, Kacha, Belbek, Chernaya). *Geopolit. ecogeodynamics regions* 4 (3), 78–87.

Tabunshchik, V., Gorbunov, R., and Gorbunova, T. (2022). Anthropogenic transformation of the River Basins of the northwestern slope of the Crimean Mountains (the crimean Peninsula). *Land* 11, 2121. doi:10.3390/land11122121

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., et al. (2016). Generation of the 30 M-mesh global digital surface model by ALOS PRISM. International Archives of the Photogrammetry. *Remote Sens. Spatial Inf. Sci.* 41 (XLI-B4), 157–162. doi:10.5194/isprs-archives-XLI-B4-157-2016

Tang, Q., and Lettenmaier, D. P. (2012). 21st century runoff sensitivities of major global river basins. *Geophys. Res. Let* 39 (6), L06403. doi:10.1029/2011GL050834

Thomas, J., and Prasannakumar, V. (2015). Comparison of basin morphometry derived from topographic maps, ASTER and SRTM DEMs: An example from Kerala, India. *Geocarto Int.* 30 (3), 346–364. doi:10.1080/10106049.2014.955063

Timchenko, Z. V. (2000). *Assessment of the geoecological state of water resources of small rivers (on the example of small rivers of the northern macroscline of the Crimean Mountains), V.I.* Simferopol: Vernadsky Taurida National University.

Timchenko, Z. V., Tabunshchik, V. A., and Zelentsova, M. G. (2020). The characteristics of the dzhankoy region rivers and dzhankoy town okrug of the republic of the crimea. *IOP Conf. Ser. Earth Environ. Sci.* 548 (5), 052038. doi:10.1088/1755-1315/548/5/052038

Trofimov, A. A., and Filippova, A. V. (2014). Assessment of the accuracy of the SRTM DEM based on topographic surveys. *Geoprofi* 6, 13–17.

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., and Kmoch, A. (2020). Vertical accuracy of freely available global digital elevation models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sens.* 12 (21), 3482. doi:10.3390/rs12213482

Uysal, M., Toprak, A. S., and Polat, N. (2015). DEM generation with UAV Photogrammetry and accuracy analysis in Sahitler hill. *Measurement* 73, 539–543. doi:10.1016/j.measurement.2015.06.010

Vanham, D., and Bidoglio, G. (2014). The water footprint of agricultural products in European river basins. *Environ. Res. Let* 9 (6), 064007. doi:10.1088/1748-9326/9/6/064007

Vermaat, J. E., Broekx, S., Van Eck, B., Guy, E., Hellmann, F., Jean Luc De Kok, et al. (2012). Nitrogen source apportionment for the catchment, estuary, and adjacent coastal waters of the river scheldt. *Ecol. Soc.* 17 (2), 30. doi:10.5751/ES-04889-170230

Viel, J. A., Rosa, K. K. D., and Mendes Junior, C. W. (2020). Avaliação da Acurácia Vertical dos Modelos Digitais de Elevação SRTM, ALOS World 3D e ASTER GDEM: Um Estudo de Caso no Vale dos Vinhedos, RS – brasil. *Rev. Bras. Geogr. Física* 13 (5), 2255–2268. doi:10.26848/rbgf.v13.5.p2255-2268

Vlasova, A. N. (2012). The management of river basins and the Black Sea coastal zone by landscape planning instruments (The Crimea, the Voron river basin as an example). *Turkish J. Fish. Aquatic Sci.* 12 (5), 535–537.

Vlasova, A. (2011). The estimation of ecological state of the Salgir river basin. *Sci. Notes V.I. Vernadsky Taurida Natl. Univ. Geogr.* 24 (1), 66–71.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophys. Res. Let* 44 (11), 5844–5853. doi:10.1002/2017GL072874

Yeritsian, H. H. (2013). Comparison of the DEMS generated from 1:50000, 1: 100000 and 1:200000 topo-maps with the SRTM DEM data. *Proc. NAS RA - Earth Sci.* 66 (1), 39–47.

Yermolaev, O., Mukharamova, S., and Vedeneeva, E. (2021). River runoff modeling in the European territory of Russia. *Catena* 203, 105327. doi:10.1016/j.catena.2021.105327

Yuan, F., Repse, M., Leith, A., Rosenqvist, A., Milcinski, G., Moghaddam, N. F., et al. (2022). An operational analysis ready radar backscatter dataset for the african continent. *Remote Sens.* 14, 351. doi:10.3390/rs14020351

Zhang, K., Gann, D., Ross, M., Robertson, Q., Sarmiento, J., Santana, S., et al. (2019). Accuracy assessment of ASTER, SRTM, ALOS, and TDX DEMs for Hispaniola and implications for mapping vulnerability to coastal flooding. *Remote Sens. Environ.* 225, 290–306. doi:10.1016/j.rse.2019.02.028

Zhao, S., Qi, D., Li, R., Cheng, W., and Zhou, C. (2021). Performance comparison among typical open global DEM datasets in the Fenhe River Basin of China. *Eur. J. Remote Sens.* 54 (1), 145–157. doi:10.1080/22797254.2021.1891577

frontiers | Frontiers in Environmental Science

TYPE Brief Research Report
PUBLISHED 19 July 2023
DOI 10.3389/fenvs.2023.1187592

OPEN ACCESS

EDITED BY
Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Jian Xu,
Chinese Academy of Sciences (CAS), China
Simone Lolli,
National Research Council (CNR), Italy

*CORRESPONDENCE
Truong Xuan Ngo,
truonggnx@vnu.edu.vn

RECEIVED 16 March 2023
ACCEPTED 04 July 2023
PUBLISHED 19 July 2023

CITATION
Ngo TX, Phan HDT and Nguyen TTN (2023), Development of ground-level NO_2 models in Vietnam using machine learning and satellite observations with ancillary data.
Front. Environ. Sci. 11:1187592.
doi: 10.3389/fenvs.2023.1187592

Development of ground-level NO_2 models in Vietnam using machine learning and satellite observations with ancillary data

Truong Xuan Ngo*, Hieu Dang Trung Phan and Thanh Thi Nhat Nguyen

Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

In this study, the aim was to create daily ground-level NO_2 maps for Vietnam spanning from 2019 to 2021. To achieve this, various machine learning models (including the Mixed Effect Model, Neural Network, and LightGBM) were utilized to process satellite NO_2 tropospheric columns from Ozone Monitoring Instrument (OMI) and TROPOMI, as well as meteorological and land use maps and ground measurement NO_2 data. The LightGBM model was found to be the most effective, producing results with a Pearson r of 0.77, RMSE of 7.93 μg/m^3, and Mean Relative Error (MRE) of 42.6% compared to ground truth measurements. The annual average NO_2 maps from 2019–2021 obtained by the LightGBM model for Vietnam were compared to a global product and ground stations, and it was found to have superior quality with Pearson r of 0.95, RMSE of 2.27 μg/m^3, MRE of 9.79%, based on 81 samples.

KEYWORDS

Sentinel 5p, OMI, ground-level NO_2 model, machine learning, Vietnam

1 Introduction

Air pollution poses a significant threat to the environment and human health in many countries. In Vietnam, Nitrogen dioxide (NO_2) is recognized as a particularly important air pollutant. To monitor and manage the levels of NO_2 and other harmful pollutants such as $PM_{2.5}$, PM_{10}, SO_2, and O_3, the Ministry of Natural Resources and Environment (MONRE) has implemented automatic and continuous monitoring systems. However, the current monitoring of NO_2 in Vietnam is limited due to the lack of representative monitoring stations across the country. In recent times, modeling techniques utilizing data from monitoring stations, satellite imagery (remote sensing), and auxiliary sources have gained widespread acceptance in generating spatial NO_2 information. This approach provides additional data to supplement the readings from monitoring stations, thus providing insights into the distribution of NO_2 concentrations on a larger scale, especially in regions without monitoring stations. The NO_2 satellites used for this purpose include the Ozone Monitoring Instrument (OMI), Global Ozone Monitoring Experiment–2 (GOME-2), SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), and TROPOspheric Monitoring Instrument (TROPOMI).

Many studies have been conducted globally to map NO_2 using satellite imagery. For instance, Larkin et al. (2017) used a land use regression (LUR) model to estimate global NO_2

levels in 2011 with a resolution of 100 × 100 m. They incorporated model and satellite data/model data from SCIAMACHY, GOME-2, and GEOS Chem, as well as land cover features such as vegetation index, tree cover, traffic, *etc.*, and monitoring station data from 58 countries. The model's performance varied depending on the region, with the coefficient of determination (R^2) ranging from 0.42 in Africa to 0.67 in South America. In North America, Europe, and Asia, the R^2 value was approximately 0.52, which is consistent with the global average (0.54) (Larkin et al., 2017). To further enhance the accuracy of NO_2 mapping, a study conducted by Anenberg et al. (2022) estimated the global average annual NO_2 levels from 1990 to 2020 at a resolution of 1 × 1 km. This study used Land Use Regression (LUR) incorporating OMI NO_2 and MERRA2-reanalysis data. Results indicate that the new NO_2 concentration data is more precise than that of Larkin's study in rural areas, with a Pearson r of 0.58 and a Root mean square error (RMSE) of 2.26 (ppb) (Anenberg et al., 2022). The results of this study have important implications for public health, as they were able to estimate the NO_2-attributable pediatrics asthma incidence using the improved NO_2 concentration data. Paraschiv examined the relationship between OMI data and monitoring stations across Europe during the period of 2005–2014. Their findings indicate a Pearson r value ranging from 0.53 to 0.86 (Paraschiv et al., 2017). Hyung Joo Lee and colleagues (2014) developed a mixed-effect model (MEM) to estimate daily NO_2 concentrations in New England, United States from 2005–2010. Their model was based on various data sources, including station data, tropospheric column NO_2 (OMI), historical land use data such as population density, traffic, topography, as well as meteorological data such as temperature and wind speed. They evaluated the model using a 10-fold cross-validation (CV) method and found an R^2 value of 0.79, indicating good model performance (Lee and Koutrakis, 2014). In the mentioned studies, OMI NO_2 satellite data is commonly used to estimate NO_2 maps.

Recently, some studies have been conducted using TROPOMI satellite data (the most recently launched satellite with high resolution data) with Machine Learning models and auxiliary data to estimate ground-level pollutant concentrations (e.g., NO_2, O_3). A study by Kang et al. (2021) estimated ground-level NO_2 and O_3 with a resolution of 6 × 6 km at East Asia using NO_2 data from the TROPOMI satellite, other satellite products (Landcover, Aerosol Optical Depth - AOD, Digital Elevation Model - DEM), meteorological data from models, and auxiliary data (road density, population density). Several different machine learning models were experimented, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). XGBoost showed better results when estimating NO_2 with a 10-fold cross-validation R^2 of 0.7 and RMSE of 4.75 ppb. Long et al. (2022) map daily ground-level NO_2 concentrations in China at a resolution of 0.05° using machine learning models based on decision trees (Decision Tree, Gradient Boost Decision Tree, Random Forest, Extra-Trees). They found that the Extra-Trees model incorporating spatial and temporal information performed exceptionally well in estimating ground-level NO_2 concentrations, achieving a cross-validation R^2 of 0.81 and an RMSE of 3.45 μg/m^3 in test datasets (Long et al., 2022). Wang et al. (2022) used Random Forest to estimate the daily maximum 8-hour average ground-level ozone concentration at a 10 km spatial resolution in California. They utilized TROPOMI's total ozone column combined with ozone profile information retrieved by the Ozone Monitoring Instrument (OMI) and auxiliary data (meteorological, land use). Their model achieved an overall 10-fold CV R^2 of 0.84 and an RMSE of 0.0059 ppm. In another study, Grzybowski et al. (2023) employed various data sources, including Sentinel-5P, meteorological data, and other ancillary data, to estimate ground NO_2 levels in Poland. Among the methods used, the random forest (RF) model emerged as the most accurate, with mean absolute error (MAE) values of 3.4 μg/m^3 and 3.2 μg/m^3 for the hourly and weekly estimates, respectively. The corresponding mean absolute percentage error (MAPE) values were 37% and 31%, indicating relatively moderate deviations from the true values (Grzybowski et al., 2023). The tree-based model demonstrates strong estimation capabilities in air pollution estimation problems using remote sensing and auxiliary data.

Currently, there are no studies on nationwide NO_2 estimation in Vietnam utilizing satellite images and multi-source data. However, a study conducted in 2015 developed daily $PM_{2.5}$ maps for Vietnam from 2010–2014 using a multivariable regression model (Nguyen et al., 2015). Recently, a study provided a long-term daily $PM_{2.5}$ map for Vietnam from 2012–2020 using mixed effect models based on ground $PM_{2.5}$ measurements, integrated satellite Aerosol Optical Depth (AOD), meteorological and land use maps (Ngo et al., 2023). The daily mean $PM_{2.5}$ maps have high validation results with ground $PM_{2.5}$ measurements, achieving a Pearson r of 0.87, R^2 of 0.75, RMSE of 11.76 μg/m^3, and MRE of 36.57% on a total of 13,886 data samples.

This study aimed to develop daily ground-level NO_2 maps with a resolution of 1 × 1 km over Vietnam using satellite images and multi-source data from 2019–2021. The NO_2 tropospheric columns were derived from OMI and TROPOMI satellite devices, and different models such as Mixed Effect Model, Neural Network, and LightGBM were tested. Although the models are not new, this is the first study to experimentally construct a high-resolution NO_2 map for the entire territory of Vietnam based on satellite data. Various machine learning models were experimented to find the optimal model that fits the data in Vietnam. The NO_2 maps hold promise in providing useful information on NO_2 distribution across Vietnam, supporting decision-making and policies to reduce NO_2 pollution and improving public health.

2 Materials

2.1 Measurement data

The hourly ground measurements of NO_2 were collected from monitoring stations in Vietnam. Vietnam is situated in the East of the Indochina peninsula, at the heart of Southeast Asia, with its land area covering 331,236 km^2, stretching from (8°27′N, 102°8′E) to (23°23′N, 109°27′E). The country is divided into six distinct economic zones, namely, the Northern Midlands and Mountains,

Red River Delta (RRD), North Central Coast and South Central Coast, Central Highlands, South East, and Mekong River Delta (MRD) as illustrated in Supplementary Figure S1.

The Northern Center for Environmental Monitoring (NCEM), which operates under the Vietnam Administration of Environment (VEA) under MONRE, is responsible for air pollution monitoring in Vietnam. As of 2021, over 90 stations have been installed across the country, with most of them located in the Red River Delta (RRD) region. These stations measure various pollutants such as NO_2, PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, as well as meteorological variables like temperature, humidity, and wind speed. Hourly NO_2 concentration ($\mu g/m^3$) data from 74 stations were collected between 2019–2021 in this study, with poor quality data stations removed. The distribution of ground stations is illustrated in Supplementary Figure S1.

2.2 Satellite data

In order to monitor air pollution at stations on a national scale, satellite images are also used which has a larger coverage than the traditional monitoring method. The development of satellite technology can solve the problem of monitoring air pollution on a large scale. For this study, we utilized two satellite based NO_2 tropospheric column products, namely, OMI (Ozone Monitoring Instrument) (Levelt et al., 2006) and TROPOMI (TROPOspheric Monitoring Instrument) (Veefkind et al., 2012), to estimate NO_2 concentrations at ground level over Vietnam.

TROPOMI, launched in October 2017, is a satellite instrument on board the Copernicus Sentinel-5 Precursor satellite (S5P). It measures air quality, ozone, ultraviolet radiation, and aids in climate forecasts with high spatial resolution. TROPOMI provides daily and global coverage of multiple trace gases (such as NO_2, CO, SO_2, CH_4, CH_2O, O_3) and aerosol properties. Prior to Sentinel-5P, NASA's OMI on the Aura satellite had been observing the ozone layer and atmospheric pollutant gases, including NO_2, since October 2004. However, the daily OMI NO_2 product has a lower spatial resolution (13 × 24 km) compared to the more detailed NO_2 product from TROPOMI (3.5 × 5.5 km).

Both of OMI and TROPOMI data were obtained from the Multi-Decadal Nitrogen Dioxide and Derived Products from Satellites (MINDS) program (Lamsal et al., 2022a; Lamsal et al., 2022b). The goal of this project is to adapt OMI operating algorithms to other satellite devices, and to create and store consistent multi-satellite Level 2 and Level 3 NO_2 products. They adapt their well-validated OMI NO_2, cloud, and geometry-dependent surface reflectivity retrieval algorithms to satellite instruments that include SCIAMACHY, GOME-2, TROPOMI. The adaptation of OMI algorithms for these satellite data aims to provide consistent and long-term records suitable for analyzing global trends in NO_2. OMI MINDS NO_2 and TROPOMI MINDS NO_2 were both downloaded from NASA's open source (https://disc.gsfc.nasa.gov/). The data are listed in Supplementary Table S1.

2.3 Meteorological data

Meteorological parameters are the factors that have an important influence on the concentration of NO_2 pollutant over time. For example, high temperature can accelerate photochemical reactions thereby reducing NO_2 concentration; high relative humidity increases the conversion rate from NOx to secondary aerosols thereby also reducing NO_2 concentrations. In this study, we utilized meteorological maps generated by the Weather Research and Forecasting (WRF) model, which employed input data from the fifth generation of ECMWF reanalysis (ERA-5) obtained from (https://cds.climate.copernicus.eu) during 2019–2021. The spatial resolution of the input data was 0.25 ° × 0.25 ° with hourly temporal resolution. The meteorological data of the ERA-5 was used as the initial and boundary conditions for the simulation in the WRF model. The WRF configuration was set up with two nested domains over Vietnam, with spatial resolution of 15 and 5 km respectively. The output data of the model was meteorological maps (including Temperature, Humidity, WindSpeed, Planetary Boundary Layer Height - PBLH) with a frequency of 4 images/day at 0, 6, 12, 18 h (GMT+0) and a spatial resolution of 5 × 5 km. The data are listed in Supplementary Table S1.

2.4 Land use data

Land use factors are closely associated with the sources of emissions. For instance, regions characterized by high traffic density tend to exhibit elevated smog emissions from vehicles, leading to higher concentrations of NO_2. Conversely, areas covered with vegetation generally experience lower pollution levels compared to urbanized areas. In this study, we utilized the following data: normalized difference vegetation index (NDVI) map, road map. The data are listed in Supplementary Table S1.

The NDVI product used in this study is generated from Terra MODIS satellite images through the MOD13Q1 product, Collection 6, level 3, which has a spatial resolution of 250 m and a temporal resolution of 16 days (Didan, 2015). NDVI maps provide spatially and temporally consistent observations of vegetation status in the study area. In this study, we collected MOD13Q1 product during 2021 from NASA open source (https://search.earthdata.nasa.gov/search).

The road map used in this study was obtained from the latest OpenStreetMap (OSM) data in 2022, available in vector format and comprising road shapes. OSM is a community-driven mapping service that is freely accessible and open to the public. OSM widely employed in various applications within the geosciences, earth observation, and environmental sciences. OSM offers global map objects, including data types such as nodes (representing points on Earth), ways (polyline representations of road objects, buildings, *etc.*), relations (establishing relationships between objects), and tags (containing object-related information) (Vargas-Munoz et al., 2021).

3 Methods

This study developed daily NO_2 maps using a method shown in Figure 1. The input data included NO_2 data from monitoring stations, NO_2 tropospheric column density from satellites, meteorological maps from the WRF model, NDVI maps, and road maps. These data were preprocessed and integrated to

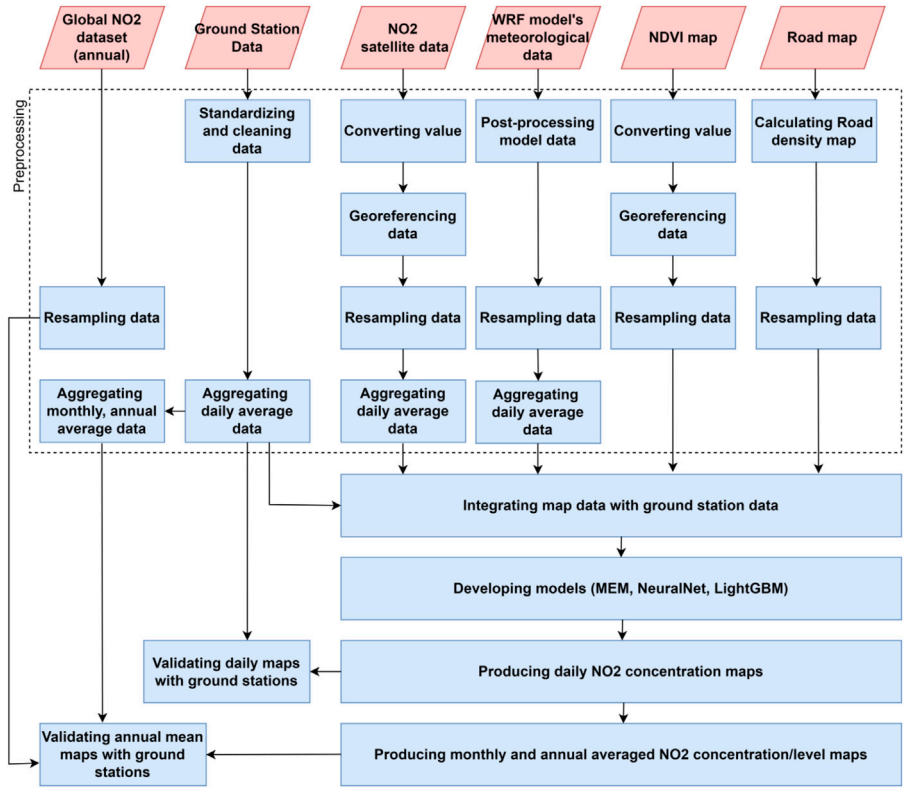


FIGURE 1
The methodological approach for estimating ground-level NO_2.

create a training dataset, which was used to develop statistical models for generating the daily NO_2 map. The daily NO_2 maps were then aggregated into monthly and annual averages, and validated using station observations and compared with the global NO_2 product.

3.1 Preprocessing data

The preprocessing of the monitoring station data, satellite images and ancillary data was similar to what we did for $PM_{2.5}$ pollutant published recently (Ngo et al., 2023). NO_2 concentration data from monitoring stations were standardized in uniform structure. After that, the data was cleaned and removed outliers. The process of removing outliers was carried out in the following steps: 1) Eliminating outliers by threshold. NO_2 observations with values exceeding 300 μg/m3 or less than 1 μg/m3 were discarded). 2) Using statistical methods to find outliers (too high/too low) compared to measured data in the neighboring period (±15 days). 3) Using the statistical method to find outliers (too high/too low) compared to the measured data in the neighboring period (±15 h), find out the outliers compared to the measured data measured at neighboring stations. 4) Finding outliers where the value does not change over a long period of time (Wu et al., 2018). These outliers were manually rechecked for accuracy. Subsequently, the hourly data were aggregated into daily, monthly, and annual averages for the purpose of data integration and modeling.

Multi-source satellite data, which are NO_2 tropospheric column density data from OMI, TROPOMI products and NDVI from the MOD13Q1 product, have different format, temporal and spatial resolutions. Preprocessing is required to convert satellite data into the same format and to project them in the same spatial grid. The preprocessing steps for the NO_2 and NDVI satellite images involve value extraction and transformation (converting value), geo-referencing, and resampling. Value extraction and transformation is the process of extracting related data layers and re-computing the values based on metadata information such as offset and scale factor of data. Geo-referencing means correlating the internal coordinate system of a map or an aerial image to a geographic coordinate system. In order to integrate multi-source data, a grid with uniform coverage and spatial resolution was defined. The grid covers the entire territory of Vietnam based on the WGS84 reference system and has cell size of 1 × 1 km. The satellite data were resampled and projected on this grid using the nearest resampling method for images with spatial resolution greater than 1 km (i.e., OMI, TROPOMI, meteorological maps) and the average resampling method for satellite images with resolution less than 1 km (i.e., MODIS NDVI, population density map). The GDAL tool was used to perform the above processes (GDAL, 2022). All the maps were then aggregated into daily maps for further calculation.

Quality flag bands were used to filter out low-quality pixels from the satellite products (OMI NO_2 and TROPOMI NO_2) to ensure the accuracy of the data. The bands used for filtering include "VcdQualityFlag" (even integer), "CloudFraction" (<0.3), and "qa_value" (>0.75), as recommended in previous studies (Lamsal et al., 2022a; Lamsal et al., 2022b). After quality control, the OMI and TROPOMI data were averaged on a daily basis to create a daily satellite combined dataset with a common grid (1 × 1 km grid).

The WRF model provides meteorological data in NetCDF format. The Unified Post Processing (UPP) Toolkit (NCEP UPP, 2022) was used to process the WRF model output data. UPP, which was developed at the National Center for Environmental Prediction (NCEP), has the capability of calculating various fields and interpolate them at different pressure levels from output data of the WRF model. We used the UPP tool to calculate temperature maps, humidity maps at 2 m height, planetary boundary layer height maps, wind speed at 10 m. Then, those data were resampled on the standard grid in order to be consistent with other satellite image products in the study area. These meteorological maps were then aggregated into daily mean maps for modeling.

NDVI, a MOD13Q1 product from Terra MODIS, was preprocessed similarly to those described for NO_2 maps, which were value extraction and transformation, geo-referencing, and resampling. The road map data was in vector format (shapefile), containing road lines and line characteristics. In order to use this feature as input of the model, the line density calculation was applied to convert the data into raster format (grids). It calculates a magnitude-per-unit area from polyline features, which fall within a radius around each cell (pixel). The radius is set approximately 1 km. The output image was then applied the nearest neighbor resampling method using the gdalwarp tool to get the same grid as the other maps.

3.2 Integrating data

Once the maps and station data were preprocessed, they were combined to create the training dataset. The aim was to establish the connection between the values on the maps and the observed NO_2 at the ground level. To ensure compliance with spatial and time constraints, the following measures were taken:

- Spatial constraint: The map data was extracted at the exact location of the ground station.
- Time constraint: The map data and ground-based NO_2 observations were synchronized by calculating the daily average values.

3.3 Modeling and validation

This study tested three different models: mixed effect model, neural network, and LightGBM. The MEM model has been widely used in the past to estimate pollution using satellite imagery and multi-source data. Recently, tree-based models have shown good results in estimating NO_2 maps. Therefore, in this study, we selected two machine learning models (MEM and LightGBM) to compare their performance. Additionally, we also wanted to experiment with a deep learning model. However, CNN-based models were not suitable for the current dataset, as complex deep learning models may not be suitable for sparse and limited data. Hence, we chose to experiment with a neural network model with multiple hidden layers and compared it with traditional machine learning models.

These models were fed with input parameters including NO_2 tropospheric column density (combined OMI and TROPOMI), meteorological data (humidity, PBLH), land cover (NDVI), and road density. Temperature and Wind Speed was not included in the input parameters due to its potential to create significant errors in estimating NO_2 concentrations in areas where ground monitoring stations are not installed in Vietnam. In other words, due to the uneven distribution of stations, the learned characteristics from the training dataset may not accurately reflect the patterns in areas without stations. For example, in mountainous regions with rocky terrain and dense forests (where there are no monitoring stations), the estimated pollution levels may appear higher than in flatland areas (with multiple monitoring stations, representing high emission areas).

The mixed effects model (MEM) is a type of land-use regression (LUR) model that consists of both fixed and random effect components. The formula for this model can be expressed as:

$$NO_{2\,i,j} = \sum_{k=1}^{N} \alpha_k X_{k,i,j} + (\alpha + \beta) \quad (1)$$

Where $NO_{2\,i,j}$ represents the estimated NO_2 concentration at spatial location j on day i. $X_{k,i,j}$ refers to the kth parameter at location j on day i, where N is the total number of parameters used in the model. The α_k, α coefficients denote the fixed effect component, which includes the slope and intercept of input parameters. The β coefficient represents the random effect of the intercept that varies from day to day.

LightGBM is a popular gradient boosting tree algorithm (Ke et al., 2017) used in machine learning. It utilizes a group of weak learners to improve the performance of the model. The regressor is optimized by adjusting hyper-parameters, such as the number of trees, the maximum tree depth, and learning rate, through the use of a grid search technique. The goal of this process is to improve the model's accuracy and reduce errors.

Neural network is a powerful method for modeling the complex and nonlinear relationships between inputs and outputs, which makes it suitable for studying atmospheric chemistry processes. It usually includes input, output, and hidden layers in its architecture (Nielsen, 2018). In this study, the neural network architecture was customized to fit the dataset size in terms of features and samples. During training and testing, the optimizer/learning rate, metric, and epochs were adjusted to optimize the performance of the model.

To assess the quality of the models, statistical indicators were used to compare the estimated NO_2 levels from the model with the actual NO_2 observations recorded at ground stations. The 10-fold CV method was employed to evaluate the performance of the model. After being trained and validated, the model was utilized to produce daily NO_2 concentration maps with a spatial resolution of 1 × 1 km. To evaluate its accuracy, the daily maps were compared with

TABLE 1 Models' evaluation results.

	Model	N	Pearson r	RMSE (μg/m^3)	MRE (%)
All data	Mixed Effect Model	9,027	0.66	9.39	54.01
	Neural Network	9,027	0.57	10.25	61.63
	LightGBM	9,027	0.87	6.28	34.65
10 Fold CV	Mixed Effect Model	903	0.56	10.46	59.29
	Neural Network	903	0.55	10.43	62.64
	LightGBM	903	0.77	7.93	42.6

ground station measurements using both temporal (daily mean) and spatial (pixel value extracted at station locations) constraints. In addition, to provide a more comprehensive analysis, daily ground measurements and NO_2 maps were aggregated into monthly and annual averages. The annual mean of our NO_2 maps was compared with the global NO_2 product (Anenberg et al., 2022) for the same study area, which provides annual average NO_2 datasets from 1990–2020 using a LUR method. The comparison involved evaluating the annual averages of our maps and the global product against ground station measurements of NO_2 taken in Vietnam from 2019–2021.

To compare and evaluate the models and maps, various statistical indicators were utilized, including the Pearson correlation coefficient (r), Root Mean Square Error (RMSE), and Mean Relative Error (MRE).

$$\text{Pearson r} = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - x_t)^2} \quad (3)$$

$$\text{MRE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_t - x_t|}{y_t}.100\% \quad (4)$$

Here, x_t, y_t represent the estimated values from the model (or extracted from the map) and the measured values at the ground station, respectively. $\bar{x}$ and $\bar{y}$ are the respective average values of the two data series.

4 Results and discussion

4.1 Model validation

Supplementary Table S2 presents the selected parameters for each model. For the MEM model, the model structure has been presented in Section 3.3 and no parameters need to be adjusted. With the NN network, due to the small input dataset size (9,027 samples and 5 features), we designed a small size neural network consisting of 1 input layer, 3 hidden layers including 16 nodes, 32 nodes, 16 nodes, respectively. Adam optimizer was selected with the learning rate of 0.001. The metric used was mean squared error (MSE) and the epochs was set to 200. With the LightGBM model, through the grid search technique, we selected a set of parameters for the model which presented in the Supplementary Table S2.

Table 1 shows the evaluation results after setting up and training the models. Among the experimental models, the LightGBM model achieved the best performance, with a Pearson correlation coefficient of 0.87, RMSE of 6.28 μg/m^3, and MRE of 34.65%. In contrast, the MEM and Neural Network models had poorer quality. The LightGBM model also demonstrated superior performance in the 10-fold CV, with a Pearson correlation coefficient of 0.77, RMSE of 7.9 μg/m^3, and MRE of 42.6%. Based on these results, we selected the LightGBM model to estimate the daily NO_2 maps for Vietnam from 2019–2021, which were then aggregated into monthly and annual average maps.

4.2 Map validation

A comparison was made between the daily NO_2 maps and ground station measurements during the period of 2019–2021. The scatter plot depicted in Supplementary Figure S2 supports the findings presented in Table 1 regarding the model evaluation. The daily maps had a high correlation with the ground station observations, with Pearson r at 0.87, RMSE at 6.28 μg/m^3, MRE at 34.65% based on 9,027 samples. However, the evaluation results varied by stations as presented in Supplementary Table S3. Pearson r varied from 0.27 to 0.88 with lower values at stations in Vung Tau, Long An (SE and MRD region) and higher values in Bac Ninh, Quang Ninh, Ha Noi (RRD region). The RMSE varied from 2.1 to 10.1 μg/m^3. The stations with low RMSE values were located across regions, while stations with high RMSE were mostly located in Ha Noi, Bac Ninh, Quang Ninh (RRD). Furthermore, some stations located in the same province had highly different evaluation results, such as Bac Ninh, Hai Duong, Quang Ninh (RRD) and Gia Lai (Central Highland), indicating the need for further investigation.

Annual average NO_2 maps were created by aggregating daily NO_2 data from 2019 to 2021, as illustrated in Figure 2. The maps reveal that NO_2 was predominantly concentrated in the Red River Delta region in the North, along the North Central Coast, and in the Ho Chi Minh city area in the South. These regions are critical economic centers of Vietnam with high population density, heavy traffic, numerous industrial parks, and factories that generate significant NO_2 emissions. Across the country, the annual average NO_2 concentration varied from 4.4 to 36 μg/m^3 in 2019, 4.2 to 32.8 μg/m^3 in 2019 and 5.3 to 40.1 μg/m^3 in 2021. Notably, the national average concentration remained relatively stable between 2019 and 2021, indicating a persistent NO_2 pollution problem in Vietnam. Despite the implementation of social distancing measures

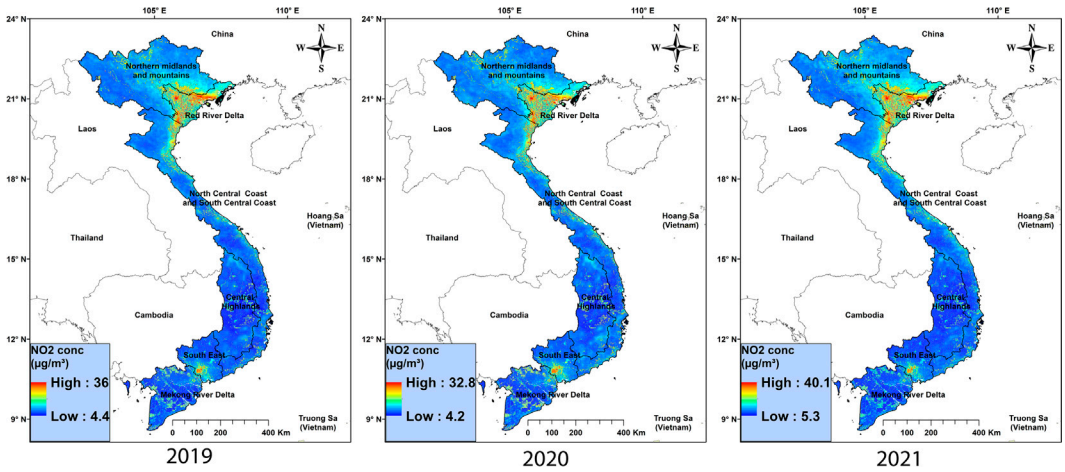


FIGURE 2
Annual mean ground-level NO_2 maps from 2019 to 2021.

TABLE 2 Comparison of validation results for ours and the global annual mean maps to ground station values.

Time	Study	N	Pearson *r*	RMSE (μg/m³)	MRE (%)
2019–2020	(Anenberg et al., 2022)	33	0.27	13.3	57.4
2019–2020	This study	33	0.95	2.1	8.6
2019–2021	This study	81	0.95	2.27	9.79

in response to the COVID-19 pandemic in 2019 and 2020 in Vietnam, there was not a significant variation in the annual mean NO_2 levels measured at stations. This lack of variation resulted in no significant changes in the annual NO_2 maps over the years (see Table 2; Supplementary Figure S3).

In Supplementary Figure S3, a detailed comparison is presented between the annual average NO_2 concentration maps for the years 2019, 2020, and 2021, and the ground stations located in Vietnam. It is noteworthy that the number of stations used for annual map assessment is less than that used for daily map assessment. This is because, when aggregating daily data into an annual average, any station that did not have more than 50% of the data for the year was discarded and not used for evaluation. Furthermore, in 2019, only three stations were evaluated, whereas this number increased to 30 in 2020 and to 48 in 2021. The difference between the annual maps and the ground stations varied from −2.5 μg/m³ (Quang Ninh - RRD) to 0.6 μg/m³ (Ha Noi - RRD) in 2019; −3.96 μg/m³ (Bac Ninh - RRD) to 4.6 μg/m³ (Ha Noi - RRD) in 2020; −4.12 μg/m³ (Ha Noi–RRD) to 8.13 μg/m³ (Bac Ninh–RRD) in 2021.

To ensure a thorough assessment, we compared the quality of our annual maps from 2019 to 2021, not only against ground stations, but also against the annual global product (2019–2020) developed by Anenberg et al. (2022). Table 2 displays the findings. Our annual maps showed markedly superior quality in comparison to both the global annual maps and the ground stations. Specifically, we achieved a Pearson correlation coefficient of 0.95, an RMSE of 2.1 μg/m³, and an MRE of 8.6%, while the global annual maps achieved only a Pearson r of 0.27, an RMSE of 13.3 μg/m³, and an MRE of 57.4%. Additionally, our map from 2019 to 2021 had a Pearson *r* of 0.95, an RMSE of 2.27 μg/m³, an MRE of 9.79%, and 81 samples, indicating the high quality of the annual NO_2 maps in this study and the potential of this approach to develop NO_2 maps from multi-satellite images over Vietnam.

5 Conclusion

In this study, daily NO_2 maps at 1 × 1 km over Vietnam were created using OMI and TROPOMI satellite images as well as auxiliary data from 2019–2021. Three models were experimented, including MEM, NN, and LightGBM, with LightGBM proving to have the best quality (Pearson *r* of 0.87, RMSE of 6.28 μg/m³, MRE of 34.65%). The LightGBM model was used to generate the daily NO_2 maps, which were validated against ground stations and found to be accurate. However, the quality of the maps varied by station, with Pearson r ranging from 0.27 to 9.88 and RMSE ranging from 2.1 to 10.1 μg/m³ between 2019–2021. The daily maps were then combined to produce monthly and yearly average maps. Our annual average map was compared to a global product and ground stations, and it was found to have superior quality with Pearson r of 0.95, RMSE of 2.27 μg/m³, MRE of 9.79%, and 81 samples. This is the first study on constructing NO_2 concentration maps in Vietnam using multi-source satellite data. The study encountered challenges such as uneven distribution of monitoring stations in the research area and limitations posed by cloud coverage on NO_2 satellite data (OMI, TROPOMI). Further exploration of these issues is needed in future research to enhance the quality of the maps.

Data availability statement

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

Author contributions

TXN: methodology, software, validation, formal analysis, writing—original draft, and visualization. HP: writing—review and editing. TTNN: conceptualization, methodology, validation, writing—original draft, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2019.331.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2023.1187592/full#supplementary-material

References

Anenberg, S. C., Mohegh, M., Goldberg, L., Kerr, H. K., Brauer, B., Burkart, K., et al. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: Estimates from global datasets. *Lancet Planet. Health* 6 (1), e49–e58. doi:10.1016/S2542-5196(21)00255-2

Didan, K. (2015). "MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006," in *Distributed by NASA EOSDIS land processes DAAC* (United States: United States Geological Survey). doi:10.5067/MODIS/MOD13Q1.006

GDAL (2022). GDAL documentation 2022. Available at: https://gdal.org/programs/gdalwarp.html.

Grzybowski, P. T., Markowicz, M., and Musiał, J. P. (2023). Estimations of the ground-level NO2 concentrations based on the sentinel-5P NO2 tropospheric column number density product. *Remote Sens.* 15 (2), 378. doi:10.3390/rs15020378

Kang, Y., Choi, H., Im, I., Park, S., Shin, M., Song, C-K., et al. (2021). Estimation of surface-level NO2 and O3 concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* 288, 117711. doi:10.1016/j.envpol.2021.117711

Ke, G., Qi, M., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). "LightGBM: A highly efficient gradient boosting decision tree." in *Advances in neural information processing systems*. Long Beach, CA: Curran Associates, Inc.

Lamsal, L. N., Krotkov, N. A., Marchenko, S. V., Joiner, J., Oman, L., Alexander, V., et al. (2022a). *OMI/Aura NO2 tropospheric, stratospheric and total columns MINDS 1-orbit L2 swath 13 Km x 24 km*. Greenbelt, Maryland: Goddard Earth Sciences Data and Information Services Center GES DISC. NASA Goddard Space Flight Center.

Lamsal, L. N., Krotkov, N. A., and Marchenko, S. V. (2022b). *TROPOMI/S5P NO2 tropospheric, stratospheric and total columns MINDS 1-orbit L2 swath 5.5 Km x 3.5 km*. Greenbelt, Maryland: Goddard Earth Sciences Data and Information Services Center. NASA Goddard Space Flight Center.

Larkin, A., Geddes, J. A., Martin, R. V., Xiao, Q., Liu, Y., Marshall, J. D., et al. (2017). Global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* 51 (12), 6957–6964. doi:10.1021/acs.est.7b01148

Lee, H. J., and Koutrakis, P. (2014). Daily ambient NO2Concentration predictions using satellite ozone monitoring instrument NO2Data and land use regression. *Environ. Sci. Technol.* 48 (4), 140204134232009–140204134232011. doi:10.1021/es404845f

Levelt, P. F., van den OordVan Den Oord, H. J., Dobber, M. R., Malkki, A., Huib Visser, H., Johan de Vries, J., et al. (2006). The ozone monitoring instrument. *IEEE Trans. Geosci. Remote Sens.* 44 (5), 1093–1101. doi:10.1109/TGRS.2006.872333

Long, S., Wei, X., Zhang, F., Zhang, R., Xu, J., Wu, K., et al. (2022). Estimating daily ground-level NO2 concentrations over China based on TROPOMI observations and machine learning approach. *Atmos. Environ.* 289, 119310. doi:10.1016/j.atmosenv.2022.119310

NCEP UPP (2022). NCEP unified Post processing system (UPP). Available at: https://dtcenter.org/community-code/unified-post-processor-upp.

Ngo, T. X., Pham, H. V., Phan, H. D. T., Nguyen, A. T. N., To, H. T., and Nguyen, T. T. N. (2023). A daily and complete PM2.5 dataset derived from Space observations for Vietnam from 2012 to 2020. *Sci. Total Environ.* 857, 159537. doi:10.1016/j.scitotenv.2022.159537

Nguyen, T., Bui, H. Q., Pham, H. V., Luu, H. V., Man, C. D., Pham, H. N., et al. (2015). Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study. *Environ. Res. Lett.* 10 (9), 095016. doi:10.1088/1748-9326/10/9/095016

Nielsen, M. A. (2018). *Neural networks and deep learning*. Oxford: Determination Press.

Paraschiv, S., Constantin, D. E., Paraschiv, S. L., and Constantin, M. (2017). OMI and ground-based *in-situ* tropospheric nitrogen dioxide observations over several important European cities during 2005–2014. *Int. J. Environ. Res. Public Health* 14 (11). doi:10.3390/ijerph14111415

Vargas-Munoz, J. E., Srivastava, S., Tuia, D., and Falcao, A. X. (2021). OpenStreetMap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience Remote Sens. Mag.* 9 (1), 184–199. doi:10.1109/MGRS.2020.2994107

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., et al. (2012). TROPOMI on the esa sentinel-5 precursor: A gmes mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* 120, 70–83. doi:10.1016/j.rse.2011.09.027

Wang, W., Liu, X., Bi, J., and Liu, Y. (2022). A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environ. Int.* 158, 106917. doi:10.1016/j.envint.2021.106917

Wu, H., Tang, X., Wang, Z., Wu, L., Lu, M., Wei, L., et al. (2018). Probabilistic automatic outlier detection for surface air quality measurements from the China national environmental monitoring network. *Adv. Atmos. Sci.* 35 (12), 1522–1532. doi:10.1007/s00376-018-8067-9

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 31 August 2023
DOI 10.3389/fenvs.2023.1240106

OPEN ACCESS

EDITED BY
Prof-Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Calogero Schillaci,
Joint Research Centre, Italy
Thang Ha,
Hue University, Vietnam
Shamsollah Ayoubi,
Isfahan University of Technology, Iran

*CORRESPONDENCE
Tiago G. Morais,
tiago.g.morais@tecnico.ulisboa.pt

RECEIVED 14 June 2023
ACCEPTED 18 August 2023
PUBLISHED 31 August 2023

CITATION
Morais TG, Jongen M, Tufik C, Rodrigues NR, Gama I, Serrano J, Gonçalves MC, Mano R, Domingos T and Teixeira RFM (2023), Satellite-based estimation of soil organic carbon in Portuguese grasslands.
Front. Environ. Sci. 11:1240106.
doi: 10.3389/fenvs.2023.1240106

Satellite-based estimation of soil organic carbon in Portuguese grasslands

Tiago G. Morais[1]*, Marjan Jongen[1], Camila Tufik[2], Nuno R. Rodrigues[3], Ivo Gama[3], João Serrano[4], Maria C. Gonçalves[5], Raquel Mano[6], Tiago Domingos[1] and Ricardo F. M. Teixeira[1]

[1]MARETEC—Marine, Environment and Technology Centre, LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, [2]Centro de Investigação em Agronomia, Alimentos, Ambiente e Paisagem (LEAF), Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon, Portugal, [3]Terraprima—Serviços Ambientais, Sociedade Unipessoal, Samora Correia, Portugal, [4]Mediterranean Institute for Agriculture, Environment and Development (MED), Universidade de Évora, Évora, Portugal, [5]Instituto Nacional de Investigação Agrária e Veterinária (INIAV), Laboratório de Solos, Oeiras, Portugal, [6]Instituto Nacional de Investigação Agrária e Veterinária (INIAV), Laboratório Químico Agrícola Rebelo da Silva (LQARS), Lisbon, Portugal

Introduction: Soil organic carbon (SOC) sequestration is one of the main ecosystem services provided by well-managed grasslands. In the Mediterranean region, sown biodiverse pastures (SBP) rich in legumes are a nature-based, innovative, and economically competitive livestock production system. As a co-benefit of increased yield, they also contribute to carbon sequestration through SOC accumulation. However, SOC monitoring in SBP require time-consuming and costly field work.

Methods: In this study, we propose an expedited and cost-effective indirect method to estimate SOC content. In this study, we developed models for estimating SOC concentration by combining remote sensing (RS) and machine learning (ML) approaches. We used field-measured data collected from nine different farms during four production years (between 2017 and 2021). We utilized RS data from both Sentinel-1 and Sentinel-2, including reflectance bands and vegetation indices. We also used other covariates such as climatic, soil, and terrain variables, for a total of 49 inputs. To reduce multicollinearity problems between the different variables, we performed feature selection using the sequential feature selection approach. We then estimated SOC content using both the complete dataset and the selected features. Multiple ML methods were tested and compared, including multiple linear regression (MLR), random forests (RF), extreme gradient boosting (XGB), and artificial neural networks (ANN). We used a random cross-validation approach (with 10 folds). To find the hyperparameters that led to the best performance, we used a Bayesian optimization approach.

Results: Results showed that the XGB method led to higher estimation accuracy than the other methods, and the estimation performance was not significantly influenced by the feature selection approach. For XGB, the average root mean square error (RMSE), measured on the test set among all folds, was 2.78 g kg^{-1} (r^2 equal to 0.68) without feature selection, and 2.77 g kg^{-1} (r^2 equal to 0.68) with feature selection (average SOC content is 13 g kg^{-1}). The models were applied to obtain SOC content maps for all farms.

Discussion: This work demonstrated that combining RS and ML can help obtain quick estimations of SOC content to assist with SBP management.

KEYWORDS

remote sensing, satellite, cross-validation, features selection, sown biodiverse pasture

1 Introduction

Soil systems are intricate networks of both organic and inorganic matter with varying chemical and physical attributes that can differ from site to site, or even within the same site. These systems also serve as the primary carbon reservoirs on land, with a capacity to store roughly 80% of all organic carbon, totalling an estimated 2,400 Pg of carbon (PgC)—more than three times the amount found in the atmosphere (Jobbágy and Jackson, 2000; Chappell et al., 2016). The level of soil organic carbon (SOC) present is heavily influenced by soil management practices, soil properties, and climatic conditions, with significant spatial differences that pose a challenge when estimating terrestrial carbon stocks and fluxes (Giardina et al., 2014; Doetterl et al., 2015; Koven et al., 2017). In terms of preserving SOC and other essential ecosystem services, grasslands rank among the most significant terrestrial ecosystems (Egoh et al., 2016; Bardgett et al., 2021). However, SOC estimation in grassland ecosystems is challenging due to factors such as the high spatial and temporal variability of SOC, heterogeneous distribution within soil profiles and the fact that methods for SOC estimation are often destructive and time-consuming (Angelopoulou et al., 2019; Xiao et al., 2019). Remote sensing (RS) and machine learning (ML) models have the potential to improve the accuracy and certainty of SOC estimation in grassland ecosystems.

RS data is often used in providing explanatory variables for estimating SOC using ML methods (Angelopoulou et al., 2019), especially as spectral sensors have improved significantly in recent decades, with enhanced spatial and temporal resolutions. Consequently, RS data from satellites (such as Landsat 7/8 and Sentinel-2) and unmanned aerial vehicles (UAVs) have led to a rise in applications for monitoring SOC in croplands and grasslands (Zheng et al., 2004; Mariano et al., 2018; Sun et al., 2021). Vegetation indices, have been widely used to estimate SOC (Xu et al., 2008; Ullah et al., 2012; Davids et al., 2018), but there are limitations and uncertainties associated with their use (Zhao et al., 2014; Ali et al., 2016). More recently, individual spectral bands, sometimes in combination with VIs, have been used to indirectly estimate SOC (Wang et al., 2021; Zepp et al., 2021; Pan et al., 2022). RS data is often combined with other covariates such as terrain and climatic variables to improve the estimation (Mallik et al., 2020; Gardin et al., 2021; Wang et al., 2022).

In recent years, there has been an increased interest in using ML methods for estimating SOC or soil organic matter (SOM) (Pezzuolo et al., 2017; Angelopoulou et al., 2019; Odebiri et al., 2021; Biney, 2022; Chan et al., 2023). ML methods are automated techniques that look for hypotheses to explain data and can be applied to any learning task. Commonly used models to estimate SOC/SOM include random forests (RF) and artificial neural networks (ANNs) (Lamichhane et al., 2019). These models have demonstrated their capacity to enhance SOC estimation by reducing the error between the ground-measured SOC/SOM values and the estimates generated by the models (e.g., Ladoni et al., 2010; Pouladi et al., 2019; Zepp et al., 2021; Wang et al., 2022). Further, some ML methods such as RF have also demonstrated higher performance in estimating SOC than geospatial models (Veronesi and Schillaci, 2019). Estimations of SOC/SOM content at high spatial resolutions (<50 m) have significantly improved in the past decades (Angelopoulou et al., 2019). While ML methods are predominantly associated with the use of satellite data, there has been a limited number of studies exploring other remote sensing sources with higher spatial resolution, such as UAVs (Angelopoulou et al., 2019). Satellite data sources remain the most commonly used as they offer advantages such as short revisit times and medium spatial resolution (Xiao et al., 2019). However, most applications developed to estimate SOC/SOM content are still specific to the particular land cover systems in which they were trained and validated. For highly specific land use systems that can be a problem, as existing models were never trained with system-specific data.

Sown biodiverse permanent pastures rich in legumes (SBP) are one example of such unique grassland/pasture systems. SBP have been implemented since the 1960 s in Portugal to boost pasture yields and increase animal stocking rates (Teixeira et al., 2015; Morais et al., 2022). This system involves sowing a combination of up to 20 legume and grass species or cultivars that provide high-quality animal feed. In addition to the direct benefits of this system, such as increased forage production, a major co-benefit is soil carbon sequestration, as noted by Moreno et al. (2021) and Teixeira et al. (2011). To assist with compliance to the Kyoto Protocol goals under the Agriculture, Forestry and Other Land Uses activities, the Portuguese Carbon Fund provided support for the installation and maintenance of SBP between 2009 and 2014. Payments were made to over 1,000 farmers based on predetermined sequestration factors that were established from data gathered during previous studies, rather than on carbon content increases that were measured on the farm (Teixeira et al., 2011; APA, 2018). Thus, there is a lack of indirect methods that can be broadly applied and are specifically tailored to SBP systems, hindering effective carbon management of this unique pasture system.

In the present research, we employed a combination of RS data and various ML techniques to estimate SOC content at a depth of 20 cm in SBP. We collected data from Sentinel-1 and Sentinel-2 satellites during two periods, August and the closest date to soil sampling. Five VIs were extracted from the RS data, along with various climatic, soil, terrain, and other auxiliary variables. Two variable selection methods were used, one utilizing all variables and the other using the sequential feature selection (SFS) approach to measure multicollinearity among input variables and select the most relevant ones for the SOC estimation. We evaluated the performance of the models using a random cross-validation approach with 10 folds. The resulting models were then used to estimate SOC and generate SOC content maps for the sampled farms' entire sites.

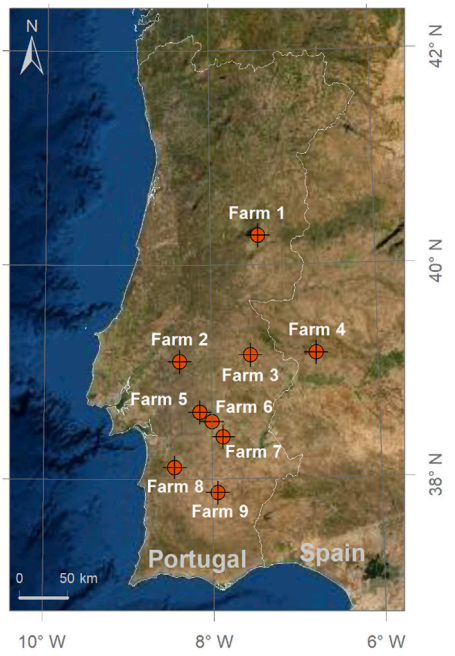


FIGURE 1
Location of the nine sampled farms used in this work. Farm 4 is the only one in Spain, all other farms being in Portugal.

2 Material and methods

2.1 Study area and soil sampling design

Data from nine different farms were used in this work: eight farms in Portugal (Farms 1, 2, 3, 5, 6, 7, 8, and 9) and one in Spain (Farm 4). They are located across latitudes and longitudes ranging respectively between 37°50′ and 40°30′N and 6°80′ and 8°30′W (Figure 1). The size of surveyed farms ranges between 26 ha (Farm 8) and 42 ha (Farm 6). All farms are in the hot-summer Mediterranean climate region, according to the Köppen climate classification system (Rubel and Kottek, 2010; IPMA, 2018).

According to the European Soil Database (ESDAC, 2003), the nine sampled farms are characterized by five different soil types: Dystric Cambisol (Farms 1 and 4), Orthic Podzol (Farms 2, 3, and 5), Eutric Cambisol (Farms 6 and 8), Rhodo-Chromic Luvisol (Farm 7) and Ferric Luvisol (Farm 9). Regarding dominant parent material, there are six different types: granite (Farms 1 and 6), diorite (Farms 3 and 5), acid regional metamorphic rocks (Farms 7 and 9), river terrace sand or gravel (Farm 2), (meta-) shale/argillite (Farm 4) and sandstone (Farm 8).

In total, four production years were covered in this study (between 2017-18 and 2020-21). The number of production years covered and the number of samples per production year vary between farms. For example, Farm 1 was sampled in all four production years, but Farm 9 was only sampled in one production year (2018-19). Additionally, considering only Farm 1, in the first year, 40 plots/locations were sampled, but in the following years, more samples were collected, with 2018-19 having the highest number of samples (75 samples). The total number of collected samples and collection years are summarized in Table 1. In each farm, the selection of sampling locations was carefully made to minimize any potential influence of trees and rocks on the measured SOC content. Due to the significantly different tree densities across the sampling locations, achieving an equal number of sampling locations per farm was not feasible.

Soil sampling took place in the period between September and May. They were collected using two different methods: 1) manual collection and 2) mechanical collection. This was expressed in the analysis as an auxiliary binary variable. In both collection methods, samples were collected in the 0–20 cm topsoil layer, which is the reference depth in the LUCAS Soil project conducted by the European Soil Data Centre (ESDAC)—Joint Research Centre (JRC) (Orgiazzi et al., 2018). Manual collection used an auger (2 cm diameter), while mechanical collection used a Wintex 2000 soil sampler installed on a utility terrain vehicle. Each soil sample was composed of four sub-samples that were pooled and mixed to achieve uniformity. All soil samples were air-dried and passed through a 2 mm stainless steel sieve. SOC content was calculated using the soil fractions after an elemental analysis performed after a combustion at 1050°C. In all soil samples, inorganic carbon removal was performed prior to the total SOC quantification. All values of SOC presented here are expressed in grams of SOC per kg of dry soil.

2.2 Data collection and preprocessing

In this study, we used RS data, climate, terrain, and soil data to model SOC content. All data was obtained from Google Earth Engine (GEE), which reduced data processing time and storage space. GEE is a cloud-based platform that allows users to access and process massive amounts of geospatial data. The platform includes a catalogue of over 600 petabytes of satellite imagery, aerial imagery, and other geospatial datasets. GEE enables users to analyse data to track changes over time, map trends, and quantify differences on the Earth's surface. For example, the complete Sentinel-2 database is available. Table 3 summarizes all the data used, including their sources, variable names, and spatial resolution. In total, 49 input variables were considered.

For all data used, we applied "min-max" normalization (i.e., values were normalized between 0 and 1). Each input was subjected to individual and independent data normalization, without any dependence on the other inputs. This was done to increase the learning rate and ensure faster convergence as models with large weights tend to be unstable and suffer from poor performance during learning and sensitivity to input values, the latter resulting in higher generalization error (Bishop, 1995; Goodfellow et al., 2016).

In order to understand the relationship between the data used and the measured SOC content, we calculated a Spearman's rank correlation (Spearman, 1904). This is a non-parametric measure of monotonic statistical dependence between two variables, and it does not make any assumptions about the distribution of the variables.

2.2.1 Remotely sensed data collection

The RS data were obtained from the Sentinel-1 and Sentinel-2 missions. We used the Sentinel-1 C-band Level-1 Ground Range Detected images provided by GEE, which were acquired on a

10.3389/fenvs.2023.1240106

TABLE 1 Description of the collected soil samples per farm and production year.

Farm	Year	Number of collection periods	Number of samples	Sample collection method
Farm 1 (28 ha)	2017-18	3	40	Manual—40
	2018-19	5	75	Manual—28
				Mechanical - 47
	2019-20	3	58	Manual—24
				Mechanical—34
	2020-21	3	64	Manual—22
				Mechanical—42
Farm 2 (27 ha)	2019-20	1	35	Mechanical—35
Farm 3 (29 ha)	2017-18	2	32	Manual—32
	2018-19	4	71	Manual—24
				Mechanical—47
	2019-20	3	57	Manual—24
				Mechanical—33
	2020-21	2	43	Manual—12
				Mechanical—31
Farm 4 (34 ha)	2018-19	2	24	Manual—24
Farm 5 (34 ha)	2018-19	4	74	Manual—24
				Mechanical—50
	2019-20	3	58	Manual—24
				Mechanical—34
	2020-21	3	52	Manual—24
				Mechanical—28
Farm 6 (42 ha)	2017-18	3	39	Manual—39
	2018-19	3	72	Manual—15
				Mechanical—57
	2019-20	3	57	Manual—24
				Mechanical—33
	2020-21	3	51	Manual—24
				Mechanical—27
Farm 7 (35 ha)	2018-19	1	12	Manual—12
	2019-20	1	33	Mechanical—33
	2020-21	1	30	Mechanical—30
Farm 8 (26 ha)	2018-19	3	28	Manual—28
	2019-20	3	51	Manual—22
				Mechanical—29
	2020-21	3	53	Manual—24
				Mechanical—29
Farm 9 (30 ha)	2018-19	1	12	Manual—12

(Continued on following page)

TABLE 1 (*Continued*) Description of the collected soil samples per farm and production year.

Farm	Year	Number of collection periods	Number of samples	Sample collection method
Total		**63**	**1,121**	**Manual—502**
				Mechanical—619

In Bold are the sum of the lines per column.

TABLE 2 Calculation formula for the vegetation indices used in this paper. NDVI, normalized difference vegetation index; NDWI, normalized difference water index; SR, simple ratio; SAVI, soil-adjusted vegetation index; OSAVI, optimized soil-adjusted vegetation index.

Vegetation indices	Formula
NDVI	$\frac{NIR-Red}{NIR+Red}$
NDWI	$\frac{Green-NIR}{Green+NIR}$
SR	$\frac{NIR}{Red}$
SAVI	$1.5\frac{NIR-Red}{NIR+Red+0.5}$
OSAVI	$1.16\frac{NIR-Red}{NIR+Red+0.16}$

descending orbit in Interferometric Wide swath mode (IW). The imagery in GEE consists of Level-1 Ground Range Detected (GRD). We utilized the VV and VH polarization bands, and the intensity cross-ratio (CR) VV/VH was also calculated. Sentinel-2 is a two-satellite constellation mission (Sentinel-2A and Sentinel-2B), which carries a wide-swath multispectral imager with 13 spectral bands. The image resolutions are 10 m (Blue, Green, Red, and Near Infrared bands), 20 m (three Vegetation Red Edge bands, Narrow NIR band, and two shortwave-infrared bands), and 60 m (Coastal aerosol, Water vapour, and SWIR-Cirrus bands). We used Level-2A data products, i.e., bottom of atmosphere (BOA) reflectance images obtained from Level-1C products. Bands 1 (coastal aerosol), 9 (water vapour), and 10 (SWIR-Cirrus) were excluded as they are specific to atmospheric characterization and not land surface monitoring. Besides the individual bands, we used spectral data to calculate five vegetation indices (Table 2): the normalized difference vegetation index (NDVI) (Tucker, 1979), normalized difference water index (NDWI) (Gao, 1996), simple ratio (SR), soil-adjusted vegetation index (SAVI) (Huete, 1988) and optimized soil-adjusted vegetation index (OSAVI) (Rondeaux et al., 1996).

Regarding the Sentinel-1 and Sentinel-2 data, for each band or vegetation index, we considered data from two periods. First, we considered a composite image of the available images for the period between August 1st and August 31st. This composite image aims to capture the spectral reflectance of the bare soil. Second, we also considered data from Sentinel-1 and Sentinel-2 from the closest date to the soil collection date. This aims to capture the inter-yearly variation of SOC between the period when the soil was bare and the collection date, when the soil was covered by vegetation.

For the period when the soil is almost bare in the SBP system, i.e., during August, we considered a composite image of the available Sentinel-1 and Sentinel-2 images for the period between 1st August and 31st August. The composite image in August captures the spectral reflectance of the bare soil, and the image closest to the soil collection period captures the influence of vegetation on SOC. We also removed pixels masked as clouds and cloud shadow using the "pixel_qa" band from Sentinel-2 data obtained from GEE. Additionally, we also used the available image closest to each soil collection period. All the individual bands and the vegetation indices were calculated and downloaded using GEE.

2.2.2 Climate, soil and terrain data collection

The mineralization and accumulation of SOC are highly dependent on climate, specifically soil temperature and moisture (Rey et al., 2005; Thornton et al., 2009). Therefore, we used data from the Global Land Data Assimilation System (GLDAS—Rodell et al., 2004) for these variables. The data available in GLDAS is on a daily basis and we used both soil temperature and moisture on the collection date. We also included soil data to characterize SOC, such as clay, sand, silt content and soil pH (H_2O). Soil data was obtained from SoilGrids (Hengl et al., 2017). SOC is also influenced by terrain characteristics (Rogge et al., 2018) and thus we used data from NASA EOSDIS Land Processes DAAC (NASA, 2020) and Theobald et al. (2015) for the Digital Elevation Model (DEM), the Continuous Heat-Insolation Load Index (CHILI), the Multi-Scale Topographic Position Index (mTPI) and Topographic Diversity (topoDivers). CHILI captures the effects of insolation and topographic shading on evapotranspiration (calculated by the insolation at early afternoon, sun altitude equivalent to the equinox). mTPI distinguishes ridge from valley forms (calculated by the elevation at each location subtracted by the mean elevation within a neighborhood). Finally, topoDivers represents the variety of temperature and moisture conditions available to species as local habitats (calculated by mTPI and soil moisture). All data was calculated and downloaded using GEE.

2.2.3 Auxiliary data

We also considered six additional auxiliary variables: the number of days since the beginning of the production year (counting from 31st August), the number of days between the closest Sentinel-2 image and the soil sampling date, the number of days between the closest Sentinel-1 image and the soil sampling, the collection method (manual or mechanical) the year, and the month.

2.3 Modelling and mapping soil organic carbon

2.3.1 Feature selection

In this study, we used a long list of independent variables (49 inputs) to estimate SOC content. However, in practice not all of those variables might be relevant for estimating SOC. To address this, we used a two-step approach: 1) first, all input variables were included in the estimation of SOC, then 2) we applied SFS and retrained the algorithm with a subset of

variables. The SFS approach involves adding features in an automated and iterative manner to form a feature subset. At each iteration, the best feature to add or remove is chosen based on the cross-validation score of the model validation procedure. Then, after applying SFS, we obtained a subset of the input data that has the most relevant variables for estimating SOC. This method allowed us to identify and select only the pertinent variables that are crucial for accurately estimating SOC content within the dataset.

2.3.2 Regression methods

The SOC content was modelled using four regression methods: multiple linear regression (MLR—Barbur et al., 1994), random forest (RF—Breiman, 2001), extreme gradient boosting (XGBoost- XGB—Chen and Guestrin, 2016) and artificial neural network (ANN—Rumelhart et al., 1986). To optimize the regression models, we used Bayesian optimization with 100 initializations to find the best hyperparameters for each method. The methods and their respective hyperparameter option spaces are described in detail in the next section. All methods were implemented on Python 3.8.4, using multiple toolboxes. For MLR regression and RF, we used the scikit-learn 0.24 toolbox (https://github.com/scikit-learn/scikit-learn). For XGB, we used the xgboost 1.4.2 toolbox (https://github.com/dmlc/xgboost). For ANN, keras 2.9 was used to construct the ANN architecture and TensorFlow 2.7 as the backend for keras (https://github.com/keras-team/keras; https://github.com/tensorflow/tensorflow). To prepare the data, we used Numpy 1.18.5 (https://github.com/numpy/numpy) and Pandas 1.0.4 (https://github.com/pandas-dev/pandas). The Bayesian optimization was performed using the scikit-optimizer 0.8.1 (https://github.com/scikit-optimize/scikit-optimize).

MLR was the simplest method used in this study. It fits a linear equation to the observed data using the relationship between all independent variables and a dependent variable, using a least squares fit. Decision trees/forests, such as RF, is a learning method that creates multiple decision trees and fits the trees to training data. In a RF, the value of the response variable can change across the trees in the forest. However, within each individual tree, the predicted variable does not change in each leaf. This is because each tree is built using the same set of predictor variables and the same splitting criteria, resulting in consistent splits at each node of the tree. One advantage of RF over other bagging models is its ability to produce nearly uncorrelated predictions due to the random features, producing predictions with low variance. For optimization, we tested various options involving the number of estimators, the minimum number of samples per leaf, the maximum depth, the error function, the maximum number of features/inputs in each split, and the use of a bootstrap approach.

XGB is a newer method, proposed in 2016, that is based on gradient boosting tree methods. It trains by making predictions sequentially and combining weak predictive tree models, learning from the obtained errors. XGB has significant improvements to traditional gradient boost methods, namely, in terms of performance, parallelization, distributed computing, and computational time. For optimization, various options such as the number of estimators, the learning rate, the maximum depth of the trees, and L1 and L2 regularization were considered.

An artificial neural network (ANN) is a multi-layer network structure that consists of an input layer with a set of input/explanatory variables, an output layer containing the dependent/objective variable, and one or more hidden layers with nodes or artificial neurons. Each hidden layer receives a signal, processes it through a transfer function, and passes the processed signal to neurons connected to it in the following layer. In order to optimize the hyperparameters of the ANN, we considered one or two hidden layers, the number of neurons in each hidden layer (between 50 and 10,000 with intervals of 50), the learning rate (between 0.01 and 1 with intervals of 0.015), and the activation function (which can be "elu," "relu" or "sigmoid").

2.3.3 Validation approach and accuracy assessment

We used a random cross-validation (CV) method, considering 10 folds, in order to have an appropriate measure of the estimation error. The dataset was split into 10 approximately equal portions. In each fold, a different portion of the data set was used to train the models (i.e., 9/10 of total samples) and the remaining 1 part (hold-out samples) was used as the test set. The performance of each model was measured in the hold-out samples in each fold. This procedure was applied similarly to all regression models used.

The performance of the obtained models was assessed in the test sets of the k-fold approach using four metrics: the root mean squared error (RMSE), the relative RMSE (rRMSE), the ratio of performance to deviation (RPD) and the coefficient of determination (r^2). The mathematical formula of the metrics are

$$RMSE = \sqrt{\frac{1}{n}\sum_{n=1}^{N}(c-\hat{y}_i)^2}$$

$$rRMSE = \frac{RMSE}{\bar{y}}$$

$$RDP = \frac{\sigma}{RMSE}$$

$$r^2 = 1 - \frac{\sum_{n=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{n=1}^{N}(y_i-\bar{y})^2}$$

where n is the number of observations, y_i is the observed value, and $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the observed values and σ is the standard deviation of the observed values.

3 Results

3.1 Analysis of measured soil organic carbon

For the farms with data available for more than 1 year, there was a tendency for the observed SOC content to increase with time (Figure 2). This pattern is clearly visible in Farm 1, which had an average SOC of 12.73 g kg^{-1} in 2017-18 and 16.87 g kg^{-1} in 2020-21. From the second to the third year, there was a 25% increase in SOC (from 1.92 g kg^{-1}–2.40 g kg^{-1}) and, between the third and fourth year, there was a 10% increase in SOC (from 2.40 g kg^{-1}–2.63 g kg^{-1}). Farm 7 had the highest mean SOC (15.72 g kg^{-1}) and Farm 9 had the lowest mean SOC (5.89 g kg^{-1}).

Additionally, the mean SOC content was 13.12 g kg^{-1}. The lowest observed SOC content was 4.70 g kg^{-1} (Farm 9 in 2018-

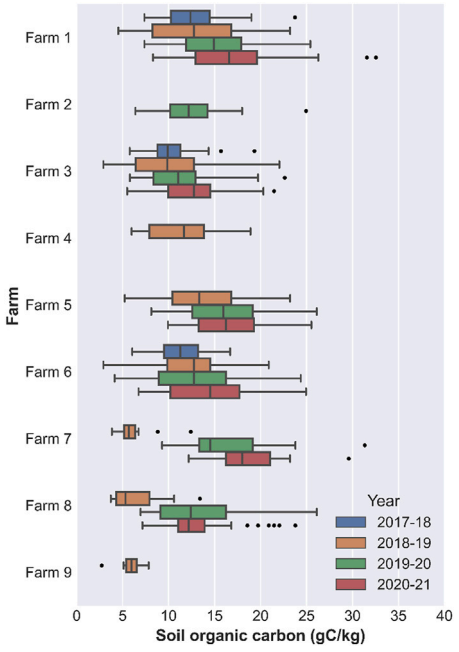


FIGURE 2
Boxplot of the soil organic carbon (SOC) content for the nine sampled farms in the four sampled production years.

19), and the maximum observed SOC content was 32.54 g kg^{-1} (Farm 1 in 2020-21). A positive correlation was observed between the number of samples per farm and the variation of SOC. Farm 1 was the farm with the highest variation of SOC. It had an interquartile distance (considering all years) of 8.30 g kg^{-1}. Farm 1 was also the farm with the highest number of soil samples (237). On the other hand, Farm 9, which had the lowest number of samples (12 samples), had the lowest interquartile distance, only 1.14 g kg^{-1}. From the nine sampled farms, only one (Farm 4) is in Spain, but it has similar SOC content distribution as the other Portuguese farms. The average SOC content in Farm 4 is 13.10 g kg^{-1} (min: 6.03 g kg^{-1}; max: 19.40 g kg^{-1}) and the average SOC in the Portuguese farms is 13.6 g kg^{-1} (min: 4.70 g kg^{-1}; max: 32.54 g kg^{-1}).

Although two sampling methods (manual and mechanical) were used for sample collection, the observed SOC content between the two methods was very similar. Specifically, the samples collected within the same farm using both methods show a high level of similarity (less than 7% differences with no observable bias), with any observed differences likely attributable to the typical spatial variation within the farm.

The Spearman rank correlation between observed SOC content and the input variables ranged between −0.61 and 0.32 (Figure 3). The lowest correlation corresponded to the correlation between SOC content the auxiliary dummy variable for manual or mechanical soil sampling (−0.61) and the highest correlation of SOC content was with the year (0.32). Analyzing the average correlation in absolute value, per type of input (according to the "Type" column in Table 3), auxiliary variables had the highest correlation (mean: 0.34), followed by climatic variables (mean: 0.22), and by terrain variables (mean: 0.14); the remaining average correlations were lower than 0.10. Despite the low correlations, about 80% (40 out of 49 input variables) were significantly correlated with SOC content, 37 variables at a significance level of 5% and 3 variables at 10% significance level.

In the composite image of August, all bands were strongly and significantly correlated with each other (average correlation of 0.65); however, the correlation between bands in the Sentinel-2 image closest to the collection date was significantly lower (average correlation of 0.35). Vegetation indices, as expected, were strongly and significantly correlated with the Sentinel-2 imagery that was used to calculate them, i.e., vegetation indices in August are strongly correlated with the composite Sentinel-2 imagery. There were also strong correlations between location variables (latitude and longitude) and soil variables (sand, silt, and pH) and the DEM.

3.2 Estimation of soil organic carbon

The feature selection procedure using SFS selected only 24 out of the 49 input variables considered in this work, representing approximately 48% of the total number of inputs. The selected inputs covered all the "Process Categories" defined in Table 2. The remote sensing imagery variables selected were Bands 2 and 12 from Sentinel-2 in August, Bands 3, 4, 7, 8, and 8 A from Sentinel-2 at the closest date, and VV from Sentinel-1 at the closest date. The vegetation indices selected were NDVI and NDWI in August, as well as NDVI, SR, SAVI, and OSAVI at the closest date. The selected climatic variable was soil temperature. The soil variables selected were silt content and pH. The terrain variables considered were the DEM and the mTPI. Additionally, the auxiliary variables selected were the number of days since August, the number of days from the closest Sentinel-2 imagery, and the month of the year. Lastly, both location variables, latitude and longitude, were also selected.

Among the regression methods used, XGB had the lowest estimation error for both feature selection approaches, as can be seen in Table 4 for the metrics of RMSE, rRMSE, RPD, and r^2. A general trend is that more complex models (RF, XGB, and ANN) outperform simpler models (MLR) in predicting SOC content in SBP systems. When comparing the regression methods, the mean RMSE of XGB was, on average, 52% lower than the mean RMSE of the other methods in the training sets and 11% lower than the other methods in the test sets. Similar trends can be observed in the other estimation error metrics. For example, the difference between MLR (the method with the highest RMSE) and XGB was 72% in the training sets (MLR: 3.10 g kg^{-1}; XGB: 0.87 g kg^{-1}—considering the approach without feature selection), and the difference was 18% in the test sets (MLR: 3.27 g kg^{-1}; XGB: 2.69 g kg^{-1}). Further, decision tree methods (RF and XGB) have a lower estimation error than the other methods MLR, ANN). The RF and XGB regression methods had similar estimation errors in the test sets, but XGB performed better than RF in the training sets. MLR was also the regression method with the lowest variation of the RMSE between training and test sets, only 6% (considering the approach without feature selection). The estimation error between the training set and test set in the other methods always had an increase higher than 50%, e.g., for the ANN, the difference was about 56%. The XGB was the method with the highest error increase, considering the RMSE, it more than doubled in the test set in relation to the training set, but even so, it was lower than in other methods.

Using the feature selection approach, where only 24 out of the total 49 inputs were used, did not significantly influence the

FIGURE 3
Spearman's rank correlation between the soil organic carbon and the considered input variables. The input variables are: 22 individual bands from Sentinel-2 (11 in August and 11 in closest date), 2 individual bands from Sentinel-1 (1 in August and 1 in closest date), 10 vegetative indices (5 in August and 5 in closest date), SOC proxies, soil variables, terrain variables, auxiliar variables and location variables. Variable names are explained in Table 3.

estimation error in the test sets for all regression methods. For example, considering XGB, the RMSE with feature selection was almost the same with all variables or with the selected variables (without selection: 2.78 g kg^{-1}; with selection: 2.77 g kg^{-1}). Nevertheless, in the training error, feature selection reduced the RMSE in RF and XGB (about 13%) and increased the RMSE of MLR and ANN (about 6%). This result highlights the efficacy of the feature selection approach in identifying the most relevant input variables for estimating SOC content. By accomplishing these dual objectives, the feature selection process enhances the convergence of the training procedure and ultimately improves the fitting performance of the RF and XGB models.

Considering XGB, there was no significant change in the estimation error between the two feature selection approaches. Figure 4 presents the estimated SOC versus the observed SOC when each sample is left on the test set using the approach with feature selection (using a hexagonal binning plot). As can be seen in Table 4, the estimation errors in the test sets were good, particularly in the region with the highest point density, i.e., between 10 and 15 g kg^{-1}. In this region, the RMSE in the test sets decreased by about 20% (2.19 g kg^{-1}). However, there was a non-significant overestimation of the observed SOC between 7 and 12 g kg^{-1}. Additionally, there was a noticeable underestimation of the measured SOC in the highest values (higher than 20 g kg^{-1}), which corresponds to the range of values with fewer observations.

In the XGB model with SFS, the VV feature (from Sentinel-1) had the highest importance (about 35%) in the obtained results. It was followed by the month of the year, latitude, and longitude. The Sentinel-2 bands in August (Bands 2 and 12) had the lowest contribution to the estimated SOC (less than 2%). Vegetation indices also had a greater relevance for SOC estimation than the individual satellite bands (each Vegetation Index at the closest date has a feature relevance of about 5%, and individual bands are lower than 3%). The terrain variables with the highest contribution are DEM and mTPI with an importance of 3% and 4%, respectively. All the soil input data has an accumulated importance lower than 7%.

3.3 Application at field-level

The obtained models can be used to estimate SOC for entire parcels in the farms. As an example of the application, Figure 5

TABLE 3 Description of the variables used to model soil organic carbon, including type of data, sources, variable and spatial resolution.

Type	Process category	Source	Variable	Spatial resolution (m)
Dynamic	Remote sensing imagery	Sentinel-1	Single polarisation VV	10
		Sentinel-2)	Band 2 (Blue)	10
			Band 3 (Green)	10
			Band 4 (Red)	10
			Band 5 (Vegetation red edge)	20
			Band 6 (Vegetation red edge)	20
			Band 7 (Vegetation red edge)	20
			Band 8 (NIR)	10
			Band 8A (Vegetation red edge)	20
			Band 11 (SWIR)	20
			Band 12 (SWIR)	20
	Vegetation indices		Normalized difference vegetation index (NDVI)	10
			Normalized difference water index (NDWI)	10
			Simple ratio (SR)	10
			Soil-adjusted vegetation index (SAVI)	10
			Optimized soil-adjusted vegetation index (OSAVI)	10
	Climatic	GLDAS	Soil moisture	27 km
			Soil temperature	27 km
Static	Auxiliary variables	-	Number of days since the beginning of the production year	—
			Number of days since last Sentinel-2 image	—
			Number of days since last Sentinel-1 image	—
			Year	—
			Month of the year	—
			Collection method	—
	Soil	GridSoils	Clay content	250
			Sand content	250
			Silt content	250
			Soil pH H_2O	250
	Terrain	NASA EOSDIS Land Processes DAAC	Digital elevation model (DEM)	30
		Theobald et al. (2015)	Continuous Heat-Insolation Load Index (CHILI)	90
			Multi-Scale Topographic Position Index (mTPI)	270
			Topographic diversity (topoDivers)	270
	Location	—	Latitude	—
			Longitude	—

depicts the spatial representation of SOC in the 9 sampled farms. This figure was obtained for the day of 29 May 2021, using the dynamic input data for that day, namely, the climatic data, Sentinel-2 imagery, and vegetation indices. Sentinel-1 imagery was not available for the same date, so we used Sentinel-1 imagery for the closest date, i.e., 27 May 2021. All the other input data is static, so it was not influenced by the date. The model used was the XGB model with the feature selection approach.

The trends observed in SOC between farms in Figure 2 are also verified when the XGB model was applied to the entire farm. For example, Farms 1, 5, and 7 had the highest mean SOC in the year 2020–2021 in both observed and predicted values. Farm 8 was the

TABLE 4 Estimation accuracy of the soil organic carbon in the training and test set of the cross-validation approach, for all using each of the machine learning (ML) methods and for the two features selection approach. Metrics presented: considering mean root mean squared error (RMSE), relative RMSE (rRMSE), ratio of performance to deviation (RPD) and r squared (r^2). MLR, Multiple linear regression; RF, Random forests; XGB, XGBoost; ANN, Artificial neural network. The model with the highest performance is in bold.

Without features selection									
ML method	Mean observed	Training				Test			
		RMSE	rRMSE	RPD	r^2	RMSE	rRMSE	RPD	r^2
MLR	13.12	3.10	0.24	1.59	0.60	3.27	0.25	1.51	0.55
RF		1.23	0.09	4.00	0.94	2.85	0.22	1.73	0.66
XGB		**0.87**	**0.07**	**5.66**	**0.97**	**2.78**	**0.21**	**1.81**	**0.68**
ANN		1.90	0.14	2.59	0.89	2.97	0.23	1.66	0.64
With features selection									
ML method	Mean observed	Training				Test			
		RMSE	rRMSE	RPD	r^2	RMSE	rRMSE	RPD	r^2
MLR	13.12	3.34	0.25	1.47	0.54	3.40	0.26	1.45	0.52
RF		1.05	0.08	4.69	0.95	2.83	0.22	1.74	0.64
XGB		**0.76**	**0.05**	**5.60**	**0.98**	**2.77**	**0.21**	**1.80**	**0.68**
ANN		1.98	0.15	2.49	0.85	3.06	0.23	1.61	0.59

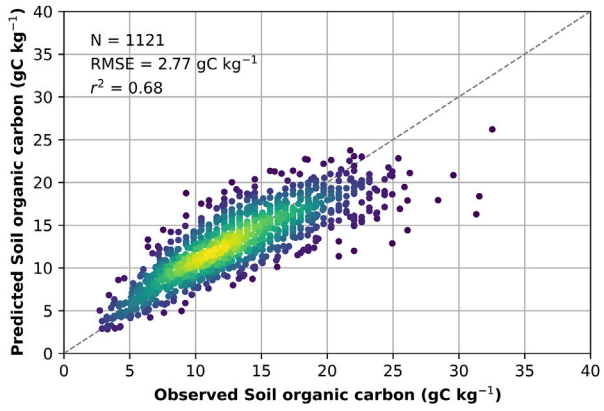


FIGURE 4
Estimated versus observed soil organic carbon (SOC) using the best model (XGBoost) in the features selection approach (i.e., only using 24 features).

farm with the highest spatial variation (standard deviation (SD) of 1.34 g kg^{-1}) and Farm 2 had the lowest spatial variation (SD: 0.74 g kg^{-1}). The minimum predicted SOC was also in Farm 2 (7.56 g kg^{-1}) and the highest predicted SOC was in Farm 8 (18.80 g kg^{-1}). Farm 2 had the lowest predicted SOC, 7.56 g kg^{-1}, but this farm was not sampled in the production year 2020-2021. However, there are other aspects that vary from the observed data. For example, in the observed date, in the production year of 2020-2021, Farm 1 has the highest SOC (32.54 g kg^{-1}) and the highest predicted SOC was at Farm 8, 18.80 g kg^{-1} in the predicted results. Nevertheless, the highest observed SOC at Farm 1 was in January (on January 16), which is significantly far from the date of May 29. Between January and May, soil temperature increases and soil moisture decreases, which supports SOC mineralization.

4 Discussion

This study demonstrated that more complex models (such as RF, XGB, and ANN) perform better in predicting SOC content in SBP systems in Portugal and Spain compared to simpler models like MLR (Liu et al., 2011; Ali et al., 2016). Complex models are capable of capturing complex, high-dimensional relationships between dependent and explanatory variables, which simple models cannot achieve. Two feature selection approaches were used to evaluate the performance impact. Our findings indicate that using all 49 input variables or a subset of just 24 (48%) yields comparable estimation performance in both training and testing phases. Moreover, the remaining variables encompassed almost all data categories that affect SOC content, including remote sensing, climatic, soil, and terrain characteristics.

Over the last decade, there has been a substantial increase in the number of combined applications that utilize satellite RS and ML to estimate SOC or SOM content. To investigate the extent of this increase, we conducted a very simple search in the Google Scholar database on 10 January 2023, specifically focusing on papers that estimated SOC content in pastures or grasslands using satellite RS. We utilized the search string: "(soil organic matter" OR "soil organic carbon") AND "remote sensing" AND "satellite" AND "regression" AND "machine learning" AND ("grassland" OR "pasture"), which resulted in 2,110 hits. Of these, 30% (688 hits) were from 2022 to 50% (1,080 hits) were from 2021. However, upon sorting the results by relevance according to

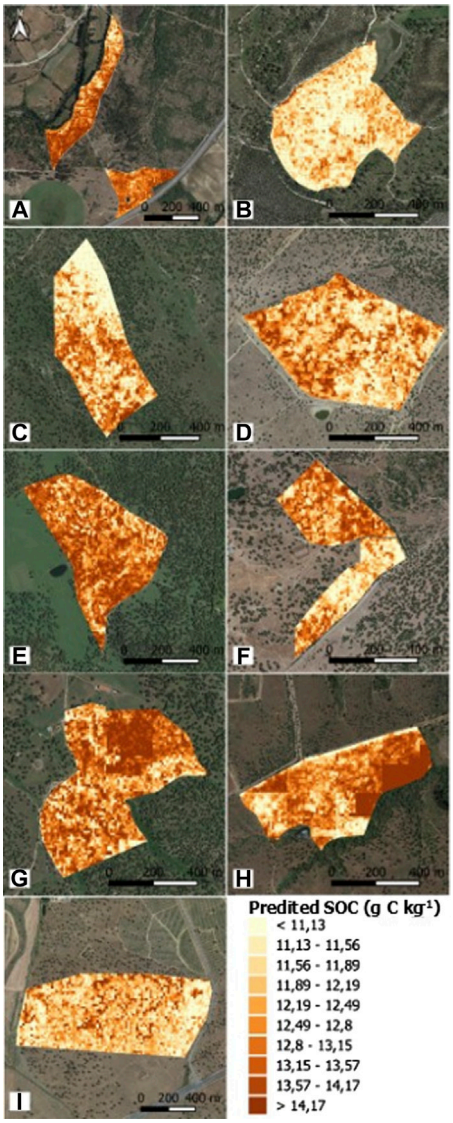


FIGURE 5
Spatial representation of the predicted SOC in the 9 sampled farms using the best model (XGBoost) in the features selection approach (i.e., only using 12 inputs). These results were obtained using the Sentinel-2 image of May 29 and Sentinel-1 image of 27 May 2021. **(A)** Farm 1; **(B)** Farm 2; **(C)** Farm 3; **(D)** Farm 4; **(E)** Farm 5; **(F)** Farm 6; **(G)** Farm 7; **(H)** Farm 8; **(I)** Farm 9.

Google Scholar, none of the first 50 hits were focused on grassland or pasture systems as the present paper does. This analysis is by no means a thorough review of the literature and surely depicts incomplete results, but shows that grassland systems remain under analysed and, in particular, this is the first study of this nature focusing on SBP.

This paper achieved better estimation performance for SOC content in grasslands and pastures compared to many other papers in the literature. For instance, Zhou et al. (2021) obtained an r^2 of 0.47 in their best model using a cross-validation approach for Switzerland's multiple land use/cover systems, whereas the highest r2 obtained in this study was 0.70. Hamzehpour et al. (2019) predicted SOC stock in a sub-region of Iran and achieved an r^2 of 0.44, while Wu et al. (2019) predicted SOC content in a sub-region of China using various machine learning regression models, and their best model, XGB, had an r^2 of 0.74, which was similar to the r^2 obtained in this paper. Similarly, Keskin et al. (2019) estimated total soil carbon in a sub-region of the United States of America using multiple regression models, and the best model was a RF with an r^2 of 0.72 in the validation set. Notably, decision trees consistently outperformed other simpler or more complex methods (such as ANNs) in all the studies that used different regression methods. In this study, extreme gradient boosting (XGB) demonstrated superior performance compared to the other models. Specifically, the XGB model, along with other decision tree-based models, outperformed artificial neural networks (ANN). There are several plausible reasons for this observation. Firstly, XGB models tend to be less reliant on extensive fine-tuning of hyperparameters, potentially contributing to their improved performance, as suggested by the results (Memon et al., 2019; Shwartz-Ziv and Armon, 2022).

In this study, we observed that the estimation accuracy for the highest SOC values was significantly lower than that for low-medium values. This trend has been observed in other studies that estimated SOC, as well as in the estimation of other variables in croplands and grasslands, among others (Castaldi et al., 2018). The normal frequency distribution of the data on SOC is the cause of this limitation since the dataset is dominated by mid-range values. To overcome this limitation, quantile regression methods based on the approach used in this study can be employed, such as quantile RF. Quantile regression models the relationship between independent variables and specific percentiles of the dependent variable, which is an improvement over regression methods that represent the mean increase in the response function produced by one unit increase in the associated independent variables. In fact, recent studies have applied these regression methods to SOC estimation (Lombardo et al., 2018; Kasraei et al., 2021; Zhao et al., 2021). In the future, the application of these methods should be tested to confirm if the estimation performance increases significantly.

In addition, the number of observations per farm can also influence results. It has been observed that the model tends to achieve a better fit when applied to farms with a larger number of samples compared to those with a smaller number of samples. For instance, Farm 1 consists of a total of 237 samples, while Farm 2 comprises only 35 samples. Consequently, the model is more likely to exhibit improved performance in capturing the specific characteristics associated with Farm 1 rather than Farm 2. The imbalance in the number of observations across farms may also impact the generalization error when applying the model to other locations. However, considering that the characteristics of the different farms are not significantly different, we do not anticipate that the obtained model would yield highly inaccurate estimations of SOC content for the sample used here. The effectiveness of the model when applied to other SBP farms should be assessed in future research work.

Here, we developed a rapid and cost-effective indirect method for the purpose of expedite mapping of SOC in SBP farms. This represents a significant improvement compared to the approach proposed by Morais et al. (2021), which relied on data from *in situ* field spectrometry and only replaced the laboratory analysis. In terms of results, the obtained r^2 value (0.68) is lower than the value

previously reported by Morais et al. (0.80). However, it is important to note that our method is solely based on remote sensing data and therefore applicable to multiple farms and regions without the need for repeated field work and laboratory analysis.

In this study, we used RS data from Sentinel-1 and Sentinel-2, which offer significantly higher spatial resolution compared to other spatially explicit variables. The inclusion of Sentinel-1 and Sentinel-2 data allowed us to capture fine-scale spatial variations within individual parcels or farms. Conversely, other static data sources with lower spatial resolution lacking the capability to capture intricate spatial variations within parcels primarily facilitated the assessment of regional variation. Additionally, remote sensing data provided a distinct advantage by enabling us to capture of temporal variations across different years, as they were the only data sources exhibiting temporal variability over time. Despite achieving good performance in our study, there is potential for improvement by enhancing the quality of climatic and soil data. It is important to note that the SFS method, while not affecting SOC estimation performance, may be influenced by the spatial resolution of the input data. SFS excluded soil temperature and soil moisture as explanatory variables, probably due to the course scale of the data sources available. However, those variables are vital in regulating microbial activity, nutrient availability, and overall soil health. The same was true of some climate variables, which had a spatial resolution of 27 km, which may be insufficient for depicting intra-farm variations.

RS data derived from Sentinel-1 and Sentinel-2 present a significantly elevated spatial resolution in comparison to other spatially explicit variables. The utilization of Sentinel-1 and Sentinel-2 data enables the capture of intricate spatial variations within individual parcels or farms. Conversely, static data sources with diminished spatial resolution predominantly facilitate the assessment of regional variations, as they lack the ability to capture the detailed spatial nuances within parcels. Moreover, remote sensing data proffers the distinct advantage of capturing temporal variations across different years, rendering it the sole data source characterized by temporal variability over time. In fact, this procedure of using multiple data sources with multiple spatial and temporal resolutions is frequently used to characterize different land cover systems (Zhang et al., 2016; Venter and Sydenham, 2021), namely, to estimate SOC content, e.g., Venter et al. (2021). Nevertheless, enhancing the spatial resolution of the data with low spatial resolution could potentially improve the estimation performance of SOC content. For example, in this study, the soil data used had a spatial resolution of 250 m. It is not expected that soil characteristics such as sand, clay, and silt fractions would vary significantly within the same farm. Consequently, the variables that contributed the most to explaining SOC content were the ones that had the higher resolutions, such as those measured or calculated from Sentinel-1 and Sentinel-2 data. Increasing the spatial resolution of coarse soil-specific data could enhance the fine variation of SOC content and help address some of the variance unexplained by our model.

The obtained models in this study have a spatial resolution of 10 m, which is the lowest resolution among all the spatialized data used, including Sentinel-1 data and the red, green, and blue bands of Sentinel-2. However, even this resolution may not be sufficient to capture all the spatial variability of pasture systems such as SBP. To enhance the spatial resolution of RS data from satellites, UAVs can be utilized. UAVs can have a spatial resolution of a few centimeters, providing a significant improvement in spatial resolution. For instance, a 5 cm resolution UAV would yield 100,000 pixels in a 10 × 10 m pixel of Sentinel-2. UAVs are currently preferred for agricultural land characterization due to their affordability and ease of operation. Nonetheless, UAV data has a significantly lower spatial coverage, lower spectral resolution, and potentially lower temporal coverage than satellite data (Colomina and Molina, 2014; Vilar et al., 2020). Moreover, the quality of UAV data can be negatively impacted by factors such as sun elevation angle, diffuse sunlight, and shadow effects of objects such as trees (De Luca et al., 2019). Rather than completely replacing satellite data with UAV data, it is more beneficial to use them in combination to minimize estimation errors. For instance, Maimaitijiang et al. (2020) improved the estimation of biomass characteristics by integrating RGB UAV data with Sentinel-2 data.

In this paper, we used individual bands from the Sentinel-1 satellite. Nevertheless, recent research has proposed a technique to merge two Sentinel-1 image products of complementary polarimetric information (HH/HV and VH/VV) to derive pseudo-polarimetric features (Braun and Offermann, 2022). Despite some inaccuracies, the polarimetric features turned out to improve potential land cover mapping compared with backscatter intensities and dual-polarization features of the input products alone. However, such a technique has not yet been tested in regression problems to estimate SOC content. Alternatively, synthetic-aperture radar data from other satellites could provide different bands and wavelengths (Moreira et al., 2013). Data with different wavelengths and frequencies also have different penetration power, spatial resolution, sensitivity to surface roughness, and sensitivity to atmospheric effects (Moreira et al., 2013; Paek et al., 2020; Le et al., 2021). The C-band used in Sentinel-1 refers to the microwave frequency range between 4 to 8 GHz (Gigahertz) in the electromagnetic spectrum (ESA, 2022). It is one of the most commonly used bands in SAR remote sensing due to its favourable characteristics, namely: moderate penetration capabilities, meaning it can penetrate through vegetation and light to moderate rainfall; good spatial resolution allowing the detection of small to medium-sized features on the Earth's surface; sensitivity to surface roughness variations, which makes it useful for monitoring changes in ocean waves, soil moisture, and snow cover; and is less affected by atmospheric conditions like clouds and precipitation compared to higher-frequency bands (e.g., X-band or Ku-band) (Monti-Guarnieri et al., 2017; ESA, 2022). Another frequency band that is commonly used is the P band, for example, used in ALOS (Advanced Land Observing Satellite) PALSAR (Phased Array type L-band Synthetic Aperture Radar), which is in the microwave frequency range between 0.3 to 1 GHz (Gigahertz) in the electromagnetic spectrum. The P-band has higher penetration than the C-band. Due to its lower frequency, P-band SAR typically has a coarser spatial resolution compared to higher-frequency bands like the C-band. P-band SAR is also less sensitive to surface roughness compared to C-band SAR, but it is relatively less affected by atmospheric conditions (Li et al., 2019; Minh et al., 2021). Other bands with higher frequency (e.g., X-band) have higher spatial resolution but lower penetration capacity (Zhou et al., 2020). Thus, in the future, approaches that combine alternative/complementary SAR data should be tested to improve the characterization of land cover systems, such as grasslands.

Here we used several vegetation indices (NDVI, NDWI, SR, SAVI, and OSAVI) as well as the raw data for the bands used to calculate them. The fact that the bands are used nonlinearly takes away some of the explanatory power of the indices. However, because the indices were more important than the individual

bands in our results, exploring additional indices may offer valuable insights into SOC content estimation. For example, the Normalized Difference Red/Green Redness Index and the Dark Green Color Index that utilize both red and green bands have been previously used to estimate SOC content in agricultural soils (Heil et al., 2022). These and other alternative indices could potentially complement the existing ones and enhance the accuracy of SOC estimation.

In this study, we did not perform an assessment of bare soil pixels, which is a common practice in other research studies (Bhunia et al., 2017; Castaldi, 2021). Typically, bare soil pixels are determined using vegetation indices calculated from individual bands of Sentinel-2, such as NDVI and normalized burn ratio 2 (NBR2) (Castaldi, 2021). This process involves defining a threshold for the vegetation indices, and pixels with lower values than the threshold are classified as bare. However, the number of bare soil pixels can vary significantly depending on the chosen thresholds. For instance, Castaldi (2021) observed that reducing the NBR2 threshold from 0.2 to 0.05 in Northeastern Germany croplands led to a decrease in the percentage of Sentinel-2 pixels classified as bare soil from over 25% to about 10%. Additionally, this method requires the removal of data points that do not meet the defined thresholds. For these reasons, we chose not to use this approach. Instead, we utilized data not only near the sampling date but also data from August when the soil is mostly bare in well-managed SBP systems. Incorporating observations from August allows us to capture the soil's characteristics when it is bare, while observations near the sampling date enable us to indirectly evaluate the effect of vegetation on SOC.

The models that we developed lack a formal representation of the processes that occur in soil and influence SOC content, such as an equation for SOC mineralization that process-based models possess (Morais et al., 2019). Unlike data-driven models, process-based soil models consider biogeochemical processes formulated based on mathematical-ecological theory (Coleman et al., 1997; Liu et al., 2011). These models' equations are often derived from statistical relationships, which can be improved by incorporating data-driven modeling approaches. Combining the benefits of both data-driven models (such as those used in this study) and process-based modeling is critical for developing more robust models in the future. One approach is to replace process-based models' rate modifiers with ML models. Tsai et al. (2021) have done this successfully to predict soil moisture and streamflow.

The models derived in this study have the potential to retrospectively estimate SOC content since 2015 when Sentinel-2 data was initiated. Consequently, a considerable amount of data can be generated that can be employed in other models. Process-based models, such as those that evaluate soil sinks and emissions of carbon and nitrogen and their impacts on environmental concerns, can benefit significantly from longer data series (Prado et al., 2006; Morais et al., 2018; Teixeira et al., 2019).

5 Conclusion

This work combined multiple data types from different sources with ML methods in order to estimate SOC content of SBP in Portugal and Spain. The most relevant variables that are known to influence SOC content and change, such as climatic, soil, and terrain characteristics, were combined with RS imagery. The most relevant variables from the full set of independent (or input) data were selected using an SFS approach. This approach reduced the number of variables to 24 (instead of 49) but maintained the overall accuracy of the best model: without feature selection, the root mean squared error (RMSE) was 2.78 g kg-1 (on the test set) and with feature selection, the RMSE was 2.77 g kg-1. XGB was the model with the highest estimation performance, using a cross-validation approach.

SOC content plays a significant role in plant growth and characteristics. Nevertheless, the type of models developed in this work are still infrequently used as a farm management tool, despite the fact that they are powerful tools that could increase incomes and/or reduce costs. Based on the best models, SOC content can be approximately estimated throughout the year, even when the soil is covered by plants, and with that, advisors can inform farmers to perform practices to improve soil quality for plant and animal production.

Data availability statement

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

Author contributions

TM, TD, and RT contributed to the conceptualization and methodology of the study. MJ, CM, NR, IG, and JS performed field investigations. MG and RM were responsible for lab analysis of the soil samples. TM, TD, and RT contributed to data analysis and interpretation of the results. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Fundação para a Ciência e Tecnologia through projects "GrassData—Development of algorithms for identification, monitoring, compliance checks and quantification of carbon sequestration in pastures" (DSAIPA/DS/0074/2019), "LEAnMeat—Lifecycle-based Environmental Assessment and impact reduction of Meat production with a novel multi-level tool" (PTDC/EAM-AMB/30809/2017), and CEECIND/00365/2018 (RT). This work was also supported by FCT/MCTES (PIDDAC) through project LARSyS—FCT Pluriannual funding 2020-2023 (UIDP/EEA/50009/2020), by COMPETE 2020, FEDER through project GreenBeef– "GreenBeef: Towards carbon neutral Angus beef production in Portugal "(POCI-01-0247-FEDER-047050) and by Programa de desenlvolvimento rural (PDR 2020) through "GO SOLO: Avaliação da dinâmica da matéria orgânica em solos de pastagens semeadas biodiversas através do desenvolvimento de um método de monitorização expedito e a baixo custo" (PDR2020-101-031243).

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

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., and Green, S. (2016). Satellite remote sensing of grasslands: from observation to management. *J. Plant Ecol.* 9, 649–671. doi:10.1093/jpe/rtw005

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., and Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sens.* 11, 676. doi:10.3390/rs11060676

APA (2018). *Portuguese national inventory report on greenhouse gases, 1990 - 2018*. Amadora, Portugal: Portuguese Environmental Agency.

Bardgett, R. D., Bullock, J. M., Lavorel, S., Manning, P., Schaffner, U., Ostle, N., et al. (2021). Combatting global grassland degradation. *Nat. Rev. Earth Environ.* 2, 720–735. doi:10.1038/s43017-021-00207-2

Bhunia, G. S., Kumar Shit, P., and Pourghasemi, H. R. (2017). Soil organic carbon mapping using remote sensing techniques and multivariate regression model. *Geocarto Int.* 1, 215–226. doi:10.1080/10106049.2017.1381179

Biney, J. K. M. (2022). Verifying the predictive performance for soil organic carbon when employing field Vis-NIR spectroscopy and satellite imagery obtained using two different sampling methods. *Comput. Electron. Agric.* 194, 106796. doi:10.1016/J.COMPAG.2022.106796

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Braun, A., and Offermann, E. (2022). Polarimetric information content of Sentinel-1 for land cover mapping: an experimental case study using quad-pol data synthesized from complementary repeat-pass acquisitions. *Front. Remote Sens.* 3. doi:10.3389/frsen.2022.905713

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Castaldi, F., Chabrillat, S., Jones, A., Vreys, K., Bomans, B., and van Wesemael, B. (2018). Soil organic carbon estimation in croplands by hyperspectral remote APEX data using the LUCAS topsoil database. *Remote Sens.* 10, 153. doi:10.3390/rs10020153

Castaldi, F. (2021). Sentinel-2 and landsat-8 multi-temporal series to estimate topsoil properties on croplands. *Remote Sens.* 13, 3345. doi:10.3390/RS13173345

Chan, C. K., Gomez, C. A., Kothikar, A., and Baiz-Villafranca, P. M. (2023). Satellite-based carbon estimation in scotland: AGB and SOC. *Land* 12, 818. doi:10.3390/land12040818

Chappell, A., Baldock, J., and Sanderman, J. (2016). The global significance of omitting soil erosion from soil organic carbon cycling schemes. *Nat. Clim. Chang.* 6, 187–191. doi:10.1038/nclimate2829

Chen, T., and Guestrin, C. (2016). "XGBoost: A scalable tree boosting system," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA, August 2016. doi:10.1145/2939672.2939785

Coleman, K., Jenkinson, D. S., Crocker, G. J., Grace, P. R., Klír, J., Körschens, M., et al. (1997). Simulating trends in soil organic carbon in long-term experiments using RothC-26.3. *Geoderma* 81, 29–44. doi:10.1016/S0016-7061(97)00079-7

Colomina, I., and Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 92, 79–97. doi:10.1016/j.isprsjprs.2014.02.013

Davids, C., Karlsen, S. R., Ancin, M., and Jorgensen, M. (2018). "UAV based mapping of grassland yields for forage production in northern Europe Sustainable meat and milk production from grasslands," in Proceedings of the 27th General Meeting of the European Grassland Federation, Cork, Ireland, June 2018, 845–847.

De Luca, G., Silva, N. J. M., Cerasoli, S., Araújo, J., Campos, J., et al. (2019). Object-based land cover classification of cork oak woodlands using UAV imagery and orfeo ToolBox. *Remote Sens.* 11, 1238. doi:10.3390/rs11101238

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., et al. (2015). Soil carbon storage controlled by interactions between geochemistry and climate. *Nat. Geosci.* 8, 780–783. doi:10.1038/ngeo2516

Egoh, B. N., Bengtsson, J., Lindborg, R., Bullock, J. M., Dixon, A. P., Rouget, M., et al. (2016). "The importance of grasslands in providing ecosystem services," in *Routledge handbook of ecosystem services* (New York, NY: Routledge), 421–441. doi:10.4324/9781315775302-37

ESA (2022). Sentinel-1 - missions - Sentinel online - Sentinel online. *Eur. Sp. Agency*. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-1 (Accessed June 26, 2023).

ESDAC (2003). *European Soil Database (distribution version v2.0)*. Brussels (Belgium): European Commission Joint Research Centre.

Gao, B. C. (1996). Ndwi - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi:10.1016/S0034-4257(96)00067-3

Gardin, L., Chiesi, M., Fibbi, L., and Maselli, F. (2021). Mapping soil organic carbon in Tuscany through the statistical combination of ground observations with ancillary and remote sensing data. *Geoderma* 404, 115386. doi:10.1016/j.geoderma.2021.115386

Giardina, C. P., Litton, C. M., Crow, S. E., and Asner, G. P. (2014). Warming-related increases in soil CO2 efflux are explained by increased below-ground carbon flux. *Nat. Clim. Chang.* 4, 822–827. doi:10.1038/nclimate2322

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning (adaptive computation and machine learning series)*. Cambridge, England: MIT Press.

Hamzehpour, N., Shafizadeh-Moghadam, H., and Valavi, R. (2019). Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. *CATENA* 182, 104141. doi:10.1016/J.CATENA.2019.104141

Heil, J., Jörges, C., and Stumpe, B. (2022). Fine-scale mapping of soil organic matter in agricultural soils using UAVs and machine learning. *Remote Sens.* 14, 3349. doi:10.3390/rs14143349

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12, e0169748. doi:10.1371/journal.pone.0169748

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. doi:10.1016/0034-4257(88)90106-X

IPMA (2018). Climate normals. Available at: http://www.ipma.pt/en/index.html (Accessed January 18, 2018).

Jobbágy, E. G., and Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* 10, 423–436. doi:10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

Kasraei, B., Heung, B., Saurette, D. D., Schmidt, M. G., Bulmer, C. E., and Bethel, W. (2021). Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. *Environ. Model. Softw.* 144, 105139. doi:10.1016/J.ENVSOFT.2021.105139

Keskin, H., Grunwald, S., and Harris, W. G. (2019). Digital mapping of soil carbon fractions with machine learning. *Geoderma* 339, 40–58. doi:10.1016/J.GEODERMA.2018.12.037

Koven, C. D., Hugelius, G., Lawrence, D. M., and Wieder, W. R. (2017). Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nat. Clim. Chang.* 7, 817–822. doi:10.1038/nclimate3421

Ladoni, M., Bahrami, H. A., Alavipanah, S. K., and Norouzi, A. A. (2010). Estimating soil organic carbon from soil reflectance: A review. *Precis. Agric.* 11, 82–99. doi:10.1007/s11119-009-9123-3

Lamichhane, S., Kumar, L., and Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* 352, 395–413. doi:10.1016/j.geoderma.2019.05.031

Le, N. N., Pham, T. D., Yokoya, N., Ha, N. T., Nguyen, T. T. T., Tran, T. D. T., et al. (2021). Learning from multimodal and multisensor earth observation dataset for improving estimates of mangrove soil organic carbon in Vietnam. *Int. J. Remote Sens.* 42, 6866–6890. doi:10.1080/01431161.2021.1945158

Li, W., Tong, Q., Xu, L., Ji, P., Dong, F., Yu, Y., et al. (2019). "The P-band SAR satellite: opportunities and challenges," in Proceedings of the 2019 6th Asia-Pacific Conf. Synth. Aperture Radar, (APSAR), Xiamen, China, November 2019. doi:10.1109/APSAR46974.2019.9048581

Liu, D. L., Chan, K. Y., Conyers, M. K., Li, G., and Poile, G. J. (2011). Simulation of soil organic carbon dynamics under different pasture managements using the RothC carbon model. *Geoderma* 165, 69–77. doi:10.1016/j.geoderma.2011.07.005

Lombardo, L., Saia, S., Schillaci, C., Mai, P. M., and Huser, R. (2018). Modeling soil organic carbon with Quantile Regression: dissecting predictors' effects on carbon stocks. *Geoderma* 318, 148–159. doi:10.1016/J.GEODERMA.2017.12.011

Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., and Fritschi, F. B. (2020). Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sens.* 12, 1357. doi:10.3390/RS12091357

Mallik, S., Bhowmik, T., Mishra, U., and Paul, N. (2020). Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data. *Geocarto Int.* 37, 2198–2214. doi:10.1080/10106049.2020.1815864

Mariano, D. A., Santos, C. A. C. dos, Wardlow, B. D., Anderson, M. C., Schiltmeyer, A. V., Tadesse, T., et al. (2018). Use of remote sensing indicators to assess effects of drought and human-induced land degradation on ecosystem health in Northeastern Brazil. *Remote Sens. Environ.* 213, 129–143. doi:10.1016/J.RSE.2018.04.048

Memon, N., Patel, S. B., and Patel, D. P. (2019). "Comparative analysis of artificial neural network and XGBoost algorithm for PolSAR image classification," in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics* (Berlin, Germany: Springer). doi:10.1007/978-3-030-34869-4_49

Minh, D. H. T., Ngo, Y. N., and Lê, T. T. (2021). Potential of P-band SAR tomography in forest type classification. *Remote Sens.* 13, 696. doi:10.3390/RS13040696

Monti-Guarnieri, A., Giudici, D., and Recchia, A. (2017). Identification of C-band radio frequency interferences from sentinel-1 data. *Remote Sens.* 9, 1183. doi:10.3390/RS9111183

Morais, T. G., Jongen, M., Tufik, C., Rodrigues, N. R., Gama, I., Fangueiro, D., et al. (2022). Characterization of Portuguese sown rainfed grasslands using remote sensing and machine learning. *Precis. Agric.* 24, 161–186. doi:10.1007/s11119-022-09937-9

Morais, T. G., Teixeira, R. F. M., and Domingos, T. (2019). Some croplands can potentially accumulate more soil carbon than forests and grasslands: implications of detailed global modelling. *PLoS One* 14, e0222604. doi:10.1371/journal.pone.0222604

Morais, T. G., Teixeira, R. F. M., Rodrigues, N. R., and Domingos, T. (2018). Characterizing livestock production in Portuguese sown rainfed grasslands: applying the inverse approach to a process-based model. *Sustainability* 10, 4437. doi:10.3390/su10124437

Morais, T. G., Tufik, C., Rato, A. E., Rodrigues, N. R., Gama, I., Jongen, M., et al. (2021). Estimating soil organic carbon of sown biodiverse permanent pastures in Portugal using near infrared spectral data and artificial neural networks. *Geoderma* 404, 115387. doi:10.1016/J.GEODERMA.2021.115387

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., and Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* 1, 6–43. doi:10.1109/MGRS.2013.2248301

Moreno, G., Hernández-Esteban, A., Rolo, V., and Igual, J. M. (2021). The enduring effects of sowing legume-rich mixtures on the soil microbial community and soil carbon in semi-arid wood pastures. *Plant Soil* 465, 563–582. doi:10.1007/s11104-021-05023-7

Nasa, J. P. L. (2020). *NASADEM Merged DEM Global 1 arc second V001*.

Odebiri, O., Odindi, J., and Mutanga, O. (2021). Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102389. doi:10.1016/j.jag.2021.102389

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernández-Ugalde, O. (2018). LUCAS soil, the largest expandable soil dataset for europe: A review. *Eur. J. Soil Sci.* 69, 140–153. doi:10.1111/EJSS.12499

Paek, S. W., Balasubramanian, S., Kim, S., and de Weck, O. (2020). Small-satellite synthetic aperture radar for continuous global biospheric monitoring: A review. *Remote Sens.* 12, 2546. doi:10.3390/RS12162546

Pan, Y., Zhang, X., Liu, H., Wu, D., Dou, X., Xu, M., et al. (2022). Remote sensing inversion of soil organic matter by using the subregion method at the field scale. *Precis. Agric.* 23, 1813–1835. doi:10.1007/s11119-022-09914-2

Pezzuolo, A., Dumont, B., Sartori, L., Marinello, F., De Antoni Migliorati, M., and Basso, B. (2017). Evaluating the impact of soil conservation measures on soil organic carbon at the farm scale. *Comput. Electron. Agric.* 135, 175–182. doi:10.1016/J.COMPAG.2017.02.004

Pouladi, N., Møller, A. B., Tabatabai, S., and Greve, M. H. (2019). Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma* 342, 85–92. doi:10.1016/j.geoderma.2019.02.019

Prado, A. d., Brown, L., Schulte, R., Ryan, M., and Scholefield, D. (2006). Principles of development of a mass balance N cycle model for temperate grasslands: an Irish case study. *Nutr. Cycl. Agroecosyst.* 74, 115–131. doi:10.1007/s10705-005-5769-z

Rey, A., Petsikos, C., Jarvis, P. G., and Grace, J. (2005). Effect of temperature and moisture on rates of carbon mineralization in a Mediterranean oak forest soil under controlled and field conditions. *Eur. J. Soil Sci.* 56, 589–599. doi:10.1111/j.1365-2389.2004.00699.x

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 85, 381–394. doi:10.1175/BAMS-85-3-381

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., and Heiden, U. (2018). Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sens. Environ.* 205, 1–17. doi:10.1016/j.rse.2017.11.004

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55, 95–107. doi:10.1016/0034-4257(95)00186-7

Rubel, F., and Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorol. Z.* 19, 135–141. doi:10.1127/0941-2948/2010/0430

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi:10.1038/323533a0

Shwartz-Ziv, R., and Armon, A. (2022). Tabular data: deep learning is not all you need. *Inf. Fusion* 81, 84–90. doi:10.1016/j.inffus.2021.11.011

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15, 72–101. doi:10.2307/1412159

Sun, S., Zuo, Z., Yue, W., Morel, J., Parsons, D., Liu, J., et al. (2021). Estimation of biomass and nutritive value of grass and clover mixtures by analyzing spectral and crop height data using chemometric methods. *Comput. Electron. Agric.* 192, 106571. doi:10.1016/J.COMPAG.2021.106571

Teixeira, R. F. M., Barão, L., Morais, T. G., and Domingos, T. (2019). BalSim: A carbon, nitrogen and greenhouse gas mass balance model for pastures. *Sustainability* 11, 53. doi:10.3390/su11010053

Teixeira, R. F. M., Domingos, T., Costa, A. P. S. V., Oliveira, R., Farropas, L., Calouro, F., et al. (2011). Soil organic matter dynamics in Portuguese natural and sown rainfed grasslands. *Ecol. Modell.* 222, 993–1001. doi:10.1016/j.ecolmodel.2010.11.013

Teixeira, R. F. M., Proença, V., Crespo, D., Valada, T., and Domingos, T. (2015). A conceptual framework for the analysis of engineered biodiverse pastures. *Ecol. Eng.* 77, 85–97. doi:10.1016/j.ecoleng.2015.01.002

Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., and Albano, C. M. (2015). Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PLoS One* 10, e0143619. doi:10.1371/JOURNAL.PONE.0143619

Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., et al. (2009). Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: Results from an atmosphere-ocean general circulation model. *Biogeosciences* 6, 2099–2120. doi:10.5194/bg-6-2099-2009

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: harnessing the scaling effects of big data in geoscientific modeling. *Nat. Commun.* 12, 5988–6013. doi:10.1038/s41467-021-26107-z

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. doi:10.1016/0034-4257(79)90013-0

Ullah, S., Si, Y., Schlerf, M., Skidmore, A. K., Shafique, M., and Iqbal, I. A. (2012). Estimation of grassland biomass and nitrogen using MERIS data. *Int. J. Appl. Earth Obs. Geoinf.* 19, 196–204. doi:10.1016/J.JAG.2012.05.008

Venter, Z. S., Hawkins, H. J., Cramer, M. D., and Mills, A. J. (2021). Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. *Sci. Total Environ.* 771, 145384. doi:10.1016/J.SCITOTENV.2021.145384

Venter, Z. S., and Sydenham, M. A. K. (2021). Continental-scale land cover mapping at 10 m resolution over europe (ELC10). *Remote Sens.* 13, 2301. doi:10.3390/rs13122301

Veronesi, F., and Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecol. Indic.* 101, 1032–1044. doi:10.1016/j.ecolind.2019.02.026

Vilar, P., Morais, T. G., Rodrigues, N. R., Gama, I., Monteiro, M. L., Domingos, T., et al. (2020). Object-based classification approaches for multitemporal identification and monitoring of pastures in agroforestry regions using multispectral unmanned aerial vehicle products. *Remote Sens.* 12, 814. doi:10.3390/rs12050814

Wang, B., Gray, J. M., Waters, C. M., Rajin Anwar, M., Orgill, S. E., Cowie, A. L., et al. (2022). Modelling and mapping soil organic carbon stocks under future climate change in south-eastern Australia. *Geoderma* 405, 115442. doi:10.1016/j.geoderma.2021.115442

Wang, X., Han, J., Wang, X., Yao, H., and Zhang, L. (2021). Estimating soil organic matter content using sentinel-2 imagery by machine learning in shanghai. *IEEE Access* 9, 78215–78225. doi:10.1109/ACCESS.2021.3080689

Wu, T., Luo, J., Dong, W., Sun, Y., Xia, L., and Zhang, X. (2019). Geo-object-based soil organic matter mapping using machine learning algorithms with multi-source geo-spatial data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12, 1091–1106. doi:10.1109/JSTARS.2019.2902375

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., et al. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sens. Environ.* 233, 111383. doi:10.1016/j.rse.2019.111383

Xu, B., Yang, X. C., Tao, W. G., Qin, Z. H., Liu, H. Q., Miao, J. M., et al. (2008). MODIS-based remote sensing monitoring of grass production in China. *Int. J. Remote Sens.* 29, 5313–5327. doi:10.1080/01431160802036276

Zepp, S., Heiden, U., Bachmann, M., Wiesmeier, M., Steininger, M., and van Wesemael, B. (2021). Estimation of soil organic carbon contents in croplands of bavaria from scmap soil reflectance composites. *Remote Sens.* 13, 3141. doi:10.3390/rs13163141

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., and Liu, G. (2016). Application of synthetic NDVI time series blended from landsat and MODIS data for grassland biomass estimation. *Remote Sens.* 8, 10. doi:10.3390/rs8010010

Zhao, F., Xu, B., Yang, X., Jin, Y., Li, J., Xia, L., et al. (2014). Remote sensing estimates of grassland aboveground biomass based on modis net primary productivity (NPP): A case study in the xilingol grassland of northern China. *Remote Sens.* 6, 5368–5386. doi:10.3390/rs6065368

Zhao, W., Wu, Z., and Yin, Z. (2021). "Estimation of soil organic carbon content based on deep learning and quantile regression," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, July 2021, 3717–3720. doi:10.1109/igarss47720.2021.9553418

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., et al. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sens. Environ.* 93, 402–411. doi:10.1016/J.RSE.2004.08.008

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., et al. (2021). Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between sentinel-2, sentinel-3 and landsat-8 images. *Sci. Total Environ.* 755, 142661. doi:10.1016/j.scitotenv.2020.142661

Zhou, Y., Wang, P., Chen, Z., Zhao, Q., Wang, W., Zhang, L., et al. (2020). Very high resolution SAR imaging with DGPS-supported airborne X-band data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 3605–3617. doi:10.1109/JSTARS.2020.3004013

frontiers | Frontiers in Ecology and Evolution

TYPE Methods
PUBLISHED 17 October 2023
DOI 10.3389/fevo.2023.1250971

OPEN ACCESS

EDITED BY
Samy Ismail Elmahdy,
United Arab Emirates University, United Arab Emirates

REVIEWED BY
Chao Chen,
Suzhou University of Science and Technology, China
Rui Xu,
Chang'an University, China

*CORRESPONDENCE
Yulong Dong
1982433418@qq.com

RECEIVED 30 June 2023
ACCEPTED 06 October 2023
PUBLISHED 17 October 2023

CITATION
Chen Y, Dong Y, Wang Y, Zhang F, Liu G and Sun P (2023) Machine learning algorithms for lithological mapping using Sentinel-2 and SRTM DEM in highly vegetated areas.
Front. Ecol. Evol. 11:1250971.
doi: 10.3389/fevo.2023.1250971

Machine learning algorithms for lithological mapping using Sentinel-2 and SRTM DEM in highly vegetated areas

Yansi Chen[1], Yulong Dong[1]*, Yunchen Wang[2,3], Feng Zhang[4], Genyuan Liu[1] and Peiheng Sun[1]

[1]Center for Geophysical Survey, China Geological Survey, Langfang, China, [2]School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, China, [3]Shaanxi Satellite Application Technology Center for Natural Resources, Shaanxi Institute of Geological Survey, Xi'an, China, [4]Center for Applied Geology Research, China Geological Survey, Sichuan, China

Lithological mapping in highly vegetated areas using remote sensing techniques poses a significant challenge. Inspired by the concept of "geobotany", we attempted to distinguish lithologies indirectly using machine learning algorithms (MLAs) based on Sentinel-2 and SRTM DEM in Zhangzhou City, Fujian Province. The study area has high vegetation cover, with lithologies that are largely obscured. After preprocessing such as cloud masking, resampling, and median image synthesis, 17 spectral bands and features from Sentinel-2 and 9 terrain features from DEM were extracted. Five widely used MLAs, MD, CART, SVM, RF, and GBDT, were trained and validated for lithological mapping. The results indicate that advanced MLAs, such as GBDT and RF, are highly effective for nonlinear modeling and learning with relative increases reaching 8.18%~11.82% for GBDT and 6.36%~10% for RF. Compared with optical imagery or terrain data alone, combining Sentinel-2 and DEM significantly improves the accuracy of lithological mapping, as it provides more comprehensive and precise spectral characteristics and spatial information. GBDT_Sen+DEM utilizing integrated data achieved the highest classification accuracy, with an overall accuracy of 63.18%. This study provides a case study for lithological mapping of areas with high vegetation cover at the local level. This also reinforces the idea that merging remote sensing and terrain data significantly enhances the precision and reliability of the lithological mapping methods.

KEYWORDS

GBDT, RF, lithological mapping, highly vegetated areas, Sentinel-2, SRTM DEM

1 Introduction

Lithological mapping of highly vegetated areas is an essential part of geological surveys and mineral resource exploration (Lu et al., 2021). On the one hand, challenges related to low comprehensive utilization of mineral resources, uneven regional development, ecological degradation, and environmental pollution in mineral resource exploitation underscore the importance of addressing these issues to achieve sustainable development in the mining industry (Wang and Li, 2020; Xie, 2020). On the other hand, the spatial distribution of rocks contributes to the extraction of mineral alteration information, selection of target areas within mining regions, and acquisition of multiscale structural control information (Shuai, 2022). Therefore, it provides theoretical support for delineating prospective mineralization zones and holds significant importance for geological prospecting and evaluation (Wang and Liu, 2020). However, traditional geological mapping encounters significant challenges in areas with high vegetation cover, including fieldwork difficulties, high costs, lengthy time requirements, and limited accessibility to certain areas.

Remote sensing technology enables efficient and large-scale identification of rock types due to its fast speed, low cost, and efficiency in terms of time and labor (Carli and Sgavetti, 2011; Pour and Hashim, 2014). However, rocks are largely obscured by soil in densely vegetated areas, with studies indicating that vegetation cover of only 10% or more can effectively conceal surface information (Siegal and Goetz, 1977; Ager and Milton, 2012). Extracting weak rock-type information using remote sensing is a major challenge in geological applications (Chen et al., 2012). "Geobotany" offers an important approach to address this challenge, enabling indirect rock-type identification by considering the relationship between vegetation and underlying substrates (Grebby et al., 2011). Rock types influence the composition of aboveground plant communities in two ways: nutrient provision to plants (Landeweert et al., 2001; Hahm et al., 2014), and the weathering depth, degree, and increased porosity of rocks, which impact water storage potential (Schwinning, 2010; Klos et al., 2018). Hahm's research showed that plant communities in different regions are correlated with hydrology and rock types, even under similar climatic conditions (Hahm et al., 2019). Do Amaral mapped three geological phases using indicator species (do Amaral et al., 2018). Qiao's research showed a significant impact of rock types on the spatiotemporal pattern changes in vegetation (Qiao et al., 2020).

High-resolution optical and radar remote sensing data, along with terrain information, are valuable for extracting rock-type information from densely vegetated areas. In the past decade, medium-resolution remote sensing imagery such as Landsat series and ASTER, has been extensively employed for rock type mapping in vegetated areas (Knepper, 1989; Langford, 2015; Han et al., 2021; Zeng et al., 2023). It establishes a strong foundation for rock-type identification by offering cost-effective, wide coverage, high spatial resolution (Chen et al., 2022; Zou et al., 2022), valuable indications of vegetation and rock-soil information, rich surface information and a small mixed pixel effect (Meroni et al., 2021). Sentinel-2 imagery is a new and freely accessible dataset that offers a high spatial resolution. It has been widely used in fields like geology, agriculture, and urban studies. Equipped with a multispectral sensor, it can capture visible and near-infrared spectral data making it an invaluable resource for lithological identification research. Radar is highly sensitive to surface physical characteristics, particularly the C-band response related to vegetation biomass, structure, and soil conditions. Digital Terrain Models (DTMs) can qualitatively and quantitatively reveal terrain variations reflecting subtle changes in rock types. The erosion and weathering resistance of rocks fundamentally shapes the terrain, and rocks with higher resistance to erosion form steeper terrain sections under similar conditions (Snyder et al., 2000; Montgomery, 2001). In stable landscapes, terrain slope indices can represent the erodibility of different rock types (Mills, 2003; Gallen, 2018). Radar and terrain data partially overcome the limitations of optical remote sensing in identifying and classifying rock types in vegetation- and shallow-covered areas (Gloaguen et al., 2019). In addition, compared to traditional remote sensing data, hyperspectral data can provide richer spectral information, thus reflecting the spectral response characteristics of different lithologies more accurately. This can provide effective technical support for geological exploration and mineral resource development (Chen L. et al., 2023).

An appropriate algorithm is one of the key factors contributing to achieving satisfactory classification results. Machine learning algorithms such as maximum likelihood (ML) (Grebby et al., 2011), partial least squares discriminant analysis (PLSDA) (Lu et al., 2021), support vector machine (SVM) (Othman and Gloaguen, 2014; Bachri et al., 2019), and random forest (RF) (Han et al., 2021) have been extensively used for rock classification in vegetation-covered areas because of the rapid advancement of machine learning. In Grebby's study, airborne multispectral imagery and laser scanning data were used to map rock types in the Troodos ophiolite. The self-organizing map algorithm achieved the highest accuracy (72.7%) among the algorithms utilized (Grebby et al., 2011). Othman and Gloaguen improved the lithologic map of the Mawat ophiolite complex in northeastern Iraq using an SVM classifier based on ASTER multispectral data, landform features, and texture data with overall accuracy (OA) of 79.28% (Othman and Gloaguen, 2014). Lu used multiple algorithms, including PLSDA, SVM, k-nearest neighbors and Bayesian, combined with Sentinel-1 and SRTM data, to map rock distribution of Huludao City in Liaoning Province, China. The highest accuracy of 0.444 was achieved using the PLSDA (Lu et al., 2021). Otele updated lithological mapping in the southern region of Cameroon using Landsat 7 imagery and a multilayer perceptron neural network and achieved an accuracy of 53.01% (Otele et al., 2021). Zeng compared the accuracy of KNN, MLC, and SVM algorithms in classifying basalt using Landsat 5 and ASTER imagery. The SVM model with Landsat 5 achieved the highest accuracy of 70.92% (Zeng et al., 2023). These methods establish a solid basis for the rapid classification of rocks using "geobotany" principles in remote sensing. Furthermore, HAN utilized the RF method to map Quaternary rock (including Pleistocene gravel, Holocene sand, Holocene clay, and Holocene gravel) in vegetation-covered areas of Vietnam based on multiple remote sensing data sources, achieving OA of 80.99% (Han et al.,

2021). This highlights the potential of the RF algorithm in geological mapping.

This study aims to achieve lithological mapping in areas with dense vegetation using Sentinel-2 and SRTM DEM data combined with MLAs. The study focuses on Zhangzhou City in Fujian Province as an example. We tested the performance of five popular MLAs to determine which one works best for this task. We carefully compared and analyzed the accuracy of the classification results achieved by each algorithm, gaining valuable insights into their effectiveness. Furthermore, we systematically evaluated the accuracy of classification results obtained from Sentinel-2 alone, DEM alone, and the combination of Sentinel-2 and DEM to better understand their impact on lithological mapping. This research contributes to the advancing field of geospatial science and remote sensing.

2 Study area and data

2.1 Study area

The study area is located around Daxi Town, Pinghe County, Zhangzhou City, Fujian Province. The specific administrative division map, true-color composite image, and DEM are shown in Figure 1. It is situated in a complex region with intersecting Nanshan Mountain Range and the Second Complex Uplift Belt of the Xinhua-Xia series. It has undergone multiple crustal movements and has exhibited complex tectonic faults. The primary geological formations consist of Lower Cretaceous, Jurassic strata, and Yanshanian intrusion rocks. The main rock types in this area include Rhyolite tuff (RhyT), Yingan tuff (YinT), Rhyolite (Rhy), sandstone (San), and granite (Gra). Sedimentary deposits consist mainly of conglomerate, gravel, and sandy soil layers (CGS).

The area is characterized by dense vegetation growth, with a forest coverage rate of 73.2% as of 2021. It exhibits a variety of plant community types and a complex hierarchical structure, resembling the vibrant landscapes of the South Asian tropical rainforests. The main vegetation types include coniferous forests, broad-leaved forests, mixed forests of conifers and broad-leaved trees, bamboo forests, shrubs, and grass slopes. The bedrock is mostly covered by a few outcrops (Shi and Wang, 2014). Additionally, in highly vegetated areas, the surface layer is heavily weathered, and the soil layer is thick, posing challenges for rock-type identification using remote sensing techniques.

2.2 Data and preprocessing

The Sentinel satellite carries the Multispectral Instrument (MSI), capturing imagery data in visible, near-infrared, and shortwave infrared bands. It provides a spatial resolution ranging from 10 m to 60 m (band details in Table 1). With a 5-day revisit period at the equator, it covers a swath width of 290 km (Chen et al., 2021). The S2 MSI Level-2A products offer preprocessed bottom-of-atmosphere reflectance images, including geometric correction, orthorectification, image registration, radiometric calibration, and atmospheric correction. A total of 78 images for the year 2021 with cloud coverage of less than 10% in the study area were obtained, and

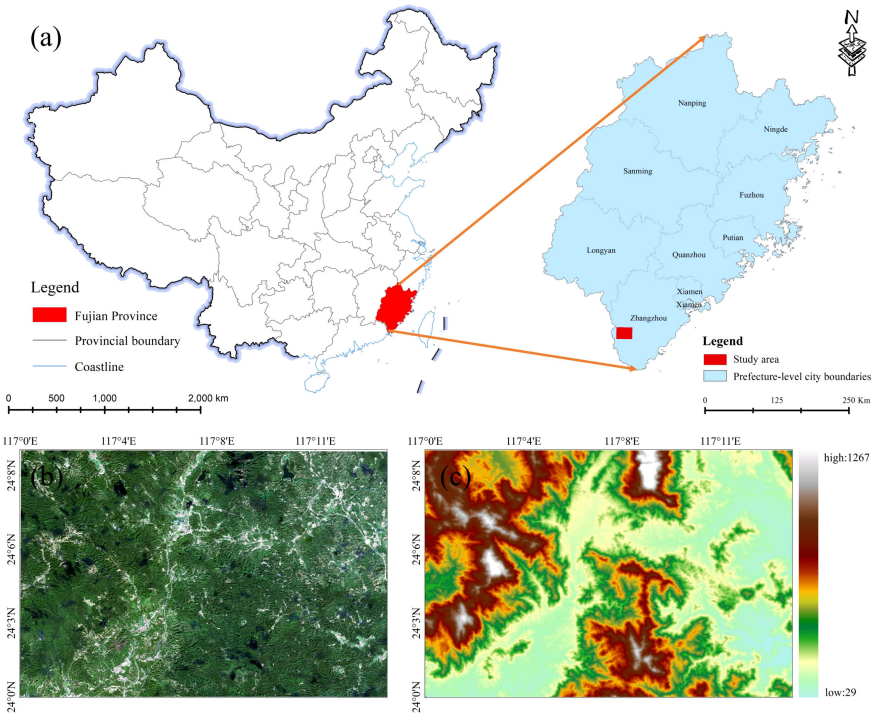


FIGURE 1
The study area for **(A)** a specific location in the administrative map of China, **(B)** Sentinel-2 tru color composite image and **(C)** Digital Elevation Model (DEM).

TABLE 1 Sentinel-2 band and spectral characteristics.

Band or index	Central wavelength/ formula	Resolution (meter)
B1	443.9nm (S2A)/442.3nm (S2B)	60m
B2	496.6nm (S2A)/492.1nm (S2B)	10m
B3	560nm (S2A)/559nm (S2B)	10m
B4	664.5nm (S2A)/665nm (S2B)	10m
B5	703.9nm (S2A)/703.8nm (S2B)	20m
B6	740.2nm (S2A)/739.1nm (S2B)	20m
B7	782.5nm (S2A)/779.7nm (S2B)	20m
B8	835.1nm (S2A)/833nm (S2B)	10m
B8A	864.8nm (S2A)/864nm (S2B)	20m
B9	945nm (S2A)/943.2nm (S2B)	60m
B11	1613.7nm (S2A)/1610.4nm (S2B)	20m
B12	2202.4nm (S2A)/2185.7nm (S2B)	20m
EVI	$2.5 \times (B8 - B4)/(B8 + 6 \times B4 - 7.5 \times B2 + 1)$	
NDBI	$(B12 - B4)/(B12 + B4)$	
NDWI	$(B3 - B4)/(B3 + B4)$	
LSWI	$(B4 - B11)/(B4 + B11)$	
GCVI	$(B8/B3) - 1$	

cloud masking using the QA60 band was performed to eliminate cloud effects. Bands B5, B6, B7, B8A, B11, and B12 were resampled to 10-meter spatial resolution using bilinear interpolation, and then the median composite was applied to the images.

The Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) is a digital representation of terrain elevation obtained using radar equipment carried by space shuttles, covering over 80% of the Earth's land surface. SRTM DEM data is acquired by emitting radar beams from a space shuttle towards the Earth's surface and measuring the returning signals. This technique bypasses cloud cover and vegetation obstruction to acquire surface elevation data. The SRTM DEM 30 m dataset was resampled to 10 m using bilinear interpolation to match the spatial resolution of the optical imagery.

2.3 Ground reference data

The reference data includes the regional geological map from the First Survey Team of Fujian Provincial Geological Survey Bureau in 1982 and lithology distribution maps obtained through a combination of vegetation suppression methods and manual visual interpretation by professional experts. We adopted the approach used in the lithological distribution map to classify land surface cover types and collected pixel-level sample data for 7 land cover classes based on the regional geological map: CGS (118), San (106), Rhy (112), RhyT (121), YinT (121), Gra (110), and water (39).

3 Methods

3.1 Method system construction

The main idea for pixel-wise rock classification is based on "geobotany". It uses differences in surface reflectance between vegetation communities to indirectly identify underlying rock types by considering terrain and landform characteristics. To achieve this, we construct five classical machine learning algorithms: minimum distance algorithm (MD), classification and regression trees (CART), support vector machine (SVM), random forest (RF), and gradient boosting decision tree (GBDT) (refer to Figure 2).

The following research plan aims to reduce computational burden and accelerate the production of optimal lithological spatial distribution maps. The available geological data is divided into training and validation sets at a 7:3 ratio to assess the model's generalization ability, which evaluates its performance on new data. Then, five classification algorithms are trained and optimized using the preprocessed Sentinel-2 remote sensing image and a sample dataset. The optimized models are validated and evaluated for accuracy using validation samples. Subsequently, the optimal model is used to classify rock types and generate a lithological map based on Sentinel-2 data. In the subsequent studies based on SRTM DEM (alone) and combination of Sentinel-2 with SRTM DEM, at least two classifiers that performed well in the previous step are used for validation and accuracy evaluation. Similarly, the optimal model is used to generate lithological maps based on SRTM DEM (alone) and combined with Sentinel-2 and SRTM data.

3.2 Feature extraction

B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12, were selected from Sentinel-2. It should be noted that B1 represents the aerosol band and B9 represents the water vapor band, therefore, they were not used in this study. Additionally, five spectral features were computed, which are crucial for rock or water body identification: enhanced vegetation index (EVI), green chlorophyll vegetation index (GCVI), normalized difference building index (NDBI), normalized difference water index (NDWI), and Land Surface Water Index (LSWI). This study employed EVI to indicate vegetation growth status, as it is well known that normalized difference vegetation index (NDVI) can saturate areas with high vegetation cover (Huete et al., 1997). GCVI is sensitive to chlorophyll and can be used to identify agricultural areas (Huete et al., 2002). NDBI is useful for identifying built-up areas (Benbahria et al., 2018). NDWI and LSWI are vegetation indices that are highly sensitive to surface water (Jeong et al., 2012).

For SRTM DEM, six terrain features, including elevation (E), slope, aspect, vertical curvature, horizontal curvature, and Gaussian curvature, were obtained using terrain analysis algorithms proposed by Florinsky (2016) and Safanelli (Safanelli et al., 2020). Additionally, surface roughness (SR), high integral (HI), and surface index (SI) were calculated within a 3 × 3 window,

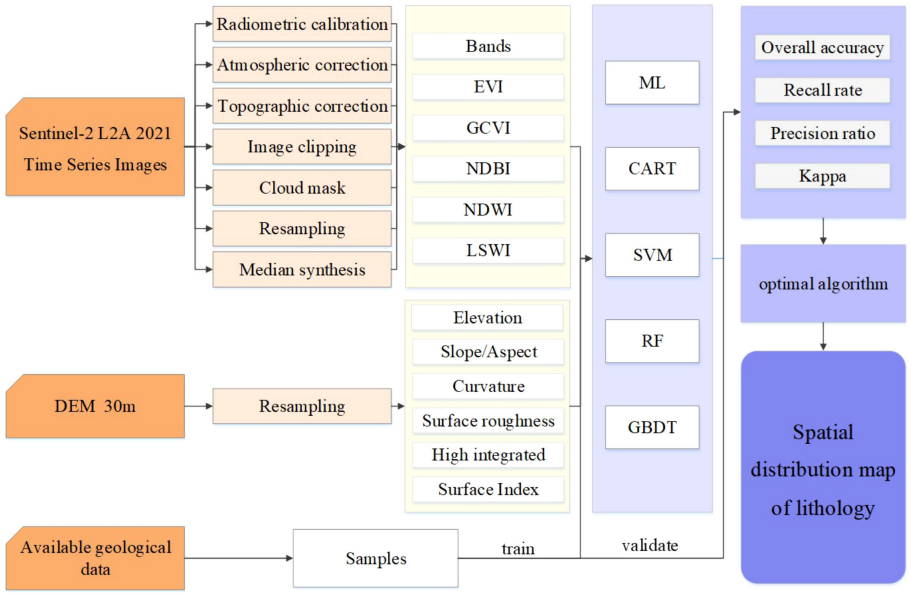


FIGURE 2
The flowchart. "DEM" for digital elevation model, "EVI" for enhanced vegetation index, "GCVi" for green chlorophyll vegetation index, "NDBI" for normalized difference building index, "NDWI" for normalized difference water index, and "LSWI" for land surface water index.

providing indications of terrain features associated with rock type information.

$$SR = 1/cos(slop)$$

$$HI = (E_{mean} - E_{min})/(E_{max} - E_{min})$$

$$SI = \left(\frac{HI - HI_{min}}{HI_{max}}\right) \times \left(\frac{H - H_{min}}{H_{max}}\right) - \left(\frac{SR - (1 + SR_{min})}{SR_{max}}\right)$$

Where E is extracted directly from the DEM. *E_mean*, *E_max*, and *E_min* represent the mean, maximum, and minimum values within the moving window, respectively. *SR* is the ratio of grid surface area to projected area, which is used to quantify tectonic landform changes. A higher *SR* value indicates a more severe degree of regional deformation (Han et al., 2021). *SI* is a new efficient index that simultaneously depicts the preserved and eroded portions of the landscape (Andreani et al., 2014).

3.3 Classification algorithms for lithological mapping

The Minimum Distance (MD) algorithm is widely used for classification and pattern recognition (Wacker and Landgrebe, 1972). It assigns input data to the closest category based on computed distances between categories, using methods such as Euclidean, Manhattan, and Minkowski distances. In the context of lithological mapping, this algorithm can be used to infer the lithology of specific locations based on their proximity to known geological features or data points (Pal et al., 2020). MD is simple, easy to implement, and well-suited for problems with distinct category boundaries.

Classification and regression trees (CART) is a tree-based algorithm for classification and regression (Friedl and Brodley, 1997; Pal and Mather, 2003). It recursively splits and evaluates input data to build a tree-like model, where nodes represent features, branches represent feature values, and leaf nodes represent final results. Applying CART to lithological mapping involves systematic interpretation of geological features for accurate lithological classification. Visualizing the decision tree offers insights into hierarchical feature divisions that lead to lithological categorization, enhancing our understanding of how different features influence prediction accuracy (Serbouti et al., 2022). CART has a simple structure, making it easy to understand, interpret, and generate decision-making rules. It's worth noting that the effectiveness of the CART decision tree method relies on the quality and relevance of selected features, as well as the representation of distinct lithological classes in the dataset (Lewis, 2000).

Support vector machine (SVM) is a non-parametric classifier widely used for binary and multi-class classification tasks (Pal and Mather, 2005; Mountrakis et al., 2011). It aims to minimize structural risk by identifying an optimal hyperplane in the feature space that maximizes the margin between samples of different classes. SVM can be used for lithologic classification by selecting training and testing pixels, training the SVM classifier using a kernel function, and mapping the input data (such as spectra, textures, topography, gamma-ray and land temperature) (Othman and Gloaguen, 2014; Harris and Grunsky, 2015; Yang et al., 2018; Chen C. et al., 2023) into a higher-dimensional space to find a hyperplane that separates the different classes. SVM has the advantage of handling high-dimensional data, nonlinear classification problems, and small sample sizes (Shebl and Csamer, 2021; Shebl et al., 2021). However, it also has limitations

in terms of selecting appropriate kernel functions, computational intensity, and sensitivity to noise and outliers (Othman and Gloaguen, 2017).

Random forest (RF) works by creating multiple decision trees on randomly selected subsets of the data and then combining the results to make a final prediction (Pal, 2005; Belgiu and Drăguţ, 2016). It has significant advantages with the ability to combine multiple remote sensing and data sources in lithology mapping as it improves its generalization ability by randomly selecting input or input combinations at each node (Breiman, 2001). It is especially effective for processing high-dimensional and noisy input data and can overcome the interference of vegetation coverage, thereby improving the accuracy of lithological mapping (Harris and Grunsky, 2015; Bachri et al., 2019). However, caution should be exercised when fine-tuning parameters for optimal outcomes and effectively managing computational expenses, particularly when dealing with substantial datasets. The risk of overfitting due to an abundance of trees or noisy data should be considered, along with its limited efficacy with imbalanced datasets (Guo et al., 2022).

The Gradient Boosting Decision Tree (GBDT) is an iterative ensemble learning algorithm that constructs a strong prediction model (Yang et al., 2018; Xu et al., 2020). It builds multiple decision trees iteratively, using the residual between the current predicted value and the true label to train each tree. Each iteration adjusts the predicted value to approximate the true label. The final prediction is obtained by combining the predictions of all the trees. Although the algorithm is rarely used in lithology mapping in high vegetation cover areas, studies have shown that it is effective in handling large datasets and high-dimensional feature spaces, and it is robust to noise, outliers, and missing data (Lemercier et al., 2012; Zhou et al., 2020; Cai et al., 2022).

3.4 Accuracy assessment

Accuracy assessment objectively evaluates the performance of remote sensing algorithms and models, providing a reliable foundation for remote sensing applications and decision-making (Hay, 1988). It validates the accuracy of tasks like image classification and object detection, helping to determine the reliability and feasibility of the results. The Confusion Matrix is a tabular representation used to evaluate the performance of a classification model (Comber et al., 2012; Salmon et al., 2015). Therefore, we calculate four evaluation metrics using the Confusion Matrix to assess the classification results and optimize the model: overall accuracy (OA), recall (R), precision (P), and Kappa coefficient.

4 Results

4.1 Model tuning

Model tuning optimizes the performance and generalizability of the machine learning model by adjusting its parameters or hyperparameters, improving its effectiveness in addressing real-world problems. By systematically adjusting parameters, the model can better adapt to data patterns and characteristics. Proper parameter settings also reduce computational resources required for training and prediction. Through tuning, structure of the model and parameter selection are optimized, resulting in reduced computational costs and time consumption.

The sample dataset was randomly divided into training and validation sets at a 7:3 ratio for model training and tuning based on Sentinel-2 imagery. To determine the best settings for our models, we test them by using different parameters in the validation set. We begin by defining a range of parameter values and then try different values within that range. Finally, we select the parameter values that make our models perform the best on the validation set. We can show this process on a chart, where the horizontal line represents different parameter choices, and the vertical line shows the performance of the model (typically, model performance metrics such as OA). The tuning results for each model are shown in Figure 3. The MD algorithm achieved optimal performance using the Mahalanobis distance metric. The CART was tuned with a maximum of 50 nodes and a minimum of 5 nodes. The gamma coefficient of SVM model was set to 20, and the Cost parameter was set to 10. The RF utilized 210 trees and a minimum of 3 leaf nodes. The GBDT employed 80 trees and a maximum of 70 leaf nodes.

In machine learning and data modeling, the optimal parameters of a model typically change with different datasets and feature combinations. Once we have identified at least two promising classifiers based on Sentinel-2, we apply the same tuning methodology to both the DEM data and the combination of Sentinel-2 and DEM data to ensure optimal model performance.

4.2 Lithology classification using Sentinel-2 image

The model accuracy and validation accuracy of five typical machine learning algorithms were summarized (as shown in Figure 4). The SVM algorithm had higher model accuracy but lower validation accuracy, possibly indicating overfitting due to linear inseparability of rock categories and limited sample size. MD and CART performed poorly, with model accuracy below 70% and validation accuracies below 0.4, suggesting limitations in handling complex structural information. RF and GBDT showed better performance as ensemble algorithms with strong learning capabilities with OA 46.82% and 45%, and Kappa 0.371 and 0.349 respectively.

Figure 5 illustrates rock-type accuracy achieved by the five machine learning algorithms. The “water” class consistently exhibits high accuracy (close to 100%) across all models, indicating the effectiveness of the algorithms. In terms of class accuracy, the RF model demonstrates notable improvements in precision for “RhyT”, “San”, and “YinT” classes, as well as increased recall for “RhyT” and “YinT” classes. For instance, precision of San improved by 13.1% to 20%, and recall of “RhyT” increased by 10% to 36.6%. The GBDT achieved overall high accuracy through

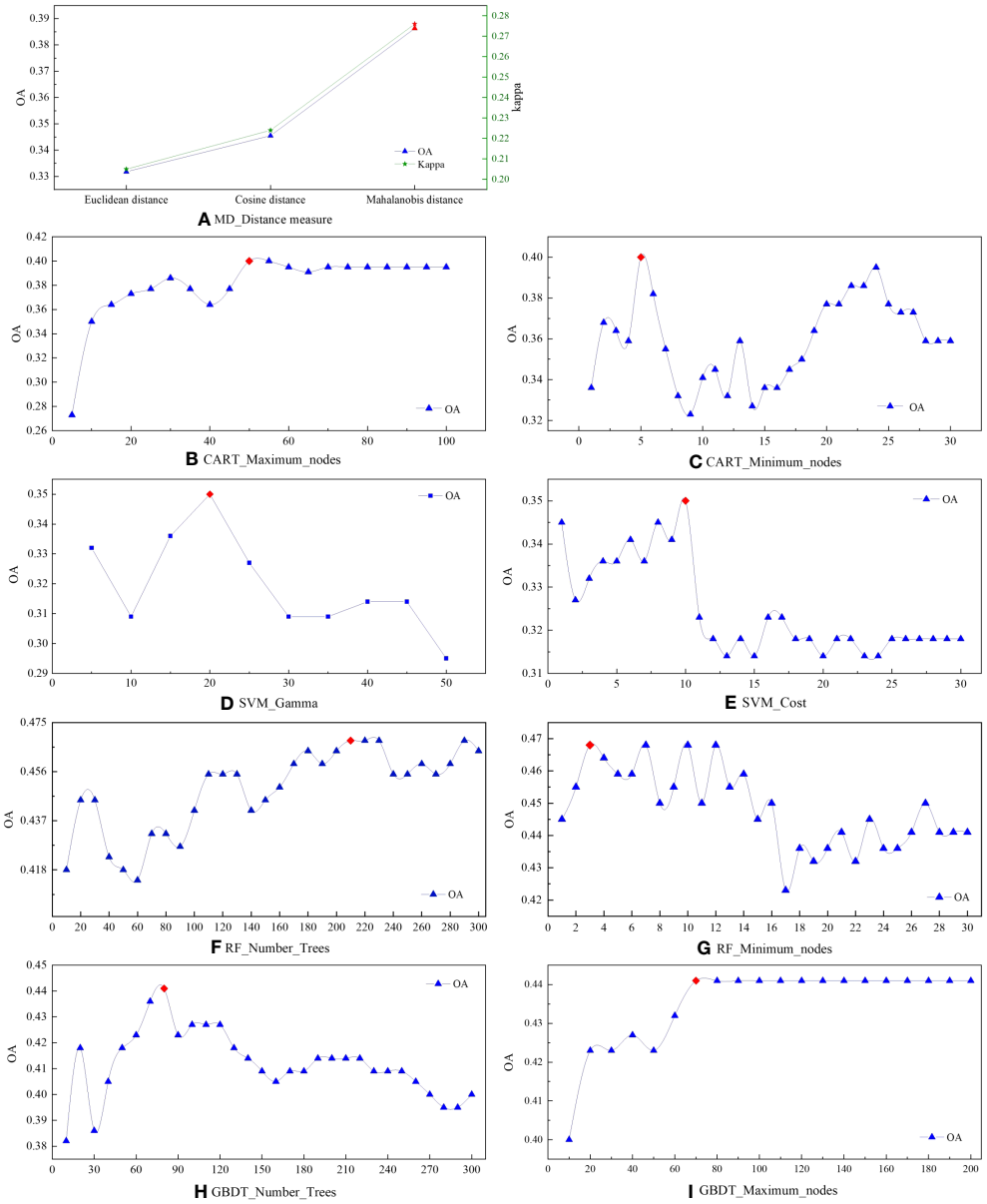


FIGURE 3
Adjust the parameters for **(A)** the distance measure method of MD, **(B)** the maximum and **(C)** minimum nodes of Cart, **(D)** gamma and **(E)** cost of SVM, **(F)** the number of trees and **(G)** minimum nodes of RF, **(H)** the number of trees and **(I)** maximum nodes of GBDT using validation samples. Each subscription below the figure is labeled as "Algorithms Parameter" to indicate the specific parameters used for each classifier.

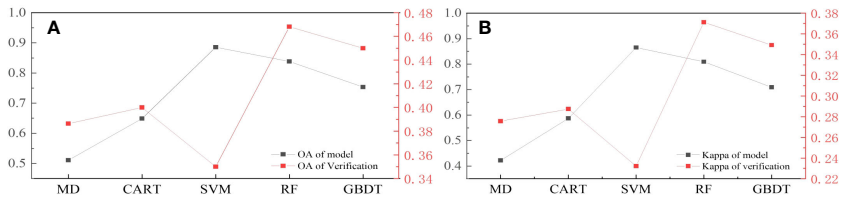


FIGURE 4
(A) Overall accuracy (OA) and **(B)** Kappa of five classifiers in lithological mapping. The black dots represent model accuracy, whereas the red dots represent model validation accuracy.

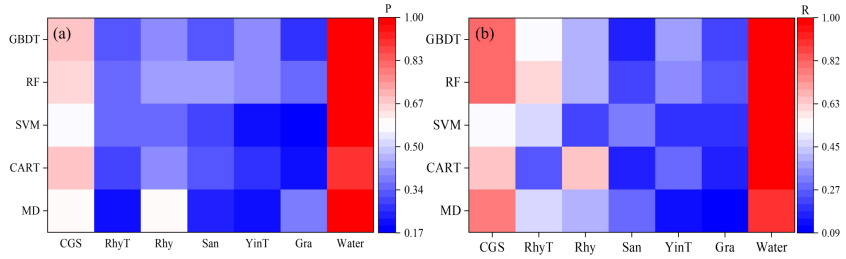


FIGURE 5
(A) Precision (P) and **(B)** recall (R) of rock types using five classifiers (MD, CART, SVM, RF and GBDT) based on Sentinel-2.

significant improvements in precision and recall for "CGS" and "San". Specifically, precision and recall of "CGS" increased by 1.3% to 11.8% and 2.7% to 29%, respectively.

4.3 Lithology classification using SRTM DEM

Using RF and GBDT algorithms for lithology classification achieves higher accuracy based on the accuracy assessment results (see Figures 6, 7). The overall classification accuracy is approximately 49% with a kappa coefficient of approximately 0.4. When comparing individual rock types, both algorithms show higher accuracy for "CGS" and "Rhy". "CGS" plays a significant role in terrain morphology and exhibits a certain relationship with terrain features. "Rhy" is commonly found in uplift zones associated with tectonic landforms.

The lithology classification map generated using SRTM DEM (Digital Elevation Model) data shows higher accuracy than Sentinel-2 imagery. The overall improvement in lithology classification accuracy is mainly attributed to better classification of "Rhy" and "San", despite a slight decrease in the accuracy of classifying "RhyT" and "Gra". This can be explained by the strong correlation between spatial distribution and landform morphology. "San" is commonly found in low-lying areas, while "Rhy" is associated with volcanic activity and its distribution relates to volcanic topography. However, it should be noted that the accuracy of "water" based on SRTM DEM is significantly lower than that of Sentinel-2 imagery due to the limited information about water bodies in DEM, which primarily focuses on surface terrain elevation. As a result, there may be errors or omissions in the parts of the lithology classification that involve water bodies, as it does not capture the characteristics of water flow from higher to lower elevations. Overall, the utilization of DEM improves the accuracy of lithological classification by capturing the relationship between lithological variations and landform morphology.

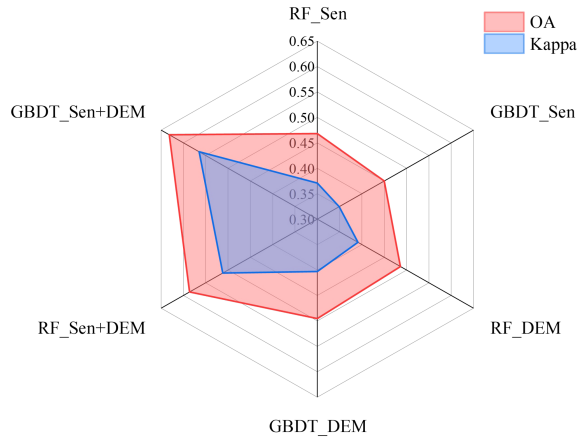


FIGURE 6
Overall accuracy (OA) and Kappa of the RF and GBDT in lithological classification based on Sentinel-2 (alone), DEM (alone), and combined Sentinel-2 and DEM data. "Sen" is the abbreviation for Sentinel-2.

4.4 Lithology classification using combined Sentinel-2 and SRTEM DEM

The RF and GBDT are also utilized in a lithology mapping study that combines Sentinel-2 imagery and SRTM DEM data. The extraction of lithological information can be greatly improved by integrating optical remote sensing data with terrain morphology features, resulting in more comprehensive and accurate spatial information. Both algorithms (RF and GBDT) show a significant improvement in accuracy compared to Sentinel-2 (alone) or SRTM DEM (alone), achieving OA of approximately 60% (see Figure 6). The GBDT demonstrates a more precise lithology classification, achieving OA of 63.18% and a Kappa of 0.565. Compared with RF, GBDT improved OA and Kappa by 4.54% and 0.053, respectively.

The joint utilization of Sentinel-2 imagery and SRTM DEM, along with advanced machine learning algorithms like GBDT, significantly improves the classification accuracy of various lithological categories, as depicted in Figure 7. The performance in classifying "CGS", "San" and "Rhy" is notably enhanced. Moreover, the classification of "water" achieves optimal results.

Based on the above study, lithological mapping was performed using RF for Sentinel-2 (alone), GBDT for DEM (alone) and the combination of Sentinel-2 and DEM (Figure 8). In the lithological map based on Sentinel-2, the limited distinguishability of rock types other than "CGS" and the presence of significant "speckle" phenomenon may be attributed to the optical sensor limitations of Sentinel-2, which are affected by cloud cover, atmospheric interference, and vegetation obstruction. In contrast, DEM data

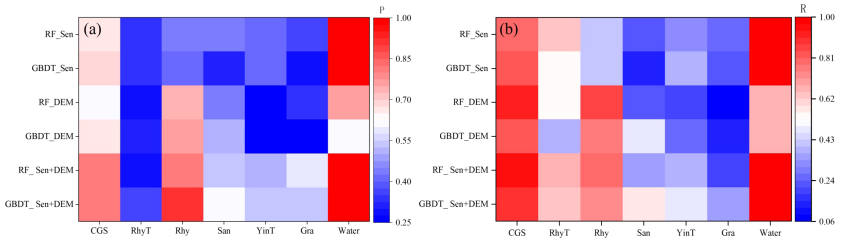


FIGURE 7
(A) Precision (P) and **(B)** recall (R) of rock types using RF and GBDT based on Sentinel-2 (alone). DEM (alone), and combined Sentinel-2 and DEM.

offers continuous surface elevation and terrain information, resulting in a clearer representation of lithological distribution. The combination of Sentinel-2 imagery and DEM allows for leveraging their respective strengths, leading to more comprehensive and accurate rock-type classification results.

5 Discussion

Using SRTM DEM data for lithological classification provides more accurate results than Sentinel-2 imagery, particularly for "Rhy" and "San". However, "water" requires additional data sources or methods to enhance accuracy. Sentinel-2 imagery offers rich spectral information for analyzing rock characteristics such as color and reflectance. On the other hand, SRTM DEM provides elevation data for terrain morphology revealing surface undulations and morphological features complementing lithological classification. The accuracy and reliability of lithological classification are substantially improved through the integration of these data sources, meticulous consideration of spectral and terrain features, and the utilization of state-of-the-art machine learning algorithms.

Choosing the correct classification algorithm is crucial for achieving satisfactory results in land cover classification using remote sensing data. In this study, we evaluated the performance of five classical machine learning algorithms for rock identification in densely vegetated areas. Our findings indicate that complex algorithms like RF and GBDT outperformed the others, consistent with previous research (He et al., 2015; Othman and Gloaguen, 2017). RF exhibited robustness, stability, and the ability to handle feature selection and outliers. GBDT demonstrated strong fitting and generalization capabilities, making it suitable for capturing complex nonlinear relationships.

Overlaying the classification result map with geological contour maps allows for visual analysis of the spatial distribution of rock types (Figure 8). In this study, we observed consistency between the classification results and existing geological maps for "GCS", and "Rhy", indicating a good match. Fine sand and siltstone were primarily found in valley areas, corresponding to the gentle topographic features of valleys. These areas are often impacted by

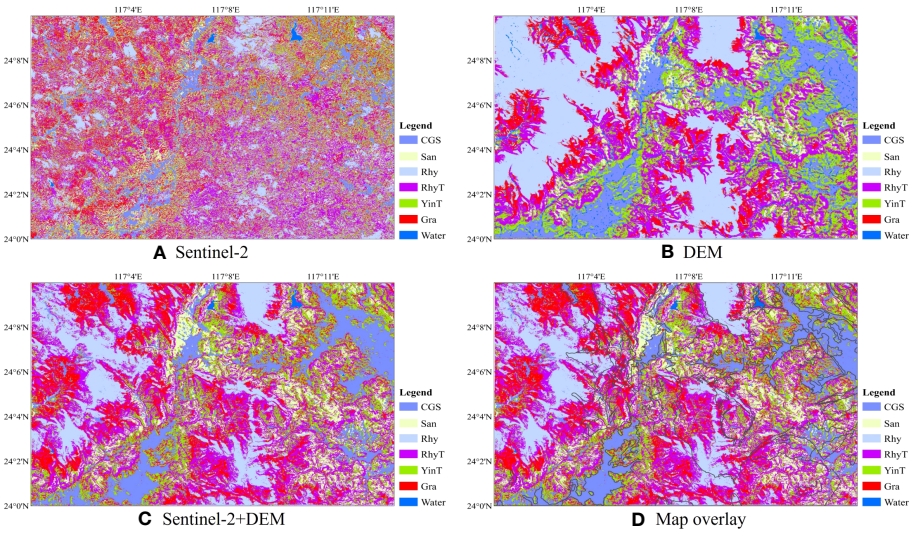


FIGURE 8
Lithological mapping of the study area from **(A)** Sentinel-2 (alone), **(B)** DEM (alone), **(C)** combined Sentinel-2 and DEM, and **(D)** overlay of the rock classification map and existing geological contour map.

human activities for agricultural, residential, and industrial purposes, leading to modified surface cover types. The mixing of the other three rock types is prominent. However, further observation shows that the classification performance of granite and andesitic tuff in the northern region is consistently high. In the southern region, although there is a mixing phenomenon in the rhyolite area, rhyolite remains dominant. These findings are valuable for studying the geological features, distribution of rock types, and geological evolution in the study area.

In areas with dense vegetation, the presence of subsurface rock types can potentially affect the distribution and characteristics of vegetation (Ott, 2020). However, the growth and distribution of vegetation are influenced by various factors such as soil type, moisture levels, light intensity, and climate conditions (Yang et al., 2021). The complex interactions among these factors make the relationship between vegetation and subsurface rock types complex. In our study, we used maps created from field surveys as a reference for our sample data. Although we tried to avoid including samples from border regions to reduce potential errors due to geographic bias, there may still be some mistakes in our pixel-based sample data, including issues with the representativeness of the sample and measurement accuracy. Additionally, we didn't thoroughly analyze how rock types correlate with vegetation and terrain features. This oversight could reduce the certainty of our research results, especially if the connections between these factors are weak or unclear. To address this, in our future research, we will focus on a detailed correlation analysis to better understand these relationships. This will improve the accuracy of rock classification and provide more reliable tools and data support for geological research, resource exploration, and related fields. As well as,future research should consider incorporating more extensive on-site validation efforts to confirm the accuracy of our classification results.

Moreover, the high variability within rock classes and similarities between different rock types (Otele et al., 2021) contribute to the complexity of vegetation and terrain features, making it challenging to directly infer subsurface rock information solely based on vegetation indices. Recent advancements in deep learning algorithms have significantly enhanced their application in remote sensing for land cover classification (Sun et al., 2022). Particularly, in areas with dense vegetation cover, deep learning algorithms have demonstrated higher accuracy in identifying rock types (Otele et al., 2023; Pan et al., 2023). These algorithms leverage the ability to learn from extensive image data, enabling them to explore and capture intricate relationships and feature representations among different land cover classes (Dimitrovski et al., 2023). They effectively address the challenges posed by vegetation interference and complex land cover backgrounds, resulting in improved classification accuracy and stability. Additionally, deep learning algorithms exhibit a certain level of generalization capability, performing well across different regions and datasets (Yasir et al., 2023). These findings provide valuable insights and pave the way for future research in this field.

6 Conclusions

Through the integration of optical remote sensing imagery and terrain data, coupled with the utilization of advanced algorithms, the potential to discern various rock types becomes attainable. This integration significantly enhances the precision and dependability of lithological mapping within regions characterized by dense vegetation cover. When juxtaposed with the individual use of optical imagery or terrain data, the amalgamation of these two datasets for rock classification purposes exhibits a synergistic effect, enriching the informational representation of rock types.

Furthermore, the employment of sophisticated Machine Learning Algorithms (MLAs) adeptly harnesses and delves into the wealth of feature information stemming from these diverse data sources. These algorithms, characterized by their resilient nonlinear modeling and learning capabilities, enable the capture of intricate relationships within land cover. This is achieved through the extensive use of sample data, ultimately resulting in improved accuracy and stability in lithology classification.

It's worth noting that the GBDT and RF algorithms employed in this study exhibit robustness and resilience, rendering them applicable across various research regions and for the categorization of rock types involving different combinations of data types. This holds significant implications for geological surveys and mineral exploration.

Looking ahead to future research endeavors, our focus will remain on deepening our understanding of the interplay between rock types and vegetation as well as terrain features. This ongoing exploration aims to further enhance the accuracy of rock classification. Additionally, the application of deep learning algorithms presents an intriguing avenue, particularly in addressing the challenges posed by rock classification in areas with dense vegetation cover.

Data availability statement

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

Author contributions

Conceptualization, YD and YW. Methodology, YC, GL and YD. Software, YC. Investigation, PS. Writing—original draft preparation, YC. Writing—review and editing, YC, FZ and YW. Visualization, YC. Supervision, YD and GL. Funding acquisition, PS. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by China Geological Survey Project (project number DD20230591), Science and Technology Innovation Fund of Command Center of Integrated Natural Resources Survey Center (KC20220020) and Soft Science Research Project of Xi'an Science and Technology Plan (22RKYJ0044).

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

Ager, C. M., and Milton, N. M. (2012). Spectral reflectance of lichens and their effects on the reflectance of rock substrates. *Geophysics* 52, 898–906. doi: 10.1190/1.1442360

Andreani, L., Stanek, K. P., Gloaguen, R., Krentz, O., and Domínguez-González, L. (2014). DEM-based analysis of interactions between tectonics and landscapes in the Ore Mountains and Eger Rift (East Germany and NW Czech Republic). *Remote Sens.* 6, 7971–8001. doi: 10.3390/rs6097971

Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A. C., and Benbouziane, A. (2019). Machine learning algorithms for automatic lithological mapping using remote sensing data: A case study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *ISPRS Int. J. Geo-Information* 8, 248. doi: 10.3390/ijgi8060248

Belgiu, M., and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. photogrammetry Remote Sens.* 114, 24–31. doi: 10.1016/j.isprsjprs.2016.01.011

Benbahria, Z., Sebari, I., Hajji, H., and Smiej, M. F. (2018). "Automatic mapping of irrigated areas in mediteranean context using landsat 8 time series images and random forest algorithm," in *Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*. Valencia, Spain: Institute of Electrical and Electronics Engineers 7986–7989. doi: 10.1109/IGARSS.2018.8517810

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cai, W., Wei, R., Xu, L., and Ding, X. (2022). A method for modelling greenhouse temperature using gradient boost decision tree. *Inf. Process. Agric.* 9, 343–354. doi: 10.1016/j.inpa.2021.08.004

Carli, C., and Sgavetti, M. (2011). Spectral characteristics of rocks: Effects of composition and texture and implications for the interpretation of planet surface compositions. *Icarus* 211, 1034–1048. doi: 10.1016/j.icarus.2010.11.008

Chen, C., Liang, J., Xie, F., Hu, Z., Sun, W., Yang, G., et al. (2022). Temporal and spatial variation of coastline using remote sensing images for Zhoushan archipelago, China. *Int. J. Appl. Earth Observation Geoinformation* 107, 102711. doi: 10.1016/j.jag.2022.102711

Chen, Y., Hou, J., Huang, C., Zhang, Y., and Li, X. (2021). Mapping maize area in heterogeneous agricultural landscape with multi-temporal Sentinel-1 and Sentinel-2 images based on random forest. *Remote Sens.* 13, 2988. doi: 10.3390/rs13152988

Chen, C., Liang, J., Yang, G., and Sun, W. (2023). Spatio-temporal distribution of harmful algal blooms and their correlations with marine hydrological elements in offshore areas, China. *Ocean Coast. Manage.* 238, 106554. doi: 10.1016/j.ocecoaman.2023.106554

Chen, S., Liu, Y., Yang, Q., Zhou, C., and Zhao, L. (2012). Lithology classification of vegetated area by satellite hyperspectral remote sensing. *J. Jilin Univ. (Earth Sci. Edition)* 42, 1959–1965. doi: 10.13278/j.cnki.jjuese.2012.06.010

Chen, L., Sui, X., Liu, R., Chen, H., Li, Y., Zhang, X., et al. (2023). Mapping alteration minerals using ZY-1 02D hyperspectral remote sensing data in coalbed methane enrichment areas. *Remote Sens.* 15, 3590. doi: 10.3390/rs15143590

Comber, A., Fisher, P., Brunsdon, C., and Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* 127, 237–246. doi: 10.1016/j.rse.2012.09.005

Dimitrovski, I., Kitanovski, I., Kocev, D., and Simidjievski, N. (2023). Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS J. Photogrammetry Remote Sens.* 197, 18–35. doi: 10.1016/j.isprsjprs.2023.01.014

do Amaral, C. H., de Almeida, T. I. R., de Souza Filho, C. R., Roberts, D. A., Fraser, S. J., Alves, M. N., et al. (2018). Characterization of indicator tree species in neotropical environments and implications for geological mapping. *Remote Sens. Environ.* 216, 385–400. doi: 10.1016/j.rse.2018.07.009

Florinsky, I. (2016). *Digital terrain analysis in soil science and geology* (Pushchino, Russia: Academic Press). doi: 10.1016/C2010-0-65718-X

Friedl, M. A., and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61, 399–409. doi: 10.1016/S0034-4257(97)00049-7

Gallen, S. F. (2018). Lithologic controls on landscape dynamics and aquatic species evolution in post-orogenic mountains. *Earth Planetary Sci. Lett.* 493, 150–160. doi: 10.1016/j.epsl.2018.04.029

Gloaguen, R., Fuchs, M., Khodadadzadeh, M., Ghamisi, P., and Lorenz, S. (2019). "Multi-source and multi-scale imaging-data integration to boost mineral mapping," in *Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium*. Valencia, Spain: Institute of Electrical and Electronics Engineers. doi: 10.1109/IGARSS.2019.8900426

Grebby, S., Naden, J., Cunningham, D., and Tansey, K. (2011). Integrating airborne multispectral imagery and airborne LiDAR data for enhanced lithological mapping in vegetated terrain. *Remote Sens. Environ.* 115, 214–226. doi: 10.1016/j.rse.2010.08.019

Guo, S., Yang, C., He, R., and Li, Y. (2022). Improvement of lithological mapping using discrete wavelet transformation from sentinel-1 SAR data. *Remote Sens.* 14, 5824. doi: 10.3390/rs14225824

Hahm, W. J., Rempe, D. M., Dralle, D. N., Dawson, T. E., Lovill, S. M., Bryk, A. B., et al. (2019). Lithologically controlled subsurface critical zone thickness and water storage capacity determine regional plant community composition. *Water Resour. Res.* 55, 3028–3055. doi: 10.1029/2018WR023760

Hahm, W. J., Riebe, C. S., Lukens, C. E., and Araki, S. (2014). Bedrock composition regulates mountain ecosystems and landscape evolution. *Proc. Natl. Acad. Sci.* 111, 3338–3343. doi: 10.1073/pnas.1315667111

Han, S., Shuai, S., Guo, W., and Yang, P. (2021). Automatic classification method of quaternary lithology in vegetation cover area combining spectral, textural, topographic, geothermal, and vegetation. doi: 10.3233/ATDE210216

Harris, J., and Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Comput. geosciences* 80, 9–25. doi: 10.1016/j.cageo.2015.03.013

Hay, A. (1988). The derivation of global estimates from a confusion matrix. *Int. J. Remote Sens.* 9, 1395–1398. doi: 10.1080/01431168808954945

He, J., Harris, J., Sawada, M., and Behnia, P. (2015). A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic. *Int. J. Remote Sens.* 36, 2252–2276. doi: 10.1080/01431161.2015.1035410

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. doi: 10.1016/S0034-4257(02)00096-2

Huete, A., Liu, H., Batchily, K., and Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* 59, 440–451. doi: 10.1016/S0034-4257(96)00112-5

Jeong, S., Kang, S., Jang, K., Lee, H., Hong, S., and Ko, D. (2012). Development of Variable Threshold Models for detection of irrigated paddy rice fields and irrigation timing in heterogeneous land cover. *Agric. Water Manage.* 115, 83–91. doi: 10.1016/j.agwat.2012.08.012

Klos, P. Z., Goulden, M. L., Riebe, C. S., Tague, C. L., O'Geen, A. T., Flinchum, B. A., et al. (2018). Subsurface plant-accessible water in mountain ecosystems with a Mediterranean climate. *Wiley Interdiscip. Reviews: Water* 5, e1277. doi: 10.1002/wat2.1277

Knepper, D. H. Jr. (1989). "Mapping hydrothermal alteration with Landsat thematic mapper data," in *Remote Sensing in Exploration Geology: Golden, Colorado to Washington, DC, June 30–July 8, 1989*, Washington, America, mineral deposits of North America. vol. 182. , 13–21. doi: 10.1002/9781118669877.ch3

Landeweert, R., Hoffland, E., Finlay, R. D., Kuyper, T. W., and van Breemen, N. (2001). Linking plants to rocks: ectomycorrhizal fungi mobilize nutrients from minerals. *Trends Ecol. Evol.* 16, 248–254. doi: 10.1016/S0169-5347(01)02122-X

Langford, R. L. (2015). Temporal merging of remote sensing data to enhance spectral regolith, lithological and alteration patterns for regional mineral exploration. *Ore Geology Rev.* 68, 14–29. doi: 10.1016/j.oregeorev.2015.01.005

Lemercier, B., Lacoste, M., Loum, M., and Walter, C. (2012). Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma*, 171, 75–84. doi: 10.1016/j.geoderma.2011.03.010

Lewis, R. J. (2000). "An introduction to classification and regression tree (CART) analysis," in *Proceedings of the Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, Vol. 14.

Lu, Y., Yang, C., and Meng, Z. (2021). Lithology discrimination using Sentinel-1 dual-pol data and SRTM data. *Remote Sens.* 13, 1280. doi: 10.3390/rs13071280

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., et al. (2021). Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and-2. *Remote Sens. Environ.* 253, 112232. doi: 10.1016/j.rse.2020.112232

Mills, H. H. (2003). Inferring erosional resistance of bedrock units in the east Tennessee mountains from digital elevation data. *Geomorphology* 55, 263–281. doi: 10.1016/S0169-555X(03)00144-2

Montgomery, D. R. (2001). Slope distributions, threshold hillslopes, and steady-state topography. *Am. J. Sci.* 301, 432–454. doi: 10.2475/ajs.301.4-5.432

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS J. photogrammetry Remote Sens.* 66, 247–259. doi: 10.1016/j.isprsjprs.2010.11.001

Otele, C. G. A., Onabid, M. A., and Assembe, P. S. (2023). *Design and Implementation of an Automatic Deep Stacked Sparsely Connected Auto-encoder (ADSSCA) Neural Network Architecture for Lithological Mapping under thick Vegetation using Remote Sensing.* doi: 10.21203/rs.3.rs-2537926/v1

Otele, C. G. A., Onabid, M. A., Assembe, P. S., and Nkenlifack, M. (2021). Updated lithological map in the Forest zone of the Centre, South and East regions of Cameroon using multilayer perceptron neural network and Landsat images. *J. Geosci. Environ. Prot.* 9, 120–134. doi: 10.4236/gep.2021.96007

Othman, A. A., and Gloaguen, R. (2014). Improving lithological mapping by SVM classification of spectral and morphological features: The discovery of a new chromite body in the Mawat ophiolite complex (Kurdistan, NE Iraq). *Remote Sens.* 6, 6867–6896. doi: 10.3390/rs6086867

Othman, A. A., and Gloaguen, R. (2017). Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq. *J. Asian Earth Sci.* 146, 90–102. doi: 10.1016/j.jseaes.2017.05.005

Ott, R. F. (2020). How lithology impacts global topography, vegetation, and animal biodiversity: A global-scale analysis of mountainous regions. *Geophysical Res. Lett.* 47, e2020GL088649. doi: 10.1029/2020GL088649

Pal, M. (2005). Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26, 217–222. doi: 10.1080/01431160412331269698

Pal, M., and Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* 86, 554–565. doi: 10.1016/S0034-4257(03)00132-9

Pal, M., and Mather, P. M. (2005). Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* 26, 1007–1011. doi: 10.1080/01431160512331314083

Pal, M., Rasmussen, T., and Porwal, A. (2020). Optimized lithological mapping from multispectral and hyperspectral remote sensing images using fused multi-classifiers. *Remote Sens.* 12, 177. doi: 10.3390/rs12010177

Pan, T., Zuo, R., and Wang, Z. (2023). Geological mapping via convolutional neural network based on remote sensing and geochemical survey data in vegetation coverage areas. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens* 16, 3485–3494. doi: 10.1109/JSTARS.2023.3260584

Pour, A. B., and Hashim, M. (2014). ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration. *SpringerPlus* 3, 130–130. doi: 10.1186/2193-1801-3-130

Qiao, Y., Chen, H., and Jiang, Y. (2020). Quantifying the impacts of lithology on vegetation restoration using a random forest model in a karst trough valley, China. *Ecol. Eng.* 156, 105973. doi: 10.1016/j.ecoleng.2020.105973

Safanelli, J. L., Poppiel, R. R., Ruiz, L. F. C., Bonfatti, B. R., Mello, F., Rizzo, R., et al. (2020). Terrain analysis in google earth engine: A method adapted for high-performance global-scale analysis. *ISPRS Int. J. Geo-Information* 9, 400. doi: 10.3390/ijgi9060400

Salmon, B. P., Kleynhans, W., Schwegmann, C. P., and Olivier, J. C. (2015). "Proper comparison among methods using a confusion matrix," in *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Milan, Italy: Institute of Electrical and Electronics Engineers, 3057–3060. doi: 10.1109/IGARSS.2015.7326461

Schwinning, S. (2010). The ecohydrology of roots in rocks. *Ecohydrology: Ecosystems land Water process interactions Ecohydrogeomorphology* 3, 238–245. doi: 10.1002/eco.134

Serbouti, I., Raji, M., Hakdaoui, M., El Kamel, F., Pradhan, B., Gite, S., et al. (2022). Improved lithological map of large complex semi-arid regions using spectral and textural datasets within google earth engine and fused machine learning multi-classifiers. *Remote Sens.* 14, 5498. doi: 10.3390/rs14215498

Shebl, A., and Csamer, A. (2021). Stacked vector multi-source lithologic classification utilizing Machine Learning Algorithms: Data potentiality and dimensionality monitoring. *Remote Sens. Applications: Soc. Environ.* doi: 10.1016/j.rsase.2021.100643

Shebl, A., Csamer, A., and Abdellatif, M. (2021). Lithological mapping enhancement by integrating Sentinel 2 and gamma-ray data utilizing support vector machine: A case study from Egypt. *International Journal of Applied Earth Observation and Geoinformation.* San Francisco, CA, USA, 105 doi: 10.1016/j.jag.2021.102619

Shi, C., and Wang, X. (2014). Extraction and prospecting prediction of remote sensing geological structure anomaly information in Pinghe Vegetated area, Fujian Province. *J. Geology* 38, 464–469. doi: 10.3969/j.issn.1674-3636.2014.03.464

Shuai, S. (2022). *Study on lithology classification method based on multi-level information Fusion of multi-source Remote Sensing Data: A case study of basic-ultrabasic rocks in Bijiashan area, Beishan* (Hubei: PhD, China University of Geosciences).

Siegal, B. S., and Goetz, A. F. (1977). Effect of vegetation on rock and soil type discrimination. *Photogrammetric Eng. Remote Sens.* 43, 191–196. doi: 10.1016/0031-8663(77)90007-2

Snyder, N. P., Whipple, K. X., Tucker, G. E., and Merritts, D. J. (2000). Landscape response to tectonic forcing: Digital elevation model analysis of stream profiles in the Mendocino triple junction region, northern California. *Geological Soc. America Bull.* 112, 1250–1263. doi: 10.1130/0016-7606(2000)112<1250:LRTTFD>2.0.CO;2

Sun, W., Ren, K., Meng, X., Yang, G., Xiao, C., Peng, J., et al. (2022). MLR-DBPFN: A multi-scale low rank deep back projection fusion network for anti-noise hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2022.3146296

Wacker, A., and Landgrebe, D. (1972). Minimum distance classification in remote sensing. *LARS Tech. Rep.* 25.

Wang, Z., and Li, D. (2020). Analysis on comprehensive utilization of geological prospecting and mineral resources in China. *China Resour. Compr. Utilization* 3, 109–111.

Wang, L., and Liu, F. (2020). Analysis on Eco geological conditions of mineral resources exploration and development in North Junggar metallogenic belt, Xinjiang. *World Nonferrous Metals* 2.

Xie, Y. (2020). Current situation and improvement of Chinese legislation on mineral resources protection – comment on the innovation research on Chinese legislation on mineral resources protection. *Min. Res. Dev.* 40, 170.

Xu, S., Liu, S., Wang, H., Chen, W., Zhang, F., and Xiao, Z. (2020). A hyperspectral image classification approach based on feature fusion and multi-layered gradient boosting decision trees. *Entropy* 23, 20. doi: 10.3390/e23010020

Yang, L., Shen, F., Zhang, L., Cai, Y., Yi, F., and Zhou, C. (2021). Quantifying influences of natural and anthropogenic factors on vegetation changes using structural equation modeling: A case study in Jiangsu Province, China. *J. Cleaner Production* 280, 124330. doi: 10.1016/j.jclepro.2020.124330

Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., and Jia, A. (2018). Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method. *Remote Sens.* 10, 185. doi: 10.3390/rs10020185

Yasir, M., Jianhua, W., Shanwei, L., Sheng, H., Mingming, X., and Hossain, M. (2023). Coupling of deep learning and remote sensing: a comprehensive systematic literature review. *Int. J. Remote Sens.* 44, 157–193. doi: 10.1080/01431161.2022.2161856

Zeng, L., Li, T., Huang, H., Zeng, P., He, Y., Jing, L., et al. (2023). Identifying Emeishan basalt by supervised learning with Landsat-5 and ASTER data. *Front. Earth Sci.* 10. doi: 10.3389/feart.2022.1097778

Zhou, K., Zhang, J., Ren, Y., Huang, Z., and Zhao, L. (2020). A gradient boosting decision tree algorithm combining synthetic minority oversampling technique for lithology identification. *Geophysics* 85, WA147–WA158. doi: 10.1190/geo2019-0429.1

Zou, Z., Chen, C., Liu, Z., Zhang, Z., Liang, J., Chen, H., et al. (2022). Extraction of aquaculture ponds along coastal region using U2-net deep learning model from remote sensing images. *Remote Sens.* 14. doi: 10.3390/rs14164001

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 03 November 2023
DOI 10.3389/fevo.2023.1283281

OPEN ACCESS

EDITED BY
Samy Ismail Elmahdy,
United Arab Emirates University,
United Arab Emirates

REVIEWED BY
Yulong Guo,
Henan Agricultural University, China
Mohamad Awad,
Lebanese University, Lebanon

*CORRESPONDENCE
Yuhao Jin
yuhao.jin@scau.edu.cn
Peitong Cong
slxyky@scau.edu.cn

RECEIVED 25 August 2023
ACCEPTED 19 October 2023
PUBLISHED 03 November 2023

CITATION
Shen W, Jin Y, Li G and Cong P (2023) Analyzing the response distribution of DO concentration and its environmental factors under the influence of typhoon rain events with remote sensing.
Front. Ecol. Evol. 11:1283281.
doi: 10.3389/fevo.2023.1283281

Analyzing the response distribution of DO concentration and its environmental factors under the influence of typhoon rain events with remote sensing

Weiping Shen, Yuhao Jin*, Gengying Li and Peitong Cong*

College of Water Conservancy and Civil Engineering, South China Agricultural University, Guangzhou, China

Typhoon rain events are important factors that trigger changes in dissolved oxygen concentrations in watersheds. The direction of the typhoon driving force is clear, but the mode of action and mechanism are complex. Moreover, quantifying the relationship between these actions and dissolved oxygen is challenging. This study collected measured data from water quality monitoring and remote sensing during the 2022 typhoon rain events. By analyzing the changes in typhoon driving factors and dissolved oxygen (DO) concentrations in water under various typhoon storms, extended MOORA plus the full multiplicative form (MULTIMOORA), Multiscale Geographic Weighted Regression (MGWR), and spatial autocorrelation analysis were used to evaluate the response of DO concentration. Furthermore, the effects of the atmospheric environment under the influence of human activities on the response distribution of the urban water environment were analyzed. The results of the study showed that under the effect of a typhoon with higher rainfall intensity, the response of DO concentration in the water body of the river in the center of the city was better. However, the response of DO concentration in the water body at the mouth of the sea had a tendency to become worse. Under the influence of typhoon rain events with smaller intensity, the scouring effect of rainwater dominated, and the DO concentration response in the water body had a tendency to become worse. The analysis of spatial heterogeneity under the influence of human activities showed that the ranking values of DO concentration response in rivers in the city area of Zhongshan, under the influence of typhoon rain events, were positively correlated with the distribution of ozone (O_3) concentration and sulfur dioxide (SO_2) concentration in the eastern, central, and western parts of Zhongshan. Conversely, it was negatively correlated with the distribution of O_3 concentration and SO_2 concentration in the northern and southern parts of Zhongshan. Based on the research results, we constructed a technique to evaluate the response of dissolved oxygen concentration during the typhoon transit period, which can provide an indicator reference for urban managers in water environment management.

KEYWORDS

typhoon events, dissolved oxygen, urban, environment, river

1 Introduction

Water environmental pollution has become one of the three major water problems in cities because urbanization has changed the conditions for the formation of surface and subsurface runoff, affecting the water cycle as well as the spatial and temporal distribution of water (Qi-ting et al., 2005). Deterioration of water quality, reduction of water quantity, and degradation of water ecology are important impediments to the sustainable development of urbanization (Yu et al., 2018). Studies have shown that nonpoint source pollutants from typhoon rain events are one of the major pollutants in urban surface water and have serious impacts on surface water quality (Fu et al., 2021). Under the subtropical oceanic monsoon climate, an average of seven typhoons land and affect the Greater Bay Area every year, with high intensity and frequency. These typhoons not only cause serious casualties and economic losses but also have a significant impact on the urban water environment conditions (Liu et al., 2009; Chen et al., 2020; Cui et al., 2022). Water scarcity and deterioration of water quality pose threats to human health and survival, making them major challenges for existing freshwater resources (Mishra et al., 2021). In order to achieve the United Nations 2030 Sustainable Development Goals (SDGs) and to deepen the understanding and management of water security, it is exploratory and valuable to assess the risk of the water environment. Multi-criteria decision analysis (MCDA) is one of the most widely used methods in the environmental decision-making process (Mardani et al., 2017).

Globally, environmental risk assessment is an effective measure to prevent and control environmental events. Typhoon rain events have received attention from researchers as an important causative factor in triggering environmental events (Han et al., 2019). Liang Huanhuan et al. (2016) conducted a hierarchical study on the risk of groundwater contamination at 37 hazardous waste landfills using the MCDA model based on the idea of contamination source-pathway-receptor risk throughout the whole process control. Cabrera and Lee (2019) used multi-criteria decision analysis to assess typhoon-induced flood-prone risk areas in Davao Oriental, Philippines, by integrating various indicators such as rainfall and elevation. The role of hierarchical analysis (AHP) as well as ratio weighting (RW) in determining indicator weights was also compared, and the AHP model was found to perform better in calculating the importance of indicators. Gao et al. (2020) combined the spatial analysis method of AHP and geographic information system (GIS) to conduct a comprehensive weighted risk assessment based on the spatial and temporal cumulative patterns of typhoon-induced flooding disasters in Guangdong Province as the research object. Guangdong Province was classified into six levels of risk zones based on the integrated typhoon disaster risk, and the indirect economic impacts were further analyzed on this basis. Wang et al. (2022) synthesized the risk, exposure, and vulnerability of three typhoon hazard chains, constructed a comprehensive typhoon hazard risk indicator for the Greater Bay Area, divided the risk level of typhoon hazard chains in important towns and cities as well as the comprehensive risk level, and verified the validity of the assessment methodology of the comprehensive risk indicator for typhoon hazards. In multi-attribute decision analysis, MULTIMOORA method has simple calculation, short time and strong robustness, and is widely used in process selection (Lixia et al., 2020; Qiong et al., 2021; Lei et al., 2022), location selection (Yuan-hua et al., 2023), risk assessment (Shenghua et al., 2019; Zhang et al., 2022) and other fields(Hafezalkotob et al., 2019).

DO concentration in water bodies plays a crucial role in the survival of aquatic organisms as well as the balanced development of ecosystems, and is one of the important indicators for evaluating the quality of the water environment (Songbai et al., 2017; Espinosa-Diaz et al., 2021). Under the influence of typhoon rain events, nitrogen dioxide (NO_2), organic particulate matter, and other pollutants generated by human activities will be diluted by rainwater into the water environment, which affects the dissolved oxygen concentration of the water body under the effect of eutrophication and mineralization of organic particulate matter (Xiu-qin et al., 2019; Yufeng et al., 2023). Some scholars have conducted research on the potential influence mechanism between temperature, wind speed, rainfall, and DO concentration, and concluded that there is a positive correlation between DO concentration and temperature, wind speed, and rainfall (Xiaoran et al., 2013; Chen et al., 2016). Huang Weihui et al. (2021) and Huang Yuling et al. (2022) studied the natural factors such as geographical conditions on the concentration of saturated dissolved oxygen, and concluded that the concentration of dissolved oxygen was greatly affected by altitude and barometric pressure. In 2019, Jiayang Zhang (Zhang and Chen, 2019) assessed the risk of flooding disaster caused by typhoon rainstorms, and their evaluation indexes included wind speed, rainfall, and elevation. Ji-Myong Kim et al. (2020) selected maximum wind speed and distance as evaluation indicators in the vulnerability analysis of typhoons in Korea. Therefore, in this study, DEM, total rainfall, maximum rainfall in a single day, distance from typhoon landfall, and atmospheric $PM_{2.5}$ concentration and NO_2 concentration were comprehensively selected as evaluation indicators.

The process of urbanization is the transformation of rural territories into urban territories, involving the reshaping of natural landscapes, and it represents the most significant manifestation of human activities affecting hydrological systems (Liu et al., 2004). Nutrient export due to human activities has become a major cause of eutrophication and other ecological hazards in water bodies (Howarth, 2008; Howarth et al., 2011). Nutrient export due to human activities has become a major cause of eutrophication and other ecological hazards in water bodies. To comprehensively consider the source of non-point source pollution from human activities, Miao Jin-Dian et al. (2021) used NANI and NAPA models to analyze the spatial and temporal variation characteristics and driving factors of nitrogen and phosphorus in the Hangzhou section of the Qiandao Lake Basin. Fan Hongxiang et al. (2021) investigated the extent of the contribution of human activities to the change of water age in the lake area of Poyang Lake by coupling a deep learning network and a traditional two-dimensional hydrodynamic model. Yisong Zhao et al. (2022) took

Dianchi Lake as the research object and discussed the response of surface temperature and lake surface temperature under the influence of human activities. Therefore, it is informative for this study to discuss the effects of anthropogenic O_3 concentration and SO_2 concentration on the distribution of DO concentration response in urban waters under typhoon rain events.

Scholars at home and abroad have mostly conducted research on typhoon disaster risk assessment with the purpose of considering the economy and safety (Chaojia et al., 2022). However, relatively few typhoon disaster risk assessments have been carried out from the perspective of river water quality indicators, and there have been limited studies on the DO response of urban waters under the influence of human activities. In this paper, based on the measured data of urban water bodies during the 2022 typhoon rain events and remote sensing image data, the extended MULTIMOORA method is used to establish the correlation between typhoon drivers and the response of DO concentration in urban water bodies. Furthermore, this study aims to assess and quantify the distribution of the DO concentration response in urban water bodies. The MGWR model and spatial autocorrelation analysis were used to analyze the effects of O_3 concentration and SO_2 concentration generated under the influence of human activities on the distribution of urban water environment response. The results of the study will contribute to a more intuitive and in-depth understanding of the effects of typhoon rain events, as well as human activities, on the distribution of DO concentration response in urban waters. Local government decision-makers can formulate more effective water environmental protection policies and disaster prevention and mitigation measures based on the results of this study.

2 Materials and methods

2.1 Study area

The Guangdong-Hong Kong-Macao Greater Bay Area is located in the lower reaches of the Pearl River Basin, surrounded by mountains to the east, west, and north, and directly facing the South China Sea. It has a well-developed regional economy as well as a very high population density (Zhilin et al., 2022). Zhongshan is located in the heart of the Guangdong-Hong Kong-Macao Greater Bay Area, between latitude 22° 11'– 22° 47' north and longitude 113° 09'– 113° 46' east, with a total area of 1,783.67 km^2. The city's topography is dominated by plains, which are high in the middle and flat around the perimeter, with the plains sloping from northwest to southeast. Zhongshan City is located in the subtropical monsoon climate, abundant rainfall, the annual average rainfall is 1886 mm, the annual average inbound and transit water volume is 2662.94 billion m^3 and 2678.92 billion m^3 respectively. The disaster weather suffered by Zhongshan City mainly includes typhoons, heavy rainfall and strong convection, with high rainfall and intensity. Overall, the capacity of Zhongshan City to cope with emergencies such as pollution accidents and water quality-type water shortage conditions is relatively insufficient (Xuehua et al., 2022). The study area is shown in Figure 1.

2.2 Research date

In this study, four typhoon rain events (Chaba, Mulan, Ma-on, Nalgae) affecting Zhongshan City in 2022 were used as the study area. The extended MULTIMOORA theory was used to investigate the response of DO concentration in the water system of Zhongshan City during the transit of different typhoons. The selected indicators of the extended MULTIMOORA theory are: total rainfall, maximum rainfall in a single day, daily mean $PM_{2.5}$ concentration, daily mean NO_2 concentration, elevation, and distance from the station at the time of typhoon landfall.

Rainfall data from GPM (https://disc.gsfc.nasa.gov/; accessed on 7 June 2023) series of products. Global Precipitation Measurement (GPM) is an international satellite mission, carried out in cooperation with NASA and JAXA, which utilizes multi-sensor, multi-satellite and multi-algorithm in combination with satellite network and rain gauge inversion to obtain more accurate precipitation data. The GPM satellite carries ka-band precipitation radar and high-frequency microwave instrument, which can improve the observation of light rain and snowfall, and can provide higher spatial resolution and global coverage of precipitation observation data than TRMM satellite data.

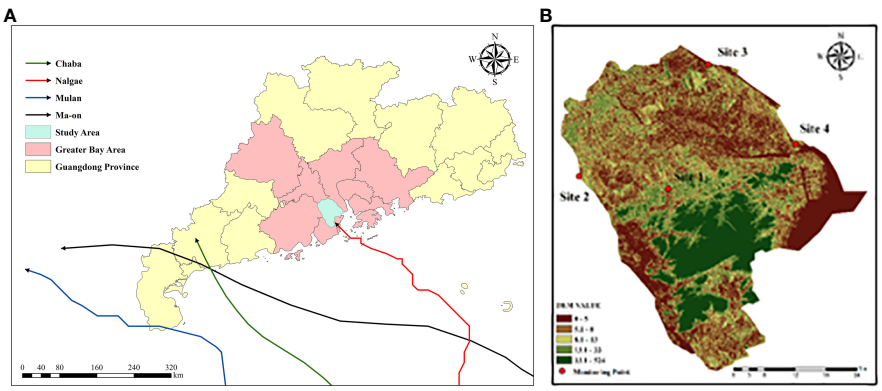


FIGURE 1
Research diagram. **(A)** Diagram of a typhoon. **(B)** Map of DEM.

Meteorological data (NO_2, SO_2, O_3) from the TROPOMI (https://s5phub.copernicus.eu/dhus/#/home; accessed on 25 June 2023) series of products. Sentinel-5P is a global atmospheric pollution monitoring satellite launched by the European Space Agency (ESA) on 13 October 2017. The TROPOMI sensor on board the satellite can effectively observe trace gas components in the atmosphere around the globe, including NO_2, O_3, SO_2, HCHO, CH_4 and CO, which are important indicators closely related to human activities, and enhance the observation of aerosols and clouds.

$PM_{2.5}$ data were obtained from Ventusky (https://www.ventusky.com/; accessed on 5 April 2023), with numerical results from FINNISH METEOROLOGICAL INSTITUTE, calculated from the SILAM model and MODIS Aqua and Terra remote sensing imagery data. DEM data were obtained from Geospatial Data Cloud (https://www.gscloud.cn/home; accessed on 26 January 2023).

The typhoon data used in this study were obtained from the Typhoon Network (http://typhoon.weather.com.cn/index.shtml; accessed on 8 March 2023). Four typhoon events (Chaba, Mulan, Ma-on, and Nalgae) affecting the study area in 2022 were used as study cases. Typhoon Chaba was generated in the South China Sea on 30 June 2022 and made landfall in Dianbai, Guangdong Province on 2 July with a landfall wind speed of 35 m/s. Chaba had an asymmetrical structure, a large circulation range of the cloud system, slow movement, a long influence time, and a wide range. Typhoon Mulan was generated in the South China Sea on 8 August 2022 and landed in Xuwen, Guangdong Province on 10 August with a landing wind speed of 23 m/s. Mulan had characteristics of, for example, a large size, a peculiar path, a short life cycle, a wide impact range, and strong local rain. Typhoon Ma-on was generated in the ocean east of the Philippines on 21 August 2022 and made landfall in Isabela Province, Philippines on 23 August and in Dianbai, Guangdong Province on 25 August. Ma-on was fast moving and had an asymmetric structure. Typhoon Nalgae was generated in the northwest Pacific Ocean on 27 October 2022 and made landfall in Catanduanes, Philippines on 29 October and Zhuhai, Guangdong Province on 3 November. Nalgae was characterised by low intensity and a loose structure with a large cloud scale.

2.3 Research method

2.3.1 MULTIMOORA method

Brauers and Zavadskas (2006) proposed a Multi-Objective-Optimization on basis of Ratio Analysis (MOORA) with discrete schemes. By constructing a decision matrix for multiple alternatives, decision makers are helped to choose the best option according to specific preference principles. The main steps are as follows:

$$Nx_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{m} x_{ij}^2}} \tag{1}$$

Where: x_{ij} is the response of alternative j to target i, $j = 1,2,\ldots,m$; m is the number of alternatives, $i = 1,2,\ldots,n$; n is the number of targets. Nx_{ij} is a dimensionless number representing the normalized response of alternative j to objective i, with the response interval at [0; 1].

According to the ratio system method, different types of targets need to be dealt with separately due to the difference between benefit and cost in the selection of targets.

$$A = \sum_{i=1}^{g} Nx_{ij} - \sum_{i=g+1}^{n} Nx_{ij} \tag{2}$$

Where: $i = 1,2,\ldots,g$ is the number of revenue targets, $i = g+1,g+2,\ldots,n$ is the number of cost class objectives. Evaluation value A under alternative j is obtained, and $Rank1$ is obtained by ranking according to the evaluation result of value A.

According to the reference point method, a maximum reference point is selected for the benefit target, whose coordinates are the largest among all responses. Select a minimal reference point for the cost class target whose coordinates are the smallest of all responses.

$$B_{ij} = \begin{cases} \left| \max_{(i)} Nx_{ij} - Nx_{ij} \right| & i < g \\ \left| \min_{(i)} Nx_{ij} - Nx_{ij} \right| & i > g \end{cases} \tag{3}$$

$$B = \max_{(j)} B_{ij} \tag{4}$$

Where: B_{ij} is the maximum reference distance of target i under alternive j, and B is the evaluation value under alternative j. $Rank2$ is obtained by ranking the results of value B.

Brauers and Zavadskas (2010) conducted a study on the robustness of multi-attribute decision making methods and pointed out that, in terms of robustness, the multi-attribute decision making method combining more decision-making methods has better effect. Therefore, the full multiplicative model is introduced into MOORA, and the full multiplicative form (MULTIMOORA) method is proposed. That is, the full multiplicative form of multiple objectives is added on the basis of the original, and the formula is as follows:

$$C = \frac{\prod_{i=1}^{g} Nx_{ij}}{\prod_{i=g+1}^{n} Nx_{ij}} \tag{5}$$

Where: $\prod_{i=1}^{g} Nx_{ij}$ is the utility of alternative j for the income objective, and $\prod_{i=g+1}^{n} Nx_{ij}$ is the utility of alternative j for the cost objective. C is the value of alternative j. $Rank3$ is obtained by ranking the results of value C.

$$minF(A,B,C) = \left(f_1(A,B,C), f_2(A,B,C) \ldots f_j(A,B,C)\right) \tag{6}$$

Finally, by summing the above three kinds of sorting, the result of comprehensive sorting is obtained according to the dominant theory.

2.3.2 Extended MULTIMOORA method

Although the MULTIMOORA method has good robustness, simple and effective calculation, and has a wide range of application scenarios. However, its shortcomings are that it does not consider the evaluation information support among experts and the non-negligible redundancy relationship between evaluation indicators, the lack of indicator weights (Jun and Shi-Hua, 2022). In this study, the subjective weight w_i is calculated by using OWA operator, the objective weight w_o is calculated by using random forest, and finally the comprehensive weight w_s is obtained based on the principle of minimum discriminative information.

The Ordered Weighted Averaging (OWA) operator is one of the most commonly used aggregation operators, which can facilitate a more versatile data fusion process. Since the introduction of the OWA operator, many researchers have delved into various methods of obtaining weights (Xu, 2005). In this study, a method of calculating combinations is employed to assign weights to the OWA operator. The formula is as follows (Yu and Ze-Shui, 2008; Guo et al., 2020). See Appendix A for the detailed calculation process.

$$w_i = \frac{\sum_{j=1}^{p} \frac{C_{p-1}^{j}}{2^{p-1}} * a_j}{\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{C_{p-1}^{j}}{2^{p-1}} * a_j} (i = 1, 2, \ldots, p, j = 0, 1, \ldots, p-1) \quad (7)$$

Random Forest is a highly flexible machine learning method that utilizes multiple decision trees to handle nonlinear data, address regression and classification problems, and perform feature selection based on its feature importance metrics. Random Forest employs the Bootstrap resampling method to extract a training set comprising 2/3 of the original samples and an Out-Of-Bag (OOB) data set consisting of 1/3 of the samples for feature importance computation (Shufang and Ruyang, 2021; Chen et al., 2023). The principle involves randomly perturbing a particular input parameter and calculating the resulting estimation error. The importance of this parameter is determined based on the magnitude of the error, where a higher importance value indicates greater significance of the parameter. The formula (LuanXiao et al., 2021; Xiao-wen et al., 2021) of objective weights is shown below. See Appendix B for the detailed calculation process.

$$w_o = \frac{IMP^i(X_{OOB}^i)}{\sum_{i=1}^{n} IMP^i(X_{OOB}^i)} \quad (8)$$

Where: $IMP^i(X_{OOB}^i)$ signifies the importance results of the feature variables.

For a discrete random variable $X = (x_1, x_2, \ldots, x_n)$, the probability distribution of x is only related to the condition δ and the condition τ. Under the condition δ, the x_k probability distribution function is $\delta(x_k)$; similarly, under the condition τ, the x_k probability distribution function is $\tau(x_k)$. Thus, the discriminatory information (Lee et al., 2019) is expressed as:

$$I[\tau(x), \delta(x)] = \sum_{k=1}^{n} \tau(x) log \frac{\tau(x)}{\delta(x)} \quad (9)$$

Since both subjective and objective weight data are discrete random variables, in order to enhance the credibility and accuracy of the composite weight for evaluation metrics with smaller errors, this study employs the Minimum Discriminant Information Principle to determine a composite weight that closely approximates both. The solution is obtained by introducing Lagrange multipliers into the equation.

$$Lag(x, \psi) = \theta(x) - \psi\delta(x) \quad (10)$$

Where: $\theta(x)$ represents the original function, and $\delta(x)$ is the constraint condition function, and ψ denotes the Lagrange multiplier.

The formula is as follows. See Appendix C for the detailed calculation process.

$$w_s = \frac{\sqrt{w_i w_o}}{\sum_{i=1}^{p} \sqrt{w_i w_o}} \quad (11)$$

Where: w_s is the combined weight of the evaluation indicators, w_i is the subjective weight of the evaluation indicators and w_o is the objective weight of the evaluation indicators.

The formula for the extended MULTIMOORA method is shown below.

$$A = \sum_{i=1}^{g} w_s Nx_{ij} - \sum_{i=g+1}^{n} w_s Nx_{ij} \quad (12)$$

$$B = \max_{(j)} \begin{cases} w_s \left| \max_{(i)} Nx_{ij} - Nx_{ij} \right| & i < g \\ w_s \left| \min_{(i)} Nx_{ij} - Nx_{ij} \right| & i > g \end{cases} \quad (13)$$

$$C = \frac{\prod_{g=1}^{i} x_{gj}^{w^s}}{\prod_{k=i+1}^{n} x_{kj}^{w^s}} \quad (14)$$

The theory of dominance is a method that can integrate multiple rankings into one ranking result based on various criteria such as dominance, equality, and transition (Brauers and Zavadskas, 2012). The extended MULTIMOORA method allows the ranking of three alternatives to be obtained, and then multiple rankings are integrated into one ranking based on the theory of dominance. Finally, the response analysis is performed based on the final ranking results.

2.3.3 Spatial autocorrelation analysis

After obtaining the extended MULTIMOORA sorting distribution data, the global Moran index (Ge et al., 2022) was used as the global spatial autocorrelation index to analyze the correlation and difference of DO concentration responses in river water bodies in Zhongshan City. The expression is as follows:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2(\sum_i \sum_j w_{ij})} \quad (15)$$

The local Moran index was further used to analyze the local spatial autocorrelation of the aggregation of DO concentration response in the river water bodies in Zhongshan City, to identify the high value aggregation area and the low value aggregation area

of the DO concentration response, in order to reflect the spatial dependence and heterogeneity of the DO concentration response. The expression of the local Moran index is:

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})}{S^2} \quad (16)$$

Where: n is the number of spatial grid cells, x_i and x_j are observations representing cell i and cell j respectively, $(x_i - \bar{x})$ is the deviation of the observation from the mean value on the i^{th} grid cell, and w_{ij} is a spatial weight matrix based on spatial k-neighbourhood relationships.

2.3.4 Multiscale geographic weighted regression

In traditional regression analysis, the relationship between the independent variable and the dependent variable is considered to be stable throughout the whole area, so the estimated regression coefficients obtained are averaged over the whole study area, which cannot respond to the real spatial characteristics of the regression parameters. Drawing on the idea of local smoothing and embedding the spatial location of the data in the regression equations, Fotheringham et al. (1998) proposed a geographically weighted regression model (Geographically weighted regression, GWR), whose expression is as follows:

$$y_i = \beta_0(u_i; v_i) + \sum_{k=1}^{p} \beta_k(u_i; v_i)x_{ik} + \varepsilon_i \quad i = 1, 2, \ldots, n \quad (17)$$

Where: y_i is the dependent variable; $\beta_0(u_i;v_i)$ is the intercept; x_{ik} is the value of the kth independent variable at the ith sampling point; $(u_i;v_i)$ is the coordinates of the sampling point, $\beta_k(u_i;v_i)$ is the kth regression parameter on the sampling point $(u_i;v_i)$; ϵ_i is the random error.

MGWR, developed from GWR, takes into account different levels of spatial heterogeneity and allows each independent variable to have an optimized bandwidth based on itself (Jin et al., 2021). Its expression is as follows:

$$y_i = \sum_{j=1}^{n} \alpha_j X_{ij} + \sum_{j=n+1}^{m} \beta_j(u_i; v_i)X_{ij} + \varepsilon_i \quad i, j = 1, 2, \ldots, n \quad (18)$$

Where: X_{ij} is the value of the independent variable; j is the number of independent variables; β_j is the regression coefficient of the global variable; β_j is the regression coefficient of the local variable.

3 Result

3.1 Analysis of DO response at different stations under the influence of the same typhoon

In order to study the DO response at different water quality automatic monitoring stations under the influence of the same typhoon. In this study, the distance from the station at the time of typhoon Ma-on's landfall, station elevation, total rainfall, maximum rainfall in a single day, average daily $PM_{2.5}$ concentration, and average daily NO_2 concentration were selected as the evaluation indexes, and the extended MULTIMOORA method was applied to carry out the study. The evaluation indicators for the different typhoons are shown in Table 1.

Five experienced experts with a long history of research in the water environment were invited to rate each indicator. The scale ranges from 1 to 10, with higher ratings indicating that the indicator has a greater impact on DO concentrations in the water system. The expert scores are shown in Table 2. The indicators were ranked in descending order according to the experts' scores, and the subjective weights w_i = (0.12,0.14,0.23,0.22,0.12,0.17) for each indicator were subsequently calculated by combining equations (7).

Using the distance from the station at the time of typhoon Ma-on landfall, station elevation, total rainfall, single-day maximum rainfall, daily average $PM_{2.5}$ concentration and daily average NO_2 concentration as inputs and the change in DO concentration as outputs, the above data were input as training samples into a random forest model to build a regression prediction model, as shown in Figure 2. In this case, the training samples are set up with a training set and a test set, which account for 80% and 20% of the number of training samples respectively. The objective weights w_o = (0.16,0.14,0.16,0.2,0.15,0.19) of each indicator were obtained by normalising the importance of each feature based on the solved one.

Finally, after coupling the subjective weight $\boldsymbol{w_i}$ and the objective weight w_o through equation (9) to equation (11), the combined weight w_s = (0.14,0.14,0.19,0.21,0.14,0.18) of each indicator is obtained.

In order to eliminate the difference in scale between each indicator, the evaluation indicator data of different sites were first standardised, and the processing results are shown in Table 3.

After standardisation of the indicator data, the indicators in Table 4 were calculated using the ratio system, the reference point approach and the full multiplication form of the extended

TABLE 1 Evaluation indicators for different stations under Typhoon Ma-on.

Typhoon	Distance from Landing Point (km)	DEM (m)	Total Rainfall (mm)	Maximum Daily Rainfall (mm)	$PM_{2.5}$ Concentration (ug/m^3)	NO_2 Concentration (ug/m^3)
Site 1	258.977	7	15.3	10.6	34	28.75
Site 2	245.974	24	18	12.3	31.792	32.458
Site 3	277.901	11	14.6	7.4	40.792	35.625
Site 4	286.133	0	40.3	27.8	46.417	41.375

10.3389/fevo.2023.1283281

TABLE 2 Expert evaluation form.

Scorer	Distance from Landing Point (km)	DEM (m)	Total Rainfall (mm)	Maximum Daily Rainfall (mm)	$PM_{2.5}$ Concentration (ug/m^3)	NO_2 Concentration (ug/m^3)
Expert 1	6	7	3	10	4	9
Expert 2	5	3	9	7	5	8
Expert 3	2	4	9	8	6	6
Expert 4	3	6	8	8	3	3
Expert 5	5	5	8	7	4	4

MULTIMOORA method via equations (12) to (14) respectively. The rating values for each indicator were determined as shown in Table 4. Where A is the rating value of the ratio system, *B* is the rating value of the reference point approach and *C* is the rating value of the full multiplication form.

As can be seen from Table 4, the response of DO concentrations in the water at different sites under the influence of Typhoon Ma-on is ranked from largest to smallest: Site 2 > Site 1 > Site 4 > Site 3. In terms of DO concentrations, the variation in DO concentrations measured before and after the typhoon at the four automatic water quality monitoring stations in the study area during Typhoon Ma-on was 0.4 (Site 2) > 0.15 (Site 1) > −0.1 (Site 4) > −2.4 (Site 3). A positive value means that the DO concentration in the water column has increased after the typhoon, indicating a positive trend; a negative value means that the DO concentration in the water column has decreased after the typhoon, indicating a negative trend; this is in line with the expert scoring and the MULTIMOORA theory of ranking. The urban pattern of Zhongshan City shows that Site 2 and Site 1 are in the more urbanised population centres, while Site 4 and Site 3 are in the less urbanised river inlets. DO concentrations are relatively low in water bodies at large population centres, and tend to increase as a result of rainfall dilution; DO concentrations are relatively high in water bodies at the mouths of less urbanised rivers, and tend to decrease as a result of rainfall scouring. Combined with the data for the selected indicators, the results of this method of ranking are considered reasonable.

3.2 Analysis of DO response in water in Zhongshan City under the influence of typhoon rain events

An in-depth understanding of the response status of DO concentration in urban water bodies under the influence of different typhoon rain events is helpful in revealing the trend of changes in the urban water environment under the influence of typhoon rain events, and in formulating strategies for controlling the water environment during the transit of typhoons. In some studies, an information system based on multi-criteria decision analysis is the preferred method because it involves multiple weighted combinations and also produces visualization results (Kut and Pietrucha-Urbanik, 2022), which is important for decision-making on the environmental risks of typhoon disasters. In this study, the extended MULTIMOORA theory is combined with geographic information system (GIS) to extract the above six evaluation index values using remote sensing image data, and the extracted index values are substituted into the formula for calculating the DO concentration response ranking under the influence of field typhoon and rainfall, and finally the DO concentration response under the influence of typhoon rain events is visualized. In this study, four typhoons (Chaba, Mulan, Ma-on, and Nalgae) affecting Zhongshan City in 2022 were visualized and analyzed separately, in which the metrics were extracted as shown in Figures 3–6, and the response of DO concentration in Zhongshan City under the influence of typhoon rain events in each scene is shown in Figure 7.

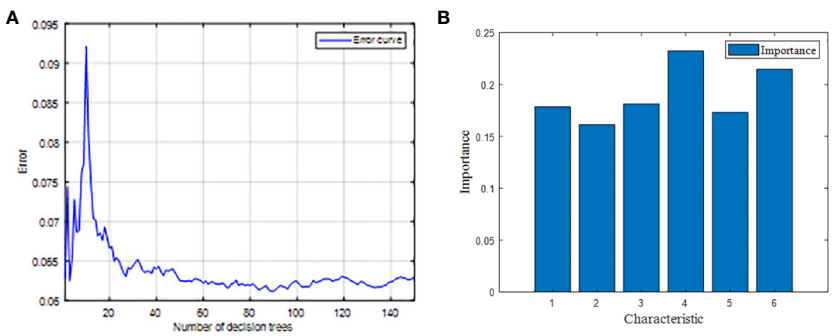


FIGURE 2
Schematic diagram of random forest regression analysis. **(A)** Diagram of the training process. **(B)** Schematic representation of the importance of features.

TABLE 3 Indicators for the different sites after standardization.

Typhoon	Distance from Landing Point (km)	DEM (m)	Total Rainfall (mm)	Maximum Daily Rainfall (mm)	$PM_{2.5}$ Concentration ($\mu g/m^3$)	NO_2 Concentration ($\mu g/m^3$)
Site 1	0.484	0.256	0.313	0.321	0.439	0.412
Site 2	0.459	0.879	0.368	0.372	0.411	0.466
Site 3	0.519	0.403	0.298	0.224	0.527	0.511
Site 4	0.534	0	0.823	0.842	0.6	0.593

Chaba was generated in the South China Sea on 30 June 2022, and landed in the coastal area of Guangdong Dianbai at 15:00 on 2 July, with landing winds reaching 35 m/s. As can be seen from Figure 3, from 2 to 4 July, the cumulative rainfall in Zhongshan City ranged from 88.1737 mm to 142.435 mm under the influence of Chaba, with the maximum single-day rainfall ranging from 39.5626 mm to 81.5451 mm. The daily average $PM_{2.5}$ concentration ranged from 15.7081 $\mu g/m^3$ to 18.8748 $\mu g/m^3$ and NO_2 concentration ranged from 0.0000963933 mol/m^2 to 0.000202383 mol/m^2.

Mulan intensified from a tropical storm in the South China Sea on 9 August 2022, and made landfall in Xuwen, Guangdong at 10:00 a.m. on 10 August, with landfall winds reaching 23 m/s. As can be seen from Figure 4, the cumulative rainfall in Zhongshan City from 9 to 11 August under the influence of Mulan ranged from 36.4076 mm to 111.974 mm, with the maximum single-day rainfall ranging from 19.5202 mm to 63.1185 mm, the daily average $PM_{2.5}$ concentration ranged from 7.16744 $\mu g/m^3$ to 16.6249 $\mu g/m^3$ and NO_2 concentration ranged from 0.0000351573 mol/m^2 to 0.00011281 mol/m^2.

Ma-on was generated on 22 August 2022 over the ocean east of the Philippines and landed on the coast of Guangdong Dianbai at 10:00 a.m. on 25 August, with landing winds reaching 33 m/s. As can be seen from Figure 5, the cumulative rainfall in Zhongshan City from 24 to 26 August under the influence of Ma-on ranged from 6.46943 mm to 20.8218 mm, with the maximum single-day rainfall ranging from 5.73675 mm to 19.8135 mm, with daily average $PM_{2.5}$ concentrations ranging from 27.5838 $\mu g/m^3$ to 46.3735 $\mu g/m^3$ and NO_2 concentrations ranging from 0.0000432654 mol/m^2 to 0.0000855731 mol/m^2.

Nalgae was generated in the northwest Pacific Ocean on 27 October 2022 and landed in Xiangzhou District, Zhuhai, Guangdong on 3 November, with landing winds reaching 16 m/s. As can be seen from Figure 6, the cumulative rainfall in Zhongshan City from 31 October to 3 November was between 5.31433 mm to 13.3876 mm under the influence of Nalgae, with the maximum single-day rainfall ranging from 2.7148 mm to 6.63383 mm, with daily average $PM_{2.5}$ concentrations ranging from 33.0633 $\mu g/m^3$ to 42.5623 $\mu g/m^3$ and NO_2 concentrations ranging from 0.000225898 mol/m^2 to 0.000561096 mol/m^2.

The lower the ranking value calculated based on the extended MULTIMOORA theory, the more likely the DO concentration in the water body at that location will respond in a good direction under the influence of the typhoon rain events; the higher the ranking value calculated based on the extended MULTIMOORA theory, the more likely the DO concentration in the water body at that location will respond in a bad direction under the influence of the typhoon rain events. As can be seen from Figure 7, the DO concentration response of the rivers in the Zhongshan city area under the influence of Chaba, Mulan and Nalgae is such that there is a tendency for the south-western water bodies to develop to the good side, while the north-eastern water bodies have a tendency to become worse. Under the influence of Ma-on, the DO concentration response in the Zhongshan municipal rivers is a trend towards worse water bodies in the south-west as well as in the north.

An accurate understanding of the regional distribution of DO concentration response in urban waters under the influence of different typhoon rain events is helpful in revealing the regional distribution of urban water environment risks under the influence of typhoon rain events, and in formulating targeted disaster prevention and mitigation efforts. In this study, the ranked values calculated by the extended MULTIMOORA theory are clustered and analyzed, as shown in Figure 8.

As can be seen from Figure 8, under the influence of Chaba, Mulan and Nalgae, the southern and northwestern parts of Zhongshan City show low-low aggregation phenomena, indicating that the response of DO concentration in this region has a tendency to develop for the better, and it is a low-risk area; the northern and northeastern parts of Zhongshan City show high-high

TABLE 4 Ratings of different stations under Typhoon Ma-on.

Typhoon	*A*	Rank1	*B*	Rank2	*C*	Rank3	SUM	Comprehensive sort
Site 1	0.095	3	0.109	2	0.62	2	7	2
Site 2	0.194	2	0.099	1	0.769	1	4	1
Site 3	0.067	4	0.13	4	0.575	3	11	4
Site 4	0.217	1	0.123	3	0	4	8	3

10.3389/fevo.2023.1283281

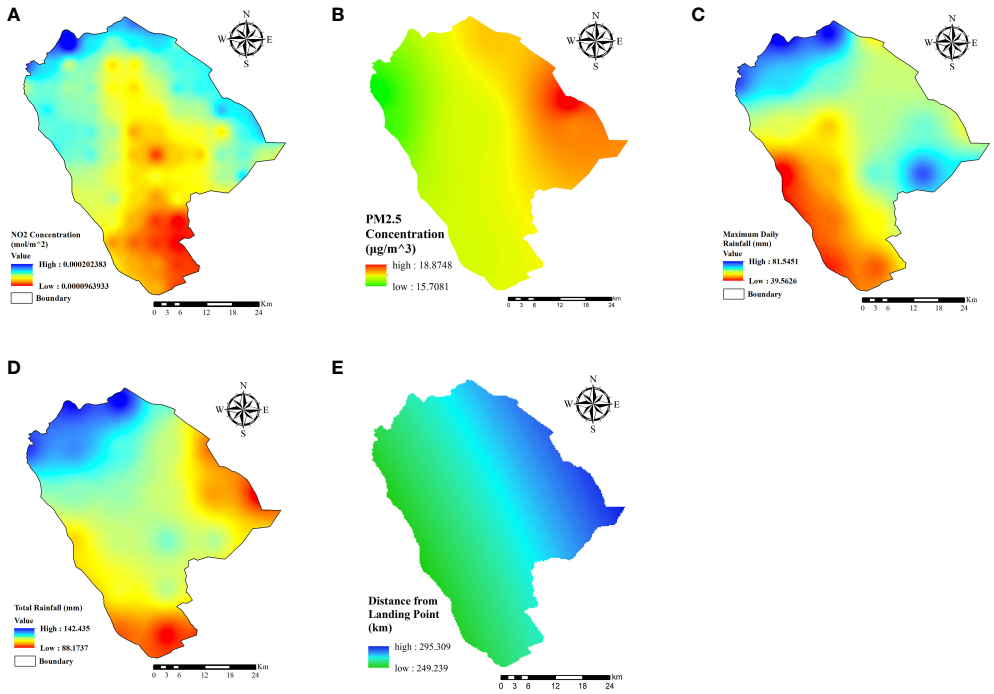


FIGURE 3
Distribution of indicators in Zhongshan under the impact of Typhoon Chaba. **(A)** Distribution of NO_2 concentration. **(B)** Distribution of $PM_{2.5}$ concentration. **(C)** Distribution of maximum rainfall in a single day. **(D)** Distribution of total rainfall. **(E)** Distance from landing site to site.

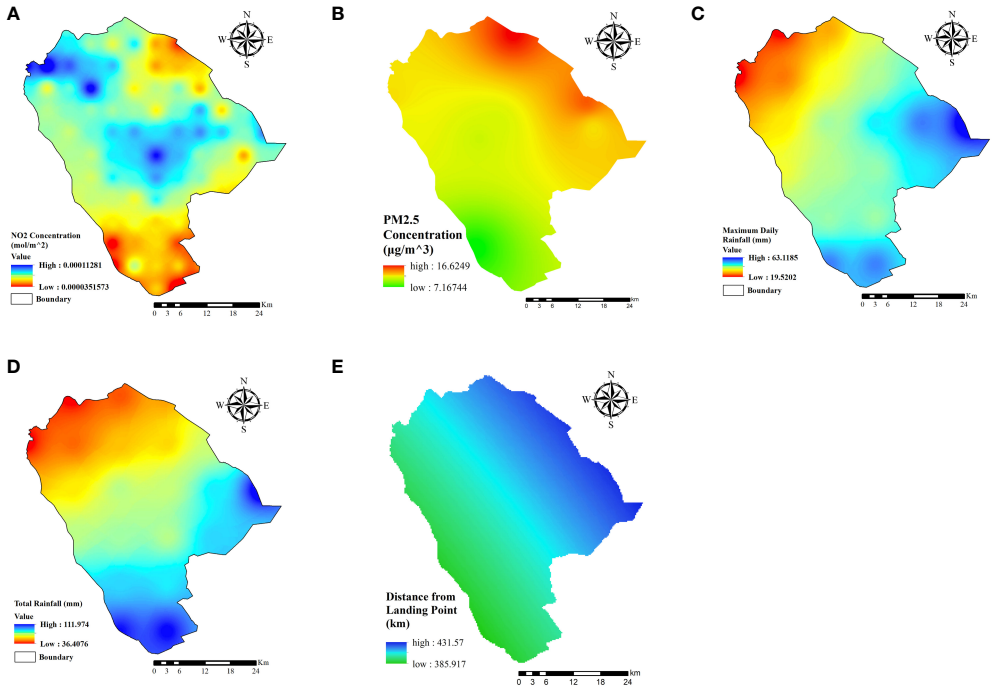


FIGURE 4
Distribution of indicators in Zhongshan under the impact of Typhoon Mulan. **(A)** Distribution of NO_2 concentration. **(B)** Distribution of $PM_{2.5}$ concentration. **(C)** Distribution of maximum rainfall in a single day. **(D)** Distribution of total rainfall. **(E)** Distance from landing site to site.

10.3389/fevo.2023.1283281

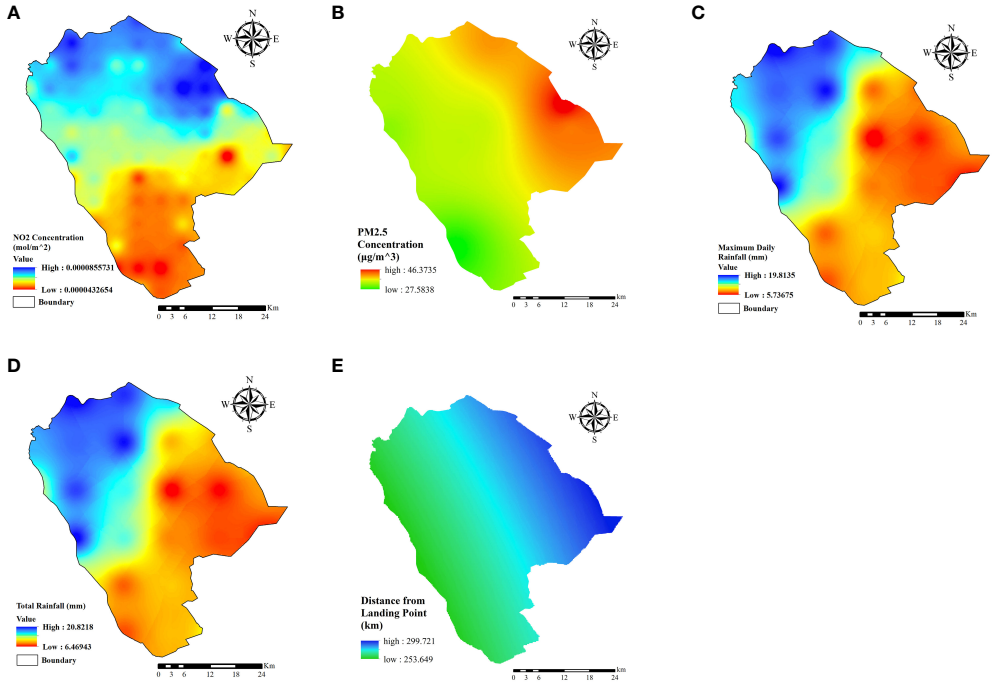


FIGURE 5
Distribution of indicators in Zhongshan under the impact of Typhoon Ma-on. **(A)** Distribution of NO_2 concentration. **(B)** Distribution of $PM_{2.5}$ concentration. **(C)** Distribution of maximum rainfall in a single day. **(D)** Distribution of total rainfall. **(E)** Distance from landing site to site.

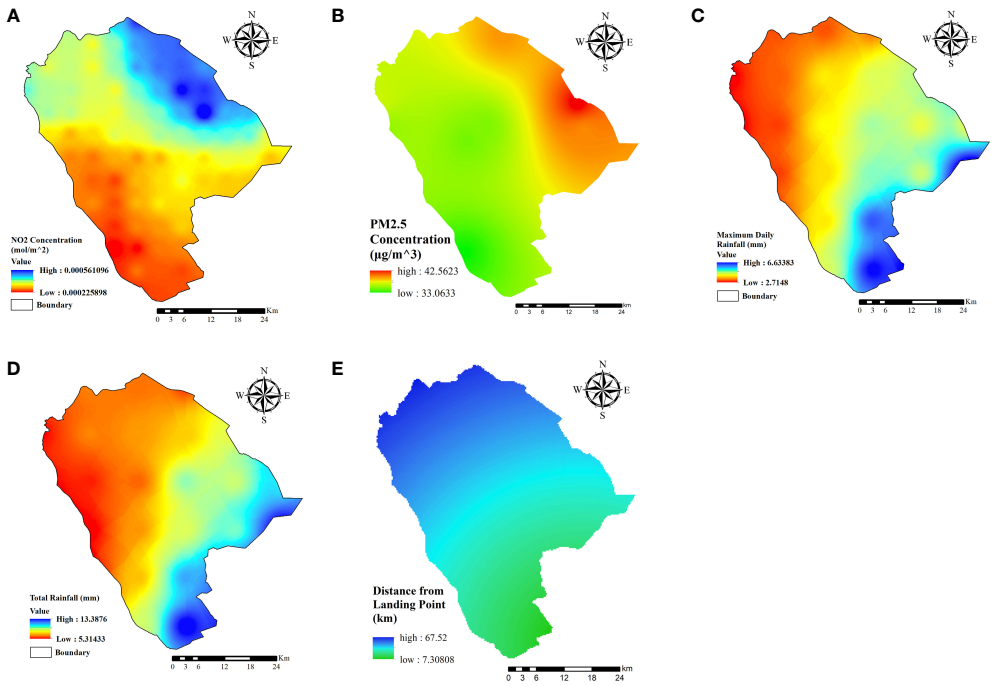


FIGURE 6
Distribution of indicators in Zhongshan under the impact of Typhoon Nalgae. **(A)** Distribution of NO_2 concentration. **(B)** Distribution of $PM_{2.5}$ concentration. **(C)** Distribution of maximum rainfall in a single day. **(D)** Distribution of total rainfall. **(E)** Distance from landing site to site.

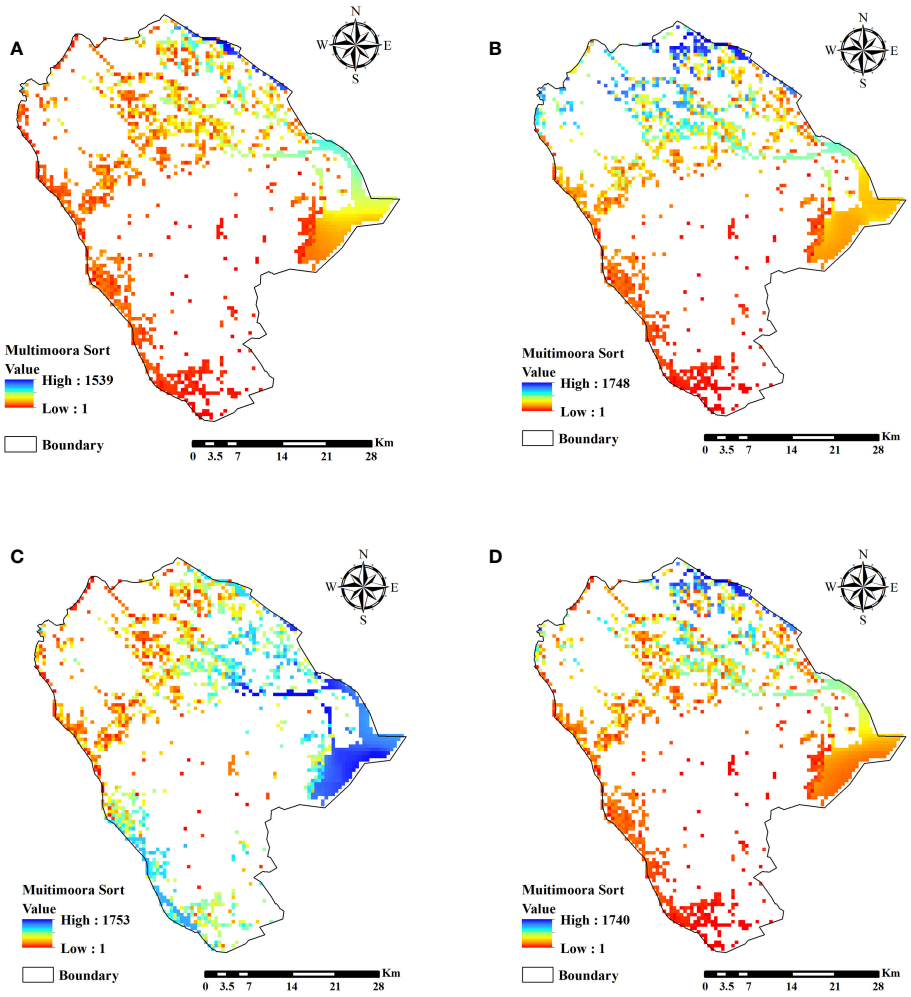


FIGURE 7
Response of DO concentration in Zhongshan City water bodies under the influence of Typhoon Rain Events. **(A)** Influenced by Chaba. **(B)** Influenced by Mulan. **(C)** Influenced by Ma-on. **(D)** Influenced by Nalgae.

aggregation phenomena, indicating that the response of DO concentration in this region has a tendency to deteriorate, and it is a high-risk area. Under the influence of Ma-on, the localized areas in the northeast and southwest of Zhongshan City showed the phenomenon of high-high aggregation, indicating that the response of DO concentration in this region has a tendency to become worse and is a high-risk area; the phenomenon of low-low aggregation in the northwest of Zhongshan City indicates that the response of DO concentration in this region has a tendency to develop in a better direction and is a low-risk area.

3.3 Spatial heterogeneity in the response of water body DO concentration under the influence of typhoon rain events

Human activities are the main drivers of atmospheric pollution (Shuping et al., 2016). SO_2 and O_3 are important gases affecting atmospheric quality, of which O_3 is a very important greenhouse gas with an uneven global distribution, which is significantly affected by human activities (Bing and Hua, 2014). SO_2 mainly originates from human activities such as industrial activities, fossil fuel combustion and biomass combustion (Jie et al., 2011), therefore SO_2 can be used to characterize the intensity of human activities. Spatial heterogeneity in population distribution has been suggested (Zhipeng et al., 2022), and thus the same spatial heterogeneity exists in SO_2 concentration distribution and O_3 concentration distribution driven by anthropogenic intensity. In this study, the MGWR model was used to explore the correlation between the O_3 concentration distribution and SO_2 concentration distribution driven by the intensity of human activities and the DO concentration response ranking values of urban streams, and the calculated results are shown in Figure 9. (a), (c), (e), (g), and (b), (d), (f), (h) are the spatial distributions of the coefficients of influence of the O_3 concentration distributions and the SO_2 concentration distributions on the sorted values of the DO concentration response for Chaba, Mulan, Ma-on, and Nalgae, respectively.

As shown in Figure 9, there is a certain correlation between the DO concentration response ordering values and the O_3 concentration distribution and SO_2 concentration distribution. In terms of O_3

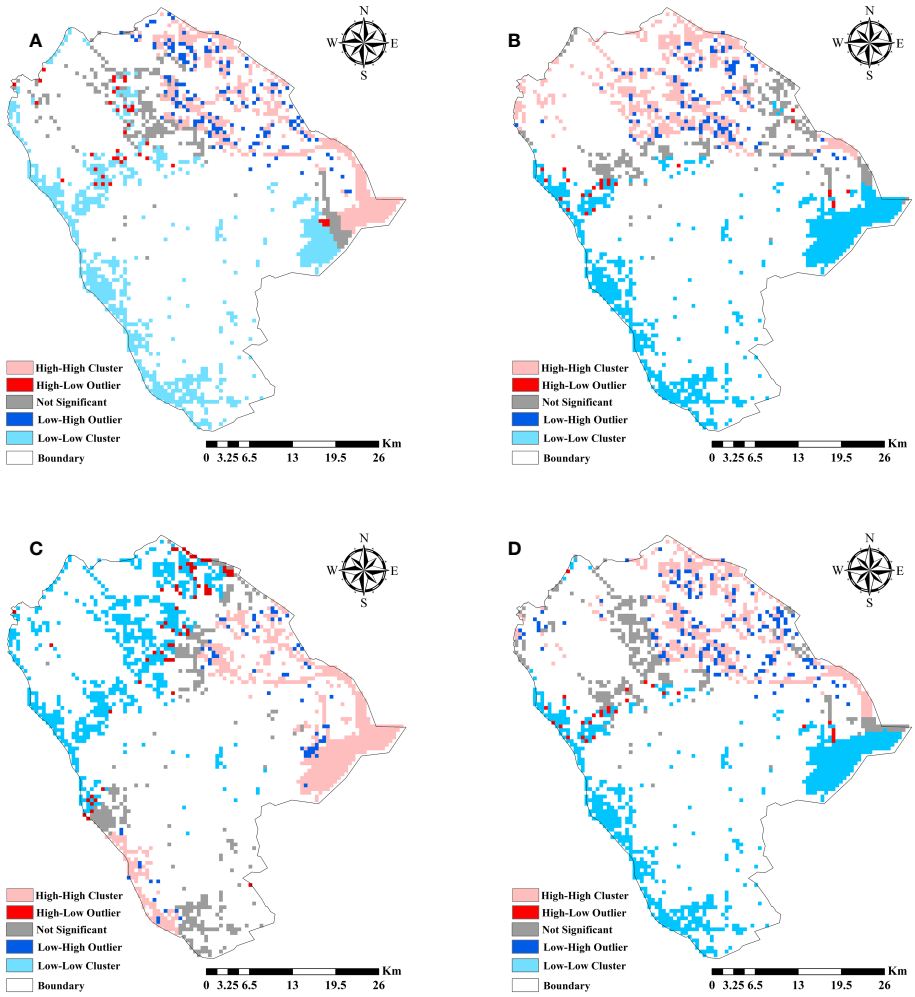


FIGURE 8
Cluster Analysis of DO Concentration Response in Zhongshan City Water Bodies under the Impact of Typhoon Rain Events. **(A)** Influenced by Chaba. **(B)** Influenced by Mulan. **(C)** Influenced by Ma-on. **(D)** Influenced by Nalgae.

concentration distribution, under the influence of the four typhoon events, the maximum positive correlation between DO concentration response ranking values and O_3 concentration distribution occurred in the eastern part of Zhongshan City, the minimum positive correlation occurred in the western part of Zhongshan City, and the maximum negative correlation occurred in the northern part of Zhongshan City. In terms of SO_2 concentration distribution, under the influence of the four typhoon rains, the maximum positive correlations between the ranked DO concentration response values and the SO_2 concentration distribution occurred in the eastern and western parts of Zhongshan City, the minimum negative correlations occurred in the southern part of Zhongshan City, and the maximum negative correlations occurred in the northern part of Zhongshan City.

4 Discussion

In recent years, a series of research results have been achieved in both environmental risk assessment and typhoon disaster risk assessment. And in the existing studies, economy and safety are mostly taken as the assessment objectives. In this study, an attempt is made to introduce the expanded MULTIMOORA theory in multi-criteria decision analysis to carry out the environmental risk assessment of typhoon disaster with the response of DO concentration in the water body as the assessment objective. Meanwhile, combining the measured water quality data and corresponding public information, the results of this study are considered reasonable, and the findings can provide some reference for the local government in formulating disaster prevention and mitigation plans and water environment control strategies during typhoons.

As can be seen in Figures 7, 8, the effect of field typhoon rains on the response of DO concentrations in urban rivers is two-sided, which is consistent with previous reports (Zhou et al., 2012; Ye et al., 2014). In general, rainfall with higher DO concentration into the river can effectively increase the DO concentration of the river (Muñoz et al., 2015), and at the same time, rainfall can also cause a large amount of oxygen-depleting compounds and organic matter to enter into the river, which can cause a rapid decrease in the DO

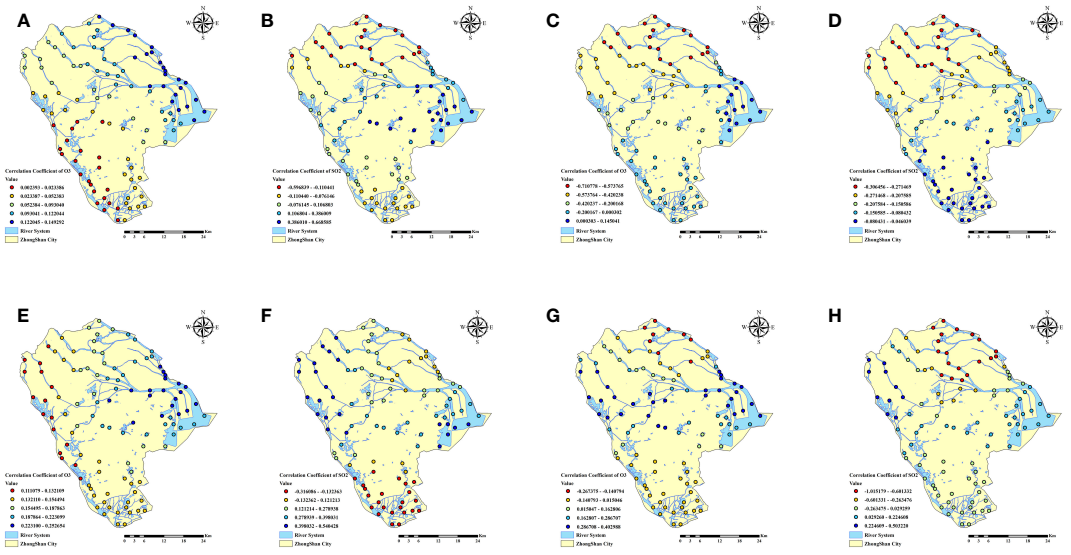


FIGURE 9
(A) Spatial distributions of the coefficients of influence of the O_3 concentration distributions on the sorted values of the DO concentration response influenced of Chaba. **(B)** Spatial distributions of the coefficients of influence of the SO_2 concentration distributions on the sorted values of the DO concentration response influenced of Chaba. **(C)** Spatial distributions of the coefficients of influence of the O_3 concentration distributions on the sorted values of the DO concentration response influenced of Mulan. **(D)** Spatial distributions of the coefficients of influence of the SO_2 concentration distributions on the sorted values of the DO concentration response influenced of Mulan. **(E)** Spatial distributions of the coefficients of influence of the O_3 concentration distributions on the sorted values of the DO concentration response influenced of Ma-on. **(F)** Spatial distributions of the coefficients of influence of the SO_2 concentration distributions on the sorted values of the DO concentration response influenced of Ma-on. **(G)** Spatial distributions of the coefficients of influence of the O_3 concentration distributions on the sorted values of the DO concentration response influenced of Nalgae. **(H)** Spatial distributions of the coefficients of influence of the SO_2 concentration distributions on the sorted values of the DO concentration response influenced of Nalgae.

content of the river. Combined with the results of the study, this paper concludes that: in the rainfall intensity of the larger typhoon rain events, rich in dissolved oxygen rainfall will play a dilution effect, increase the dissolved oxygen content in the water body of the urban water network; and in the production of convergence and the role of surface scouring to reduce in the water body at the mouth of the sea in the content of dissolved oxygen. This is consistent with existing research. Pearce and Schumann (2003) documented a 13-month period of dissolved oxygen concentration measurements in the Gamtoos Estuary, South Africa, noting that hypoxic conditions occurred throughout the estuary following a large-scale extreme rainfall event. In 2011, Mitra.A. (Mitra et al., 2011a; Mitra et al., 2011b) et al. conducted an in-situ study of hydrological parameters in the Bay of Bengal and neighboring estuaries under the influence of AILA and found that dissolved oxygen showed a decreasing trend at all the sampling points during the transit of AILA, and gradually recovered to the pre-AILA level in the water column 10 days after the end of the AILA event. Geng Ye et al. (2021) analyzed urban surface water quality under the influence of Lekima based on automatic monitoring data, and concluded that during the typhoon's transit, the DO concentration in the Jinan section of the Xiaoqing River showed an overall upward trend accompanied by small fluctuations under the combined effect of upstream inflow and surface tributary inflow during the same period. Under the action of the typhoon rain events with lower rainfall intensity, the scouring action of rainwater dominates, washing pollutants from urban impervious surfaces as well as organic matter from mountainous soils into the river water body, leading to a trend of worse overall DO concentration response in the water body. This is consistent with existing studies. Yihui et al. (2022) analyzed in detail the characteristics of the impacts of typhoon rain events on the water environment of lakes and showed that the effect of typhoon rain events would cause a decrease in the concentration of DO in the water body. In terms of rainfall intensity, the results of this paper are consistent with existing studies that different rainfall characteristics produce different runoff loads and runoff concentrations, which in turn lead to different runoff water quality (An et al., 2014).

As can be seen from Figure 9, the ranked values of DO concentration response in the rivers of Zhongshan city area under the influence of typhoon rain events show some correlation with the distribution of O_3 concentration and the distribution of SO_2 concentration. Specifically, the response rankings of DO concentration were positively correlated with the distribution of O_3 and SO_2 concentrations in the eastern, central and western parts of Zhongshan City, i.e., the larger the concentrations of O_3 and SO_2, the larger the response rankings of DO concentration, and the water quality in the river had a tendency to deteriorate under the impacts of the typhoon rain events; The response ranking of DO concentration is negatively correlated with the distribution of O_3 and SO_2 concentrations in

the north and south of Zhongshan City, i.e., the larger the concentrations of O_3 and SO_2 are, the smaller the response ranking of DO concentration is, and there is a tendency for the water quality in the river to be better under the influence of the typhoon rain events. This result is consistent with conventional knowledge: the central, western and eastern parts of Zhongshan City are densely populated urban centers, where intensive human activities provide a large amount of O_3 precursors and SO_2; while the southern and northern parts of Zhongshan City are not only set up as agroforestry ecological zones, but also as an agricultural development area in the southern part of the city, where the content of O_3 precursors and SO_2 is lower. Under the condition of similar rainfall, the DO concentration response ranking values showed positive correlation with O_3 concentration distribution and SO_2 concentration distribution in the east, center and west of Zhongshan City, and negative correlation in the north and south of Zhongshan City, which is in line with the reality.

In conclusion, the results of this study are accurate and credible, and can provide a visual reference and theoretical basis for urban managers to develop adaptive water governance and carry out targeted disaster prevention and mitigation work (Liya et al., 2022).

5 Conclusion

Characterizing the response of dissolved oxygen concentration in urban water bodies under the influence of a single typhoon is essential for city managers to make decisions on response programs and input budgets. In this study, the extended MULTIMOORA method was used to establish the correlation between typhoon drivers and the response of DO concentration in urban water bodies, and to quantify and evaluate the characteristics of the response distribution of DO concentration in urban water bodies. The results showed that the response of DO concentration in the water body of the river in the center of the city under the action of the more intense typhoon rain events was developed to be better; the response of DO concentration in the water body at the mouth of the sea tended to be worse. Under the effect of less intense typhoon rain events, the scouring effect of rainwater dominates and there is a tendency for the response of DO concentration in the water body to become worse. Meanwhile, this study used the MGWR model and spatial autocorrelation analysis to explore the impact of human activities on the distribution of urban water environment response. The results show that the response ranking values of DO concentration in rivers in Zhongshan city under the influence of the typhoon rain events are positively correlated with the distribution of O_3 concentration and SO_2 concentration in the east, center and west of Zhongshan city, and negatively correlated with the distribution of O_3 concentration and SO_2 concentration in the north and south of Zhongshan city. The results of this study can provide visual reference and theoretical support for local governments and city managers in developing adaptive water management and targeted disaster prevention and mitigation programs.

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

WS: Writing – original draft. YJ: Funding acquisition, Methodology, Writing – review & editing. GL: Funding acquisition, Resources, Writing – review & editing. PC: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (grant no. 42101422 and no. 52178209) and the Science and Technology Program of Guangzhou (grant no. 201605030009 and no. 201803020036). The project was supported by the Guangdong Engineering and Research Center for Unmanned Aerial Vehicle Remote Sensing of Agricultural Water and Soil Information.

Acknowledgments

The authors would like to thank the editors and reviewers for their useful comments, which was helpful in improving the quality of the manuscript.

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

An, L., Yuntao, G., and Liang, L. (2014). Impact of catchment and rainfall characteristics on urban runoff quality (in Chinese). *J. Tsinghua University(Science Technology)* 54, 7. doi: 10.16511/j.cnki.qhdxxb.2014.07.004

Bing, X., and Hua, Z. (2014). Main progresses in the Research on ozone (in Chinese). *Sci. Technol. Eng.* 14, 9. doi: 10.3969/j.issn.1671-1815.2014.08.019

Brauers, W. K. M., and Zavadskas, E. K. (2006). The MOORA method and its application to privatization in a transition economy. *Control Cybern.* 35, 25. doi: 10.1007/s10458-005-3783-9

Brauers, W. K. M., and Zavadskas, E. K. (2010). Project management by multimoora as an instrument for transition economies / projektŲ Vadyba su multimoora kaip priemonĖ Pereinamojo laikotarpio ūkiams. *Technol. Econ. Dev. Econ.* 16, 5–24. doi: 10.3846/tede.2010.01

Brauers, W. K. M., and Zavadskas, E. K. (2012). Robustness of MULTIMOORA: A method for multi-objective optimization. *Informatica* 23, 1–25. doi: 10.15388/Informatica.2012.346

Cabrera, J. S., and Lee, H. S. (2019). Flood-prone area assessment using GIS-based multi-criteria analysis: A case study in Davao oriental, Philippines. *Water* 11. doi: 10.3390/w11112203

Chaojia, N., Huichun, Y., Shiwen, Z., Jiawei, G., Bei, C., and Wenjiang, H. (2022). Risk assessment and sustainable development countermeasures of agricultural typhoon disaster for Hainan Island of China (in Chinese). *Trans. Chin. Soc. Agric. Eng.* 38, 10. doi: 10.11975/j.issn.1002-6819.2022.23.025

Chen, H., Zhu, Y., Zhang, Y., Chen, X., Wang, R., and Zhu, W. (2020). Cyanobacterial bloom expansion caused by typhoon disturbance in Lake Taihu China. *Environ. Sci. pollut. Res. Int.* 27, 42294–42303. doi: 10.1007/s11356-020-09292-0

Chen, R., Xue, W., Zi-Wen, W., Hao, Q., Tie-Min, M., Zheng-Guang, C., et al. (2023). Wavelength selection method of near-infrared spectrum based on random forest feature importance and interval partial least square method (in Chinese). *Spectrosc. Spectral Anal.* 43, 8. doi: 10.3964/j.issn.1000-0593(2023)04-1043-08

Chen, Z., Han-An, L., Xue-Ping, G., and Wen-Na, Z. (2016). Potential impacts of climate change on phosphorus and DO in Yuqiao reservoir (in Chinese). *Environ. Sci.* 37, 8. doi: 10.13227/j.hjkx.2016.08.014

Cui, Y., Jin, J., Huang, Q., Yuan, K., and Xu, C. (2022). A data-driven model for spatial shallow landslide probability of occurrence due to a typhoon in Ningguo City, Anhui province, China. *Forests* 13. doi: 10.3390/f13050732

Espinosa-Diaz, L. F., Zapata-Rey, Y. T., Ibarra-Gutierrez, K., and Bernal, C. A. (2021). Spatial and temporal changes of dissolved oxygen in waters of the Pajarales complex, Cienaga Grande de Santa Marta: Two decades of monitoring. *Sci. Total Environ.* 785, 147203. doi: 10.1016/j.scitotenv.2021.147203

Fotheringham, A. S., Charlton, M. E., and Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plann. A* 30, 23. doi: 10.1068/a301905

Fu, X., Liu, J., Mei, C., Luan, Q., Wang, H., Shao, W., et al. (2021). Effect of typhoon rainstorm patterns on the spatio-temporal distribution of non-point source pollution in a coastal urbanized watershed. *J. Clean Prod.* 292. doi: 10.1016/j.jclepro.2021.126098

Gao, Z., Geddes, R. R., and Ma, T. (2020). Direct and indirect economic losses using typhoon-flood disaster analysis: an application to Guangdong province, China. *Sustainability* 12. doi: 10.3390/su12218980

Ge, W., Qiang, Y., Di, Y., Teng, N., and Qian-Qian, L. (2022). Retrieval of dust retention distribution in Beijing urban green space based on spectral characteristics. *Spectrosc. Spectral Anal.* 42, 7. doi: 10.3964/j.issn.1000-0593(2022)08-2572-07

Guo, L., Wang, M., Zhong, L., and Zhang, Y. (2020). Calculation model for the mixing amount of internal curing materials in high-strength concrete based on modified MULTIMOORA. *Sci. Eng. Compos. Mater.* 27, 455–463. doi: 10.1515/secm-2020-0048

Hafezalkotob, A., Hafezalkotob, A., Liao, H., and Herrera, F. (2019). An overview of MULTIMOORA for multi-criteria decision-making: Theory, developments, applications, and challenges. *Inf. Fusion* 51, 145–177. doi: 10.1016/j.inffus.2018.12.002

Han, R., Zhou, B., An, L., Jin, H., Ma, L., Li, N., et al. (2019). Quantitative assessment of enterprise environmental risk mitigation in the context of Na-tech disasters. *Environ. Monit Assess.* 191, 210. doi: 10.1007/s10661-019-7351-1

Hongxiang, F., Ligang, X., Hua, Z., Zhao, L., Yuxian, C., Yakun, W., et al. (2021). Distinguishing the relative of climate change and anthropogenic activities on variation of water age in the Lake Poyang (in Chinese). *J. Lake Sci.* 33, 13. doi: 10.18307/2021.0419

Howarth, R. W. (2008). Coastal nitrogen pollution: A review of sources and trends globally and regionally. *Harmful Algae* 8, 14–20. doi: 10.1016/j.hal.2008.08.015

Howarth, R., Chan, F., Conley, D. J., Garnier, J., Doney, S. C., Marino, R., et al. (2011). Coupled biogeochemical cycles: eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. *Front. Ecol. Environ.* 9, 18–26. doi: 10.1890/100008

Huanhuan, L., Da, A., Yu, Y., Yue, W., Beidou, X., Minghong, W., et al. (2016). Risk ranking for groundwater pollution from hazardous waste landfills by MCDA model (in Chinese). *Res. Environ. Sci.* 29, 7. doi: 10.13198/j.issn.1001-6929.2016.01.17

Jie, J., Yong, Z., Xiao-Qiang, C., and Dan, S. (2011). Study on concentration and transport characteristics of sulfur dioxide over the Bohai Sea (in Chinese). *Acta OCEANOLOGICA Sin.* 33, 6.

Jin, Y., Zhang, H., Shi, H., Wang, H., Wei, Z., Han, Y., et al. (2021). Assessing spatial heterogeneity of factor interactions on PM2.5 concentrations in Chinese cities. *Remote Sens.* 13. doi: 10.3390/RS13245079

Jin-Dian, M., Xiao-Ming, Z., Tian-Xing, W., Yang, Z., Peng, L., and Li-Gang, Z. (2021). Temporal and spatial distribution characteristics of net nitrogen and phosphorus input from human activity: A case study of Hangzhou section of Qiandao Lake Basin (in Chinese). *China Environ. Sci.* 41, 12. doi: 10.19674/j.cnki.issn1000-6923.2021.0287

Jun, L., and Shi-Hua, L. (2022). Probabilistic hesitant Fermatean fuzzy extension MULTIMOORA method for evaluation of regional green restoration level (in Chinese). *Control Decision* 37, 11. doi: 10.13195/j.kzyjc.2021.0437

Kim, J.-M., Son, K., Yum, S.-G., Ahn, S., and Ferreira, T. (2020). Typhoon vulnerability analysis in South Korea utilizing damage record of typhoon Maemi. *Adv. Civil Eng.* 2020, 1–10. doi: 10.1155/2020/8885916

Kut, P., and Pietrucha-Urbanik, K. (2022). Most searched topics in the scientific literature on failures in photovoltaic installations. *Energies* 15. doi: 10.3390/en15218108

Lee, Y. H., J.Haberman, S., and Dorans, N. J. (2019). Use of adjustment by minimum discriminant information in linking constructed-response test scores in the absence of common items. *J. Educ. Measurement* 56, 21. doi: 10.1111/jedm.12216

Lei, G., Sihao, L., Lixia, G., Jun, W., Pingping, C., and Jiantao, Z. (2022). Optimization of PCCP welding process parameters based on improved MULTIMOORA method (in Chinese). *Trans. OF THE China WELDING INSTITUTION* 43, 8. doi: 10.12073/j.hjxb.20211018001

Liu, Y., Ke-Ming, M., Qing-Hai, G., and Jing-Zhu, Z. (2004). Impacts of the urbanization on waters non-point source pollution. *Environ. Sci.* 25, 8. doi: 10.13227/j.hjkx.2004.06.007

Liu, D., Pang, L., and Xie, B. (2009). Typhoon disaster in China: prediction, prevention, and mitigation. *Nat. Hazards* 49, 421–436. doi: 10.1007/s11069-008-9262-2

Lixia, G., Minghua, W., Yuan, T., and Yanan, Z. (2020). Performance evaluation on internal curing high-strength concrete based on improved MULTIMOORA (in Chinese). *Yangtze River* 51, 6. doi: 10.16232/j.cnki.1001-4179.2020.11.028

Liya, G., Linruo, Z., and Zhiran, X. (2022). Water security under the impact of climate change and sustainable solutions (in Chinese). *China Water Resour.* 9, 3.

LuanXiao, Z., Jinshui, L., Yunxia, Y., Kai, Z., Jiqiang, M., Caifeng, Z., et al. (2021). Quantitative seismic characterization of source rocks in lacustrine depositional setting using the Random Forest method: An example from the Changjiang sag in East China Sea basin (in Chinese). *Chin. J. Geophys.* 64, 16. doi: 10.6038 / cjg2021O0123

Mardani, A., Zavadskas, E. K., Khalifah, Z., Zakuan, N., Jusoh, A., Nor, K. M., et al. (2017). A review of multi-criteria decision-making applications to solve energy management problems: Two decades from 1995 to 2015. *Renewable Sustain. Energy Rev.* 71, 216–256. doi: 10.1016/j.rser.2016.12.053

Mishra, B., Kumar, P., Saraswat, C., Chakraborty, S., and Gautam, A. (2021). Water security in a changing environment: concept, challenges and solutions. *Water* 13. doi: 10.3390/w13040490

Mitra, A., Banerjee, K., and Sengupta, K. (2011a). Impact of AILA, a tropical cyclone on salinity, pH and dissolved oxygen of an aquatic sub-system of Indian Sundarbans. *Proc. Natl. Acad. Sci. India Section B-Biological Sci.* 81, 198–205.

Mitra, A., Halder, P., and Banerjee, K. (2011b). Changes of selected hydrological parameters in Hooghly estuary in response to a severe tropical cyclone (Aila). *Indian J. Geo-Marine Sci.* 40, 32–36.

Muñoz, H., Orozco, S., Vera, A., Suárez, J., García, E., Neria, M., et al. (2015). Relación entre ox ígeno disuelto, precipitación pluvial y temperatura: río Zahuapan, Tla xcala, México. *Water Technol. Sci. (in Spanish)* 6, 16.

Pearce, M. W., and Schumann, E. H. (2003). Dissolved oxygen characteristics of the gamtoos estuary, South Africa. *Afr. J. Mar. Sci.* 25, 99–109. doi: 10.2989/18142320309504003

Qiong, Y., Haijun, Z., Guohui, Z., Hang, L., Yueyan, L., and Zhengming, X. (2021). Research on green degree evaluation of cutting process based on MULTIMOORA (in Chinese). *Modern Manufacturing Eng.* 7, 10–16. doi: 10.16731/j.cnki.1671-3133.2021.11.002

Qi-ting, Z., Jun-Xia, M., and Chuan-Chang, G. (2005). Study on carrying capacity of urban water environment (in Chinese). *Adv. Water Sci.* 16, 6. doi: 10.14042/j.cnki.32.1309.2005.01.019

Shenghua, X., Zhengsong, C., Yonggang, C., and Qiang, H. (2019). Recognition method of airline fleet reliability based on index fuzzy segmentation and MULTIMOORA (in Chinese). *Comput. Integrated Manufacturing Syst.* 25, 8. doi: 10.13196 / j.cims.2019.02.016

Shufang, S., and Ruyang, H. (2021). Importance measure index system based on random forest (in Chinese). *JOURNAL OF Natl. UNIVERSITY OF DEFENSE Technol.* 43, 8. doi: 10.11887/j.cn.202102004

Shuping, Z., Lijian, H., Weiqi, Z., and Weifeng, L. (2016). Impact of urban population on concentrations of nitrogen dioxide (NO2) and fine particles (PM2.5) in China (in Chinese). *Acta ECOLOGICA Sin.* 36, 9. doi: 10.5846/stxb201502050292

Songbai, G., Aiguo, G., Guantao, N., Xuxu, Z., Yanpo, Z., and Yuting, H. (2017). Progress in research of hypoxia in estuaries and coastal areas in China (in Chinese). *Water Resour. Prot.* 33, 8. doi: 10.3880/j.issn.1004-6933.2017.04.010

Wang, Y., Yin, Y., and Song, L. (2022). Risk assessment of typhoon disaster chains in the Guangdong–Hong Kong–Macau greater bay area, China. *Front. Earth Sci.* 10. doi: 10.3389/feart.2022.839733

Weihui, H., Chunzi, M., Wenpan, L., Zhuoshi, H., Hanxiao, Z., and Shouliang, H. (2021). Spatial-temporal variations of dissolved oxygen and their response to global warming in China (in Chinese). *Acta Scientiae Circumstantiae* 41, 11. doi: 10.13671/j.hjkxxb.2020.0330

Xiaoran, O., Qiaohua, Z., and Yingzhu, W. (2013). A preliminary exploration of dissolved oxygen based on FVCOM in Meiliang Bay,Lake Taihu and its influence mechanism (in Chinese). *J. Lake Sci.* 25, 11.

Xiao-wen, L., Lu-Quan, W., Ya-Guang, Z., Yun-Zhao, C., Ming-Yi, W., Jun-Ping, Z., et al. (2021). Random forest retinal segmentation in OCT images based on principal component analysis (in Chinese). *Prog. Biochem. Biophys.* 48, 8. doi: 10.16476/j.pibb.2020.0270

Xiu-qin, O., Bo, W., Jian-Lin, S., Xiao, Z., Jie-Fei, W., Yong, L., et al. (2019). Atmospheric nitrogen dioxide,Nitric acid,Nitrate nitrogen concentrations,and wet and dry deposition Rates in a double Rice Region in subtropical China (in Chinese). *Environ. Sci.* 40, 8. doi: 10.13227/j.hjkx.201810252

Xu, Z. (2005). An overview of methods for determining OWA weights. *Int. J. Intell. Syst.* 20, 843–865. doi: 10.1002/int.20097

Xuehua, G., Zhihe, C., and Zhiheng, Y. (2022). Dynamic simulation and sensitivity analysis of water resources system in Zhongshan City (in Chinese). *Water Resour. Prot.* 38, 9. doi: 10.3880/j.issn.1004-6933.2022.02.015

Ye, H., Chen, C., Tang, S., Tian, L., Sun, Z., Yang, C., et al. (2014). Remote sensing assessment of sediment variation in the Pearl River Estuary induced by Typhoon Vicente. *Aquat. Ecosyst. Health Manage.* 17, 271–279. doi: 10.1080/14634988.2014.944475

Ye, G., Tian-Jun, D., Jian-Hui, X., Guo-Dong, Q., and Xue-Jun, Y. (2021). Analysis of the impact of typhoon "Lekima" Rainfall on urban surface water quality based on automatic monitoring (in Chinese). *Environ. Sci. Technol.* 34, 6. doi: 10.19824/j.cnki.cn32-1786/x.2021.0013

Yihui, Z., Yuemin, H., Zhaoliang, P., Weiping, H., and Jinge, Z. (2022). Fast-changing wind waves and their environmental effects in Lake Chaohu (in Chinese). *J. Lake Sci.* 34, 9. doi: 10.18307/2022.0322

Yu, H., Song, Y., Chang, X., Gao, H., and Peng, J. (2018). A scheme for a sustainable urban water environmental system during the urbanization process in China. *Engineering* 4, 190–193. doi: 10.1016/j.eng.2018.03.009

Yu, W., and Ze-Shui, X. (2008). A new method of giving OWA weights (in Chinese). *Mathematics Pract. Theory* 38, 11.

Yuan-hua, H., Yi-Lan, H., and Sheng-Hua, X. (2023). An evaluation model of airport fire station site selection based on entropy weight intuitionistic fuzzy extended MULTIMOORA (in Chinese). *Control Decision* 38, 9. doi: 10.13195/j.kzyjc.2021.0247

Yufeng, Z., Liu, S., Liqing, L., Hong, Z., and Wenzhong, T. (2023). Treatment of black and odorous water I : Effects of water oxygen state on speciation and bioavailability of heavy metals in sediments (in Chinese). *Acta Scientiae Circumstantiae* 43, 10. doi: 10.13671/j.hjkxxb.2022.0199

Yuling, H., Jinxin, F., Zeping, W., Mengnan, G., and Shunxin, F. (2022). Assessing requirements for dissolved oxygen in Yunnan plateau lakes (in Chinese). *J. China Institute Water Resour. Hydropower Res.* 20, 7. doi: 10.13244/j.cnki.jiwhr.20210300

Zhang, J., and Chen, Y. (2019). Risk assessment of flood disaster induced by typhoon rainstorms in Guangdong province, China. *Sustainability* 11. doi: 10.3390/su11102738

Zhang, W., Zhou, X., Wei, W., and Cheng, X. (2022). Risk assessment of water inrush in tunnels: A case study of a tunnel in Guangdong province, China. *Sustainability* 14. doi: 10.3390/su141811443

Zhao, Y., Yang, K., Luo, Y., and Yu, Z. (2022). Spatial–temporal characteristics of surface thermal environment and its effect on Lake surface water temperature in Dianchi Lake basin. *Front. Ecol. Evol.* 10. doi: 10.3389/fevo.2022.984692

Zhilin, Z., Jing, Q., Tao, C., Bensheng, H., and Ziying, F. (2022). Urban flood issues and analysis in the Guangdong-Hong Kong-Macao Greater Bay Area (in Chinese). *J. Hydraulic Eng.* 53, 10. doi: 10.13243/j.cnki.slxb.20220207

Zhipeng, G., Yuao, M., Huayi, W., and Rui, L. (2022). Urban population spatialization by considering the heterogeneity on local resident attraction force of POIs (in Chinese). *J. Geo-information Sci.* 24, 15. doi: 10.12082/dqxxkx.2022.220384

Zhou, W., Yin, K., Harrison, P. J., and Lee, J. H. W. (2012). The influence of late summer typhoons and high river discharge on water quality in Hong Kong waters. *Estuarine Coast. Shelf Sci.* 111, 35–47. doi: 10.1016/j.ecss.2012.06.004

Appendix A. OWA operator:

First, P experts are invited to rate the target Q. The rating results are then arranged in descending order, resulting in $(a_0, a_1, \ldots, a_j, \ldots, a_{p-1})$, where $a_0 > a_1 > \ldots > a_j > \ldots > a_{p-1}$. The weight y_{j+1} for a_j is determined by combinatorial calculations.

$$y_{j+1} = \frac{C_{p-1}^{j}}{2^{p-1}} \quad (j = 0, 1, \ldots, p-1) \tag{1}$$

According to the weights y_{j+1}, the target data is sequentially weighted to obtain the absolute weights of the target, denoted as $\overline{w_i}$.

$$\overline{w_i} = \sum_{j=1}^{p} y_{j+1} a_j (i = 1, 2, \ldots, p) \tag{2}$$

Hence, the subjective weights w_i calculated by the OWA operator can be obtained.

$$w_i = \frac{\overline{w_i}}{\sum_{i=1}^{p} \overline{w_i}} (i = 1, 2, \ldots, p) \tag{3}$$

Appendix B. The process of solving objective weights:

The formulae (LuanXiao et al., 2021; Xiao-wen et al., 2021) are shown below.

$$g(D \mid O) = H(D) - H(D \mid O) \tag{1}$$

$$mse_{OOBm}^{i} = mse(Y_P^i - Y)^2 - mse(Y_P - Y)^2 \tag{2}$$

$$IMP^i(X_{OOB}^i) = \frac{1}{T} \sum_{t=1}^{T} mse_{OOBm}^i \tag{3}$$

Where: $g(D|O)$ is the information gain value, $H(D)$ is the overall entropy value before branching, $H(D|O)$ is the entropy value after pre-branching in condition O. Y represents the true values, Y_P represents the predicted values, Y_P^i denotes the predicted values after modifying the feature variables, mse_{OOBm}^i represents the mean squared error of the feature variables for decision tree t_m, T is the number of decision trees, and $IMP^i(X_{OOB}^i)$ signifies the importance results of the feature variables.

Calculate the objective weights w_o as follows:

$$w_o = \frac{IMP^i(X_{OOB}^i)}{\sum_{i=1}^{n} IMP^i(X_{OOB}^i)} \tag{4}$$

Appendix C. The process of solving comprehensive weights:

For a discrete random variable $X = (x_1, x_2, \ldots, x_n)$, the probability distribution of x is only related to the condition δ and the condition τ. Under the condition δ, the x_k probability distribution function is $\delta(x_k)$; similarly, under the condition τ, the x_k probability distribution function is $\tau(x_k)$. Thus, the discriminatory information (Lee et al., 2019) is expressed as:

$$I[\tau(x), \delta(x)] = \sum_{k=1}^{n} \tau(x) log \frac{\tau(x)}{\delta(x)} \tag{1}$$

Since both subjective and objective weight data are discrete random variables, in order to enhance the credibility and accuracy of the composite weight for evaluation metrics with smaller errors, this study employs the Minimum Discriminant Information Principle to determine a composite weight that closely approximates both. The solution is obtained by introducing Lagrange multipliers into the equation.

$$Lag(x, \psi) = \theta(x) - \psi\delta(x) \tag{2}$$

Where: $\theta(x)$ represents the original function, and $\delta(x)$ is the constraint condition function, and ψ denotes the Lagrange multiplier.

Establishing the objective function as follows:

$$\begin{cases} \min\left(I[w_s, w_i] + I[w_s, w_o]\right) = \sum_{k=1}^{n} (w_s ln \frac{w_s}{w_i} + w_s ln \frac{w_s}{w_o}) \\ \sum_{k=1}^{n} w_s - 1 = 0 \end{cases} \tag{3}$$

Substituting the Lagrange multiplier yields:

$$Lag(w_s, \psi) = \sum_{k=1}^{n} (w_s ln \frac{w_s}{w_i} + w_s ln \frac{w_s}{w_o}) - \psi(\sum_{k=1}^{n} w_s - 1) \tag{4}$$

Taking the partial derivative of Equation 4 results in:

$$w_s = \frac{\sqrt{w_i w_o}}{\sum_{i=1}^{p} \sqrt{w_i w_o}} \tag{5}$$

Where: w_s is the combined weight of the evaluation indicators, w_i is the subjective weight of the evaluation indicators and w_o is the objective weight of the evaluation indicators.

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 07 November 2023
DOI 10.3389/fenvs.2023.1244238

OPEN ACCESS

EDITED BY
Jie Luo,
Zhejiang Lab, China

REVIEWED BY
Thabiso Sebolai,
Department of Agricultural Research (DAR), Myanmar
Abel Ramoelo,
University of Pretoria, South Africa

*CORRESPONDENCE
Simeneh Admasu,
simeneh2009@yahoo.com,
simeneha2009@gmail.com

RECEIVED 22 June 2023
ACCEPTED 25 October 2023
PUBLISHED 07 November 2023

CITATION
Admasu S (2023), Modeling habitat quality for rangeland ecosystem restoration in the Alledeghi Wildlife reserve, Ethiopia. *Front. Environ. Sci.* 11:1244238. doi: 10.3389/fenvs.2023.1244238

Modeling habitat quality for rangeland ecosystem restoration in the Alledeghi Wildlife reserve, Ethiopia

Simeneh Admasu[1,2]*

[1]Environmental Planning Program (EiABC), Addis Ababa University, Addis Ababa, Ethiopia, [2]GFA Consulting Group, GIZ Biodiversity, and Forests Program, Addis Ababa, Ethiopia

Assessment of habitat quality is important for the planning and management of conservation measures at a landscape level. The Alledeghi Wildlife Reserve (AWR) is an iconic wildlife conservation area because it not only contains charismatic wildlife species but also serves as a reliable livestock rangeland. The study aimed to assess habitat quality based on landscape-specific habitat threat information in the AWR using InVEST software. it is the first attempt to model the habitat quality of the landscape using expert-driven information. Six important threats were considered, namely, invasive species, bush encroachment, livestock incursion, fire, habitat destruction, and distance to roads. The quantified habitat quality was classified into low, moderate, and high. The results revealed that the quality of the habitat declined in the study area between 1998 and 2016. The high-quality habitat had a larger extent covering about 837 km^2 (57.4%) in 1998 but it was reduced by 128 km^2 (64%) during the study period. Conversely, moderate quality and low-quality habitats have increased from 78 km^2 (5.35%) in 1998 to 206 km^2 (14.12%) in 2016; and from 544 km^2 (37.3%) in 1998 to 619 km^2 (42.13%) in 2016 respectively. The decline in habitat quality was mainly associated with increased livestock incursion and expansion of invasive species which resulted in rapid land use changes. Thus, it is critical to undertake serious conservation measures to enhance biodiversity and ecosystem services in the AWR and to substantively contribute to the improved livelihood of the pastoral community.

KEYWORDS

AWR, biodiversity, ecosystem quality, InVEST, rangeland

1 Introduction

Biodiversity enhances ecosystem functionality which leads to improved ecosystem services through balancing and stabilizing ecological communities across scales (Cardinale et al., 2012; Oliver et al., 2015). The biodiversity of an area can be determined through habitat conditions whereas the importance of habitat depends on its quality (Basane and James, 2016). A particular habitat with special ecological importance is essential to the functioning of the wider ecosystem processes; such areas require extraordinary protection to safeguard the special value and vital ecosystem processes. Biological resources and the levels of biological reproduction and organism persistence have a greater effect on the ability of an ecosystem to provide living conditions for individual organisms and populations (Caro et al., 2020).

High-quality habitat is assumed as an indication of rich biodiversity (Norliyana and Mamat, 2020) and delivery of a suite of ecosystem services (Johnson, 2007; Stolton et al., 2010; Thomas et al., 2021). The state of biodiversity can be used as a basis proxy tool to measure the quality of a given habitat (Havlicek and Mitchell, 2014). Therefore, as a proxy for the condition of the state of biodiversity, quality habitat can indicate the capability of a given ecosystem to provide essential ecosystem services (Polasky et al., 2011) and as a determinant for measuring ecosystem health (Villamagna et al., 2013). The occurrence of diverse wildlife species is highly associated with the quality of habitat (Edmonds et al., 2021).

Habitat quality is an important indicator of regional ecological security (Zhu et al., 2015; Chen et al., 2016), which can reflect the level of regional biodiversity and ecosystem services (Tang et al., 2020; Zhu et al., 2020). Rigorous information on habitat quality is invaluable to making informed decisions on conservation planning and prioritization of conservation intervention strategies (Rouget et al., 2003; Baral et al., 2014; Simeneh et al., 2023) including expansion of important biodiversity areas, introduction and removal of species, and identification of principal habitat components (Basane and James, 2016) and determining of the key ecological attributes.

The landscape changes lead to corresponding modifications in the composition of the ecosystem and biodiversity (Liu et al., 2022). Further habitat quality changes affect the biodiversity and landscape pattern (Chu et al., 2018). Therefore, the occurrence of severe and complex ecological problems at landscape and species levels have a direct influence on the landscape pattern and habitat quality. Understanding the association between conservation challenges caused by land use change could provide a solution to ecological problems (Bai et al., 2019). Habitat loss consistently negatively affects species richness and population abundance (Laurance et al., 2002); and genetic diversity (Aguilar et al., 2008). The loss of critical habitats affects not only biodiversity but also directly impacts humans by decreasing the production of ecosystem services such as pollination (Potts et al., 2010), soil productivity and water provision (Bruijnzeel, 2004), and carbon storage and sequestration (Fargione et al., 2008).

The state of biodiversity, the range of habitats, and vegetation types across landscapes can be determined using the InVEST habitat quality and rarity models (Sharp et al., 2020; Liu et al., 2022). Thus, the changes in habitat quality are critical to the changes in ecosystem processes (Choudhary et al., 2021; Yang, 2021). Habitat quality monitoring provides robust information on ecological conditions and can be utilized as a basis for making habitat conservation interventions (Lin et al., 2016). Changes in habitat quality have tremendous implications for the conservation of wildlife species in savannah ecosystems (Kija et al., 2020) where the ecosystems are the principal habitats for diverse charismatic wildlife species and home to many iconic protected landscapes (Sinclair et al., 2007; Bohm and Hofer, 2018).

The Alledeghi Wildlife Reserve (AWR) is among the highly valued protected landscapes in Ethiopia which are highly pronounced with the assemblage of large mammals, but it is under severe conservation challenges, and the biodiversity endowment of the area particularly large mammals alarmingly declining (Fanuel, 2013; Simeneh et al., 2016). The important threats to biodiversity are steadily increasing such as the fast spread of invasive species, overgrazing, and bush encroachment (Almaz, 2009; Selamnesh, 2015) because of rapid Land Use Land Cover (LU/LC) changes, the ecosystem services values of the area have greatly declined (Simeneh, 2023). Moreover, urban development along the road is becoming an emerging conservation threat that will constrain the sustainability of the ecosystem (Almaz, 2009; Fanuel, 2013). Further, intensive charcoal production is well-pronounced in the entire area. Thus, this results in massive habitat destruction in the adjacent protected areas including the Awash National Park. Roadkill incidence has repeatedly occurred while wild animals are crossing the asphalted road in search of water (Simeneh et al., 2016). Fire incidence mainly in the highland forest is becoming a very common challenge for protected area management as local charcoal makers deliberately set fire to produce more charcoal.

The study hypothesizes that habitat quality declined over time in response to threat factors occurring in the study area. There is a lack of empirical studies conducted in the study area that assessed the status of the habitats to protect the values that the protected area possessed. Therefore, the novelty of this study is that it is the first attempt to model the quality of habitats of the protected landscape using expert-driven landscape threat information and analysis to indicate the state of the protected areas towards meeting its conservation goal. Therefore, this study aimed to assess the spatiotemporal changes in the quality of the habitat in the terrestrial ecosystems of AWR using InVEST software to provide a scientific basis for ecosystem planning interventions and prioritization of conservation management undertakings.

2 Methods and materials

2.1 Description of the study area

The AWR was established in the 1960 s (Hilliman, 1993). It is located in the Great Rift Valley in the northeastern region of the country between longitude 39°30′to 40°30′E and latitude 8°30′to 9°30′N, at 280 km east of Addis Ababa (Figure 1). The altitude ranges between 776 m and 2,445 m above sea level. The area is characterized by a semi-arid ecosystem with annual rainfall ranging between 400 and 700 mm. About 268 plant species and two types of ecosystems Dry evergreen montane forest and Acacia comiphora ecosystems (Addisu et al., 2017), 31 species of mammals, and over 140 avian species have been recorded (Hilliman, 1993; Fanuel, 2013) in the AWR. The most common wild animals inhabiting the reserve include the Grevy zebra (Equus grevyi), Beisa oryx (Oryx beisa beisa), Soemmering's gazelle (Gazella soemmering), Gerenuk (*Litocranius walleri*) and lesser kudu (Tragelaphus imberbis) (Hilliman, 1993).

The major vegetation types in and around the reserve include grasslands, bushland, woodland, riverine forests, and highland forests (Almaz, 2009). The grassland plain stretching from the center of the reserve to the northwest was mainly occupied by grasses and occasionally with other herbs; the dominant species include Durfu (Chrysopogon plumulosus), Isisu (Chrysopogon schoenan) and Malif (Andropogon canaliculatus) (Almaz, 2009; Selamnesh, 2015). However, the rapid encroachment of shrub species and the rapid spread of invasive Prosopis juliflora and shrubs such as Combretum aculatum, Merua oblongflora, and Terminalia species have affected the grass species and the extent

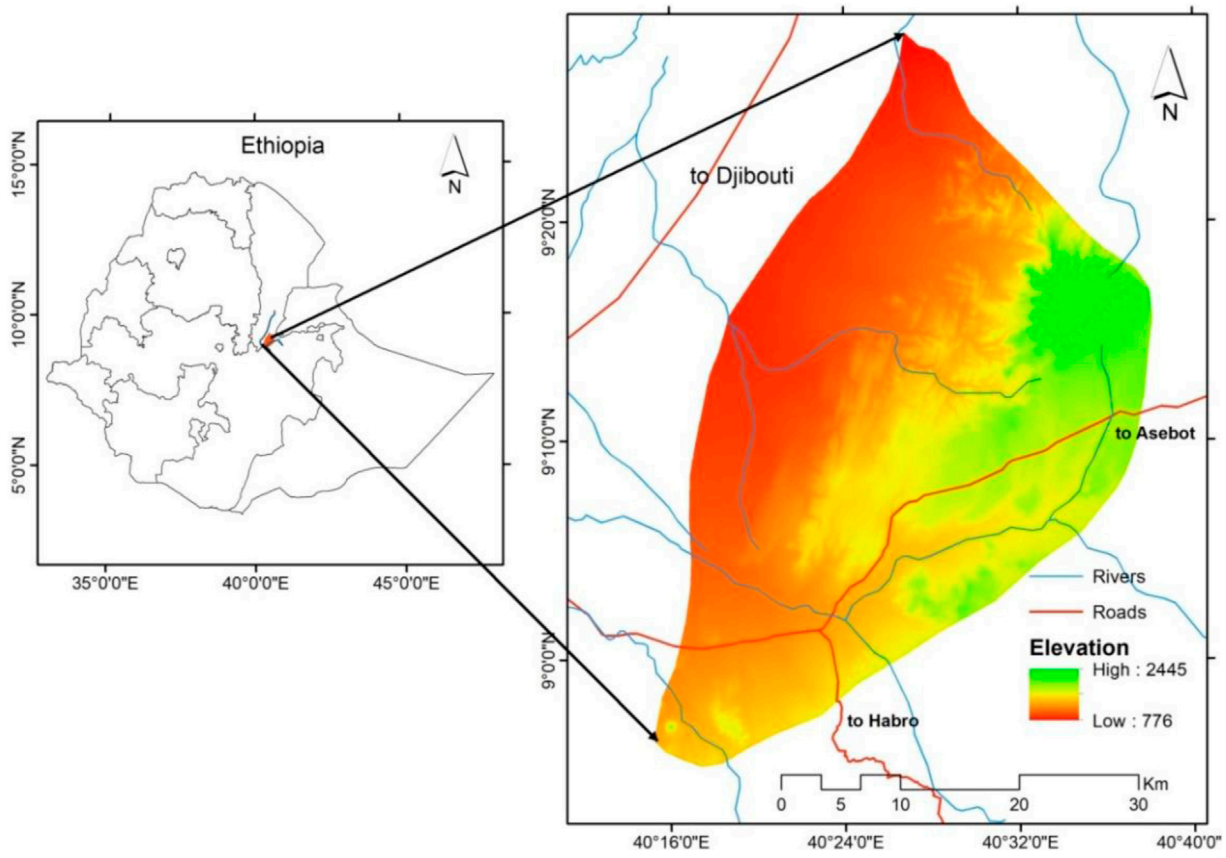


FIGURE 1
Map of the Alledighe landscape.

of the grassland habitat (Selamnesh, 2015; Simeneh, 2023). The bushland is an extensively increasing habitat type that possesses an assemblage of trees and shrubs (Simeneh, 2023). The habitat is mainly occurring in the southern, eastern, and northern edges of the landscape dominated by Acacia senegal (Almaz, 2010; Selamnesh, 2015). The common woody plant species in the AWR include Acacia tortilis, Acacia mellifera, Balanitis aegyptiaca, Cadaba, and Grewia species. The eastern mountainous section of the landscape is characterized by dense highland forest, common plant species include Cordia africana, Croton macrostachyus, Erythrina abyssinica, Juniperus procera, Olea europaea, Podocarpus falcatus, Pouteria altissima and Rhus vulgaris (Almaz, 2010). The riverine forests are a unique ecosystem and are important for the wild animals of the landscape. It is limited to seasonal streams and river courses where the water table is high.

2.2 Materials

2.2.1 Application of tools to assess habitat quality

Habitat quality can be assessed based on measured species diversity or through the analysis of the evolution of the habitat by parameter substitution (Andrus et al., 2021). In general, comparing observations to a standardized list of criteria can be used to assess the quality of a given habitat (Machado, 2004), and more recently the standardized modeling tool, particularly the Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) has been largely applied to measure the quality of habitat at various scales (Sharp et al., 2020). In this study, we assessed the state of habitats using InVEST software 3.11 version. The quantified habitat quality was classified by natural breaks into three classes (low, moderate, and high) (Kija et al., 2020).

2.2.2 The InVEST habitat quality model

The InVEST model provides good research methods and perspectives (Romero-Calcerrada and Luque, 2006; Terrado et al., 2016; Abreham et al., 2020). The model incorporates land use and biodiversity threats information to produce habitat quality maps. It uses the spatial extent of habitat quality as a proxy of biodiversity within the landscape, based on the proximity of the habitat to human-dominated land use and the intensity of disturbance caused by the land use (Sharp et al., 2020). The model considers that LU/LC with higher habitat quality is relatively intact and capable of supporting increased biodiversity and a lower habitat quality score indicates reduced biodiversity support and denotes a degraded landscape (Baral et al., 2014). The model is dependent on the relative impact of threats to the habitat, the distance between the threat sources and the habitat, and the sensitivities of the specific habitats to any possible threats, leading to habitat degradation (Sharp et al., 2020) (Table 1).

TABLE 1 Description of data input for the habitat quality model in InVEST.

Input	Description
Land use land cover	GIS raster dataset, with a numeric LULC code for each cell. The LULC raster obtained from Simeneh (2022) in the area of interest was used. The LULC codes must match the codes for the sensitivity of land cover types to each threat
Threat data	A CSV table of all threats needed to be considered in the model. The table contains information on each threat's relative importance or weight and its impact across space. Each row is a degradation source. Each column contains a different attribute of each degradation source and must be named as THREAT, MAX-DIST, WEIGHT, and DECAY.
Threat raster	GIS raster files with the distribution and intensity of each threat showing each of them affecting the habitat. However, the techniques applied for each threat raster can vary according to the data types. The threat maps should cover the area of interest and buffer the width of the greatest maximum threat distance. Each cell in the raster contains a value that indicates the density or presence of a threat within it. All threats should be measured on the same scale and units
Habitat types and sensitivity of each habitat to threats	A CSV table of LULC types contains information on whether a habitat is identified (absence/presence of habitat) or not and their specific sensitivity to each threat. Sensitivity values range from 0 to 1, where 0 represents no sensitivity to a threat and 1 represents the greatest sensitivity (Polasky et al., 2011). Sensitivity scores can be determined using expert knowledge and the AHP method (Hamere et al., 2021)
Half saturation constant (k)	The scaling parameter (or constant) of 0.5 is the default for the InVEST model. The InVEST model uses a half-saturation curve to convert habitat degradation scores to habitat quality scores (Sharp et al., 2020). It is determined as an inverse relationship between the degradation and habitat quality scores. It helps with the visual representation of heterogeneity in quality across the landscape

TABLE 2 Ecological habitat quality input data used for InVEST habitat quality model in the AWR (1998, 2016).

Threats	Maximum distance (km)	Weight	Decay	LULC types				
				BL	HF	GL	RF	WL
				Habitat suitability score				
				1	1	1	1	1
				Habitat sensitivity to threats				
Invasive species	1	0.25	Exponential	1	0.1	1	1	1
Habitat destruction	2	0.25	Exponential	0.75	0.75	0.5	0.5	0.75
Livestock incursion	2	0.15	Exponential	1	0.5	1	1	1
Bush encroachment	3	0.05	Exponential	1	0.5	1	0.75	0.75
Fire	1	0.05	Linear	0.2	0.5	0.2	0.2	0.2
Distance to road	1	0.1	Linear	0.5	0.75	1	0.5	0.2

There are three key inputs to be considered for habitat quality mapping in InVEST model. First, the suitability of each LU/LC type (H_j) for providing habitat for biodiversity; second, anthropogenic threats that originate at pixel x (r_x) affecting habitat quality; and third, the sensitivity of each LU/LC type to each threat (Table 1). For this study, six biodiversity threats were identified in the study area by following the approach of Terrado et al. (2016) and Wu et al. (2014). These were invasive species, bush encroachments, livestock incursion, fire, habitat destruction, and distance to roads (Table 2). The significance (weight) of each threat was prioritized based on the ecological and threat monitoring activities with two senior ecologists and five park rangers of the AWR between 3rd—4th December 2021 and the AHP method was applied to prioritize conservation threats following the approach by Terrado et al. (2016) and Wu et al. (2014) (Table 2; Figure 2).

The total threat level in a grid cell x with LU/LCj is calculated as the relative habitat suitability score (H_j), from 0 to 1, where 1 indicates the highest suitability to species has been assigned to LU/LC types (Sharp et al., 2020). The last input of the model is the sensitivity of habitat type to different threats; helps to account for the differentiated impacts of threats to different habitats. The impacts of the threats on the habitat are determined by 1) the effect of the threat over space (i_{rxy}); 2) the relative weight of each threat's importance compared to the others (w_r), and 3) the relative sensitivity of each habitat to each threat (S_{jr}). The stress level D_{xj} of grid x with land-use type j is calculated as follows (Sharp et al., 2020).

$$i_{rxy} = 1 - \left(\frac{d_{xy}}{d_{rmax}}\right) \ldots\ldots\ldots\ldots\ldots\ldots \text{if linear} \ldots\ldots\ldots \quad (1)$$

$$i_{rxy} = \exp\left(-\left(\frac{2.99}{d_{rmx}}\right)d_{xy}\right) \ldots\ldots\ldots\ldots \text{if exponential} \ldots\ldots\ldots \quad (2)$$

Where, d_{xy} is the linear distance between grid cells x and y, and d_{rmax} is the maximum effective distance of threats r's across space.

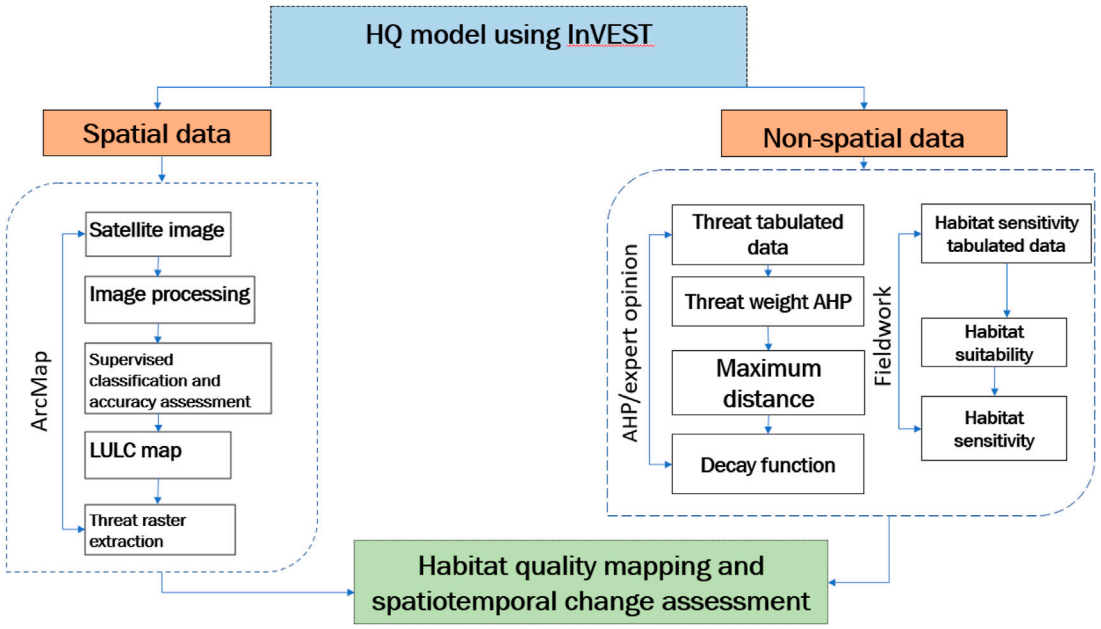


FIGURE 2
Flowchart showing methodological steps followed in the study.

$$D_{xj} = \sum_{r=1}^{R} \sum_{y=1}^{Yr} \left(\frac{w_r}{\sum_{r=1}^{R} w_r} \right) r_x i_{rxy} \theta_x S_{jr} \quad (3)$$

where R is the number of threat factors, y_r is the set of grid cells on r's map, w_r is the relative effect of each threat, θ_x is the level of accessibility to a grid cell x, and S_{jr} is the relative sensitivity of each habitat type to each threat.

The results of the model range from 0 to 1, with 1 representing the highest level of habitat quality (Sharp et al., 2020). The impacts of the threat on habitat decrease as the distance from the degradation sources increases, threats with higher destructive values (on the scale of 0-1) have higher impacts and the more sensitive a habitat type is to a threat (higher S_{jr}), the more degraded the habitat type could be by the threat.

Habitat quality is the environmental level that the ecological environment provides for the survival of individual organisms and populations. It is a continuous variable with a numerical range from low to high. The higher the quality of the habitat, the more stable the ecological structure and function of the patch. The way and intensity of human land use determines the quality of the habitat, and the more intense the land use, the more pronounced the decline in habitat quality (Almpanidou et al., 2014). Habitat quality was calculated based on the degree of habitat degradation, and the habitat quality score decreased with increasing habitat degradation score. The calculation formula for habitat quality is as follows:

$$Q_{xj} = H_j \left[1 - \left(\frac{D_{xj}^2}{D_{xj}^2 + k^z} \right) \right] \quad (4)$$

Where, Qxj is the habitat quality of grid cell x in land cover type j; Hj is the habitat suitability of land cover type j; D^2 xj is the level of habitat threat for grid cell x in land cover type j; k is the half-saturation factor, which is generally taken as half of the maximum value of D^2 xj; and x is a constant.

The information obtained from expert judgment and AHP was verified by undertaking field assessments.

2.3 Data preparation and input for the InVEST habitat quality model

The data inputs (spatial and non-spatial) are required to run the InVEST habitat quality model (Figure 2). Thus, LU/LC maps, threat sources, and impacts, habitat types, habitat sensitivity to each threat, and half-saturation constant were the required inputs (Sharp et al., 2020). The information on LU/LC was obtained from the previous study made by Simeneh (2023) in the study area. All the required inputs such as LU/LC maps of the respective years (1998–2016), threat sources and impacts, habitat types, and habitat sensitivity to each threat were loaded to run the habitat quality model. Finally, habitat quality maps for each respective year were produced; the final habitat quality maps were classified into three classes (low, moderate, and high).

3 Results and discussions

The result revealed an overall habitat quality reduction during the study period (Table 3; Figure 3). The model showed that the ecosystem was dominated by a high-quality habitat of 837 km^2 (57.4%) followed by a moderate-quality habitat of 544 km^2 (37.3%), and a low-quality habitat of 78 km^2 (5.35%) in 1998. In the subsequent 18 years (1998–2016), the low-quality and moderate-quality habitats increased by 128 km^2 (62%), and 75 km^2 (12%) respectively, while the high-quality habitats decreased by 203 km^2

TABLE 3 Habitat quality changes in the AWR using the InVEST habitat quality model (1998, 2016).

Habitat quality	Study period				Change (km^2)	% change	Trend
	1998		2016				
	Area (km^2)	%	Area (km^2)	%			
Low	78	5.35	206.00	14.12	128.00	62.14	Increasing
Moderate	544	37.29	619.00	42.43	75.00	12.12	Increasing
High	837	57.37	634.00	43.45	−203.00	−32.02	Decreasing
Total	1,459	100.00	1,459.00	100.00			

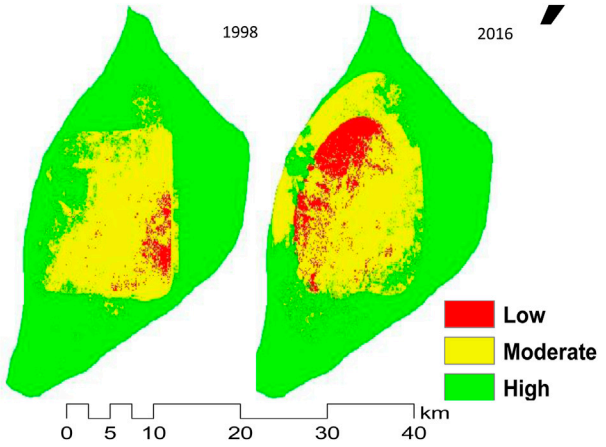


FIGURE 3
Spatial distribution of habitat quality in the AWR (1998, 2016).

(32%). The extent of high-quality habitat largely declined during the study period from 837 km^2 (57.4%) to 534 km^2 (43.45%). Conversely, moderate-quality, and low-quality habitats have increased from 78 km^2 (5.35%) in 1998 to 206 km^2 (14.12%) in 2016; and from 544 km^2 (37.3%) in 1998 to 619 km^2 (42.13%) in 2016 respectively.

This study was the first to assess habitat quality using InVEST model and expert-driven approach in Ethiopia's highly valued protected landscape. Thus, the study provides robust information that can be used for threat reduction planning and management intention in the study landscape. The habitat quality changes in the study area were highly associated with increased livestock incursion and expansion of invasive species resulting in severe changes in the healthy functioning of the ecosystems. The quality of habitat influences wildlife species diversity, density, distribution, and movement patterns in landscapes (Zhang et al., 2019; Dai et al., 2018). The decline in habitat quality is mainly attributed to increased conservation threats including the incursion of livestock and human interactions into wildlife habitats (Carter et al., 2014). Free grazing activities have an adverse negative impact on habitat quality (Su et al., 2020). Likewise, the quality of the habitat has been significantly declining particularly the grassland habitat was deteriorated by massive livestock incursion in the study landscape, which is a common prolonged problem in protected areas of Ethiopia (Mekbeb et al., 2022). Similar results were reported by Kija et al. (2020) that habitat quality has largely deteriorated by anthropogenic activities and land use policy changes in the Greater Serengeti Ecosystem of Tanzania.

Overall, high–quality 1998 became a moderate-quality and low-quality habitat during 1998–2016. The loss of habitat quality is well pronounced in the grassland habitat of the protected area which is the preferred feeding and breeding habitat for charismatic ungulate species and other wild animals of the AWR. The swift spread of invasive species coupled with livestock grazing and habitat destruction significantly affects the grassland habitat of the protected area. The savannah grassland habitat is the most preferred and suitable habitat for the charismatic plain animal of the reserve but under severe pressure, particularly invasive *Prosophis juliflora* in the grassland habitat is the principal conservation challenge for the protected area management. The communities are reliant upon livestock rearing and natural resources due to a lack of alternatives leading to overgrazing; unmanaged grazing practices are resulting in significant degradation of principal ecological habitats such as the grassland habitat in many protected area systems in Ethiopia. Due to the high livestock density in the area, the grassland habitat of the landscape has encountered severe grazing practices year after year. This has led to the deterioration of grassland habitat quality and a reduction in the capacity to provide forage for grassland-reliant wild animal species.

Land use and land cover changes can be taken as the prime factors for changes in habitat quality in the study area during the study period. The low-quality habitat has slightly shifted from the center of the highland forest to the center of the landscape which is occupied by the grassland habitat of the AWR (Figure 3); this is mainly due to the spread of invasive species, livestock incursion, and the closeness to the tarmac road. The highland forest has been unwisely utilized for various purposes mainly for charcoal production, however, improved management intervention in the highland forest contributed to the management of illegal activities thus the habitat has rapidly been restored (Simeneh, 2023). Conversely, the grassland habitat was largely converted into low habitat quality as unrestricted grazing led to reducing the quality of grassland habitat. The woodland, riverine forest, and partially bushland habitats have been unchanged in terms of quality and maintained high habitat quality during the study period. According to Fanuel (2013), the study landscape has lost about 52% of its quality to conserve the larger charismatic herbivores of the landscape. Similarly, this finding showed that only 634 km^2

(43.45%) of the landscape sustains its high quality to possess charismatic species of the landscape. This indicates that the protected landscape is losing its quality habitats to possess the endangered iconic species. Maintaining high-quality habitats could enhance the stability of ecosystem structure and function and the quick recovery potential of habitats after disturbance (Schwarz et al., 2017; Wu et al., 2017).

Unlike most areas of Ethiopia in which invasive species were spread due to the road access; in the Afar area, where the landscape is located; Prosopis juliflora was introduced mainly for water and soil conservation and to support livestock forage in the dry season in the late 1970 s and early 1980 s (Ayanu et al., 2014; Kebede and Coppock, 2015; Hailu et al., 2019). Further, additional plantations were made between the 1980 s and 1990 s as shade and wind protection trees in villages, and the raw material was used for firewood fencing, and building materials (Ayanu et al., 2014). Livestock has been identified as the principal vector for the rapid spread of invasive species and the invasion become a serious problem that started rapidly invading the rangeland (Hailu et al., 2019). The invasion could significantly affect the ecosystem services and livelihood of pastoralist communities by reducing biodiversity, grazing land, and water supply (Shackleton et al., 2014).

4 Limitations of the study

The assessment of habitat quality using the InVEST model has been successfully employed for the maintenance of biodiversity and is invaluable for the management of the landscape and land-use planning (Sharp et al., 2020) but inadequate information about the spatial and temporal distribution of species across the protected landscape (Stephen et al., 2011) is the major limitations of the InVEST habitat quality model. It is, therefore, important to conduct a field-based habitat suitability assessment to obtain ecologically valid and robust information on the distribution of quality habitat and species abundance across the landscape (Nagendra et al., 2013).

5 Conclusion

This study has assessed the quality of habitat using expert-driven landscape habitat threat information in InVEST software in the most iconic but greatly threatened protected landscape of the AWR in Ethiopia. Assessing a landscape's habitat quality has greater implications for the larger rangeland ecosystem management since it directly impacts landscape structure and spatial pattern. The most important factors for the decline of habitat quality include livestock incursion and the expansion of invasive species. Therefore, it can be concluded that the continued fast spread of invasive species and bush encroachment in the critical feeding and breeding habitat can largely influence the biodiversity and ecosystem services of AWR. Thus, it is critical to undertake serious conservation measures to maintain the ecosystem's integrity and halt biodiversity loss. Further, the boundary of the AWR has not been clearly defined and has not been legally gazetted yet. Therefore, the findings of this study can be used to redefine the landscape boundary to encompass the most critical habitat under legal protection.

Data availability statement

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

Author contributions

SA designed and conceptualized the research collected relevant data, analyzed data, and wrote the draft manuscript.

Acknowledgments

The authors are thankful for the kind cooperation of ecological monitoring experts of the AWR and EWCA and rangers of AWR.

Conflict of interest

Author SA was employed by GFA Consulting Group.

Publisher's note

References

Abreham, B., Noszczyk, T., Soromessa, T., and Elias, E. (2020). The InVEST habitat quality model associated with land use/cover change in the winike watershed in the omo-gibe basin, southwest Ethiopia. *Remote Sens.* 12, 1103.

Addisu, A., Ali, S., and Ahmed, E. (2017). Floristic composition and community analysis of woody plants in hallideghie wildlife reserve, north-east Ethiopia. *Ethiop. J. Environ. Stud. Manag.* 10 (7), 917–930.

Aguilar, R., Quesada, M., Ashworth, L., Herrerias, D. Y., and Lobo, J. (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Mol. Ecology17* 17, 5177–5188. doi:10.1111/j.1365-294x.2008.03971.x

Almaz, T. (2009). *Sustaining the Alledeghi grassland of Ethiopia: influences of pastoralism and vegetation change. (PhD dissertation)*. Logan: Utah State University.

Almpanidou, V., Mazaris, A. D., Mertzanis, Y., Avraam, I., Antoniou, I., Pantis, J. D., et al. (2014). Providing insights on habitat connectivity for male brown bears: a combination of habitat suitability and landscape graph-based models. *Ecol. Model* 286, 37–44. doi:10.1016/j.ecolmodel.2014.04.024

Andrus, R. A., Martinez, A. J., Jones, G. M., and Meddens, A. J. H. (2021). Assessing the quality of fire refugia for wildlife habitat. *Ecol. Manage* 482, 118868. doi:10.1016/j.foreco.2020.118868

Ayanu, Y., Jentsch, A., Müller-Mahn, D., Rettberg, S., Romankiewicz, C., and Koellner, T. (2014). Ecosystem engineer unleashed: Prosopis juliflora threatening ecosystem services? *Reg. Environ. Chang.* 15, 155–167. doi:10.1007/s10113-014-0616-x

Bai, L., Xiu, C., Feng, X., and Liu, D. (2019). Influence of urbanization on regional habitat quality: a case study of Changchun City. *Habitat Int.* 93, 102042. doi:10.1016/j.habitatint.2019.102042

Baral, H., Keenan, R. J., Sharma, S. K., Stork, N. E., and Kasel, S. (2014). Spatial assessment and mapping of biodiversity and conservation priorities in a heavily modified and fragmented production landscape in north-central Victoria, Australia. *Ecol. Indic.* 36, 552–562. doi:10.1016/j.ecolind.2013.09.022

Basane, C., and James, G. (2016). Habitat assessment for ecosystem services in South Africa. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* 12 (4), 242–254. doi:10.1080/21513732.2016.1217935

Bohm, T., and Hofer, H. (2018). Population numbers, density and activity patterns of servals in savannah patches of Odzala-Kokoua National Park, Republic of Congo. *Afr. J. Ecol.* 56, 841–849. doi:10.1111/aje.12520

Bruijnzeel, L. A. (2004). Hydrological functions of tropical forests: not seeing the soil for the trees? *Agric. Ecosyst. Environ.* 104, 185–228. doi:10.1016/j.agee.2004.01.015

Cardinale, B. J., Emmett Duffy, J., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). Michel loreau, James B. Grace, anne larigauderie, diane S. Srivastava and shahid naeem (2012). Biodiversity loss and its impact on humanity. *Nature* 486, 11148.

Carter, N. H., Viña, A., Hull, V., McConnell, W. J., Axinn, W., Ghimire, D., et al. (2014). Coupled human and natural systems approach to wildlife research and conservation. *Ecol. Soc.* 19, 43. doi:10.5751/es-06881-190343

Chen, Y., Qiao, F., and Jiang, L. (2016). Effects of land use pattern change on regional scalehabitat quality based on InVEST modelda case study in Beijing. *Acta Sci. Nat. Univ. Pekin.* 52, 553–562.

Choudhary, A., Deval, K., and Joshi, P. K. (2021). Study of habitat quality assessment using geospatial techniques in Keoladeo National Park, India. *Environ. Sci. Pollut. Res.* 28, 14105–14114. doi:10.1007/s11356-020-11666-3

Chu, L., Sun, T., Wang, T., Li, Z., and Cai, C. (2018). Evolution and prediction of landscape pattern and habitat quality based on CA-markov and InVEST model in hubei section of three gorges reservoir area (TGRA). *Sustainability* 11, 3854. doi:10.3390/su10113854

Claudia, C., João Carlos, M., Cunha, P. P., and Teixeira, Z. (2020). Ecosystem services as a resilience descriptor in habitat risk assessment using the InVEST model. *Ecol. Indic.* 115, 106426. doi:10.1016/j.ecolind.2020.106426

Dai, L., Li, S., Lewis, B. J., Wu, J., Yu, D., Zhou, W., et al. (2018). The influence of land use change on the spatial–temporal variability of habitat quality between 1990 and 2010 in Northeast China. *J. For. Res.* 30, 2227–2236. doi:10.1007/s11676-018-0771-x

Edmonds, N. J., Al-Zaidan, A. S., Al-Sabah, A. A., Le Quesne, W. J. F., Devlin, M. J., Davison, P. I., et al. (2021). Kuwait's marine biodiversity: qualitative assessment of indicator habitats and species. *Mar. Pollut. Bull.* 163, 111915. doi:10.1016/j.marpolbul.2020.111915

Fargione, J., Hill, J., Tilman, D., Polasky, S., and Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. *Land Clear. Biofuel Carbon Debt* 319 (5867), 1235–1238. doi:10.1126/science.1152747

Hailu, S., Bewket, W., Alamirew, T., Zeleke, G., Demel, T., Ketema, B., et al. (2019). Implications of land use/land cover dynamics and Prosopis invasion on ecosystem service values in Afar Region, Ethiopia. *Sci. Total Environ.* 675, 354–366. doi:10.1016/j.scitotenv.2019.04.220

Havlicek, E., and Mitchell, E. A. D. (2014). "Soils supporting biodiversity," in *Interactions in soil: promoting plant growth*.

Hilliman, J. C. (1993). *Compendium of wildlife conservation information: information on wildlife conservation areas. NYZS and EWCO*. Addis Ababa.

Johnson, M. D. (2007). Measuring habitat quality: a review. *Condor* 109, 489–504. doi:10.1093/condor/109.3.489

Kebede, F. (2013). *Ecology and Community Based Conservation of the Grevy's zebra (Equus grevyi) and the African wild ass (Equus africanus) in the Afar region, Ethiopia PhD dissertation*. Addis Ababa: Addis Ababa University.

Kebede, F., Afework, B., Moehlman, P. D., and Evangelista, P. H. (2012). Endangered Grevy's zebra in the Alledeghi Wildlife Reserve, Ethiopia: species distribution modeling for the determination of optimum habitat. *Endang Species Res.* 17, 237–244. doi:10.3354/esr00416

Kebede, T., and Coppock, L. (2015). Livestock-mediated dispersal of Prosopis juliflora imperils grasslands and the endangered Grevy's zebra in Northeastern Ethiopia. *Rangel. Ecol. Manag.* 68, 402–407. doi:10.1016/j.rama.2015.07.002

Kija, H. K., Ogutu, J. O., Mangewa, L. J., Bukombe, J., Verones, F., Graae, B. J., et al. (2020). Spatio-temporal changes in wildlife habitat quality in the greater Serengeti ecosystem. *Sustainability* 12, 2440. doi:10.3390/su12062440

Laurance, W. F., Lovejoy, T. E., Vasconcelos, H. L., Bruna, E. M., Didham, R. K., Stouffer, P. C., et al. (2002). Ecosystem decay of Amazonian Forest fragments: a 22-year investigation. *Conserv. Biology16* 16, 605–618. doi:10.1046/j.1523-1739.2002.01025.x

Lin, Q., Mao, J., Wu, J., Li, W., and Yang, J. (2016). Ecological security pattern analysis based on InVEST and least-cost path model: a case study of Dongguan water village. *Sustainability* 8, 172. doi:10.3390/su8020172

Liu, S., Liao, Q., Xiao, M., Zhao, D., and Huang, C. (2022). Spatial and temporal variations of habitat quality and its response of landscape dynamic in the three gorges reservoir area, China. *Int. J. Environ. Res. Public Health* 19, 3594. doi:10.3390/ijerph19063594

Machado, A. (2004). An index of naturalness. *J. Nat. Conservation* 12, 95–110. doi:10.1016/j.jnc.2003.12.002

Mekbeb, T., Asefa, A., and Delelegn, Y. (2022). *National biodiversity threat assessment: ranking major threats impacting Ethiopia's biodiversity*. Addis Ababa, Ethiopia: IUCN.

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., et al. (2013). Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecol. Indic.* 33, 45–59. doi:10.1016/j.ecolind.2012.09.014

Norliyana, A., and Mamat, M. (2020). Spatial analysis model assessing habitat quality of selangor. *IOP Conf. Ser. Earth Environ. Sci.* 549, 012049. doi:10.1088/1755-1315/549/1/012049

Oliver, T. H., Heard, M. S., Isaac, N. J. B., Roy, D. B., Procter, D., Eigenbrod, F., et al. (2015). Biodiversity and resilience of ecosystem functions. *Trends Ecol. Evol.* 11, 673–684. doi:10.1016/j.tree.2015.08.009

Polasky, S., Nelson, E., Pennington, D., and Johnson, K. A. (2011). The impact of land-use changes on ecosystem services, biodiversity and returns to Landowners: a case study in the state of Minnesota. *Environ. Resour. Econ.* 48, 219–242.

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., and Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* 25, 345–353. doi:10.1016/j.tree.2010.01.007

Romero-Calcerrada, R., and Luque, S. (2006). Habitat quality assessment using Weights-of-Evidence based GIS modelling: the case of Picoides tridactylus as species indicator of the biodiversity value of the Finnish forest. *Ecol. Model.* 196, 62–76. doi:10.1016/j.ecolmodel.2006.02.017

Rouget, M., Richardson, D. M., Cowling, R. M., Lloyd, J. W., and Lombard, A. T. (2003). Current patterns of habitat transformation and future threats to biodiversity in terrestrial ecosystems of the Cape floristic region, South Africa. *Biol. Conserv.* 112, 63–85. doi:10.1016/s0006-3207(02)00395-6

Schwarz, N., Moretti, M., Bugalho, M. N., Davies, Z. G., Haase, D., Hack, J., et al. (2017). Understanding biodiversity-ecosystem service relationships in urban areas: a comprehensive literature review. *Ecosyst. Serv.* 27, 161–171. doi:10.1016/j.ecoser.2017.08.014

Selamnesh, T. (2015). Impact of Prosopis juliflora on plant biodiversity at Alledeghi wildlife reserve and surrounding local community, Ethiopia. MSc thesis. Addis Ababa: Ababa University.

Shackleton, R. T., Le Maitre, C., Pasiecznik, N. M., and Richardson, D. M. (2014). Prosopis: a global assessment of the biogeography, benefits, impacts, and management of one of the world's worst woody invasive plant taxa. *AOB Plants* 2014 (6), plu027. doi:10.1093/aobpla/plu027

Sharp, R., Douglass, J., Wolny, S., Arkema, K., Bernhardt, J., Bierbower, W., et al. (2020). *InVEST user's guide. The natural capital project*. Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

Simeneh, A. (2023). *Assessing the impact of land use land cover changes on ecosystem services in the Alledighe Wildlife Reserve, Ethiopia: implication for rangeland ecosystem management*. Manuscript under review.

Simeneh, A., Afework, B., and Assefa, A. (2016). Population size and structure of beisa oryx and its habitat overlap with sympatric species in Alledeghi wildlife reserve, NE, Ethiopia. *Ethiop. J.Biol. Sci.* 15 (1), 37–54.

Simeneh, A., Yeshitela, K., and Argaw, M. (2023). Assessing habitat quality using the InVEST model in the Dire and Legedadi watersheds, central highland of Ethiopia. *Sustain. Environ.*, 2765–8511.

Sinclair, A. R., Mduma, S. A., Hopcraft, J. G., Fryxell, J. M., Hilborn, R., and Thirgood, S. (2007). Long-term ecosystem dynamics in the Serengeti: lessons for conservation. *Conserv. Biol.* 21, 580–590. doi:10.1111/j.1523-1739.2007.00699.x

Stephen, P., Erik Jeremy, N., Pennington, D. N., and Johnson, K. A. (2011). The impact of land-use changes on ecosystem services, biodiversity and returns to landowners: a case study in the state of Minnesota. *Environ. Resour. Econ.* 48 (2), 219–242.

Stolton, S., Mansourian, S., and Dundley, N. (2010). *Valuing protected areas. Washington (DC): world Bank`s global environment facility coordination team*. World Bank GEF Operations.

Su, X., Liu, Y., Zhou, W., and Liu, G. (2020). "An overview on status and importance of Tibetan Plateau steppe, China," in *Reference module in earth systems and environmental sciences* (Amsterdam: Elsevier).

Tang, F., Fu, M., Wang, Li, and Zhang, P. (2020). Land-use change in Changli County, China: predicting its spatio-temporal evolution in habitat quality. *Ecol. Indic.* 117, 106719. doi:10.1016/j.ecolind.2020.106719

Terrado, M., Sabater, S., Chaplin-Kramer, B., Mandle, L., Ziv, G., and Acuna, V. (2016). Model development for the assessment of terrestrial and aquatic habitat quality in conservation planning. *Sci. Total Environ.* 540, 63–70. doi:10.1016/j.scitotenv.2015.03.064

Thomas, E., Jansen, M., Chiriboga-Arroyo, F., Wadt, L. H. O., Corvera-Gomringer, R., Atkinson, R. J., et al. (2021). Habitat quality differentiation and consequences for ecosystem service provision of an amazonian hyperdominant tree species. *Front. Plant Sci.* 12, 621064. doi:10.3389/fpls.2021.621064

Villamagna, A. M., Angermeier, P. L., and Bennett, E. M. (2013). Capacity, pressure, demand, and flow: a conceptual framework for analyzing ecosystem service provision and delivery. *Ecol. Complex.* 15, 114–121. doi:10.1016/j.ecocom.2013.07.004

Wu, C.-F., Lin, Y. P., Chiang, L. C., and Huang, T. (2014). Assessing highway's impacts on landscape patterns and ecosystem services: a case study in Puli Township, Taiwan. *Landsc. Urban Plan.* 128, 60–71. doi:10.1016/j.landurbplan.2014.04.020

Wu, J., Yue, X., and Qin, W. (2017). The establishment of ecological security patterns based on the redistribution of ecosystem service value: a case study in the Liangjiang New Area, Chongqing. *Geogr. Res.* 36, 429–440.

Yang, Y. (2021). Evolution of habitat quality and association with land-use changes in mountainous areas: a case study of the Taihang Mountains in Hebei Province, China. *Ecol. Indic.* 129, 107967. doi:10.1016/j.ecolind.2021.107967

Yohannes, H., Soromessa, T., Argaw, M., and Dewan, A. (2021). Spatio-temporal changes in habitat quality and linkage with landscape characteristics in the Beressa watershed, Blue Nile basin of Ethiopian highlands. *J. Environ. Manag.* 281, 111885. doi:10.1016/j.jenvman.2020.111885

Zhang, H. B., Wu, F. E., Zhang, Y. N., Han, S., and Liu, Y. Q. (2019). Spatial and temporal changes of habitat quality in jiangsu yancheng wetland national nature reserve-rare birds of China. *Appl. Ecol. Environ. Res.* 17, 4807–4821. doi:10.15666/aeer/1702_48074821

Zhu, C., Zhang, X., Zhou, M., He, S., Gan, M., Yang, L., et al. (2020). Impacts of urbanization and landscape pattern on habitat quality using OLS and GWR models in Hangzhou, China. *Ecol. Indic.* 117, 106654. doi:10.1016/j.ecolind.2020.106654

Zhu, L., Huang, C., Liu, Q. S., and Liu, G. H. (2015). Changes of coastal zone landscape spatial patterns and ecological quality in Liaoning province from 2000 to 2010. *Resour. Sci.* 37, 1962–1972.

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 14 November 2023
DOI 10.3389/fevo.2023.1293311

OPEN ACCESS

EDITED BY
Prof-Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Marcello Vitale,
Sapienza University of Rome, Italy
Anneli Poska,
Lund University, Sweden

*CORRESPONDENCE
Jose Bermudez Castro
bermudej@mcmaster.ca
Michelle Kalamandeen
michellekalamandeen@outlook.com

RECEIVED 12 September 2023
ACCEPTED 24 October 2023
PUBLISHED 14 November 2023

CITATION
Kalamandeen M, Gulamhussein I, Castro JB, Sothe C, Rogers CA, Snider J and Gonsamo A (2023) Climate change and human footprint increase insect defoliation across central boreal forests of Canada.
Front. Ecol. Evol. 11:1293311.
doi: 10.3389/fevo.2023.1293311

Climate change and human footprint increase insect defoliation across central boreal forests of Canada

Michelle Kalamandeen[1*], Imran Gulamhussein[1], Jose Bermudez Castro[1*], Camile Sothe[2], Cheryl A. Rogers[1,3], James Snider[4] and Alemu Gonsamo[1]

[1]School of Earth, Environment and Society, McMaster University, Hamilton, ON, Canada, [2]Finite Carbon Canada, Calgary, AB, Canada, [3]Department of Geography & Environmental Studies, Toronto Metropolitan University, Toronto, ON, Canada, [4]World Wildlife Fund Canada, Toronto, ON, Canada

Anthropogenic climate change is contributing to increased insect infestation globally, leading to pest population growth, expansion of niche and geographic ranges as well as increased outbreak frequencies, resulting in economic losses and reduction in food security. In recent years, spongy moth (*Lymantria dispar dispar*), jack pine budworm (*Choristoneura pinus pinus*), large aspen tortrix (*Choristoneura conflictana*) and spruce budworm (*Choristoneura fumiferana*) caused widespread defoliation across one of Canada's most forested provinces, Ontario. Observations of such outbreaks have been limited to field sightings around Ontario, with few studies focused on mapping of outbreak occurrence across the province or exploring potential anthropogenic and climatic drivers of infestation. Using random forest probability estimates and satellite data resampled to 1 km spatial resolution from the Moderate Resolution Imaging Spectroradiometer (MODIS), we reveal greater expansion of insect defoliation across Ontario between 2018 and 2020 than previously recorded. Much of the geographic expansion of outbreaks was driven by increasing temperature, and the proximity of roads. With ongoing global warming and growing economic development, infestations will not only continue to increase across Ontario but also expand northward due to their responses to accelerated warming at higher latitudes. This expansion presents an important and alarming new challenge for forest conservation and management in Ontario, in particular, and Canada in general.

KEYWORDS

spongy moth, jack pine budworm, spruce budworm, remote sensing, Ontario Canada, insect defoliation, climate change, machine learning

1 Introduction

Understanding and predicting the consequences of climatic changes in forest ecosystems is emerging as one of the major challenges for global change scientists (Boisvenue and Running, 2006; Bonan, 2008). Globally, societies heavily rely on forests for essential services such as timber, watershed protection, and recreational and aesthetic benefits (Maroschek et al., 2009; Thom and Seidl, 2016). The effects of climate change on forests include both positive (e.g., increased forest growth from CO_2 fertilization, increased water use efficiency, and longer growing seasons) and negative responses (e.g., increases in stress and tree mortality, and changes in dynamics of forest insect) (Ayres and Lombardero, 2000; Bachelet et al., 2003; Lucht et al., 2006; Scholze et al., 2006; Lloyd and Bunn, 2007; Seidl et al., 2017; Gonsamo et al., 2017; 2021).

With current best estimates of changes in climate indicating an increase in global mean annual temperatures of 1.5°C by 2025 and 4°C by the end of the next century (Baker et al., 2018), considerable uncertainty remains in modeling how these processes will affect current and future tree growth and mortality events and forest carbon budgets (Hanson and Weltzin, 2000; Bugmann, 2001; Hollaus and Vreugdenhil, 2019). Outbreaks of forest insects are major agents of mortality and ecosystem change in forests worldwide, with climate being an important driver of changes to disturbance regimes mediated by forest insects (Pureswaran et al., 2018). Changes in climate may result in changing geographical distribution, increased overwintering, changes in population growth rates, increases in the number of generations, extension of the development season, changes in crop–pest synchrony, changes in interspecific interactions and increased risk of invasion by migrant pests (see Porter and Coon, 1991; Pureswaran et al., 2018; Lehmann et al., 2020). Although a range of responses can and should be expected, recent cases of increased tree mortality and die-offs triggered by insect infestation raise the possibility that amplified forest mortality may already be occurring in some locations in response to global climate change (Allen et al., 2010).

As a landscape-scale disturbance event, insect outbreaks also play an important role in the carbon flux in boreal forests (Kurz and Apps, 1999; Volney and Fleming, 2000). Defoliation during insect outbreaks reduces the rate of carbon accumulation by the host trees via reducing their growth. This often results in tree mortality, which abruptly increases the mass of dead organic matter where carbon is transferred to the atmosphere through decomposition (Gray et al., 2007). Increasing concentrations of carbon dioxide in the atmosphere are a major cause of global warming, creating a feedback loop for further proliferation of insect infestations.

Across Ontario, Canada, several defoliators have expanded their range in recent years causing widespread outbreaks (Liebhold et al., 1992; Régnière et al., 2009; Tobin et al., 2004; NDMNRF, 2020). Species which have had recent moderate to severe forest disturbances have included the spongy moth (*Lymantria dispar dispar*), jack pine budworm (*Choristoneura pinus pinus*), large aspen tortrix (*Choristoneura conflictana*) and spruce budworm (*Choristoneura fumiferana*) (NDMNRF, 2022). Most of these species are native to Canada (jack pine budworm, spruce budworm and large aspen tortrix) while the spongy moth is an example of an invasive defoliator that expanded its current range across Eastern United States and Southeastern Canada (Liebhold et al., 1992; Régnière et al., 2009; Tobin et al., 2004). *L. dispar* moth is native to the temperate forests of Europe and Asia (Régnière et al., 2009).

Outbreaks often follow different trajectories associated with specific ecological traits and target tree species. For instance, large aspen tortrix infestations are short-lived (2–3 years) and found earlier in the season. This species targets trembling aspen (*Populus tremuloides*), white birch (*Betula papyrifera*), willow (*Salix* spp.) and alder (*Alnus* spp.) but can also be found on balsam poplar (*Populus balsamifera*) and chokecherry (*Prunus virginiana*). Jack pine budworm periodically reaches outbreak levels every 8–10 years and prefers jack pine (*Pinus banksiana*) and other conifers such as eastern white pine (Pinus strobus), red pine (*Pinus resinosa*) and Scots pine (*Pinus sylvestris*). Spruce budworm is the most destructive pest of balsam fir (*Abies balsamea*) and white spruce (*Picea glauca*) forests in Canada, which is a historically important timber species. The larvae of spruce budworm are considered wasteful feeders, as they only eat partial needles and then move on to other needles. *L. dispar* moth infestations maintain a cyclical pattern in which exponential population growth causes major infestations for 3–4 years, with the insect then lying dormant for 7–8 years (Benoit and Lachance, 1990). The larvae of this moth are voracious folivores that can feed nearly 300 species of broadleaf and coniferous trees during their peak feeding season (Elkinton and Liebhold, 1990), significantly affecting tree growth. This combination of outbreak dynamics, broad polyphagy and long larval duration grants these species the ability to rapidly expand across forested areas such as those found in Ontario, one of Canada's most forested provinces.

Areas at risk of infestations in Canada are expected to double or triple over the next 50 years due to increasingly warmer summers, allowing the insect to complete its life cycle consistently in geographic areas that so far have been protected due to unfavorable climate. Across Canada, certain forest types seem more susceptible to infestation, such as deciduous, closed (dense) and open (thin) mixed forests (Cihlar et al., 2002). Forests across northern and eastern Canada has so far been protected from infestation by cold temperatures (Régnière et al 2009; 2012). but with increasing temperatures, the fate of these forests are unknown. Further, increasing human traffic may lead to deposition of eggs as some host plants flourish in disturbed areas (Lyons and Liebhold, 1992), triggering different dispersion pathways. However, little formal studies exist that explore probable drivers connected with insect infestation in Canada not associated with climatic studies.

Insect infestation has also been associated with other compounding ecological damages. For instance, the frass excreted by the *L. dispar* caterpillars contains a high concentration of nitrogen (N) because these defoliators are ineffective at assimilating foliar N (Lovett et al., 2002). As a result, this N-rich frass leaches into catchment basins, increasing the N concentration in lakes by an average of 0.03 mg L_1 (Woodman et al., 2021), which in turn stimulates conditions for increased microbial activity. Growing microbial populations discharges CO_2 as an element of egestion, resulting in a recurring pattern of increasing microbial activity and CO_2 accumulation near the surface of lakes

(Woodman et al., 2021). The presence of CO_2 in large quantities near lake surfaces suppresses the growth of CO_2 assimilating algae (Raven et al., 2012), further exacerbating CO_2 concentrations.

Given the magnitude of damage caused by insects and their potential to spread over the next few years, early detection and monitoring of infestations throughout Ontario is crucial. As zones immediately adjacent to infested areas become more suitable because of rising temperatures, invasion will be fast if suitable hosts are present (Régnière et al., 2009). However, detecting and monitoring infestation levels in a vast landscape as Ontario requires extensive fieldwork. Thus far, only a limited number of field observations have been conducted in the province. Such studies also miss the opportunity to assess potential drivers of infestation across the landscape. The increased frequency of infestations and their ecological and economic impacts requires the use of advanced technologies. Remote sensing is a valuable tool that provides frequent and spatially continuous data on vegetation conditions and has been previously explored for detecting forest insect infestation (Niemann and Visintini, 2005; Hollaus and Vreugdenhil, 2019; Ye et al., 2021; Romeiro et al., 2022).

Freely accessible moderate-resolution satellite datasets such as the 250 m imagery from spatial resolution Moderate Resolution Spectroradiometer (MODIS) have great potential to characterize subtle changes in forest canopies by capturing low-magnitude spectral changes in seasonal observations. Here, we make use of the MODIS satellite observations and field data to comprehensively evaluate insect infestation across Ontario, Canada, over a 3-year period (2018–2020). We specifically consider the 1) magnitude of occurrence using a random forest classification, 2) the probability or likelihood of infestation occurrence, and 3) the human and environmental drivers of infestation events. These analyses provide foundational and critical insights into shifts in potential drivers of infestation and establish a barometer for evaluating the effectiveness of government intervention and management strategies.

2 Materials and methods

Our approach involved three steps: (i) identifying infestation occurrence using ground truth data and MODIS satellite observations; (ii) estimating likelihood of infestation using random forest probability; and (iii) determining probable drivers of infestation. Figure 1 provides an overview of methods used to process and analyze satellite observations and ground data.

2.1 Study region

The study area is the province of Ontario, Canada's second-largest province, covering nearly 1 million km^2 (Figure 2) with forest ecosystems ranging from temperate to subarctic forests. In summer, temperature ranges between 30–35°C, while winter can go below −40°C. Approximately 66% of Ontario is classified as forest lands (70 million ha), including the deciduous forest of southern Ontario, the Great Lakes – St. Lawrence Forest of central Ontario, the Boreal Forest and the Hudson Bay Lowlands Forest in the north. A very small region of southern Ontario also includes Carolinian forest.

2.2 Infestation mapping

We acquired ground data of insect infestation from the Ontario GeoHub database (https://geohub.lio.gov.on.ca/documents/lio::

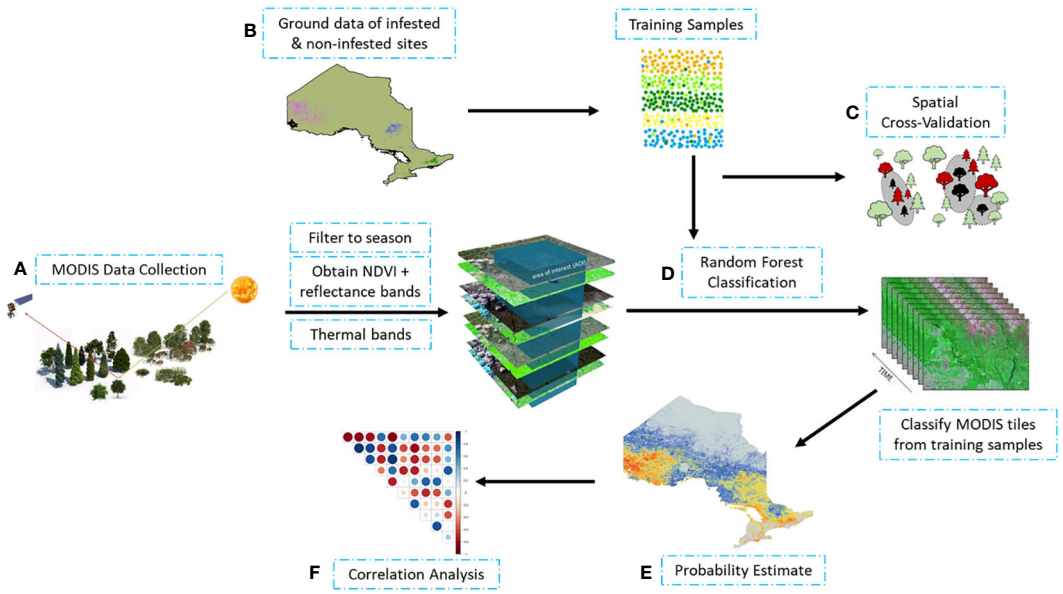


FIGURE 1
Schematic workflow of MODIS observation processing and data analysis: **(A)** processing of MODIS satellite observations; **(B)** obtaining of infested and non-infested training samples from ground measurements; **(C)** training random forest using spatial cross validation; **(D)** random forest classification using training sample to **(E)** estimate likelihood of occurrence (probability estimate); and **(F)** correlation analysis of the data.

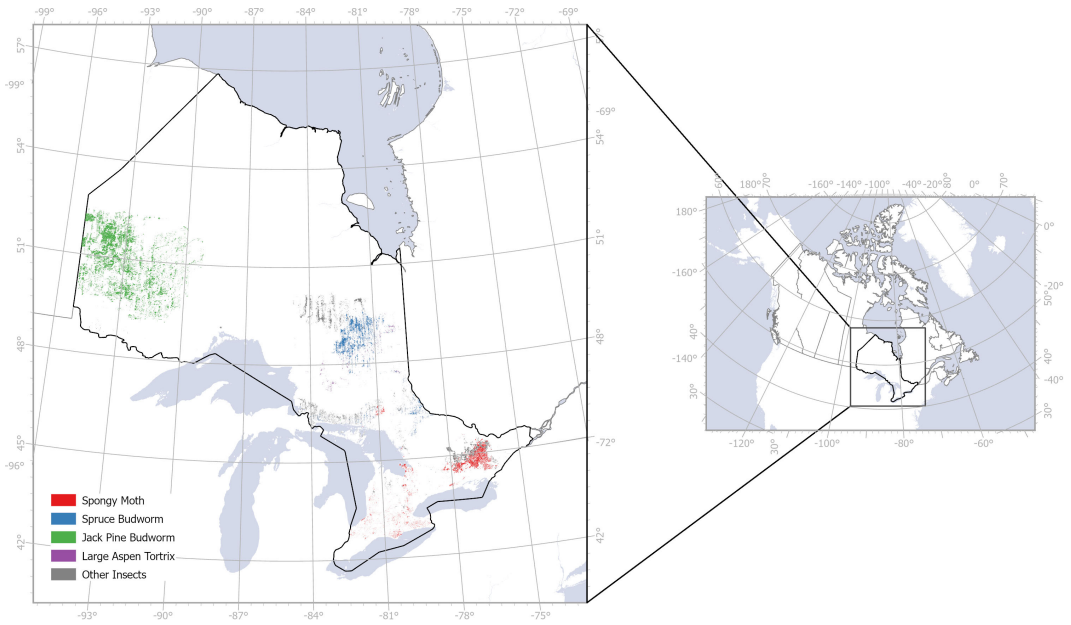


FIGURE 2
Insect infestation in Ontario, Canada from 2018 to 2020 based on field observations from Ontario GeoHub.

forest-insect-damage-event/about), filtering the infestation by year (2018–2020) and damage severity (moderate to severe, 25–100% damage; severe, 75–100% damage; and mortality ranking). We assumed that the Ontario GeoHub database is indicative of all known infested and non-infested sites, even though this may not be the case in reality. Based on paucity of data points across all study years, several species were excluded from this analysis. Selected species for this study included *L. dispar*, C. pinus, C. conflictana and C. fumiferana. Due to the scarcity of samples of some species, we treated all of them as the "defoliators" class, turning the problem into a binary classification task.

To train our model, we used covariates from the 250 m spatial resolution Moderate Resolution Spectroradiometer (MODIS) satellite observations. Surface reflectance of red and near infrared were extracted from version 6 MOD09Q1 product while the normalized difference vegetation index (NDVI) was extracted from version 6 MOD13Q1 product and thermal bands were taken from version 6.1 MOD11A1 product (Figure 3). We created bi-monthly image composites (mean and standard deviation) from the reflectance, NDVI, and thermal MODIS products during the growing season, from April to August, resampling all products to 1 km spatial resolution to match the corresponding thermal product. We noted that infestation cycles can commence anywhere between April to August annually, contingent on species and temperature fluctuations. MODIS was selected as our primary satellite data source as it offered cloud free, high temporal resolution images to compose the bimonthly datasets while accounting for the spectral variability that may occur in the trees during the infestation cycle, including the budburst of leaves in many tree species across Ontario, between April and August. We also examined data from 30 m spatial resolution Landsat-8 and 10 m spatial resolution Sentinel-2 satellites. However, these satellites did not offer sufficient cloud free images across Ontario, resulting in significant mapping gaps and artifacts (results not presented here).

To generate the infestation maps, we uploaded the ground data from the Ontario geohub to the Google Earth Engine (GEE) platform. Then, we rasterized a mask containing 29,315 and 47,130 pixels from the infested and non-infested locations, respectively. The samples were extracted from 2020, the year with the highest infestation rate, and we trained the models using the data for this reference year.

Independent sample sets for each of the studied years were also created for validation. Here, the number of samples ranged from 9,900 to 20,000 pixels, depending on infestation levels.

2.3 Random forest probability estimate

A random forest algorithm (RF) (Breiman, 2001) was trained on the training samples from ground data and MODIS observations from each year to classify MODIS pixels by likelihood of insect infestation. RF combines a large number of trees trained upon random subsets of the available labeled samples and features. Each tree contributes only one class vote to each instance, and the result is determined by the majority votes of all the forest trees (Hastie et al., 2009). In our experiments, we set the number of trees in the "forest" (*ntree*), and the number of features/predictors considered for each node in the trees (*mtry*) to 500 and 5, respectively, after performing a grid search analysis.

To avoid spatial autocorrelation among samples, we trained the RF algorithm by applying a 5-fold spatial cross-validation. In particular, we divided the pixels annotated as infected into five non-overlapping areas, where for each fold, an RF model was fitted using samples from four locations, and the pixels associated with the remaining area were employed to validate the model. To compute the probabilistic maps, all five trained RF models were

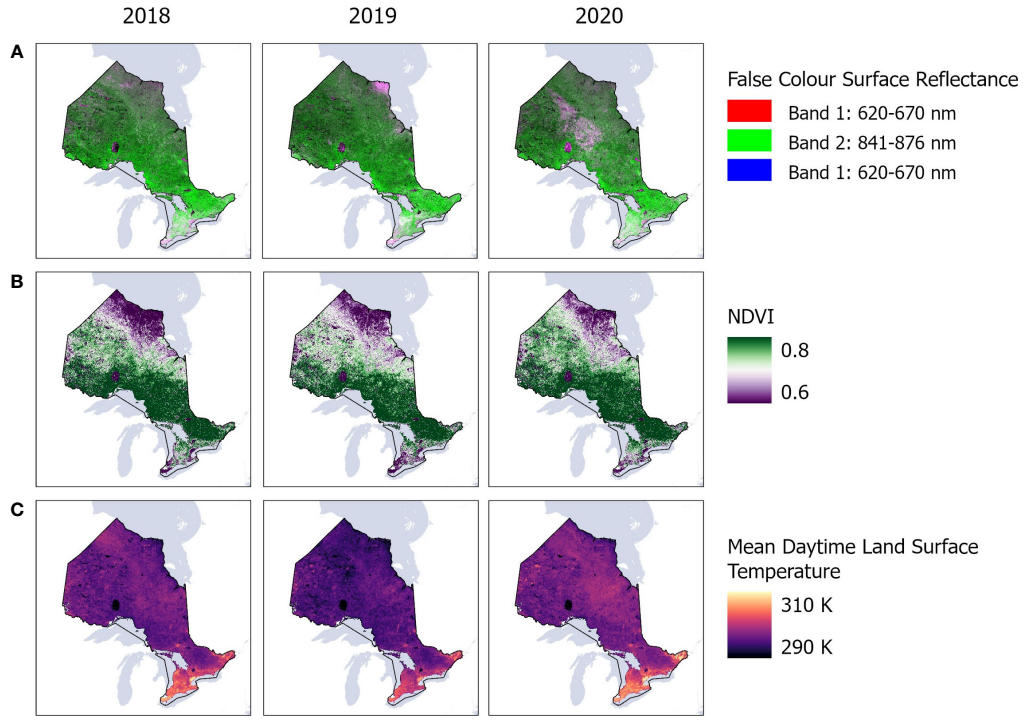


FIGURE 3
MODIS satellite observations showing mean **(A)** surface reflectance of red and near infrared bands; **(B)** normalized difference vegetation index (NDVI); and **(C)** thermal bands across Ontario, Canada between 2018 and 2020 in Google Earth Engine.

evaluated during the cross-validation phase, generating the same number of maps for 2018, 2019, and 2020. We reported the median and standard deviation of probability maps. Probability thresholds were calculated based on the percentage of pixels (Table 1), and nonforest locations were masked out using the ALOS PALSAR forest/nonforest product.

2.4 Calculation of human and environmental drivers of infestation

Using the classified maps, we tested the correlation between the occurrence of infestation and six environmental and human factors, namely temperature, precipitation, elevation, land cover, distance to roads and location of protected areas (Figure 4). Temperature and precipitation data between April and August were obtained from TerraClimate (https://www.climatologylab.org/terraclimate.html) at 4 km spatial resolution and were used as proxies to climate, which plays a significant role on the ability of defoliants to complete their life cycle (Régnière et al., 2009).

TABLE 1 Likelihood of occurrence threshold (probability estimate).

Probability threshold (%)	Description
0	No probability; associated with non-forested areas
(0–10]	Negligible probability
(10–30]	Less Likely probability
(30–60]	Likely probability
(60–90]	Moderate probability
(90–100]	High probability

A digital elevation model (DEM) obtained from Advanced Land Observing Satellite (ALOS, https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2) was used to produce an elevation map at 30 m spatial resolution. Land cover data was obtained from the Ontario GeoHub database (https://geohub.lio.gov.on.ca/documents/lio::ontario-land-cover-compilation-v-2-0/about) to mask out the non-forest areas. We also included protected areas that were obtained from Environment and Climate Change Canada database (https://www.canada.ca/en/environment-climate-change/services/national-wildlife-areas/protected-conserved-areas-database.html) as a sign of human footprint in remote areas, such as camping, logging, hiking and other associated activities, that may favor the spread of infestation. Finally, proximity to roads (data obtained from Ontario GeoHub, https://geohub.lio.gov.on.ca/datasets/mnrf::ontario-road-network-orn-road-net-element/about) was used to evaluate the direct influence of human transportation on infestation occurrence. Initially, our environmental and human variables were overlaid on our probability maps, followed by a correlation analysis to determine the likelihood of contributing factors to infestation levels.

10.3389/fevo.2023.1293311

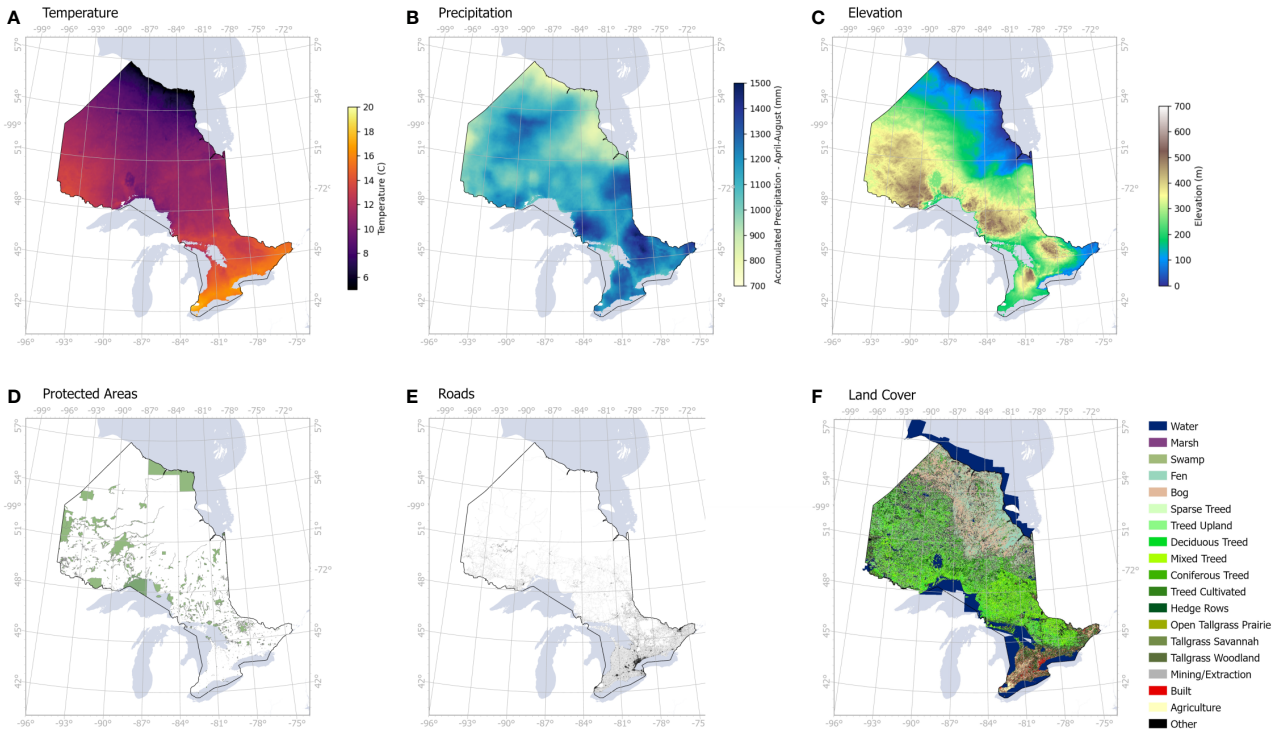


FIGURE 4
Human and environmental variables used to assess the drivers of infestation levels across Ontario: **(A)** average temperature from April to August, **(B)** accumulated precipitation from April to August, **(C)** elevation, **(D)** protected areas including all parks and conservation areas, **(E)** road networks, **(F)** land cover including coniferous and deciduous treed areas.

3 Results

3.1 Spatio-temporal occurrence of infestation

Our results reveal considerable increase in insect infestation across Ontario between 2018 and 2020 based on an RF probability using MODIS satellite observations (Figure 5A). Between 2018 and 2020, the total area classified as having a high probability of infestation across Ontario increased by >1,300%, from 2,900 km^2 to over 42,000 km^2 (Figure 6). Most of this expansion occurred in northwestern Ontario, along the border of neighboring province, Manitoba, while secondary incidences were observed in southern Ontario. Conversely, the proportion of pixels experiencing "moderate" (60–90%), "likely" (30–60%) and "less likely" (10–30%) probability thresholds decreased by 18%, 43% and 25.5% respectively during the same period (see Figure 6).

In 2018, spatial hotspots of infestation were concentrated along the northwestern Kenora, Rainy River and Thunder Bay, as well as in southern counties of Northumberland, Peterborough, Kawartha Lakes, extending into Sudbury and Timiskaming. In 2019, most of the infestation was concentrated along the northwestern part of the province. By 2020, the intensity of occurrence of hotspots increased along northwestern Ontario, extending southwesterly towards the "golden horseshoe".

3.2 Human and environmental drivers of infestation

From our analysis, recurring patterns emerged across two regions, namely southern Ontario near the golden horseshoe, and northwestern Ontario. Among these regions, large deviations in average temperature, elevation, accessibility, and vegetation type exist. However, these areas being indicated as hotspots may suggest some correlation in drivers of infestation. By determining what factors drive infestation, we may be able to (a) predict how populations will expand in the future, and (b) find appropriate methods to mitigate and/or eliminate future proliferation.

Based on correlation analysis, proximity to roads and temperature were the most important drivers of infestation occurrence across Ontario (Figure 7). There was a strong negative correlation between the Ontario provincial road network and infestation levels, i.e., the closer the roads, the higher the likelihood of infestation ($p < 0.001$). This was particularly noticeable towards northwestern Ontario. On the other hand, our analysis showed strong positive correlation with temperature, i.e., the higher the temperature increase the higher the probability of infestation ($p < 0.001$).

All other variables showed significant correlation ($p < 0.01$) albeit with smaller contributions to hotspots (see Figure 7). Precipitation emerged as having a weak negative correlation with

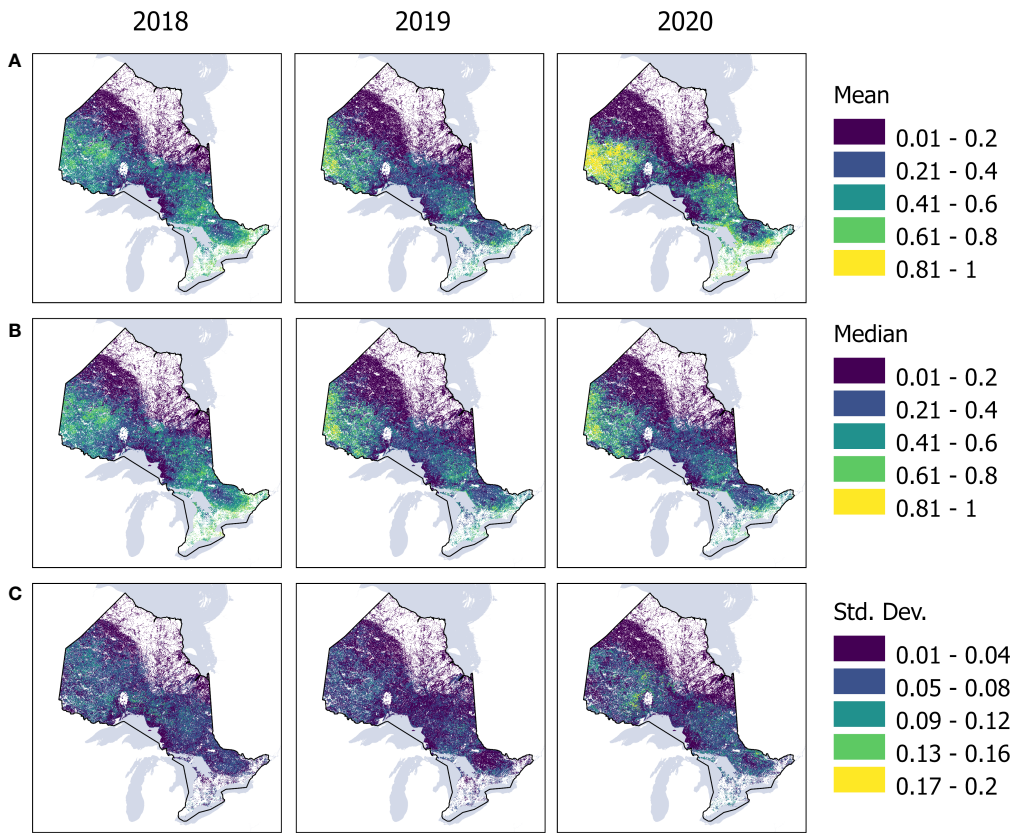


FIGURE 5
Annual **(A)** mean, **(B)** median and **(C)** standard deviation of our cross-validation maps from 2018 to 2020 using a *k*-fold of five. For mean and median, higher values indicate higher confidence of predictions (probability). Higher standard deviation values are associated with increasing uncertainty of each pixel of the model within our study region. White regions indicate non-forests.

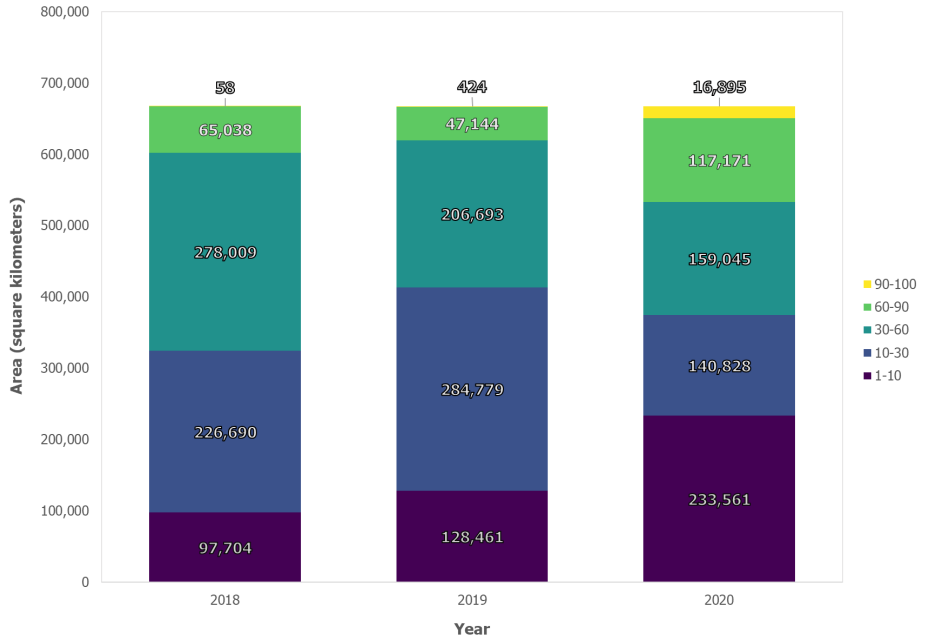


FIGURE 6
Area of annual infestation occurrence (km^2) for Ontario from 2018 to 2020 based on random forest classification.

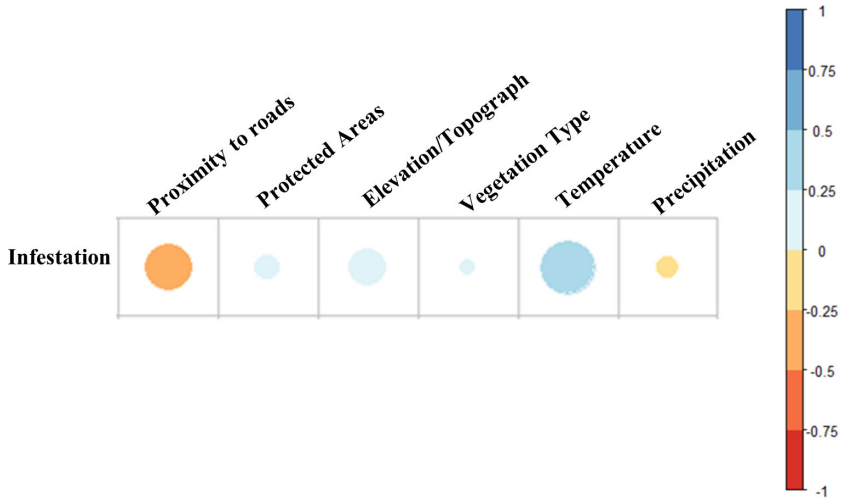


FIGURE 7
Correlation analysis of human and environmental drivers of infestation between 2018 and 2020. 0 to 1 (blue) indicates positive correlation, while 0 to −1 (orange−red) indicates negative correlation.

infestation (correlation range > −0.25), while elevation, protected areas and land cover emerged as having positive but weak correlation (correlation range < 0.25) with infestation.

3.3 Accuracy analysis

The average overall accuracy of our probability estimate was 88%, while the recall, precision, and F1 score metrics were 92%, 94%, and 93%, respectively. Recall measures the model's capability to identify the True samples, while precision quantifies the number of correct predictions. F1 score summarizes the model performance computing the harmonic mean between recall and precision. Figure 5 highlights the yearly mean, median and standard deviation of our cross-validation using a *k*-fold of five.

4 Discussion

Our findings paint a more complex picture of insect infestation than previously recorded across Ontario, Canada. While the reported observations in southern and some part of northern Ontario from government agencies have been captured in our analysis, we also documented larger, more widespread infestation across the northwest and northeast expanding into the southern part of the province.

By 2021, the Ontario government reported major outbreaks of insect infestation of *L. dispar*, *C. pinus*, *C. conflictana* and *C. fumiferana* and resulting moderate to severe defoliation in northwest, northeast and eastern districts (NDMNRF, 2022). In 2020, jack pine budworm defoliated approximately 10,658 km^2 of forest, mainly in the northwest region (NDMNRF, 2021). This was a 41% increase from 2018 (6,275 km^2, NDMNRF, 2019). Infestation levels of large aspen tortrix decreased to 228 km^2 by 2020 from 392 km^2 in 2018 and were primarily detected in Chapleau and Timmins districts in the northeast (NDMNRF, 2019; NDMNRF, 2021). Defoliation caused by spongy moth was mapped in only southern Ontario, increasing from 149 km^2 in 2018 to 5,864 km^2 by 2020 and was detected in southern districts including Northumberland, Hastings, Peterborough, and Kawartha Lakes counties (NDMNRF, 2019; NDMNRF, 2021). The area of spruce budworm defoliation increased from 1,371 km^2 in 2018 to 4,424 km^2 in 2020, mainly detected in northeast regions (NDMNRF, 2019; NDMNRF, 2021). An aggregate of 21,174 km^2 of forests were defoliated by our study species by 2020. Although the above regions were captured within our analysis, our study also revealed considerable expansion of infestation (a total area of 239,603 km^2 based on moderate to high probability estimates) in the north and south. Defoliation expanded in the districts of Cochrane, Timmins, Dryden, Sioux Lookout, Kenora, Nipigon, Fort France, Red Lake, Chapleau, Sudbury, Hearst, Kirkland Lake, North Bay, Sault Ste. Marie, southern parts of Wawa, North Bay, Parry Sound, Peterborough, Peterborough, Bancroft, Kemptville, Midhurst, Pembroke, Hastings, Aylmer, Guelp, Aurora, Northumberland, Kawartha Lakes. In 2020, defoliation was recorded in Algonquin Provincial Park, which corresponded to government reports in 2021 (NDMNRF, 2022).

The expansion of insect infestation across Ontario appears to be significantly correlated with the Ontario provincial road network, especially logging roads in the north, whereas roads in the south are used by a greater demographic of commuters (see Figure 7). Therefore, areas near roads will more likely be prone to increasing infestations. For instance, these findings confirm Benoit and Lachance's (1990) reporting of species such as *L. dispar* where although wind is the primary, natural dispersal of their caterpillars, human transportation is responsible for their long-range dispersal. Ciesla and Kruse (2009) noted that outbreaks of large aspen tortrix may be associated with recreation

sites and home sites in urban–wildland interface. Furthermore, our findings suggest that jack pine budworm and spruce budworm are the major species causing defoliation in the northwest and northeast of the province, and their expansion may be linked to proximity to roads. Government reports corroborated our findings (NDMNRF, 2019; NDMNRF, 2021) but we have found no papers which examine other drivers of infestation of these species outside of climatic analysis.

Additionally, high infestation showed a positive correlation with temperature across the province. With rising temperatures, geographical ranges of insect pests may shift or expand into new areas of suitable habitat at their northern limits and into higher elevations, and species may potentially move to secondary hosts (Williams and Liebhold, 1997; Candau and Fleming, 2011; Pureswaran et al., 2015; Hartl-Meier et al., 2017). Temperature anomalies may trigger early budburst of tree hosts, or cause late emergence of larval defoliators, leading to phenological asynchrony between species and their hosts (Jepsen et al., 2008; Régnière et al., 2012; Pureswaran et al., 2015; Jakoby et al., 2019). Species such as *C. pinus* and *C. fumiferana* are cold-blooded, with their body temperature being determined by the temperature of their surrounding environment. As temperature increases in the south, these species are extending their range farther north (Wellington et al., 1950; Greenbank, 1956; Pilon and Blais, 1961). Likewise, *L. dispar* moths cannot start their life cycle below −20°C and their life cycle is limited below 10°C (Benoit and Lachance, 1990). This may explain why they were previously unrecorded in northeast and some parts of southern Ontario. Then, suitable conditions are provided, as temperature increases, for their eggs and larvae to grow and develop, and this may explain why we saw increasing expansion in the province.

Our results also agree with predictions made by Régnière et al (2009; 2012). which predicted increasing migration northwards in Ontario for several species based on climate suitability. Régnière et al. (2009) indicated an increasing migration pattern of *L. dispar* moth infestation towards the north and east of Ontario by > 50% by 2030, and > 90% by 2070. Their predictions for 2020 are in line with our maps suggesting an infestation hotspot in western Ontario, on the border with Manitoba, an area only limited reported by ground data (see NDMNRF, 2019; NDMNRF, 2021; NDMNRF, 2022). However, our findings indicate a faster timeline for establishment of infestation across the province than predicted by Régnière et al. (2009). Additionally, Régnière et al. (2012) predicted that the distribution of spruce budworm outbreaks is likely to shift northward and towards higher elevations over the next 50 years in response to climate change but may vary considerably due to the current distribution of the insect's three main host plants (*balsam fir, white and red spruce*).

While much of this expanding infestation may be linked to climate, for propagation to occur suitable tree types must be present. According to Cihlar et al. (2002), deciduous and mixed forests are more susceptible to, for example, *L. dispar* moth infestation, which is abundant in northern parts of Ontario, but colder temperatures may have previously prevented their spread. We examined whether forest type was linked with infestation levels, but our analysis showed no correlation between infestation and forest types. This may be because caterpillars of the *L. dispar* moths seldom feed on conifer trees but may do so if populations are high and most of the leaves on deciduous trees (Figure 8), such as oaks, have already been consumed. Likewise, the distribution of host plants for spruce budworms can also be expected to change, many shifting towards higher latitudes, over the longer term (Payette, 2007), and possibly exacerbating the insect's increasing northern impact (see Régnière et al., 2012). With jack pine budworm, outbreaks patterns may be the result of inter-correlations between these abiotic variables and pollen-cone production (Nealis and Lomic, 1994).

Although we did not find notable correlation between infestation and other factors such as precipitation, elevation, and land cover, other studies from agricultural experiments have shown that increasing patterns of rainfall increase development time for certain species of insects (Chen et al., 2019); while higher elevations increase patterns of infestation for others (see Azrag et al., 2018).

Our results suggest that the protected area networks and current federal management strategies have had limited success combating the pervasive spread of insect incursion across the province within our study period. Considerable mitigation strategies have been implemented by authorities and private landowners to restrict the expansion of infestation. These strategies encompass various forest stand management techniques, such as avoiding the creation of uneven-aged or multi-storied jack pine stands, steering clear of edge stands or forest islands, and implementing accelerated harvesting and salvage harvesting. Additionally, measures like aerial spraying of bio-insecticide in specific areas, physically scraping off larvae from infested tree trunks, establishing pheromone traps on infested trees (Figure 8), applying bacterial insecticide Btk (Foray 76B) to jack pine stands, and creating physical traps like burlap and duct tape banding (Figure 8D) around tree trunks have been employed (NDMNRF, 2019; NDMNRF, 2021; NDMNRF, 2022). Despite these efforts, our results suggest that the management strategies may either need revising or expanding to regions where current practices are not implemented to account for factors such as evolving insect behavior in response to changing environmental conditions and potential resistance to control measures.

Further, as proximity to roads emerged as a major driver of infestation, monitoring and management of transportation and human activities may be required across the province. For instance, Bigsby et al. (2011) showed that regulated activities by state and federal governments in the United States correlated with decreased presence of *L. dispar* moth, which could be implemented within Ontario. Altogether, our results raise awareness of increasing expansion of infestation that national-level statistics do not capture and pose new challenges for conservation of Ontario's forests.

Additionally, incorporating indigenous knowledge in forest insect management practices, as proposed by Maloney (see Maloney, 2019), may prove instrumental in limiting outbreaks and infestation levels. To enhance the efficacy of these strategies, it may also be crucial to explore innovative approaches such as genetic modification of tree species for pest resistance and the integration of advanced remote sensing technologies for more precise monitoring and targeted intervention. Furthermore, fostering collaborations between researchers, policymakers, and industry stakeholders can lead to a more holistic and adaptive approach in tackling this persistent threat.

FIGURE 8
(A) Example of defoliation of a deciduous tree by *L. dispar* moths; **(B)** a pheromone trap placed by residents to capture moths from spreading; **(C)** captured moths in pheromone traps; **(D)** duct tape banding employed to offer a physical obstruction to the mobility of moths.

5 Conclusions

Our study unveils a nuanced understanding of insect infestation across Ontario, Canada. While government agencies captured observations in southern and northern parts of Ontario, our analysis revealed a larger and more widespread infestation level. By 2021, the Ontario government reported major outbreaks of infestation in various districts, corroborating our findings of more extensive infestation. The expansion of infestation correlated significantly with the provincial road network, particularly logging roads in the north, indicating areas near roads are more prone to increasing infestations.

Additionally, our results align with predictions of increasing migration northwards of several species based on climate suitability. Rising temperatures play a crucial role in the geographical range expansion of insect pests, further exacerbated by phenological shifts. Forest type did not emerge as a significant factor influencing infestation levels, highlighting the complexity of the interplay between insects and their host trees. While our study provides valuable insights, it is evident that current management strategies have had limited success in curbing the persistent spread of infestation. Adapting these strategies to evolving insect behavior and exploring innovation approaches, including advanced remote sensing for quick and effective monitoring, may offer more effective solutions. Collaboration between stakeholders will be instrumental in developing a holistic and adaptive approach to confront this ongoing threat.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ontario.ca/page/land-information-ontario; https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD09GQ and https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD11A1.

Author contributions

MK: Conceptualization, Supervision, Methodology, Formal Analysis, Writing – original draft, Writing – review & editing. IG: Conceptualization, Data curation, Investigation. JC: Formal Analysis, Writing – review & editing. CS: Conceptualization, Methodology, Formal Analysis, Writing – review & editing. CR:

Visualization, Writing – review & editing. JS: Resources, Writing – review & editing. AG: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2020-05,708) and the Canada Research Chairs Program (CRC-2019-00139).

Acknowledgments

We would like to thank Andrew M. Kreklewich, Ricardo Barros Lourenço and Neil Murphy for their useful insight into insect infestation across Ontario. This work was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2020-05,708), the Canada Research Chairs Program and World Wildlife Fund Canada (WWF-Canada).

Conflict of interest

Author CS was employed by the company Finite Carbon Canada.

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

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., et al. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* 259, 660–684. doi: 10.1016/j.foreco.2009.09.001

Ayres, M. P., and Lombardero, M. J. (2000). Assessing the consequences of global change for forest disturbance from herbivores and pathogens. *Sci. Total. Environ.* 262, 263–286. doi: 10.1016/S0048-9697(00)00528-3

Azrag, A. G., Pirk, C. W., Yusuf, A. A., Pinard, F., Niassy, S., Mosomtai, G., et al. (2018). Prediction of insect pest distribution as influenced by elevation: Combining field observations and temperature dependent development models for the coffee stink bug, *Antestiopsis thunbergii* (Gmelin). *PloS One* 13, e0199569. doi: 10.1371/journal.pone.0199569

Bachelet, D., Neilson, R. P., Hickler, T., Drapek, R. J., Lenihan, J. M., Sykes, M. T., et al. (2003). Simulating past and future dynamics of natural ecosystems in the United States. *Global Biogeochem. Cycles.* 17, 1045. doi: 10.1029/2001GB001508

Baker, H. S., Millar, R. J., Karoly, D. J., Beyerle, U., Guillod, B. P., Mitchell, D., et al. (2018). Higher CO_2 concentrations increase extreme event risk in a 1.5 °C world. *Nat. Climate Change* 8, 604–608. doi: 10.1038/s41558-018-0190-1

Benoit, P., and Lachance, D. (1990). *Gypsy moth in Canada: behavior and control. DPC-X-32 (Forestry Canada).*

Bigsby, K. M., Tobin, P. C., and Sills, E. O. (2011). Anthropogenic drivers of gypsy moth spread. *Biol. Invasions.* 13, 2077–2090. doi: 10.1007%2Fs10530-011-0027-6

Boisvenue, C., and Running, S. W. (2006). Impacts of climate change on natural forest productivity–evidence since the middle of the 20th century. *Global Change Biol.* 12, 862–882. doi: 10.1111/j.1365-2486.2006.01134.x

Bonan, G. B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *science* 320, 1444–1449. doi: 10.1126/science.115512

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bugmann, H. (2001). A review of forest gap models. *Clim. Change* 51, 259–305. doi: 10.1023/A:1012525626267

Candau, J.-N., and Fleming, R. A. (2011). Forecasting the response of spruce budworm defoliation to climate change in Ontario. *Can. J. For. Res.* 41, 1948–1960. doi: 10.1139/x11-134

Chen, C., Harvey, J. A., Biere, A., and Gols, R. (2019). Rain downpours affect survival and development of insect herbivores: the specter of climate change? *Ecology* 100, e02819. doi: 10.1002/ecy.2819

Ciesla, W. M., and Kruse, J. J. (2009). *Large aspen tortrix* (US Department of Agriculture, Forest Service, Pacific Northwest Region).

Cihlar, J., Beaubien, J., and Latifovic, R. (2002). Land cover of Canada 1998. Special publication, NBIOME project. Produced by the Canada Centre for Remote Sensing and the Canadian Forest Service. *Natural Resour. Canada. Available. Canada. Centre. Remote Sensing. Ottawa. Ontario.*

Elkinton, J., and Liebhold, A. (1990). Population dynamics of gypsy moth in North America. *Annu. Rev. Entomol.* 35, 571–596. doi: 10.1146/annurev.en.35.010190.003035

Gonsamo, A., Chen, J. M., Colombo, S. J., Ter-Mikaelian, M. T., and Chen, J. (2017). Global change induced biomass growth offsets carbon released via increased forest fire and respiration of the central Canadian boreal forest. *J. Geophys. Res.: Biogeosci.* 122, 1275–1293. doi: 10.1002/2016JG003627

Gonsamo, A., Ciais, P., Miralles, D. G., Sitch, S., Dorigo, W., Lombardozzi, D., et al. (2021). Greening drylands despite warming consistent with carbon dioxide fertilization effect. *Global Change Biol.* 27, 3336–3349. doi: 10.1111/gcb.15658

Gray, D., White, C., and Tomlinson, G. (2007). Increasing security and reducing carbon emissions of the us transportation sector: a transformational role for coal with biomass.

Greenbank, D. (1956). The role of climate and dispersal in the initiation of outbreaks of the spruce budworm in New Brunswick: I. The role of climate. *Can. J. Zool.* 34, 453–476. doi: 10.1139/z56-048

Hanson, P. J., and Weltzin, J. F. (2000). Drought disturbance from climate change: response of United States forests. *Sci. Total. Environ.* 262, 205–220. doi: 10.1016/S0048-9697(00)00523-4

Hartl-Meier, C., Esper, J., Liebhold, A., Konter, O., Rothe, A., and Büntgen, U. (2017). Effects of host abundance on larch budmoth outbreaks in the European Alps. *Agric. For. Entomol.* 19, 376–387. doi: 10.1111/afe.12216

Hastie, T., Tibshirani, R., and Friedman, J. (2009). Random Forests. In: *The Elements of Statistical Learning*. New York, NY: Springer Series in Statistics. Springer. doi: 10.1007/978-0-387-84858-7_15

Hollaus, M., and Vreugdenhil, M. (2019). Radar satellite imagery for detecting bark beetle outbreaks in forests. *Curr. Forestry. Rep.* 5, 240–250. doi: 10.1007/s40725-019-00098-z

Jakoby, O., Lischke, H., and Wermelinger, B. (2019). Climate change alters elevational phenology patterns of the European spruce bark beetle (*Ips typographus*). *Global Change Biol.* 25, 4048–4063. doi: 10.1111/gcb.14766

Jepsen, J. U., Hagen, S. B., Ims, R. A., and Yoccoz, N. G. (2008). Climate change and outbreaks of the geometrids *Operophtera brumata* and *Epirrita autumnata* in subarctic birch forest: evidence of a recent outbreak range expansion. *J. Anim. Ecol.* 77, 257–264. doi: 10.1111/j.1365-2656.2007.01339.x

Kurz, W. A., and Apps, M. J. (1999). A 70-year retrospective analysis of carbon fluxes in the Canadian forest sector. *Ecol. Appl.* 9, 526–547. doi: 10.1890/1051-0761(1999)009[0526:AYRAOC]2.0.CO;2

Lehmann, J., Hansel, C. M., Kaiser, C., Kleber, M., Maher, K., Manzoni, S., et al. (2020). Persistence of soil organic carbon caused by functional complexity. *Nat. Geosci.* 13, 529–534. doi: 10.1038/s41561-020-0612-3

Liebhold, A. M., Halverson, J. A., and Elmes, G. A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *J. Biogeogr.* 19 (5) 513–520. doi: 10.2307/2845770

Lloyd, A. H., and Bunn, A. G. (2007). Responses of the circumpolar boreal forest to 20th century climate variability. *Environ. Res. Lett.* 2, 045013. doi: 10.1088/1748-9326/2/4/045013

Lovett, G. M., Christenson, L. M., Groffman, P. M., Jones, C. G., Hart, J. E., and Mitchell, M. J. (2002). Insect defoliation and nitrogen cycling in forests: laboratory, plot, and watershed studies indicate that most of the nitrogen released from forest foliage as a result of defoliation by insects is redistributed within the ecosystem, whereas only a small fraction of nitrogen is lost by leaching. *BioScience* 52, 335–341. doi: 10.1641/0006-3568(2002)052[0335:IDANCI]2.0.CO;2

Lucht, W., Schaphoff, S., Erbrecht, T., Heyder, U., and Cramer, W. (2006). Terrestrial vegetation redistribution and carbon balance under climate change. *Carbon. Balance. Manage.* 1, 1–7. doi: 10.1186/1750-0680-1-6

Lyons, D. B., and Liebhold, A. M. (1992). Spatial distribution and hatch times of egg masses of gypsy moth (Lepidoptera: Lymantriidae). *Environ. Entomol.* 21, 354–358. doi: 10.1093/ee/21.2.354

Maloney, J. (2019). Insect management in Canada's forest sector: strengthening national cooperation against current and future outbreaks. *Rep. Standing. Committee. Natural Resour*.

Maroschek, M., Seidl, R., Netherer, S., and Lexer, M. (2009). Climate change impacts on goods and services of European mountain forests. *Unasylva* 60, 76–80.

NDMNRF (2019) *Ministry of Northern Development, Mines, Natural Resources and Forestry*. Available at: https://files.ontario.ca/mnrf-forest-health-conditions-report-2019-en-2020-07-31.pdf.

NDMNRF (2020) *Ministry of Northern Development, Mines, Natural Resources and Forestry*. Available at: https://files.ontario.ca/ndmnrf-forest-health-conditions-2020-en-2021-07-30.pdf.

NDMNRF (2021) *Ministry of Northern Development, Mines, Natural Resources and Forestry*. Available at: https://www.ontario.ca/files/2022-07/mnrf-srb-forest-health-conditions-ontario-2021-en-2022-07-20.pdf.

NDMNRF (2022) *Ministry of Northern Development, Mines, Natural Resources and Forestry*. Available at: https://www.ontario.ca/files/2023-05/mnrf-forest-health-conditions-2022-en-2023-05-09.pdf.

Nealis, V., and Lomic, P. (1994). Host-plant influence on the population ecology of the jack pine budworm, *Choristoneura pinus* (Lepidoptera: Tortricidae). *Ecol. Entomol.* 19, 367–373. doi: 10.1111/j.1365-2311.1994.tb00254.x

Niemann, K. O., and Visintini, F. (2005). *Assessment of potential for remote sensing detection of bark beetle-infested areas during green attack: A literature review*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC: Mountain Pine Beetle Initiative Working Paper 2005-02.

Pilon, J., and Blais, J. (1961). Weather and outbreaks of the spruce budworm in the Province of Quebec from 1939 to 1956. *Can. Entomol.* 93, 118–123. doi: 10.4039/Ent93118-2

Payette, S. (2007) Contrasted dynamics of northern Labrador tree lines caused by climate change and migration lag. *Ecology* 88, 770–780.

Porter, T., and Coon, M. (1991). Cytochrome p-450. multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.* 266, 13469–13472. doi: 10.1016/S0021-9258(18)92717-1

Pureswaran, D. S., De Grandpré, L., Paré, D., Taylor, A., Barrette, M., Morin, H., et al. (2015). Climate-induced changes in host tree–insect phenology may drive ecological state-shift in boreal forests. *Ecology* 96, 1480–1491. doi: 10.1890/13-2366.1

Pureswaran, D. S., Roques, A., and Battisti, A. (2018). Forest insects and climate change. *Curr. Forestry. Rep.* 4, 35–50. doi: 10.1007/s40725-018-0075-6

Raven, J. A., Giordano, M., Beardall, J., and Maberly, S. C. (2012). Algal evolution in relation to atmospheric CO_2: carboxylases, carbon-concentrating mechanisms and carbon oxidation cycles. *Philos. Trans. R. Soc. B.: Biol. Sci.* 367, 493–507. doi: 10.1098/rstb.2011.0212

Régnière, J., Nealis, V., and Porter, K. (2009). Climate suitability and management of the gypsy moth invasion into Canada. *Ecol. Impacts. Non-native. Invertebrates. Fungi. Terrestrial. Ecosyst.*, 135–148. doi: 10.1007/978-1-4020-9680-8_10

Régnière, J., Powell, J., Bentz, B., and Nealis, V. (2012). Effects of temperature on development, survival and reproduction of insects: experimental design, data analysis and modeling. *J. Insect Physiol.* 58, 634–647. doi: 10.1016/j.jinsphys.2012.01.010

Romeiro, J. M. N., Eid, T., Antón-Fernández, C., Kangas, A., and Trømborg, E. (2022). Natural disturbances risks in European boreal and temperate forests and their links to climate change–a review of modelling approaches. *For. Ecol. Manage.* 509, 120071. doi: 10.1016/j.foreco.2022.120071

Scholze, M., Knorr, W., Arnell, N. W., and Prentice, I. C. (2006). A climate-change risk analysis for world ecosystems. *Proc. Natl. Acad. Sci.* 103, 13116–13120. doi: 10.1073/pnas.0601816103

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., et al. (2017). Forest disturbances under climate change. *Nat. Climate Change* 7, 395–402. doi: 10.1038/nclimate3303

Thom, D., and Seidl, R. (2016). Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests. *Biol. Rev.* 91, 760–781. doi: 10.1111/brv.12193

Tobin, P. C., Sharov, A. A., Liebhold, A. A., Leonard, D. S., Roberts, A. E., and Learn, M. R. (2004). Management of the gypsy moth through a decision algorithm under the STS project. *Am. Entomol.* 50, 200–209. doi: 10.1093/ae/50.4.200

Volney, W. J. A., and Fleming, R. A. (2000). Climate change and impacts of boreal forest insects. *Agricult. Ecosyst. Environ.* 82, 283–294. doi: 10.1016/S0167-8809(00)00232-2

Wellington, W., Fettes, J., Belyea, R., and Turner, K. (1950). Physical and biological indicators of the development of outbreaks of the spruce budworm, *Choristoneura fumiferana* (Clem.) (Lepidoptera: Tortricidae). *Can. J. Res.* 28, 308–331. doi: 10.1139/cjr50d-021

Williams, D., and Liebhold, A. (1997). Latitudinal shifts in spruce budworm (Lepidoptera: Tortricidae) outbreaksand spruce-fir forest distributions with climate change. *Acta Phytopathol. Entomol. Hungarica.* 32, 205–215.

Woodman, S. G., Khoury, S., Fournier, R. E., Emilson, E. J., Gunn, J. M., Rusak, J. A., et al. (2021). Forest defoliator outbreaks alter nutrient cycling in northern waters. *Nat. Commun.* 12, 6355. doi: 10.1038/s41467-021-26666-1

Ye, S., Rogan, J., Zhu, Z., Hawbaker, T. J., Hart, S. J., Andrus, R. A., et al. (2021). Detecting subtle change from dense landsat time series: Case studies of mountain pine beetle and spruce beetle disturbance. *Remote Sens. Environ.* 263, 112560. doi: 10.1016/j.rse.2021.112560

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 23 November 2023
DOI 10.3389/fenvs.2023.1304845

OPEN ACCESS

EDITED BY
Prof-Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Anoop Kumar Shukla,
Manipal Academy of Higher Education, India
Arfan Arshad,
Oklahoma State University, United States
Ngoc Vinh Tran,
University of Michigan, United States

*CORRESPONDENCE
Thanh-Nhan-Duc Tran,
syu3cs@virginia.edu
Binh Quang Nguyen,
nqbinh@dut.udn.vn

RECEIVED 30 September 2023
ACCEPTED 13 November 2023
PUBLISHED 23 November 2023

CITATION
Tran T-N-D, Nguyen BQ, Grodzka-Łukaszewska M, Sinicyn G and Lakshmi V (2023), The role of reservoirs under the impacts of climate change on the Srepok River basin, Central Highlands of Vietnam.
Front. Environ. Sci. 11:1304845.
doi: 10.3389/fenvs.2023.1304845

The role of reservoirs under the impacts of climate change on the Srepok River basin, Central Highlands of Vietnam

Thanh-Nhan-Duc Tran[1]*, Binh Quang Nguyen[2,3]*, Maria Grodzka-Łukaszewska[4], Grzegorz Sinicyn[4] and Venkataraman Lakshmi[1]

[1]Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, VA, United States, [2]Water Resource Center, Disaster Prevention Research Institute (DPRI), Kyoto University, Kyoto, Japan, [3]Faculty of Water Resources Engineering, The University of Danang-University of Science and Technology, Da Nang, Vietnam, [4]Faculty of Building Services, Hydro and Environmental Engineering, Warsaw University of Technology, Warszawa, Poland

Forecasting streamflow is important for managing future water resources and environmental needs under the impacts of climate change. Moreover, quantifying the combined effects of future climate variations and human-made infrastructures, e.g., dams and reservoirs, poses a significant challenge. In this study, we used the Soil and Water Assessment Tool (SWAT) for a case study in the Srepok River Basin—a tributary of the Mekong River Basin. Here, we aim to reveal the impacts of various climate change scenarios and the effects of reservoir operations in this region. Our findings indicate that 1) the projected annual streamflow is anticipated to increase by a minimum of 9.2% (2046–2065) and could peak at an increase of 14.9% (2080–2099) under the highest greenhouse gas emissions, 2) Srepok 4, Srepok 3, and Buon Kuop demonstrate a higher capability for mitigating flood peaks and managing seasonal flow in the downstream floodplain, whereas Buon Tua Srah shows the least performance, and 3) reservoirs operated with annual regulation have more pronounced impacts than those regulated on a daily schedule. Our work provides i) a scientific foundation for regional stakeholders and decision-makers to develop sustainable strategies that address the combined effects of reservoir operation and future climate, and ii) it supports national authorities and officials in resolving conflicts related to transboundary rivers within the Mekong River Basin.

KEYWORDS

climate change, reservoir, SWAT, Srepok River basin, Mekong River basin

1 Introduction

Climate change, increasingly recognized as a major concern, has significant impacts on the quality and quantity of water resources (Shukla et al., 2020; Tran et al., 2021a; 2021b; Hussain et al., 2022). Previous studies have identified greenhouse gas emissions, partly resulted from urbanization (Nguyen et al., 2022), as one of the primary reasons causing global warming. These emissions are expected to lead to future increases in temperature and precipitation (IPCC, 2014; IPCC, 2019) resulting in more frequent extreme weather events (e.g., extreme heat waves, widespread floods, year-long droughts, and severe wildfires). The Representative Concentration Pathway (RCP) is a trajectory for greenhouse gas

concentration developed by the IPCC (IPCC, 2019). Numerous studies have highlighted the effects of climate change on the natural water cycle (Bolch et al., 2012; Aryal et al., 2019; Bhatta et al., 2019; Roderick et al., 2019; Tran et al., 2023b). Floods caused by extreme rainfall are expected to become more frequent across the globe (Chattopadhyay et al., 2017; Vo et al., 2018; Ivanov et al., 2021; Tran et al., 2023b), and are likely to persist even more severely in the near future. Besides, Lehner et al. (2019) observed that the flow patterns of a third of the world's 200 largest rivers have changed since the 1950s, with these alterations being directly attributable to climate change. Although it is challenging to obtain an accurate prediction of streamflow under the climate impacts, it remains a crucial undertaking for effective water resource planning, regional management, and mitigation of extreme events (Khoi and Suetsugi, 2012; Tran et al., 2022c; Tran et al., 2022e; Umar et al., 2022). In Vietnam, a number of studies have investigated the impacts of climate change at various scales (Khoi and Suetsugi, 2012; Ty et al., 2012; Huyen et al., 2017; Vo et al., 2018; Giang and Vy, 2021). However, the majority of these studies were carried out in previous decades with the neglect of important factors (e.g., man-made infrastructures), raising concerns about their reliability which is relevant to the current specifics of climate change in this region.

The Mekong River, an important transboundary river in East and Southeast Asia, flows through several countries, including China, Myanmar, Laos, Thailand, Cambodia, and Vietnam (Arias et al., 2014a) (Figure 1A). The Srepok River Basin (SRB), which is located in central Vietnam, is a significant tributary of the Mekong River Basin (MRB), annually contributing a substantial volume of water to the Mekong River (Arias et al., 2014b; Nuong et al., 2022; Chang et al., 2023) (Figure 1B). Alterations in water supply, particularly due to dam operations, can remarkably affect the region's water resources (Giang et al., 2017; Du et al., 2022; Nguyen et al., 2023a; Nguyen et al., 2023b; Bui et al., 2023; Smigaj et al., 2023). Within SRB, such changes in the water cycle may affect the hydrology of the middle and lower sections of the MRB, influencing the lives of nearly 11,000 Cambodians residing along the river and at the basin's outlet (International Rivers, 2010). Previous studies have enhanced our understanding of how climate change may impact the Srepok River as well as the 3S River (Srepok, Sesan, and Sekong) (Arias et al., 2014a; Cochrane et al., 2014; Piman et al., 2016; Huyen et al., 2017; Trang and Lakshmi, 2022; Bui et al., 2023). However, these studies have overlooked the impact of existing dams and reservoirs due to a lack of operational data, an oversight that should be considered in climate change studies. Furthermore, the critical role of the Krongbuk tributary has been largely ignored, meaning that total runoff at the SRB outlet—and by extension, regional volume assessments and disaster prevention strategies—may not be fully understood or accounted for. Furthermore, the reliability of their baseline models, which utilize historical data, is questionable due to the limited duration of the calibration and validation periods—factors that play a crucial role in deriving accurate parameters for projecting future climate scenarios. Also, the study by Piman et al. (2016), which constructed a baseline period between 1986 and 2006, or Arias et al. (2014a) from 1982 to 2005, Cochrane et al. (2014), Oeurng et al. (2016) from 1980 to 2008, and Shrestha et al. (2016) from the 1980s–2000s, have eliminated considerations of current dam operations for the SRB. This gap suggests that such studies may not fully capture the projected impacts of future climate change on this region, potentially providing decision-makers and regional planners with incomplete data.

The climate inputs for hydrological models, e.g., temperature and precipitation, can be derived using two primary methods: 1) revising records obtained from meteorological stations (Khoi and Suetsugi, 2012; Li and Fang, 2021; Raghavan et al., 2012), and 2) adjusting outputs from climate models (Vo et al., 2018; Tran et al., 2022c). General Circulation Models (GCMs), also known as Global Climate Models, are developed through assumptions and mathematical representations of the physical climate system's processes (Li and Fang, 2021). An ensemble of GCMs, created by various global organizations and institutions, can provide more accurate predictions for water resources than a single GCM (Pierce et al., 2009; Ranger et al., 2011). Accordingly, we utilized datasets from the climate change scenarios released by the Vietnam Ministry of Natural Resources and Environment (2020) in this study. These scenarios cover two timeframes: the near future (2046–2065) and the far future (2080–2099) and the ensemble would be formed based on projections from six distinct Regional Climate Models (RCMs). These include: i) the Climate-WRF (CL-WRF) model from the United States (Fita et al., 2010); ii) the Providing REgional Climates for Impacts Studies (PRECIS) model from the United Kingdom (Moberg and Jones, 2004); iii) the Conformal Cubic Atmospheric Model (CCAM) from Australia (Her, 2014); iv) the Regional Climate Model (RegCM) from Italy (NCAR, 2017); v) the Meteorological Research Institute (MRI) for Atmospheric Climate Model (AGCM/MRI) from Japan (Mizuta et al., 2012); and vi) the Rossby Centre Regional Climate (RCA3) model from Sweden (Samuelsson et al., 2011).

This study aims to reveal the impacts of dam operation and future climate scenarios on the water resources of SRB by employing the SWAT hydrological model. In general, our findings would:

(1) Refine and update previous baseline models using recent and comprehensive datasets for accurate simulations of projected future runoff, especially at the SRB's outlet;
(2) Analyze the impacts of climate change and the existing dams' operation, an aspect often overlooked in previous works; and
(3) Quantify the contributions of the SRB's main tributaries to the overall flow under different future climate conditions.

By fulfilling these objectives, our study aims to reduce the uncertainties currently associated with future hydrological projections in this region. The implications of our findings are critical for water resource management, with direct benefits for the local communities in Vietnam and Cambodia. Additionally, this work provides a solid scientific foundation for disaster prevention strategies in the lower MRB, thereby supporting stakeholders and regional authorities in making informed decisions.

2 Materials and methods

2.1 Study area

The Srepok River, a major tributary of the Mekong, has its source in the highlands of Dak Lak Province, Vietnam. It flows through Ratanakiri and Stung Treng regions before joining the

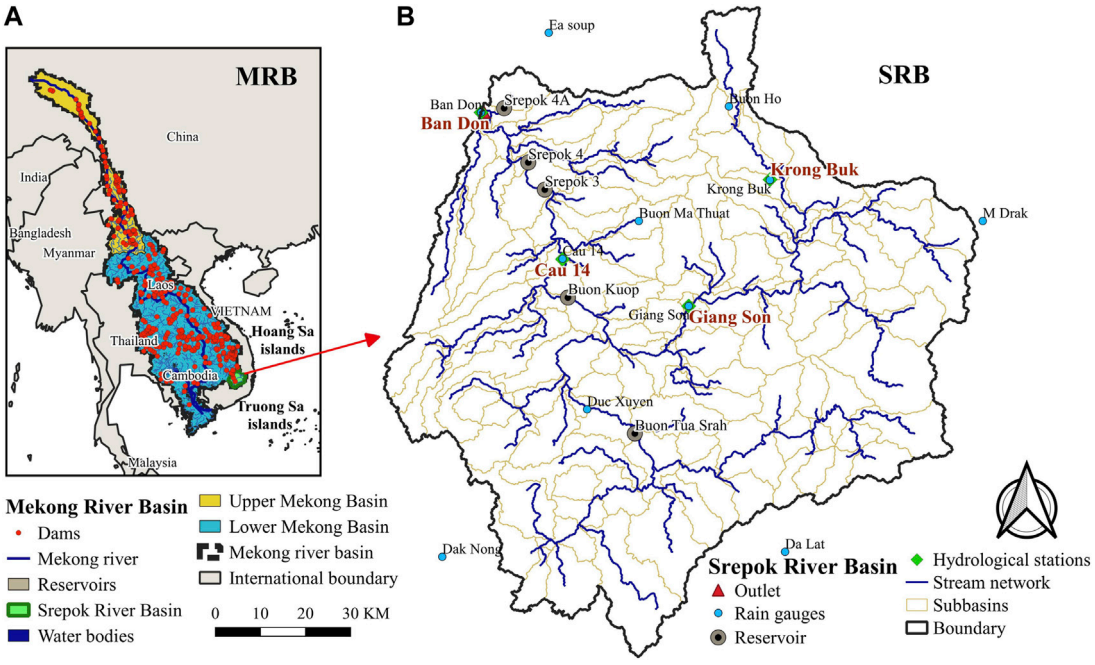


FIGURE 1
(A) Location of SRB in the MRB **(B)** SRB and the distribution of reservoirs, and stream network.

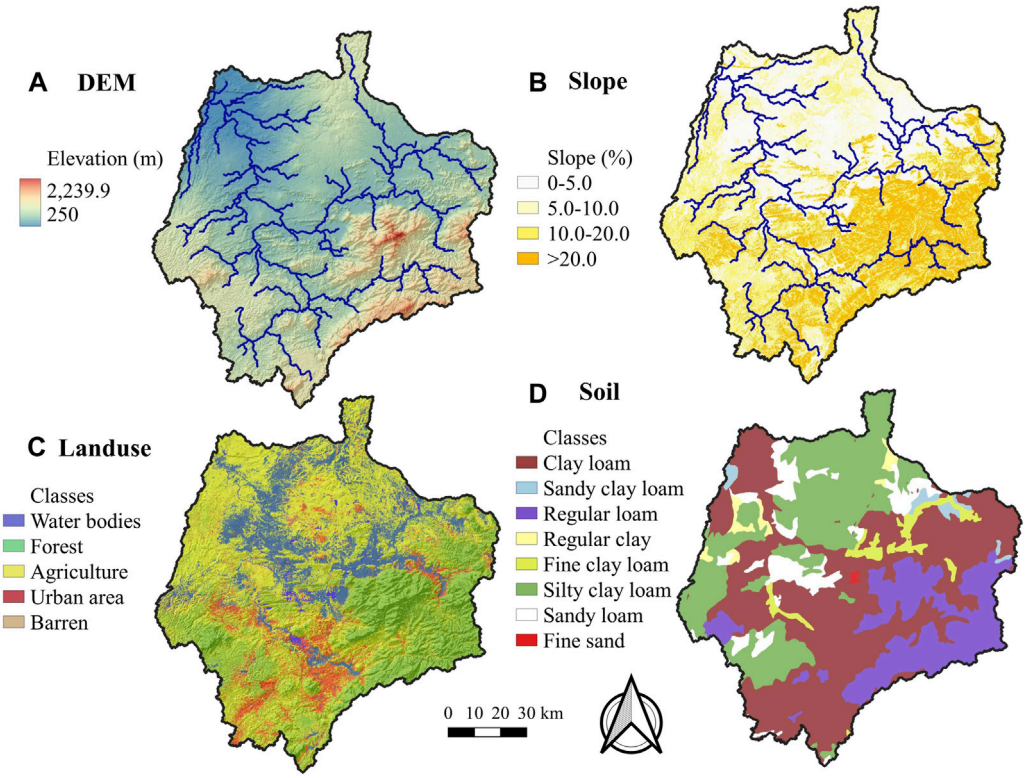


FIGURE 2
The SRB with **(A)** DEM (m) **(B)** Slope (%) **(C)** LULC, and **(D)** Soil map.

TABLE 1 Description of required inputs for SWAT in this study.

No	Name	Description	References
1	DEM	The 90-m DEM from the HydroSHEDS database with an average error of less than 3%	HydroSHEDS Core layers (V1.0) (https://www.hydrosheds.org/products/hydrosheds)
2	LULC	30-m LULC map was retrieved from SERVIR-Mekong Portal for the year 2010 (Figure 2C)	SERVIR–Mekong (https://www.landcovermapping.org/en/home/)
3	Soil	30-m resampled soil map with a scale of 1:1,000,000 was used (Figure 2D)	National Institute for Soils and Fertilizers. (2002)
4	Weather data	Daily precipitation data were obtained (1985–2018) at eleven meteorological stations: Giang Son, Buon Me Thuot, Buon Ho, M'DRak, Dak Lak, Krong Buk, Duc Xuyen, Dak Nong, Cau 14, Ban Don, and Ea So (Figure 1B) Daily maximum temperature (T_{max}) and minimum temperature (T_{min}) data were obtained from VMHA at two meteorological stations: Buon Me Thuot, and Dak Lak (Figure 1B)	Vietnam Meteorological and Hydrological Administration (VMHA) (http://kttvqg.gov.vn/)
5	Observation	Observed daily streamflow data were collected (1980–2018) at five stations: Ban Don, Cau 14, Giang Son, Duc Xuyen, and Krong Buk (Figure 1B) for SWAT model calibration and validation	VMHA (http://kttvqg.gov.vn/)
6	Reservoirs	Five dams and reservoirs have been chosen within this study, including: Buon Tua Srah with a capacity of 86 Megawatts (MW); Buon Kuop at 280 MW; Srepok 3 with 220 MW; Srepok 4 at 80 MW; and Srepok 4A at 63 MW.	MRC (2017)

Mekong River (Figures 1A, B) with the total river's length approximately 450 km. The river's initial course covers 169 km within Vietnamese territory before traveling through 281 km in Cambodia. The SRB covers an area of approximately 18,200 km^2 within Vietnam, primarily occupying the upper segment of the basin, which constitutes over 65% of the total area. This watershed is characterized by a diverse topography with the upper region's elevation ranges from 200 to over 2,240 m, with average heights transitioning from 350 m in the northwest to about 1,000 m in the southeast (Figures 2A, B). Between 1998 and 2018, the average annual precipitation recorded at main hydrological stations of SRB (e.g., Giang Son, Duc Xuyen, Cau 14, and Ban Don) were approximately 1920 mm, 1937 mm, 1704 mm, and 1,601 mm, respectively. The majority of this precipitation, exceeding 70%, occurred during the wet season (June to November), with an estimated 41% of the annual precipitation contributing to the basin's runoff. In this study, the Duc Xuyen and Giang Son regions are specifically highlighted due to their significant hydrological contributions to the SRB's outflow (Ty et al., 2011; 2012).

On the other hand, Vietnam exhibits considerable hydropower potential, notably in the northern and central highland regions. In this study, we include five primary hydropower reservoirs and associated power plants subsequent to the initial phase of Hydropower Projects in the Lower MRB, including: Buon Tua Srah, Buon Kuop, Srepok 3, Srepok 4, and Srepok 4A (Figure 1B; Table 1). As mentioned, we incorporates these infrastructures (see section 2.3.2) to assess their implications on the SRB's hydrological dynamics and to evaluate their operational performance under various climate change scenarios.

2.2 Data sets

We used the SWAT hydrological model with the necessary inputs as presented in Table 1.

2.3 Semi-distributed hydrological model SWAT

SWAT is a semi-distributed hydrological model developed and maintained by the U.S. Department of Agriculture (USDA) and the Agricultural Research Service (ARS) (Arnold et al., 2012a). In recent years, SWAT has gained popularity in both the United States and Europe, largely owing to its effectiveness in addressing a range of hydrological issues (Gassman et al., 2007; Tran et al., 2023a). Numerous studies have utilized the SWAT model to examine the impacts of various factors on streamflow and sediment loads (Ahmed et al., 2020). These factors include land use and land cover (LULC) changes (Anaba et al., 2017; Aryal et al., 2022; 2023; Tran and Lakshmi, 2022), the effects of climate change (Vo et al., 2018; Aslam et al., 2022; Shafeeque et al., 2023a; 2023b), improvements in ecosystem services (Ashrafi et al., 2022a; Arshad et al., 2022; Ashrafi et al., 2022b; Tapas et al., 2022), and the validation of satellite-based products (Arshad et al., 2021; Tran et al., 2022a; 2022b; Noor et al., 2023).

2.3.1 Parameter sensitivity analysis, model calibration, and validation

The calibration and validation of the SWAT model, as well as the assessment of parameter uncertainty using the Sequential Uncertainty Fitting procedure (SUFI-2) as the objective function, were carried out with the SWAT-CUP program (version 5.2.1) (https://www.2w2e.com/home/SwatCup), as detailed by Abbaspour et al. (2015) and Zhang et al. (2014). Model performance metrics are presented in Table 2. The objective function (Nash-Sutcliffe efficiency; NSE) was used in SWAT-CUP to calibrate and validate the SWAT model.

Where Q is the streamflow (m^3/s), m, s stand for measured and simulated, and d stands for deviation of it, i is the i^{th} measured and simulated, $\bar{Q}$ indicates the mean value and number of values is n.

A 7-year warm-up period (1985–1991) was chosen within the 33-year simulation period (1985–2018), followed by 10 years (1992–2001) for model calibration and 17 years (2002–2018) for

TABLE 2 Performance metrics for the model calibration and validation. Where *Q* is the streamflow (m^3/s), m, s stand for measured and simulated, and d stands for deviation of it, *i* is the i^{th} measured and simulated, $\bar{Q}$ indicates the mean value and number of values is *n*.

Metric equation	Optimal value	Performance evaluation criteria
$\mathbf{KGE} = \mathbf{1} - \sqrt{(\mathbf{CC}-\mathbf{1})^2 + (\frac{Q_s{}^d}{Q_m{}^d}-\mathbf{1})^2 + (\frac{\bar{Q}_s}{\bar{Q}_m}-\mathbf{1})^2}$	1	VG: KGE ≥ 1; G: 0.50 ≤ KGE ≤ 1 S: 0 ≤ KGE ≤ 0.50; NS: KGE < 0
$\mathbf{NSE} = \mathbf{1} - \frac{\sum_{i=1}^{n}(Q_m-Q_s)^2}{\sum_{i=1}^{n}(Q_m-\bar{Q}_s)^2}$	1	VG: NSE ≥0.8; G: 0.8 ≤ NSE ≤ 0.7 S: 0.5 ≤ NSE ≤ 0.7; NSE: NS ≤ 0.5
$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Q_s-Q_m)^2}{n}}$	0	VG: 0.2 ≤ RMSE ≤0.5 0 ≤ RMSE ≤ +∞; Lower is better
$\mathbf{PBIAS} = \mathbf{100}^*(\frac{\sum_{i=1}^{n}(Q_m-Q_s)}{\sum_{i=1}^{n}Q_m})$	0	VG: PBIAS ≤ ± 5; G: ± 5 ≤ PBIAS ≤ ± 10 S: ± 10 ≤ PBIAS ≤ ± 15; NS ≥ ± 15
$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert Q_m - Q_s\rvert$	0	0 ≤ MAE ≤ +∞; Lower is better
$\mathbf{R^2} = \frac{[\sum_i (Q_{m,i}-\bar{Q_m})(Q_{s,i}-\bar{Q_s})]^2}{\sum_i (Q_{m,i}-\bar{Q_m})^2 \sum_i (Q_{s,i}-\bar{Q_s})^2}$	1	VG: 1 ≥ R^2 ≥0.75; G: 0.65 ≤ R^2 ≤ 0.75 S: 0.5 0 ≤ R^2 ≤ 0.65; R^2: NS ≤ 0.5

validation. The calibration and validation of the model were performed on a daily basis, with the ideal range for the number of iterations set between 300 and 500 to balance computational efficiency and time requirements.

2.3.2 Reservoirs scenarios

This work initially focuses on the combined impacts of certain reservoirs—Buon Tua Srah, Buon Kuop, Srepok 3, and Srepok 4—on the SRB (Figure 1B). These reservoirs, previously overlooked in related studies, are included here due to their importance in assessing climate change effects. We will examine the operational impacts of each reservoir, particularly their influence on flood peaks and total runoff at the SRB outlet. To be specific, we compare a baseline scenario (natural flow without reservoirs) against four scenarios where each reservoir is independently operated, to isolate their individual impacts for better management and operational planning. The Srepok 4A reservoir, however, will not be considered in this study due to its limited capacity and primary function of serving irrigation needs for a small area, as reported by MRC (2017).

For the reservoir setup in the SWAT model, we selected the following parameters: the operational start date of the reservoir (MORES for month and IYRES for year), the reservoir's surface area at emergency spillway capacity (RES_ESA; ha), and the corresponding volume (RES_EVOL; 10^4 m^3). Additional parameters include the surface area at principal spillway capacity (RES_PSA; ha), the required volume to reach this capacity (RES_PVOL; 10^4 m^3), the initial volume of water in the reservoir (RES_VOL; 10^4 m^3), the initial sediment concentration (RES_SED; mg/L), evaporation coefficient (EVRSV), and the average daily discharge when overflowing (RES_RR; m^3/s). We also considered the non-flood season duration (IFLOD1R and IFLOD2R for start and end months, respectively), the fraction of water removed from the reservoir during non-flood season (WURTNF; m^3/s), the minimum outflow relative to principal spillway volume (OFLOWMN_FPS), and the target storage volume also relative to the principal spillway (STARG_FPS) (Arnold et al., 2012a).

Two parameters were identified as most sensitive for reservoir calibration: the hydraulic conductivity at the reservoir bottom (RES_K; mm/h) and the number of days required to reach the target storage from the current volume (NDTARGR; days). These were calibrated using methodologies described by Kim and Parajuli (2014) and Qiu et al. (2019). In addition, twenty-three parameters were chosen for model calibration, validation, and sensitivity analysis, based on *p*-value and t-Stat statistical methods as mentioned in Arnold et al. (2012b), Tuo et al. (2016), and Xu et al. (2016). These parameters' descriptions and their calibrated values are listed in Table 4.

In this study, SWAT's operating policies are based on monthly storage targets. The storage target operations are piecewise linear functions for each reservoir in the system. Each reservoir's policy is defined by parameters: a target storage value for each month (STARG) and NDTARGR (days), which are constant across all months. Daily reservoir releases ($V_{flowout}$) are calculated as follows:

$$V_{flowout} = \frac{V - V_{targ}}{ND_{targ}}$$

Where V is the volume of water stored in the reservoir, V_{targ} is the reservoir storage target, and ND_{targ} is the number of days for the reservoir to reach the target storage (Arnold et al., 2012b). Once the outflow is determined with this method, the model then adjusts the outflow to ensure that minimum and maximum discharge criteria are met. Existing target storage operations are not fully state-aware: each reservoir's release decision is based only on its own storage level and the month but does not consider any additional system state information, such as the storage at other reservoirs in the system.

2.4 Climate change projections

Future climate data were extracted from the Vietnam Climate Change Scenario dataset, which was validated by the Vietnamese Government under resolution No. 93/NQ-CP, issued on 31 October 2016. This resolution affirmed Vietnam's commitment to the Paris Agreement—a global pact within the United Nations Framework Convention on Climate Change—officially ratified on 20 July 2020. Summaries of the data from six different Regional Climate Models (RCMs), updated using the IPCC 2019 report and the 2018 Vietnam Meteorological datasets (e.g., 5 m resolution lidar DEM), are presented in Table 3. Our analysis employed the

TABLE 3 Description of RCMs, as the inputs of the Vietnam Climate Change Scenario dataset.

Model	Country	Release year	Resolution	References
clWRF	United States	2000s	0.3° x 0.3°	Fita. (2010)
PRECIS	United Kingdom	2004	0.25° x 0.25°	Moberg & Jones. (2004)
CCAM	Australia	2014	0.1° x 0.1°	Her. (2014)
RegCM	Italy	2017	0.2° x 0.2°	NCAR. (2017)
AGCM/MRI	Japan	2012	0.2° x 0.2°	Mizuta et al. (2012)
RCA3	Sweden	2010	0.5° x 0.5°	Samuelsson et al. (2011)

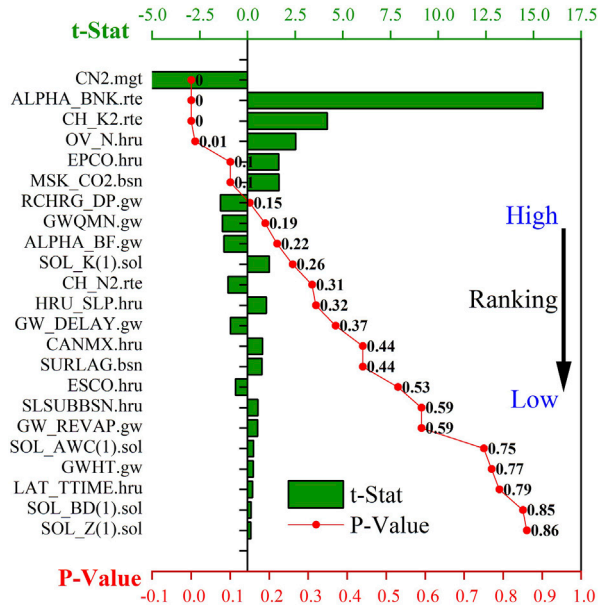


FIGURE 3
Sensitivity analysis using t-Stat and *p*-value for chosen SWAT-CUP parameters with their ranking. Parameters with a *p*-value ≤0.05 and a high t-Stat value—when |t-Stat| ≥ 1.96 were considered sensitive (Mosavi et al., 2021). This threshold indicates that, with 95% confidence, the variables have a significant effect.

RCP4.5 scenario, which projects an average annual rainfall increase of 10%–20%, and the RCP8.5 scenario, which anticipates an average annual rainfall increase of up to 40%. These scenarios correspond to the near future and far future periods, as designated by MONRE (2016).

3 Results

3.1 Model calibration and validation

Figure 3 and Table 4 present the parameter ranking, method, description, range, and fitted value for parameter sensitivity analysis using SWAT model.

The SCS runoff curve number (CN2) was found as the most sensitive parameter, followed by the base flow alpha factor for bank storage (ALPHA_BNK), the effective hydraulic conductivity in the main channel alluvium (CH_K2), Manning's "n" value for overland flow (OV_N), the plant uptake compensation factor (EPCO), the calibration coefficient used to control the impact of the storage time constant from low flow (MSK_CO2), and the deep aquifer percolation fraction (RCHRG_DP) (Figure 3). These results indicate that SRB is significantly sensitive to surface runoff parameters (defined by CN2, CH_K2, and OV_N) and EPCO. This sensitivity can be attributed to the region's dense vegetation cover. The findings are consistent with observations from previous studies by Bajracharya et al. (2018), Bhatta et al. (2019), and Li and Fang (2021). However, groundwater-related parameters, e.g., RCHRG_DP and GW_DELAY, were found to be non-sensitive in our study. This insensitivity may be due to extensive groundwater extraction for agricultural activities, which diminishes the interaction between surface and subsurface waters.

To achieve optimal calibration, we performed the model calibration and validation at multiple objective stations, including Krongbuk, Giang Son, and Duc Xuyen (Figure 1B). Figure 4 compares observed and simulated daily streamflow at A) Krongbuk B) Giang Son C) Duc Xuyen D) Cau 14, and E) Ban Don stations. A detailed summary of the model's performance metrics is presented in Figure 5. We performed these different scenarios using the same model setup and set of calibrating parameters (Table 4). The Krongbuk station, chosen as the objective station for model calibration and validation, showed a good performance, particularly during the validation phase of the second period following the operation of five reservoirs (after 2009) (Ty et al., 2011) (Figure 4A). Our findings thus address a gap in previous analyses, highlighting the capability of the Krongbuk station for model calibration purposes.

Model calibration for the Duc Xuyen region (2009–2018), revealed low values of NSE and KGE (Figure 4C). This could be explained due to a range of local factors, including variations in terrain profile, agricultural practices, and water resource management policies. Specifically, the intensive use of both surface water and groundwater for agricultural purposes in the Duc Xuyen area, as opposed to practices in the Krongbuk region, may have contributed to these results, and aligned with findings from Arias et al. (2014b), Huyen et al. (2017), and Ty et al. (2011, 2012).

Figure 5 shows the calibration and validation scenarios conducted at the Duc Xuyen, Krongbuk, and Giang Son stations. While the scenario performed at Krongbuk provided the most

TABLE 4 Summary of SWAT and reservoir parameters with their sensitivity ranked based on t-Stat values, as described in Mosavi et al. (2021).

Rank	Parameter	Method	Description	Range	Fitted-value
Flow parameters					
1	CN2	Relative	SCS runoff curve number f	35–98	65.07
2	ALPHA_BNK	Replace	Baseflow alpha factor for bank storage	0–1	0.86
3	CH_K2	Replace	Effective hydraulic conductivity in main channel alluvium	−0.01–500	388.14
4	OV_N	Replace	Manning's "n" value for overland flow	0.01–30	0.25
5	EPCO	Replace	Plant uptake compensation factor	0–1	0.31
6	MSK_CO2	Replace	Calibration coefficient used to control impact of the storage time constant from low flow	0–10	6.86
7	RCHRG_DP	Replace	Deep aquifer percolation fraction	0–1	0.30
8	GWQMN	Replace	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	0–5,000	1,403.90
9	ALPHA_BF	Replace	Baseflow alpha factor (days)	0–1	0.57
10	SOL_K	Relative	Saturated hydraulic conductivity	0–2000	19.67
11	CH_N2	Replace	Manning's "n" value for the main channel	−0.01–0.3	0.05
12	HRU_SLP	Relative	Average slope steepness	0–1	0.03
13	GW_DELAY	Replace	Groundwater delay (days)	0–500	51.80
14	CANMX	Replace	Maximum canopy storage	0–100	22.86
15	SURLAG	Replace	Surface runoff lag time	0.05–24	14.83
16	ESCO	Replace	Soil evaporation compensation factor	0–1	0.59
17	SLSUBBSN	Relative	Average slope length	10–150	69.35
18	GW_REVAP	Replace	Groundwater "revap" coefficient	0.02–0.2	0.10
19	SOL_AWC	Relative	Available water capacity of the soil layer	0–1	0.07
20	GWHT	Replace	Initial groundwater height m)	0–25	0.98
21	LAT_TTIME	Replace	Lateral flow travel time	0–180	46.76
22	SOL_BD	Relative	Moist bulk density	0.9–2.5	1.06
23	SOL_Z	Relative	Depth from soil surface to bottom of layer	0–3,500	372.97
Reservoir parameter					
1	NDTARGR	Replace	Number of days the reservoir would be filled (days)	1–365	1–30
2	RES_K	Replace	Hydraulic conductivity of the reservoir bottom (m/s)	10^{-2}–11	0.15

accurate fit for the SRB (Table 2), Giang Son station's performance was slightly less accurate, whereas the Duc Xuyen station had the least precise results. In particular, the Duc Xuyen station showed a range of KGE values from 0.45 to 0.80 across the stations of Ban Don, Cau 14, Duc Xuyen, Giang Son, and Krongbuk (Figure 5A). The least favorable outcomes were observed at the Giang Son station (NSE of 0.21 and RMSE of 96.15 m^3/s), and at the Krongbuk station (NSE of 0.05 and RMSE of 12.84 m^3/s). In addition, the calibration and validation processes at the Krongbuk station indicated improvements, particularly with respect to flood peak predictions and total streamflow volume assessment (Figure 5). The model achieved satisfactory scores here, enhancing its dependability for assessing the effects of RCPs, especially regarding the cumulative impacts from the regional reservoirs.

3.2 Quantify the flow contribution of the sub-basins

The streamflow contributions from the Giang Son and Duc Xuyen regions to the SRB outlet are quantitatively shown in Figure 6. To be specific, the Duc Xuyen region adds an extra 21.1% to the SRB's total runoff compared to the Giang Son region.

The Giang Son region contributes between 32.9% and 39.8% of the total runoff volume at the SRB outlet (Figure 6B). During the wet season, the region is responsible for 30.2% of the total streamflow, which is less than the 43.1% provided by the Duc Xuyen region, underscoring the latter's significant influence during this period (Figure 6C). The scenario changes in the dry season, where the contribution dynamics between the two regions reverse. Figure 6D

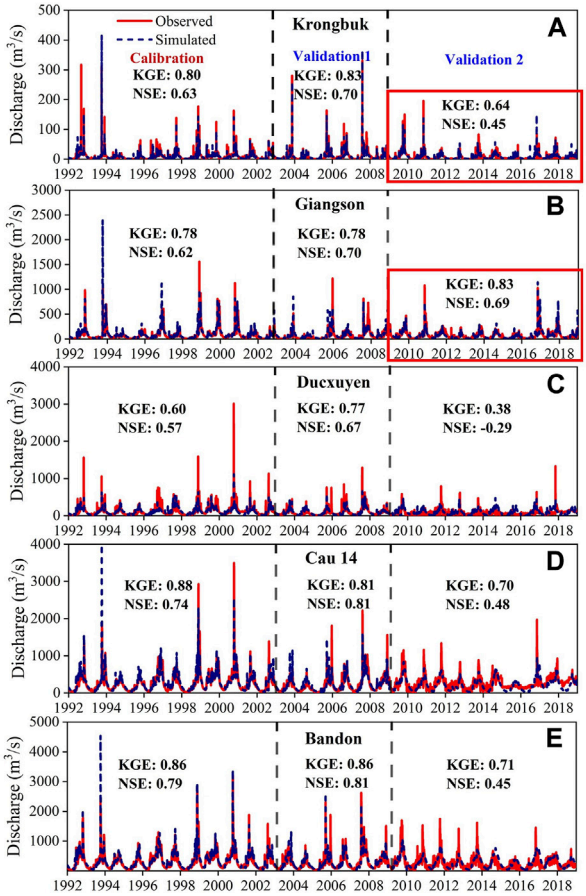


FIGURE 4
Comparisons between observed and simulated streamflow (daily scale) using the objective station–Krongbuk, for calibration period (1992–2002), first validation (without dam) (2003–2008), and second validation (with dam) (2009–2018). The results were shown at **(A)** Krongbuk **(B)** Giang Son **(C)** Duc Xuyen **(D)** Cau 14, and **(E)** Ban Don.

shows that the Giang Son region's share of the total streamflow increases to 39.2%, exceeding the 32.2% from the Duc Xuyen region. This indicates the variable influence of these regions on the SRB's hydrology across different seasons.

Figure 6C shows a notable difference in the annual streamflow patterns corresponding to the period before (pre-2009; baseline scenario without dam) and after (post-2009) the commencement of dam operations. This provides insights into the changing hydrological response that could be explained due to anthropogenic intervention (Ty et al., 2012). These observations underscore the transformative impact of reservoirs on streamflow patterns, with implications for water resource management, flood control, and ecosystem services in the SRB. Besides, the streamflow patterns in the pre-2009 period are characterized by relative stability. However, during the post-2009, when the reservoirs became operational, the runoff demonstrates marked variability, indicating the direct influence of reservoir management on the hydrology of the region. In addition, SRB experienced a noticeable 25% reduction in total runoff in the Ban Don, Cau 14, and Duc Xuyen regions (Figure 6) between 2009 and 2018. This decline is attributed to the active period of dam operation and highlights the substantial role that these structures play in modifying natural streamflow regimes.

3.3 Reservoirs' impacts on streamflow

We conducted a comprehensive assessment to evaluate the influence of individual reservoirs on the cumulative runoff at the SRB outlet (Buon Tua Srah, Buon Kuop, Srepok 3, and Srepok 4) (Figure 1B). For this, we compared the flood peak, annual, and seasonal runoff at the SRB outlet under these experimental scenarios against a baseline scenario in which all reservoirs would be inactive (see section 2.3.2) (Figure 7).

Table 5 presents the impacts of each reservoir on the runoff of the SRB, showing an increasing influence from the downstream to the upstream regions. Notably, the capacity for flood reduction correlates directly with the active storage capacity of the reservoirs. In particular, Srepok 4 demonstrates the most effective capability for mitigating flood peaks in the downstream floodplain, followed by Srepok 3 and Buon Kuop, while Buon Tua Srah exhibits the least effectiveness (Figure 7B). The least performance of Buon Tua Srah can be attributed to its location; despite having a good design capacity, it is situated in the highland and upper region of the SRB (Figure 1B). Furthermore, our results also show that the increased volume of water released during the dry season contributes to a rise in the total runoff of the SRB when compared to the baseline scenario.

Regarding the average annual runoff at the outlet of the SRB (2009–2018), the activation of each individual reservoir corresponds to a respective decrease in annual runoff: 0.26% for Buon Tua Srah, 0.13% for Buon Kuop, 0.24% for Srepok 3, and 0.08% for Srepok 4, when compared to the scenario where all these reservoirs are deactivated. Despite the overall reduction in runoff following reservoir activation, the Buon Tua Srah reservoir exhibits the most significant decrease in total runoff at the outlet of the SRB. While Buon Tua Srah is located upstream, allowing it to capture more discharge from the upper regions of the SRB, Srepok 3 is situated further downstream but possesses the highest design capacity, which allows it to store and subsequently release a substantial volume of water downstream. Further analysis revealed that the impacts of the reservoirs (Buon Kuop, Srepok 3 and 4) are particularly noticeable during the dry and wet seasons, as can be attributed to their locations in the mainstream of the Srepok River (Figure 1B; Figure 7; Table 5). Moreover, we found that reservoirs operated on an annual regulation (Buon Tua Srah, Buon Kuop, and Srepok 3) have a more pronounced and enduring impact on runoff patterns compared to those managed on a daily regulation (Srepok 4) (Table 5), and this finding therefore could be used to better regulate the runoff before and after flood and drought events.

3.4 Future predictions of climate and water changes

3.4.1 Precipitation

Figure 8 shows the rainfall variations throughout the baseline period (1992–2018) and under two future scenarios, RCP 4.5 and 8.5. The peak rainfall occurs from July to October, with September

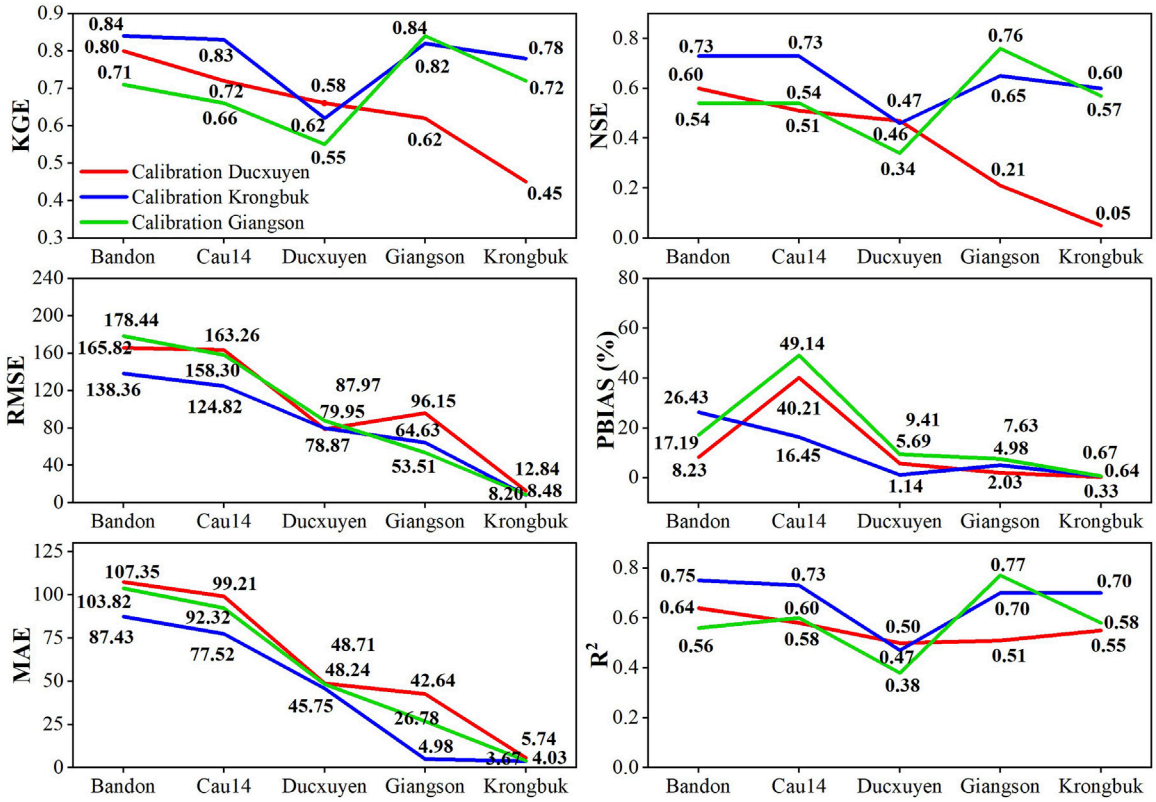


FIGURE 5
Summary of the model's calibration and validation (1992–2018) performed at different objective stations (Duc Xuyen, Krongbuk, and Giang Son).

experiencing the highest volume. During the wet season (June to November) (Nguyen and Nguyen, 2004), 63% of the annual rainfall is recorded, whereas the dry season sees a significant drop, contributing only 58% (243.45 mm) of the total accumulative volume (Figure 8A). Between 2046 and 2065, a slight increase in rainfall is projected, accounting for 3.39 mm (RCP 4.5) and 3.88 mm (RCP 8.5), compared to the baseline period (Figure 8A). In general, an upward trend rise in the annual rainfall (2080–2099) with approximately of 6.2%–7.1% increase for the 2046–2065 period and 7.7%–11.5% increase for the 2080–2099 period in rainfall compared to the historical record. We also found that these scenarios suggest average monthly rainfall increases of at least 4.21 mm–6.28 mm with the most significance expected during August and September (Figures 8B, C).

3.4.2 Temperature

Figure 8 and Figure 9 show projected increases in maximum (T_{max}), minimum (T_{min}), and average (T_{mean}) temperatures under future RCPs. The baseline data reveals the warmest months to be between March and May, with temperatures decreasing during the wet season and reaching their lowest in October (Figure 9D; Figure 9E; Figure 9F). Under the future scenarios RCP 4.5 and RCP 8.5, modest temperature rises of 1.4 °C and 1.8 °C are expected, respectively, with RCP 8.5 presenting a more pronounced increase (Figure 9E, Figure 9F, Figure 9H, and Figure 9I). The trends suggest that seasonal and annual average temperatures will continue to increase, with the period from 2080 to 2099 under RCP 8.5 anticipating the highest temperature increase.

3.4.3 Climate change impacts on streamflow

The two RCPs predict an increase in streamflow across monthly, annual, and seasonal scales. For the period between 2046 and 2065, an increase in annual streamflow is anticipated, with a range of 9.14% for RCP 4.5%–9.8% for RCP 8.5 (Figure 10). A more significant jump of 2.1%–3.2% above these levels is projected for the later period of 2080–2099. The largest streamflow increases are expected from May to September compared to the baseline scenario, with the smallest rises occurring at the start of the year in January and February. These range from an increase of 5.1% under RCP 4.5 (2046–2065) to 7.74% under RCP 8.5 (2080–2099) (Figures 10A, B). The projected average streamflow increase for 2046–2065 is approximately 29.77 m^3/s, corresponding to an increase of about 10.7% for both RCP 4.5 and RCP 8.5 from the baseline scenario (Figure 10A). Between 2080 and 2099, the rise is found at 12.10% (35.78 m^3/s) for RCP 4.5% and 15.97% (47.21 m^3/s) for RCP 8.5 (Figure 10B).

Figures 10C–E show variations in annual and seasonal streamflow, highlighting the SRB's sensitivity to seasonal shifts under different RCPs. The wet season exhibits the most notable increases in streamflow, with rises of 10.13% (RCP 4.5) and 10.92% (RCP 8.5) between 2046 and 2065. These increases are projected to grow to 12.36% and 16.8% for the respective RCPs (2080–2099)

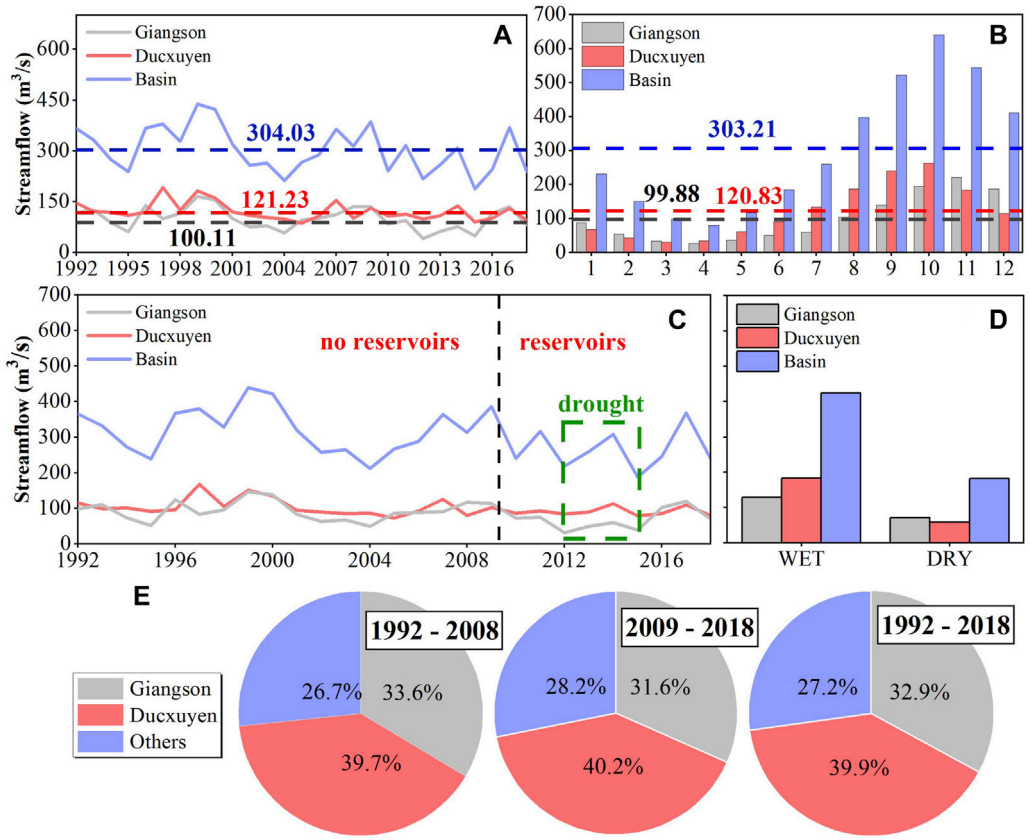


FIGURE 6
Streamflow contributions from different regions to the SRB outlet, categorized as **(A)** annual **(B)** monthly **(C)** before and after reservoirs operated **(D)** seasons, and **(E)** among different periods.

compared to the baseline scenario (Figure 10D). Conversely, the dry season's streamflow increase is more modest, at 6.81% and 7.1% for RCP 4.5 and RCP 8.5, respectively (2046–2065), and 8.3% and 10.5% (2080–2099), respectively (Figure 10E). These changes are confirmed by the monthly streamflow trends (Figures 10A, B) and align with the anticipated alterations in wet season rainfall resulted from the influences of RCP scenarios. Furthermore, average annual streamflow is projected to peak (2080–2099) under the RCP 8.5 scenario (14.9%), with the least significant rises noted at 9.2% under RCP 4.5% and 9.8% under RCP 8.5 during the 2046–2065 period (Figure 10C).

Figures 10F, G present a statistical analysis of hydrological components impacted by the RCP scenarios. Key hydrological elements, e.g., potential evapotranspiration (PET), actual evapotranspiration (ET), and groundwater levels, are all projected to trend upwards, though the differences between the RCP scenarios are not pronounced. However, the RCP 8.5 is anticipated to drive more considerable increases or decreases in streamflow and related hydrological effects.

4 Discussion

We noticed that the typical historical severe drought in the Giang Son region (2012–2013), which resulted in a large disparity between the Duc Xuyen and Giang Son regions, highlights the need for robust water conservation and allocation practices to support the agricultural sector, which is vulnerable to water scarcity during such events. Specifically, while streamflow in the Duc Xuyen region was more stable across drought periods, streamflow in the Giang Son region varied dramatically. This shows the Giang Son region's vulnerability to drought and indicates the critical need to focus on water resource management strategies, especially considering the dependency of local agriculture on reliable water supplies.

Since their commissioning (post-2009), the downstream reservoirs—Srepok 4, Srepok 3, and Buon Kuop—have demonstrated a robust capability to mitigate flood peaks in the downstream floodplain. This aligns with the operational design outlined by the MRC (2017), which specifies that mainstream reservoirs are designed with more flood prevention functions compared to upstream reservoirs (e.g., Buon Tua Srah) that are mainly used for irrigation and electricity production. The substantial impact of the Srepok 3 reservoir, with a 0.24% decrease in total runoff at the SRB's outlet, is primarily attributed to its high design capacity, which allows for the release of more water to the downstream region (Table 5). Conversely, the impact of the Buon Tua Srah reservoir (0.26% decrease in total runoff) is attributed to its location in the upstream region, which helps to collect more discharge in the upper part of the SRB.

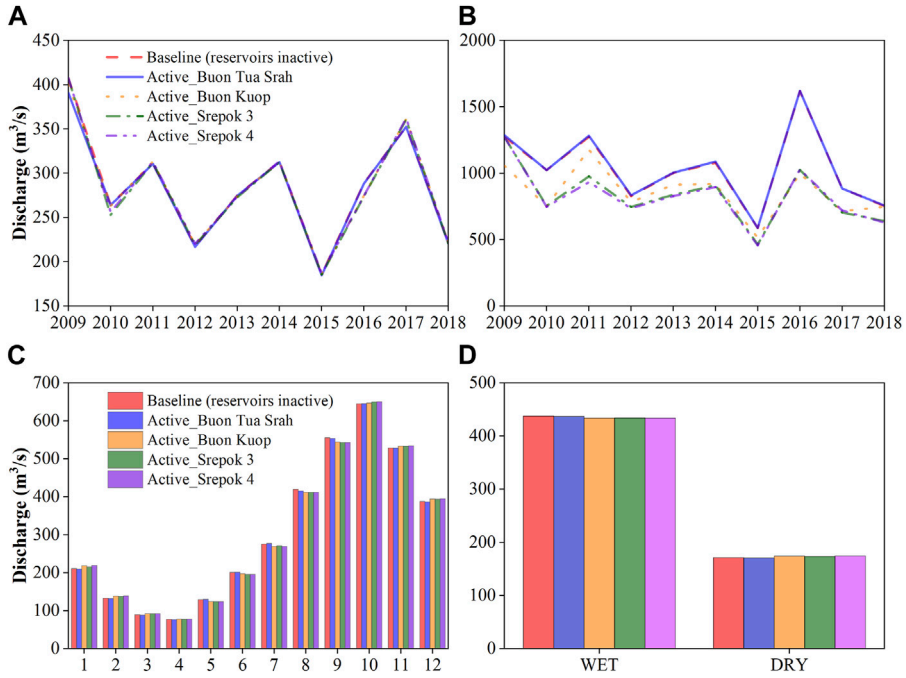


FIGURE 7
Discharge in **(A)** annual **(B)** flood peak **(C)** monthly, and **(D)** seasonal (wet, dry). Each scenario is performed with the reservoir's name, shown in the legend, corresponding to when it is activated.

TABLE 5 The reservoir's impact on streamflow at Duc Xuyen, Cau 14, and Ban Don stations, compared to the baseline (all reservoirs inactive) (2009–2018) on seasonal and annual scales.

Reservoir	Duc xuyen			Cau 14			Ban don		
	Dry season (%)	Wet season (%)	Annual (%)	Dry season (%)	Wet season (%)	Annual (%)	Dry season (%)	Wet season (%)	Annual (%)
Buon Tua Srah	−1.37	−0.42	−0.65	−0.57	−0.22	−0.32	−0.48	−0.17	−0.26
Buon Kuop	-	-	-	1.84	−0.97	−0.15	1.70	−0.84	−0.13
Srepok 3	-	-	-	-	-	-	1.20	−0.80	−0.24
Srepok 4	-	-	-	-	-	-	1.83	−0.83	−0.08

An upward trend in future climate projections indicates an approximate increase of 6.2%–7.1% for the 2046–2065 period and 7.7%–11.5% for the 2080–2099 period in rainfall, leading to an increase in the average annual streamflow by around 9.2%–11.2% and 9.8%–14.9%, under the RCP 4.5 and 8.5, respectively (Figures 8A–C, Figures 9A–C, and Figures 10A–E). To answer whether climate change impacts or reservoir operations will significantly affect the SRB's outlet, we found that it depends on several factors, which require further consideration. First, the regulations of the reservoirs presented in this study are set as constant due to the limitations of SWAT configurations for reservoir module, which can only accept a fixed setup for the entire simulation (Arnold J. et al., 2012). This thus does not reflect the actual operation of these reservoirs in reality, which vary according to different factors (e.g., agricultural activities, irrigation, groundwater extraction, local usage, *etc.*). Furthermore, when considering the impacts of climate change and the role of reservoirs, it is clear that while the current chosen reservoirs have decreased flood intensity, however, the regular operation of reservoirs does not fully reflect each reservoir's maximum flood prevention capacity. Additionally, the limitations arising from the SWAT configuration for the reservoir module mean that the findings in this study should be taken as a reference for future works using more complex numerical models, e.g., the Variable Infiltration Capacity (VIC) model (Liang et al., 1994) with more detailed experiments performed.

The two greenhouse gas emission scenarios (RCP 4.5 and 8.5) predict an average increase in precipitation, with the most significant rise expected in August and September (Figures 8B, C). The SRB is projected to experience the largest increase in annual streamflow volumes during the 2080–2099 period under the RCP 8.5 scenario. The smallest increase is anticipated under the RCP 4.5 scenario between 2046 and 2065, with the greatest increase

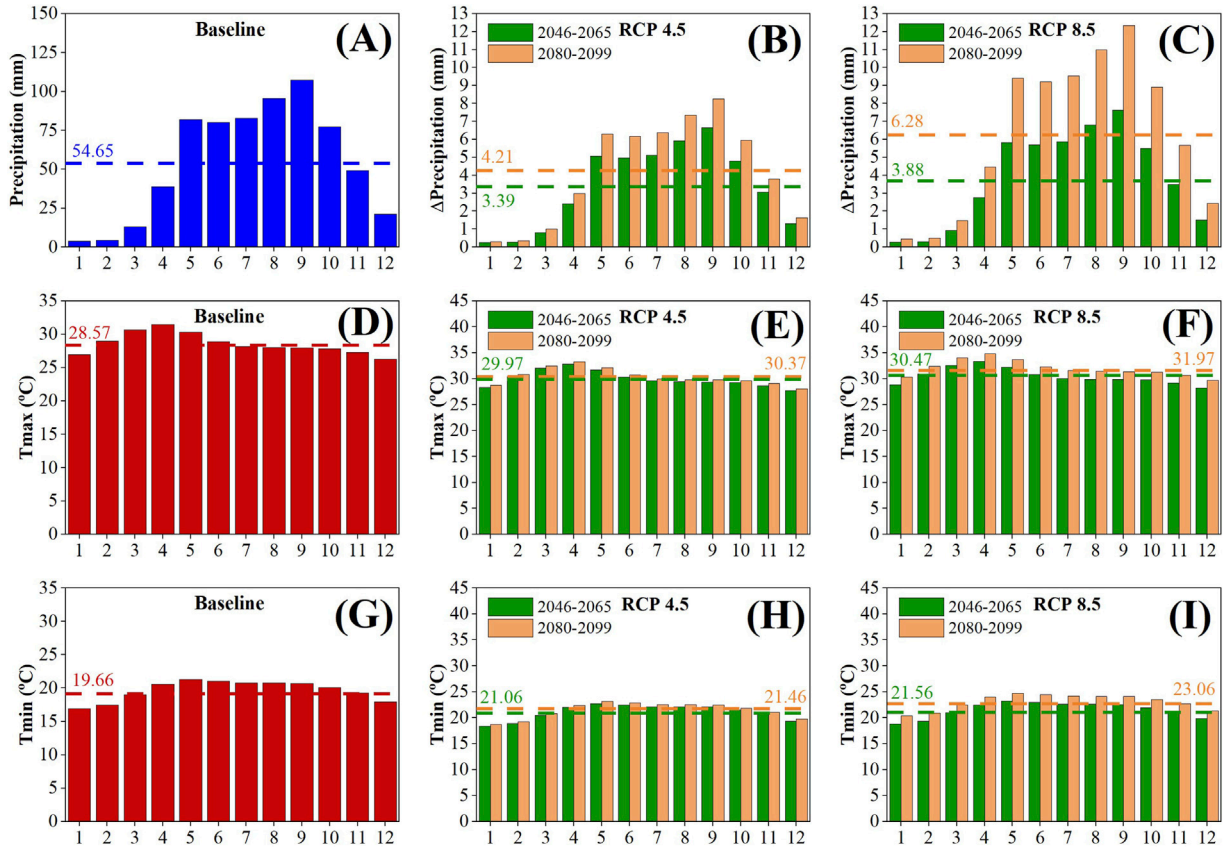


FIGURE 8
Baseline scenario for **(A)** Precipitation **(D)** T_{max} **(G)** T_{min}, and changes in monthly precipitation for **(B)** RCP4.5 **(C)** RCP8.5, monthly maximum temperature for **(E)** RCP4.5 **(F)** RCP8.5, and monthly minimum temperature for **(H)** RCP4.5, and **(I)** RCP8.5 in the 2046–2065 and 2080–2099 compared to the baseline (1992–2018). Dash line represents the mean value.

projected under RCP 8.5 from 2080 to 2099 (Figure 10C). A higher greenhouse gas emission trajectory correlates with greater increases in weather intensity and hydrological responses in terms of volume. However, we also noted that the CMIP5 models may not fully capture the actual climate characteristics due to the neglect of socioeconomic activities and environmental factors that could influence the reliability of the models' outcomes (Bourdeau-Goulet and Hassanzadeh, 2021; Chen et al., 2021). Our suggestion is to use the latest version of the CMIP model (CMIP6) which could help to decrease the uncertainty resulted from these mentioned factors.

Furthermore, various hydrological components, e.g., potential evapotranspiration (PET), actual evapotranspiration (ET), and groundwater levels, are anticipated to follow an upward trajectory. While the differences between the RCP scenarios are not markedly significant, the more intense RCP 8.5 is expected to lead to more pronounced fluctuations in streamflow and hydrological outcomes (Figures 10F, G).

In summary, the insights resulted from this study serve as an important scientific basis for stakeholders and authorities involved in water resource management and climate adaptation strategies in SRB. Strategic planning can involve the construction of additional water storage facilities or the implementation of water-saving technologies in agricultural practices. The efficacy of downstream reservoirs, e.g., Srepok 4, Srepok 3, and Buon Kuop, in mitigating flood peaks presents a case for continuing and expanding such infrastructure, particularly in the context of future projected increases in terms of precipitation and temperature. These infrastructures are not only critical for flood control and prevention but also for maintaining steady water supplies during dry periods. The clear advantage of downstream reservoirs in flood mitigation due to their design and capacity should guide future reservoir projects to maximize flood control benefits. The change in projected future precipitation and temperatures highlights the urgent need for updating regional flood risk maps, enhancing early warning systems, and preparing for more extreme weather events. Conversely, the increase in evapotranspiration and the fluctuations in groundwater levels will demand more complex agricultural and urban planning to ensure water sustainability.

Our findings suggest that proactive measures, informed by the predicted hydrological changes under various climate scenarios, are crucial for ensuring the resilience of the SRB and the communities that depend on its resources. Authorities are advised to integrate these findings into long-term planning and operational decision-making to mitigate the adverse impacts of climate variability and secure water for all uses.

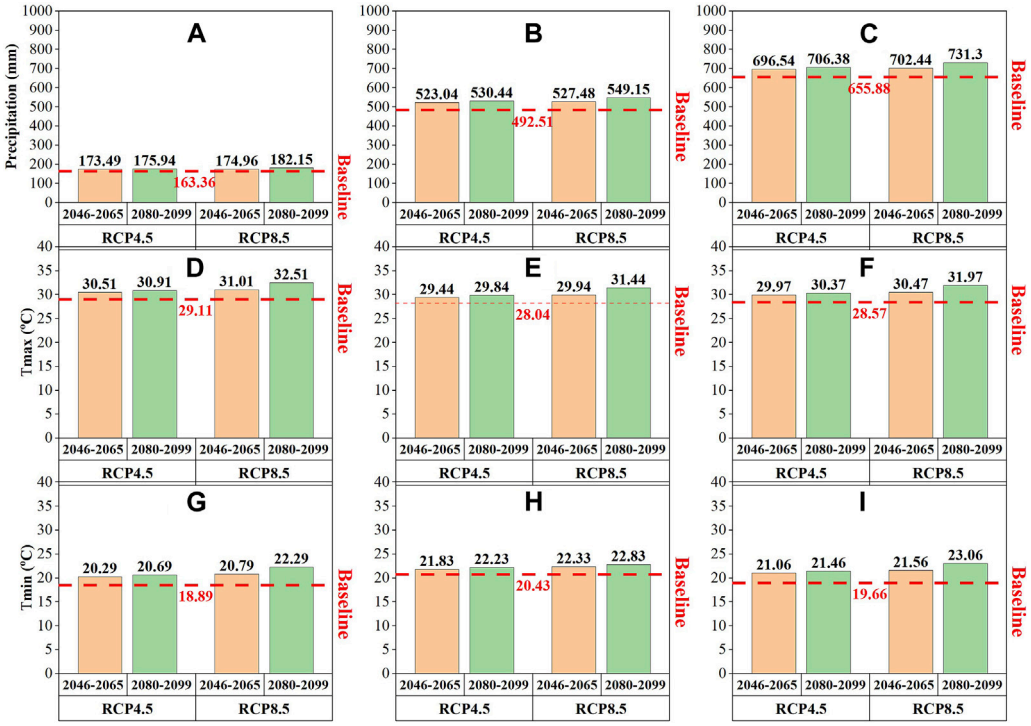


FIGURE 9
Mean changes in the dry season for **(A)** Precipitation **(D)** T_{max}, and **(G)** T_{min}, the wet season for **(B)** Precipitation **(E)** T_{max}, and **(H)** T_{min}, and the annual for **(C)** Precipitation **(F)** T_{max}, and **(I)** T_{min} in **(F)** 2046–2065 and **(G)** 2080–2099 compared to the baseline (1992–2018). Dash line represents the mean value.

5 Conclusion

In this study, we examined both the individual and combined effects of future climate change and reservoir operations on the SRB, with a focus on flood peaks and variations in seasonal and annual runoff for the near (2045–2065) and far future (2080–2099) periods. Additionally, we introduced a novel calibration approach that includes the Krongbuk region, which had been overlooked in previous research. Our key findings include:

(1) The wet season accounts for 63% of the annual precipitation, while the dry season receives 58% less precipitation. Future projections suggest an average annual increase in precipitation of about 6.2%–7.1% in the near future and up to 11.5% in the far future. Meanwhile, a maximum temperature increase of 1.8 °C is expected under the highest emission scenario compared to the historical period.

(2) Future average monthly streamflow is projected to increase by at least 10.7% (2046–2065) and could reach a maximum increase of 15.97% under the highest greenhouse gas emissions scenario (2080–2099). Variability in annual and seasonal streamflow has been indicated, highlighting the SRB's high sensitivity to seasonal shifts under varying future climate conditions. The average annual streamflow is projected to increase by 9.2% in the near future under the RCP 4.5 scenario, with an additional increase of 0.6% expected in the far future. Under the RCP 8.5 scenario, the near future could see an increase of 11.2%, with the far future anticipating a rise of 14.2%. Key hydrological components, including potential evapotranspiration, actual evapotranspiration, and groundwater levels, are also expected to rise.

(3) Climate change is predicted to have a more substantial impact on downstream streamflow than dam operations, potentially leading to significant changes in flow regimes, especially in the downstream region of the SRB. Srepok 4, Srepok 3, and Buon Kuop have demonstrated superior capabilities for mitigating flood peaks in the downstream floodplain, whereas Buon Tua Srah has shown the least effectiveness. Regarding seasonal discharge, Buon Kuop, Srepok 3, and 4 significantly influence the flow, while reservoirs operated on an annual regulation (including Buon Tua Srah, Buon Kuop, and Srepok 3) have a more pronounced and sustained impact on runoff patterns compared to those managed on a daily regulation, such as Srepok 4.

Overall, although reservoirs play a role in mitigating the impacts of climate change within the study period, the influence of climate change is expected to significantly intensify extreme hydrological events, e.g., floods and droughts, especially in critical transboundary river basins such as the SRB. These findings are crucial for implementing effective water resource management strategies in the SRB and similar basins within the MRB, as well as for increasing local awareness of water resource utilization in agriculture.

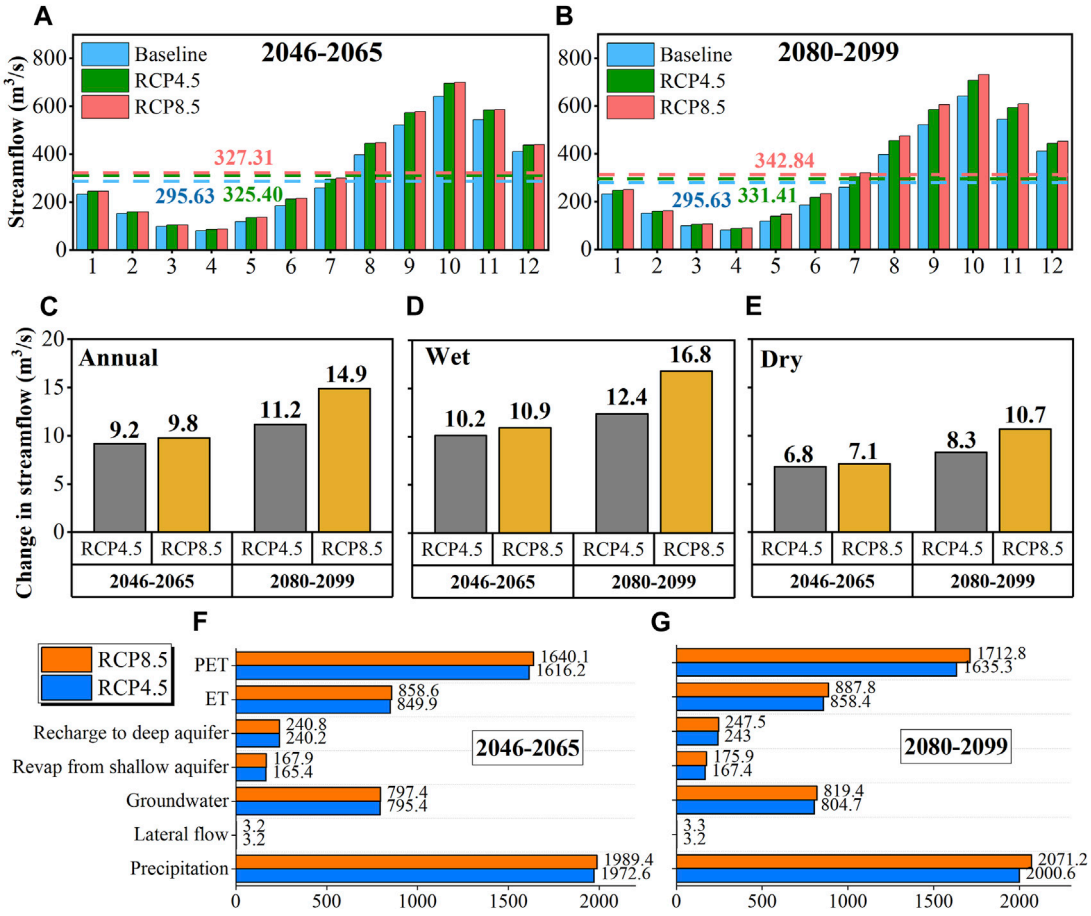


FIGURE 10
Changes in streamflow for **(A)** monthly (2046–2065) **(B)** monthly (2080–2099) **(C)** annual **(D)** wet season **(E)** dry season, and hydrological components during **(F)** 2046–2065, and **(G)** 2080–2099.

6 Limitations of the study

We recognize that this study had certain limitations, primarily due to the unavailability of daily scale dam operation data as well as the poor performance of the SWAT reservoir module that can be used only with constant setup for long-term simulation. This resulted in uncertainties related to the dam's operation during the SRB study periods and during extreme events. Besides, we did not consider the effects of deforestation caused by constructing these reservoirs that could affect the flow regimes. In addition, the climate change scenarios should be updated to CMIP6, which contains socioeconomic and human impacts. Besides, previous studies in this region have compared satellite-based precipitation with *in-situ* rain-gauge observations (Mohammed et al., 2018a; 2018b; Le et al., 2020); downscaled soil moisture (Dandridge et al., 2020); estimation of flooding using satellite data sets (Fayne et al., 2017) and droughts (Lakshmi et al., 2023); combination with socio-economic data (Tiwari et al., 2023); and land-use land cover (Spruce et al., 2020; Nguyen et al., 2022). Thus, it is necessary to implement these data sets to provide more precise results. Such improvements would improve the understanding of city and country officials concerning water resource management and hazard mitigation in the SRB and MRB.

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

T-N-DT: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. BN: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing–review and editing. MG-L: Funding acquisition, Project administration, Supervision, Writing–review and editing. GS: Funding acquisition, Project administration, Supervision, Writing–review and editing. VL: Project administration, Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study is funded by the Faculty of Building Services, Hydro and Environmental Engineering, and Warsaw University of Technology, Warszawa, Poland, under the representative of MG-Ł (Grant ID: VAT 525-000-58-34).

Acknowledgments

We also would like to express our sincere gratitude to reviewers for their constructive comments to improve the quality of our manuscript. Special acknowledgment also goes to reviewers for their constructive comments and support.

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

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., and Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. *J. Hydrology* 524, 733–752. doi:10.1016/j.jhydrol.2015.03.027

Ahmed, Z., Tran, T. N. D., and Nguyen, Q. B. (2020). "Applying semi distribution hydrological model SWAT to assess hydrological regime in Lai Giang catchment, Binh Dinh Province, Vietnam," in Proceedings of the 2nd Conference on Sustainability in Civil Engineering (CSCE'20), Islamabad, Pakistan (Pakistan: Capital University of Science and Technology). https://csce.cust.edu.pk/archive/20-404.pdf.

Anaba, L. A., Banadda, N., Kiggundu, N., Wanyama, J., Engel, B., and Moriasi, D. (2017). Application of SWAT to assess the effects of land use change in the murchison bay catchment in Uganda. *Comput. Water, Energy, Environ. Eng.* 06 (01), 24–40. doi:10.4236/cweee.2017.61003

Arias, M. E., Cochrane, T. A., Kummu, M., Lauri, H., Holtgrieve, G. W., Koponen, J., et al. (2014a). Impacts of hydropower and climate change on drivers of ecological productivity of Southeast Asia's most important wetland. *Ecol. Model.* 272, 252–263. doi:10.1016/j.ecolmodel.2013.10.015

Arias, M. E., Piman, T., Lauri, H., Cochrane, T. A., and Kummu, M. (2014b). Dams on Mekong tributaries as significant contributors of hydrological alterations to the Tonle Sap Floodplain in Cambodia. *Hydrology Earth Syst. Sci.* 18 (12), 5303–5315. doi:10.5194/hess-18-5303-2014

Arnold, J., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E., and Neitsch, S. (2012a). *Swat input data*. Res (Chapter 29). 393–406 https://swat.tamu.edu/docs/.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., et al. (2012b). SWAT: model use, calibration, and validation. *Trans. ASABE* 55 (4), 1491–1508. doi:10.13031/2013.42256

Arshad, A., Mirchi, A., Samimi, M., and Ahmad, B. (2022). Combining downscaled-GRACE data with SWAT to improve the estimation of groundwater storage and depletion variations in the Irrigated Indus Basin (IIB). *Sci. Total Environ.* 838 (April), 156044. doi:10.1016/j.scitotenv.2022.156044

Arshad, A., Zhang, W., Zhang, Z., Wang, S., Zhang, B., Jehanzeb, M., et al. (2021). Reconstructing high-resolution gridded precipitation data using an improved downscaling approach over the high altitude mountain regions of Upper Indus Basin (UIB). *Sci. Total Environ.* 784, 147140. doi:10.1016/j.scitotenv.2021.147140

Aryal, A., Shrestha, S., and Babel, M. S. (2019). Quantifying the sources of uncertainty in an ensemble of hydrological climate-impact projections. *Theor. Appl. Climatol.* 135 (1–2), 193–209. doi:10.1007/s00704-017-2359-3

Aryal, A., Tran, T. N. D., Kim, K. Y., Rajaram, H., Lakshmi, V., and Venkat), (2022). *Climate and land use/land cover change impacts on hydrological processes in the mountain watershed of gandaki River Basin, Nepal.* Chicago: AGU Fall Meeting Abstracts, H52L.0615

Aryal, A., Tran, T. N. D., Kumar, B., and Lakshmi, V. (2023). Evaluation of satellite-derived precipitation products for streamflow simulation of a mountainous himalayan watershed: a study of myagdi khola in kali gandaki. *Remote Sens.* 15(19), 47–62. doi:10.3390/rs15194762

Ashrafi, S., Kerachian, R., Pourmoghim, P., Behboudian, M., and Motlaghzadeh, K. (2022a). Evaluating and improving the sustainability of ecosystem services in river basins under climate change. *Sci. Total Environ.* 806, 150702. doi:10.1016/j.scitotenv.2021.150702

Ashrafi, S., Khoie, M. M. M., Kerachian, R., and Shafiee-Jood, M. (2022b). Managing basin-wide ecosystem services using the bankruptcy theory. *Sci. Total Environ.* 842 (June), 156845. doi:10.1016/j.scitotenv.2022.156845

Aslam, R. A., Shrestha, S., Usman, M. N., Khan, S. N., Ali, S., Sharif, M. S., et al. (2022). Integrated SWAT-MODFLOW modeling-based groundwater adaptation policy guidelines for lahore, Pakistan under projected climate change, and human development scenarios. *Atmosphere* 13 (12), 2001. doi:10.3390/atmos13122001

Bajracharya, A. R., Bajracharya, S. R., Shrestha, A. B., and Maharjan, S. B. (2018). Climate change impact assessment on the hydrological regime of the Kaligandaki Basin, Nepal. *Sci. Total Environ.* 625, 837–848. doi:10.1016/j.scitotenv.2017.12.332

Bhatta, B., Shrestha, S., Shrestha, P. K., and Talchabhadel, R. (2019). Evaluation and application of a SWAT model to assess the climate change impact on the hydrology of the Himalayan River Basin. *Catena* 181, 104082. doi:10.1016/j.catena.2019.104082

Bolch, T., Kulkarni, A., Kääb, A., Huggel, C., Paul, F., Cogley, J. G., et al. (2012). The state and fate of himalayan glaciers. *Science* 336 (6079), 310–314. doi:10.1126/science.1215828

Bourdeau-Goulet, S. C., and Hassanzadeh, E. (2021). Comparisons between CMIP5 and CMIP6 models: simulations of climate indices influencing food security, infrastructure resilience, and human health in Canada. *Earth's Future* 9 (5). doi:10.1029/2021EF001995

Bui, T. T. P., Kantoush, S., Kawamura, A., Du, T. L. T., Bui, N. T., Capell, R., et al. (2023). Reservoir operation impacts on streamflow and sediment dynamics in the transboundary river basin, Vietnam. *Hydrol. Process.* 37 (9). doi:10.1002/hyp.14994

Chang, C., Lee, H., Do, S. K., Du, T. L. T., Markert, K., Hossain, F., et al. (2023). Operational forecasting inundation extents using REOF analysis (FIER) over lower Mekong and its potential economic impact on agriculture. *Environ. Model. Softw.* 162 (August 2022), 105643. doi:10.1016/j.envsoft.2023.105643

Chattopadhyay, S., Edwards, D. R., Yu, Y., and Hamidisepehr, A. (2017). An assessment of climate change impacts on future water availability and droughts in the Kentucky River Basin. *Environ. Process.* 4 (3), 477–507. doi:10.1007/s40710-017-0259-2

Chen, C. A., Hsu, H. H., and Liang, H. C. (2021). Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the seasonal extreme precipitation in the Western North Pacific and East Asia. *Weather Clim. Extrem.* 31, 100303. doi:10.1016/j.wace.2021.100303

Cochrane, T. A., Arias, M. E., and Piman, T. (2014). Historical impact of water infrastructure on water levels of the Mekong River and the Tonle Sap system. *Hydrology Earth Syst. Sci.* 18 (11), 4529–4541. doi:10.5194/hess-18-4529-2014

Dandridge, C., Fang, B., and Lakshmi, V. (2020). Downscaling of SMAP soil moisture in the lower Mekong River Basin. *Water* 12, 56. doi:10.3390/w12010056

Du, T. L. T., Lee, H., Bui, D. D., Graham, L. P., Darby, S. D., Pechlivanidis, I. G., et al. (2022). Streamflow prediction in highly regulated, transboundary watersheds using multi-basin modeling and remote sensing imagery. *Water Resour. Res.* 58 (3), e2021WR031191–25. doi:10.1029/2021WR031191

Fayne, J., Bolten, J., Fuhrmann, S., Rice, M., Houser, P., and Lakshmi, V. (2017). Flood mapping in the lower mekong basin using MODIS observations. *Int. J. Remote Sens.* doi:10.1080/01431161.2017.1285503

Fita, L., Fernández, J., Fernández, F., and García-Díez, M. (2010). *CLWRF: WRF modifications for regional climate simulation under future scenarios*. http://www.meteo.unican.es/wiki/cordexwrf.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). The soil and water assessment Tool: historical development, applications, and future research directions. *Trans. ASABE* 50 (4), 1211–1250. doi:10.13031/2013.23637

Giang, N. T., Thao, H. T., Vinh, T. N., Duy, P., Binh, H., and Quan, V. D. (2017). Study on sediment regime changes in downstream of the Ba River, Vietnam under the impact of reservoirs system. *VNU J. Sci. Earth Environ. Sci.* 4, 127–134. doi:10.25073/2588-1094/vnuees.4217

Giang, P. Q., and Vy, T. T. (2021). Will climate change exacerbate the economic damage of flood to agricultural production? A case study of rice in ha tinh Province, vietnam. *Front. Environ. Sci.* 9 (June), 1–13. doi:10.3389/fenvs.2021.643947

Her, K., and Hawkins, E. (2014). Climate projections. *Taiwan Rev.* 64 (11), 38–60. doi:10.1079/9781780641973.0038

Hussain, A., Zaib, K., Khalil, J., Rahman, U., Shang, S., Ejaz, N., et al. (2022). Analyzing the impact of drought on agriculture: evidence from Pakistan using standardized precipitation evapotranspiration index. *Nat. Hazards* 115, 389–408. doi:10.1007/s11069-022-05559-6

Huyen, N. T., Tu, L. H., Tram, V. N. Q., Minh, D. N., Liem, N. D., and Loi, N. K. (2017). Assessing the impacts of climate change on water resources in the srepok watershed, central highland of Vietnam. *J. Water Clim. Change* 8 (3), 524–534. doi:10.2166/wcc.2017.135

International Rivers (2010). *Srepok dams*. https://archive.internationalrivers.org/campaigns/srepok-dams.

IPCC (2019). "Global warming of 1.5°C," in *Special report on global warming of 1.5°C*.

Ivanov, V. Y., Xu, D., Dwelle, M. C., Sargsyan, K., Wright, D. B., Katopodes, N., et al. (2021). Breaking down the computational barriers to real-time urban flood forecasting. *Geophys. Res. Lett.* 48, 1–12. doi:10.1029/2021GL093585

Khoi, D. N., and Suetsugi, T. (2012). Uncertainty in climate change impacts on streamflow in Be River Catchment, Vietnam. *Vietnam. Water Environ. J.* 26 (4), 530–539. doi:10.1111/j.1747-6593.2012.00314.x

Kim, H., and Parajuli, P. B. (2014). Impacts of reservoir outflow estimation methods in SWAT model calibration. *Trans. ASABE* 57 (4), 1029–1042. doi:10.13031/trans.57.10156

Lakshmi, V., Le, M.-H., Goffin, B., Besnier, J., Pham, H., Do, H., et al. (2023). Regional analysis of the 2015–16 Lower Mekong River basin drought using NASA satellite observations. *J. Hydrology, Regional Stud.* 46, 101362. doi:10.1016/j.ejrh.2023.101362

Le, M.-H., Lakshmi, V., Bolten, J., and Bui, D. (2020). Adequacy of satellite-derived precipitation estimate for hydrological modeling in vietnam basins. *J. Hydrology* 586, 124820. doi:10.1016/j.jhydrol.2020.124820

Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., and Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nat. Clim. Change* 9 (12), 926–933. doi:10.1038/s41558-019-0639-x

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* 99 (Issue D7), 14415–14428. doi:10.1029/94JD00483

Ministry of Natural Resources and Environment of The Socialist Republic of Vietnam. (2016). Approval of the Paris agreement to implement the united Nations framework convention on climate change 7, 1.

Mizuta, R., Yoshimura, H., Murakami, H., Matsueda, M., Endo, H., Ose, T., et al. (2012). Climate simulations using MRI-AGCM3.2 with 20-km grid. *J. Meteorological Soc. Jpn.* 90 (A), 233–258. doi:10.2151/jmsj.2012-A12

Moberg, A., and Jones, P. D. (2004). Regional climate model simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data 1961-1990. *Clim. Dyn.* 23 (7–8), 695–715. doi:10.1007/s00382-004-0464-3

Mohammed, I., J Bolten, R. S., and Lakshmi, V. (2018a). Improved hydrological decision support system for the Lower Mekong River Basin using satellite-based earth observations. *Remote Sens.* 10, 885. doi:10.3390/rs10060885

Mohammed, I., J Bolten, R. S., and Lakshmi, V. (2018b). Satellite observations and modeling to understand the Lower Mekong River Basin streamflow variability. *J. Hydrology* 564, 559–573. doi:10.1016/j.jhydrol.2018.07.030

MONRE (2016). Climate change and sea level rise scenarios. *Science* 294 (5547), 1379–1388.

Mosavi, A., Golshan, M., Choubin, B., Ziegler, A. D., Sigaroodi, S. K., Zhang, F., et al. (2021). Fuzzy clustering and distributed model for streamflow estimation in ungauged watersheds. *Sci. Rep.* 11, 8243–8314. doi:10.1038/s41598-021-87691-0

MRC (2017). "The council study: the study on sustainable management and development of the Mekong River Basin, including impacts of mainstream hydropower projects," in *Thematic report on the positive and negative impacts of hydropower development on the social* (Lao PDR: Environmen. Mrc), 107.

National Institute for Soils and Fertilizers (2002). *The basic information of main soil units of Vietnam*. Vietnam: The Gioi Publishers.

NCAR (2017). *Regional climate model RegCM reference manual. January*.

Nguyen, B. Q., Kantoush, S., Binh, D. V., Saber, M., Vo, D. N., and Sumi, T. (2023a). Quantifying the impacts of hydraulic infrastructure on tropical streamflows. *Hydrol. Process.* 37 (3). doi:10.1002/hyp.14834

Nguyen, B. Q., Kantoush, S., Binh, D. V., Saber, M., Vo, D. N., and Sumi, T. (2023b). Understanding the anthropogenic development impacts on long-term flow regimes in a tropical river basin, Central Vietnam. *Hydrological Sci. J.* 68, 341–354. doi:10.1080/02626667.2022.2153298

Nguyen, B. Q., Tran, T. N. D., Łukaszewska, M. G.-, Sinicyn, G., and Lakshmi, V. (2022). Assessment of urbanization-induced land-use change and its impact on temperature, evaporation, and humidity in central vietnam. *Humidity Central Vietnam* 14 (21), 3367. doi:10.3390/w14213367

Nguyen, D. N., and Nguyen, T. H. (2004). *Climate and climate resources in Vietnam (in Vietnamese)*. Vietnam: Agricultural Publishing House.

Nguyen, M. D., and Nguyen, H. T. (2013). *Drought and water depletion during the 2012-2013 dry season in the Central Highlands*. Vietnam: Daklak News. https://baodaklak.vn/channel/3461/201304/han-han-va-can-kiet-nguon-nuoc-trong-mua-kho-nam-2012-2013-o-tay-nguyen-2229396/.

Nguyen, N. T., Du, T. L. T., Park, H., Chang, C.-H., Choi, S., Chae, H., et al. (2023). Estimating the impacts of ungauged reservoirs using publicly available streamflow simulations and satellite remote sensing. *Remote Sens.* 15 (18), 4563. doi:10.3390/rs15184563

Noor, R., Arshad, A., Shafeeque, M., Liu, J., Baig, A., Ali, S., et al. (2023). Combining APHRODITE rain gauges-based precipitation with downscaled-TRMM data to translate high-resolution precipitation estimates in the indus basin. *Remote Sens.* 15 (2), 318. doi:10.3390/rs15020318

Nuong, B. T., Linh, B. K., Trang, V. Q., Huyen, P. T., Thi, B., Thao, P., et al. (2022). Applying fuzzy analytical hierarchy process to establish environmental sustainability indicators for water resources Srepok River Basin, vietnam. *VNU J. Sci. Earth Environ. Sci.* 38 (4), 63–74. doi:10.25073/2588-1094/vnuees.4877

Pathak, S., Ojha, C. S. P., Shukla, A. K., and Garg, R. D. (2019). Assessment of annual water-balance models for diverse Indian watersheds. *J. Sustain. Water Built Environ.* 5 (3). doi:10.1061/jswbay.0000881

Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. U. S. A.* 106 (21), 8441–8446. doi:10.1073/pnas.0900094106

Piman, T., Cochrane, T. A., and Arias, M. E. (2016). Effect of proposed large dams on water flows and hydropower production in the Sekong, sesan and srepok rivers of the mekong basin. *River Res. Appl.* 32 (10), 2095–2108. doi:10.1002/rra.3045

Piman, T., Lennaerts, T., and Southalack, P. (2013). Assessment of hydrological changes in the lower mekong basin from basin-wide development scenarios. *Hydrol. Process.* 27 (15), 2115–2125. doi:10.1002/hyp.9764

Pokhrel, Y., Tiwari, A., Kramer, D., Akhter, T., Tang, Q., Liu, J., et al. (2023). A synthesis of hydroclimatic, ecological, and socioeconomic data for transdisciplinary research in the Mekong. *Sci. Data* 10 (1), 283. doi:10.1038/s41597-023-02193-0

Qiu, J., Yang, Q., Zhang, X., Huang, M., Adam, J. C., and Malek, K. (2019). Implications of water management representations for watershed hydrologic modeling in the Yakima River basin. *Hydrology Earth Syst. Sci.* 23 (1), 35–49. doi:10.5194/hess-23-35-2019

Raghavan, S. V., Vu, M. T., and Liong, S. Y. (2012). Assessment of future stream flow over the sesan catchment of the lower mekong basin in vietnam. *Hydrol. Process.* 26 (24), 3661–3668. doi:10.1002/hyp.8452

Ranger, N., Hallegatte, S., Bhattacharya, S., Bachu, M., Priya, S., Dhore, K., et al. (2011). An assessment of the potential impact of climate change on flood risk in Mumbai. *Clim. Change* 104 (1), 139–167. doi:10.1007/s10584-010-9979-2

Roderick, T. P., Wasko, C., and Sharma, A. (2019). Atmospheric moisture measurements explain increases in tropical rainfall extremes. *Geophys. Res. Lett.* 46 (3), 1375–1382. doi:10.1029/2018GL080833

Samuelsson, P., Jones, C. G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., et al. (2011). The Rossby Centre regional climate model RCA3: model description and performance. *Tellus, Ser. A Dyn. Meteorology Oceanogr.* 63 (1), 4–23. doi:10.1111/j.1600-0870.2010.00478.x

Shafeeque, M., Hafeez, M., Sarwar, A., Arshad, A., Khurshid, T., Asim, M. I., et al. (2023a). Quantifying future water - saving potential under climate change and groundwater recharge scenarios in Lower Chenab Canal, Indus River Basin. *Theor. Appl. Climatol.* doi:10.1007/s00704-023-04621-y

Shafeeque, M., Luo, Y., Arshad, A., Muhammad, S., Ashraf, M., and Pham, Q. B. (2023b). Assessment of climate change impacts on glacio - hydrological processes and their variations within critical zone. *Nat. Hazards* 115 (3), 2721–2748. doi:10.1007/s11069-022-05661-9

Shukla, A. K., Ojha, C. S. P., Garg, R. D., Shukla, S., and Pal, L. (2020). Influence of spatial urbanization on hydrological components of the upper ganga River Basin, India. *J. Hazard. Toxic, Radioact. Waste* 24 (4). doi:10.1061/(asce)hz.2153-5515.0000508

Smigaj, M., Hackney, C. R., Kieu, P., Dang, V. P., Ngoc, N. T., Du, D., et al. (2023). Monitoring riverine traffic from space: the untapped potential of remote sensing for measuring human footprint on inland waterways. *Sci. Total Environ.* 860 (September 2022), 160363. doi:10.1016/j.scitotenv.2022.160363

Spruce, J., Bolten, J., Mohammed, I., Srinivasan, R., and Lakshmi, V. (2020). Mapping land use land cover change in the lower mekong basin from 1997 to 2010. *Front. Environ. Sci. Sect. Land Use Dyn.* 8. doi:10.3389/fenvs.2020.00021

Tapas, M., Etheridge, J. R., Howard, G., Lakshmi, V. V., and Tran, T. N. D. (2022). *Development of a socio-hydrological model for a coastal watershed: using stakeholders' perceptions*. Chicago: AGU Fall Meeting Abstracts, H22O–H0996.

Tran, T., Nguyen, Q. B., Tam, D., Le, L., Nguyen, T. D., Vo, N. D., et al. (2022d). Evaluate the influence of groynes system on the hydraulic regime in the ha thanh river, binh dinh Province, vietnam. In *Advances in hydroinformatics*, 241–254. doi:10.1007/978-981-19-1600-7_15

Tran, T.-N.-D., and Lakshmi, V. (2022). *The land use changes impacts on socio-economic drivers and simulation of surface and groundwater in the Eastern Shore of Virginia, the United States*. Chicago: AGU Fall Meeting Abstracts, H42D–H1270.

Tran, T.-N.-D., Le, M.-H., Zhang, R., Nguyen, B. Q., Bolten, J. D., and Lakshmi, V. (2023a). Robustness of gridded precipitation products for vietnam basins using the comprehensive assessment framework of rainfall. *Atmos. Res.* 293 (15), 106923. doi:10.1016/j.atmosres.2023.106923

Tran, T. N. D., Nguyen, B. Q., Le, M.-H., Lakshmi, V., Venkat), Bolten, J. D., et al. (2022a). *Robustness of gridded precipitation products in hydrological assessment for Vietnam river basins*. Chicago: AGU Fall Meeting Abstracts, H22M–H07.

Tran, T. N. D., Nguyen, Q. B., Nguyen, T. T., Vo, N. D., Nguyen, C. P., and Gourbesville, P. (2022c). "Operational methodology for the assessment of typhoon waves characteristics. Application to ninh thuan Province, vietnam," in *Advances in hydroinformatics*, 887–902. doi:10.1007/978-981-19-1600-7_55

Tran, T. N. D., Nguyen, Q. B., Vo, N. D., Le, M. H., Nguyen, Q. D., Lakshmi, V., et al. (2023c). Quantification of global Digital Elevation Model (DEM) – a case study of the newly released NASADEM for a river basin in Central Vietnam. *J. Hydrology Regional Stud.* 45 (October 2022), 101282. doi:10.1016/j.ejrh.2022.101282

Tran, T. N. D., Nguyen, Q. B., Vo, N. D., Marshall, R., and Gourbesville, P. (2022b). Assessment of terrain scenario impacts on hydrological simulation with SWAT model. Application to lai Giang catchment, vietnam. *Adv. Hydroinformatics*, 1205–1222. doi:10.1007/978-981-19-1600-7_77

Tran, T. N. D., Nguyen, Q. B., and Zeeshan, A. (2021a). *Application of plaxis for calculating the construction stability and soft embankment in protecting ha thanh. 2nd conference on sustainability in Civil engineering (CSCE) 2020, 202–210*. https://csce.cust.edu.pk/archive/20-613.pdf.

Tran, T.-N.-D., Nguyen, Q. B., Zhang, R., Aryal, A., Łukaszewska, M.-G., Sinicyn, G., et al. (2023b). Quantification of gridded precipitation products for the streamflow simulation on the Mekong River Basin using rainfall assessment framework: a case study for the Srepok River subbasin, central highland vietnam. *Remote Sens.* 15 (4), 10–30. doi:10.3390/rs15041030

Tran, T. N. D., Zeeshan, A., and Vo, N. D. (2021b). "Application of hydrodynamic modelling to assess the efficiency of hurricane protection measure at xom ro dike," in PHU YEN PROVINCE, VIETNAM. 2nd Conference on Sustainability in Civil Engineering (CSCE). https://csce.cust.edu.pk/archive/20-406.pdf.

Tuo, Y., Duan, Z., Disse, M., and Chiogna, G. (2016). Evaluation of precipitation input for SWAT modeling in Alpine catchment: a case study in the Adige river basin (Italy). *Sci. Total Environ.* 573, 66–82. doi:10.1016/j.scitotenv.2016.08.034

Ty, T. V., Sunada, K., and Ichikawa, Y. (2011). A spatial impact assessment of human-induced intervention on hydrological regimes: a case study in the upper Srepok river basin, central highlands of vietnam. *Int. J. River Basin Manag.* 9 (2), 103–116. doi:10.1080/15715124.2011.595720

Ty, T. V., Sunada, K., Ichikawa, Y., and Oishi, S. (2012). Scenario-based impact assessment of land use/cover and climate changes on water resources and demand: a case study in the Srepok River Basin, vietnam-Cambodia. *Water Resour. Manag.* 26 (5), 1387–1407. doi:10.1007/s11269-011-9964-1

Umar, M., Nasir, S., Arfan, K., Ammar, R., Hafiz, A., Safdar, M., et al. (2022). A modified approach to quantify aquifer vulnerability to pollution towards sustainable groundwater management in Irrigated Indus Basin. *Environ. Sci. Pollut. Res.* 29, 27257–27278. doi:10.1007/s11356-021-17882-9

Vietnam Ministry of Natural Resources and Environment (2020). *Vietnam Climate change scenario*.

Vo, N. D., Nguyen, T. H., Vu, H. C., Gourbesville, P., and Nguyen, Q. B. (2018). Effects of climate change on streamflow in kon – ha thanh river watershed. *Vietnam* 3 (September), 2233–2224. doi:10.29007/1hrc

Xu, F., Dong, G., Wang, Q., Liu, L., Yu, W., Men, C., et al. (2016). Impacts of DEM uncertainties on critical source areas identification for non-point source pollution control based on SWAT model. *J. Hydrology* 540, 355–367. doi:10.1016/j.jhydrol.2016.06.019

Zhang, P., Liu, R., Bao, Y., Wang, J., Yu, W., and Shen, Z. (2014). Uncertainty of SWAT model at different DEM resolutions in a large mountainous watershed. *Water Res.* 53, 132–144. doi:10.1016/j.watres.2014.01.018

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 15 December 2023
DOI 10.3389/fevo.2023.1276239

OPEN ACCESS

EDITED BY
Maged Marghany,
Syiah Kuala University, Indonesia

REVIEWED BY
Imen Ben Salem,
Zayed University, United Arab Emirates
Pedzisai Kowe,
Midlands State University, Zimbabwe

*CORRESPONDENCE
Jinliang Wang
jlwang@ynnu.edu.cn

RECEIVED 11 August 2023
ACCEPTED 30 November 2023
PUBLISHED 15 December 2023

CITATION
Ma J, Khromykh V, Wang J, Zhang J, Li W and Zhong X (2023) A landscape-based ecological hazard evaluation and characterization of influencing factors in Laos.
Front. Ecol. Evol. 11:1276239.
doi: 10.3389/fevo.2023.1276239

A landscape-based ecological hazard evaluation and characterization of influencing factors in Laos

Jun Ma[1,2], Vadim Khromykh[2], Jinliang Wang[1,3,4]*, Jianpeng Zhang[1,3,4], Wenjuan Li[5] and Xuzheng Zhong[1,3,4]

[1]Faculty of Geography, Yunnan Normal University, Kunming, China, [2]Faculty of Geology and Geography, Tomsk State University, Tomsk, Russia, [3]Key Laboratory of Resources and Environmental Remote Sensing for Universities in Yunnan, Kunming, China, [4]Center for Geospatial Information Engineering and Technology of Yunnan Province, Yunnan Normal University, Kunming, China, [5]Faculty of Economics and Management, Tomsk State University, Tomsk, Russia

The study of the spatiotemporal evolution of landscape ecological hazard and human and natural influences is essential for conservative management and regional sustainable development. This study applied a landscape pattern analysis method and geodetector to multi-source data for 2000, 2010, and 2020 to analyze changes in and drivers of landscape ecological hazard in Laos. The results indicated that: (1) There were more prominent changes in landscape types in Laos. Forest area decreased, whereas the areas of other landscape types increased. There was an overall steady change in the landscape patterns of Laos. Besides for significant changes in the artificial surface landscape index, landscape indices remained stable; (2) The cumulative high and extreme ecological hazard areas increased by 1,947.81 km^2, whereas the cumulative areas of low and minimal ecological hazard decreased by 8,461.8 km^2. Areas of low and moderate ecological hazard accounted for > 85% of the total area. Areas of low ecological hazard were mainly in the northwest and southeast. The area of high ecological hazard was concentrated in the central and northeastern regions. The distributions of different landscape ecological hazards in Laos during the study period were similar, with general patterns of decreasing hazard from north to south; (3) A positive Moran's I of landscape ecological hazard in Laos was obtained. While the agglomeration effect was pronounced, it decreased over time, resulting in a weakening in spatial autocorrelation. A significant positive autocorrelation was observed in the spatial distribution of landscape ecological hazard in the study area. Agglomerated areas of high and low ecological hazard were mainly concentrated in the northeast and southeast, respectively; (4) The spatiotemporal evolution of landscape ecological hazard in Laos over the last 20 years could be attributed to interactions between natural and anthropogenic influences. Natural influences were a significant driver of changes to landscape ecological hazard in Laos, with

annual precipitation and average temperature being the most significant. Anthropogenic influences, including socioeconomic factors and regional accessibility, significantly impacted local ecological deterioration in Laos.

KEYWORDS

landscape ecological hazard, landscape pattern, spatial autocorrelation, geodetector, risk management, Laos

1 Introduction

Human socioeconomic development is dependent on ecosystem services. Therefore, maintaining the integrity of natural ecosystems is essential for human survival (Munns et al., 2016). However, continuous human expansion has resulted in the degradation of natural ecosystems globally (Liu et al., 2022). Ecological hazard is as risks of external factors exerting stress on an ecosystem. These stresses can degrade the productivity, health, genetic structure, and aesthetic and economic value of ecosystems (Chen et al., 2013). Since many factors can interact to cause ecological hazard, the scope of ecological hazard can be considerable and complicated to predict (Suter, 2001). Landscape ecological hazard assessment can be used to characterize the regional-scale risks of adverse effects on ecosystems by environmental pollution, anthropogenic endeavors, or natural disasters. The importance of landscape ecological hazard assessment has gradually increased since this risk assessment method overcomes the limitation of traditional regional risk assessment by using a specific natural risk factor (Paustenbach, 2015). The ecological hazard assessment based on landscape pattern emphasizes the influence of landscape structure, spatiotemporal heterogeneity, and scale effect on ecological hazard, thereby achieving a comprehensive characterization and spatial visualization of multiple risks (Chen et al., 2013). Therefore, the study of landscape ecological hazard can act as the foundation for regional hazard prevention and can assist in the management and optimization of regional patterns in the landscape.

Studies incorporating landscape ecological hazard assessment have increased both in China and abroad. There have been many recent studies on ecological hazard in different regions, mainly focusing on watersheds (Kapustka et al., 2001), oases (Hope, 2006), coastal zones (Yanes et al., 2019), wetlands (Malekmohammadi and Rahimi Blouchi, 2014), and cities (Ran et al., 2022). These studies have widely applied the landscape pattern index (Su et al., 2012; Zhang et al., 2022), entropy (Liu et al., 2013; Gao et al., 2022), and exposure-response (Chapman et al., 1998; Bartell, 2006) methods for the construction of an ecological hazard evaluation model, which has been combined with geographical information system (GIS) spatial modeling. The scales at which these studies have been conducted have evolved from single to multiple. Some studies have assessed landscape ecological hazard and appropriate methods by determining appropriate spatial granularity (Hope, 2006). The above studies have focused on landscape ecological hazard assessment by constructing an appropriate model and spatial analysis. There remains a need for further studies on ecological risk assessment at a national scale, as well as studies on local and regional influencing factors. There also remains a need for further analyses of the cumulative effects of the overall evolution of risk on local systems (Hope, 2006).

Loas is the only landlocked country in the Mekong River Basin, and has had to confront significant challenges in recent years, including the combined effects of natural development-oriented economic growth, a growing population, and the resulting environmental pollution, land degradation, and depletion of natural resources. The study of landscape ecological hazard pattern is significant for enabling regional environmental conservation and national ecological security in that country. Concurrently, this such study can improve landscape ecology theory and augment landscape planning and management decision-making. The aim of the present study was to apply principles of landscape ecology and spatial statistical analysis to land cover data for 2000, 2010, and 2020, to construct a landscape ecological hazard index and ecological hazard assessment model for quantitatively analyzing landscape ecological hazard and the associated natural and anthropogenic influencing factors in Laos.

2 Research area and method

2.1 Study area

Laos is a landlocked country forming part of the northern Indochina Peninsula (13°56'–22° 27' N, 100°02'–107°38'E). Laos consists of 17 and one provinces and municipality, respectively, and has a total area of 236,800 km^2 (Figure 1). Laos has a complex and diverse topography, consisting predominantly (80%) of mountains and plateaus. The remaining area forms the Mekong River Valley, along which there are basins and small plains. The terrain of Laos decreases from north to south. The region falls into tropical and subtropical monsoon climate zones. The region experiences a distinct rainy season with annual average precipitation and annual average temperature of 1,250 mm–3,750 mm and 20°C–30°C, respectively. Laos has a developed water system,

with the Mekong River in the west representing the largest river in the region.

2.2 Sources of data and preprocessing

The present study used data falling into three broad categories: (1) natural environment data, (2) socioeconomic data, and (3) regional accessibility data. The present study utilized land cover data for the region to extract the landscape pattern index. The spatial resolution of these data for 2000, 2010, and 2020 was 30 m × 30 m. The present study considered eight land use types: (1) artificial, (2) bare land, (3) cultivated land, (4) forest, (5) grassland, (6) shrubland, (7) water, (8) wetlands. Temperature and precipitation data utilized in the current study were downloaded from the Google Earth Engine (GEE) platform, and these data were processed to obtain annual averages. Data for the Normalized Difference Vegetation Index (NDVI) were obtained through the GEE platform and were utilized to estimate fractional vegetation cover (Pettorelli et al., 2005). Table 1 shows the specific data sources used in the present study.

TABLE 1 Types of data used in the present study and their sources.

Type of data	Data	Unit	Data sources
Natural data	DEM	m	USGS EarthExplorer (https://earthexplorer.usgs.gov)
	Slope	°	Using DEM data extraction
	Fractional vegetation cover	%	The NDVI data downloaded from the GEE platform was obtained by correlation calculation
	Annual average temperature	°C	Level-1 and Atmosphere Archive & Distribution System (https://ladsweb.modaps.eosdis.nasa.gov)
	Annual precipitation	mm	Google Earth Engine (https://explorer.earthengine.google.com)
	Organic matter content	dg/kg	International Soil Reference and Information Centre (https://data.isric.org)
	Land use	–	Globalland30 (http://www.globallandcover.com)
Social and economic data	Population density	Per/km^2	WorldPop (https://hub.worldpop.org)
	Kilometer grid GDP	USD Billions/ km^2	Global Gridded Geographically Based Economic Data (https://sedac.ciesin.columbia.edu)
	Night light	–	Google Earth Engine (https://explorer.earthengine.google.com)
Regional accessibility data	Road	km	DIVA-GIS (https://www.diva-gis.org/gdata)
	Water bodies	km	Google Earth Engine (https://explorer.earthengine.google.com)
	Town center	km	Resource and Environment Science and Data Center (https://www.resdc.cn)
	Parks and reserves	km	Resource and Environment Science and Data Cen (https://www.resdc.cn)

2.3 Sampling method

As an essential basis for the study of landscape ecological hazard, the choice of spatial granularity will directly affect the spatial representation accuracy of landscape ecological hazard. Therefore, selecting the spatial granularity and conducting relevant scientific research is necessary when researching landscape ecology. The present study gridded the study area in ArcGIS 10.2.2 software, considering the optimal grid size for calculating the landscape ecological hazard index, the size of the study area, and the convenience of hazard index data extraction. The equal spacing sampling method was used. Typically, the division of evaluation units exceeds the average landscape patch area by 2 to 5 (Chen et al., 2013). The present study evaluated grid sizes of granularities of 10 × 10 km, 15 × 15 km, and 20 × 20 km. However, the 15 × 15 km grid was optimal and was chosen to divide Laos into 1,364 evaluation units (Figure 2) (Zhang et al., 2020).

2.4 Method used to analyze the landscape pattern index

The landscape pattern index provides dense information on landscape layout, and can utilize a single or a combination of several indices to analyze the spatial structure and evolution of the landscape. The present study selected six landscape indices to comprehensively reflect the patterns of landscape and characteristics of the study area and to reduce information redundancy. These were the landscape fragmentation, isolation, dominance, disturbance, vulnerability, and loss indices (Table 2). The moving window method using a 15-km window was applied in Fragstats 4.2 to calculate the landscape pattern index of Laos (Bai and Weng, 2023; Xu et al., 2023). The window was moved from the left, upper section of the study area, and progressed at a single-grid step. The window calculated the landscape pattern index for each grid and assigned the center grid as the landscape index of the center point of each sample plot. The current study obtained the index of landscape disturbance by superimposing the landscape, landscape isolation, landscape dominance, and landscape vulnerability indices using an expert scoring method.

2.5 Constructing the index of ecological hazard to the landscape

An ecological hazard index is an index characterizing disturbance of the landscape structure and can be used to reflect the degree of human influence on natural ecosystems. Ecological

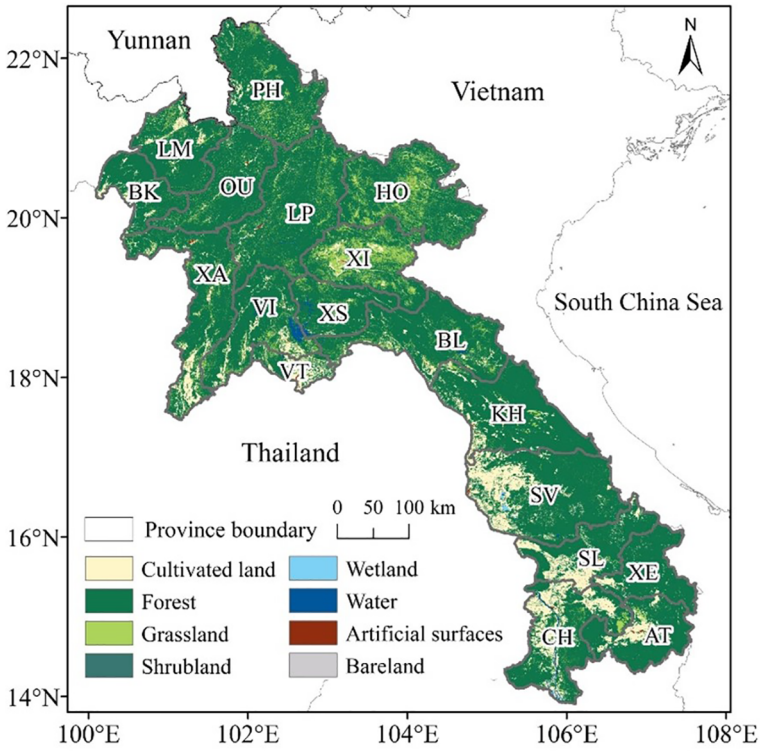


FIGURE 1
Geographical location and land cover of Laos. PH, Phongsali; LM, Louang Namtha; BK, Bokeo; OU, Oudomsai; LP, Louangphrabang; HO, Houaphan; XA, Xaignabouri; VI, Viangchan; XI, Xiangkhoang; XS, Xaisomboun; VT, Vientiane; BL, Bolikhamxai; KH, Khammouan; SV, Savannakhet; SL, Saravan; XE, Xekong; CH, Champasak; AT, Attapu.

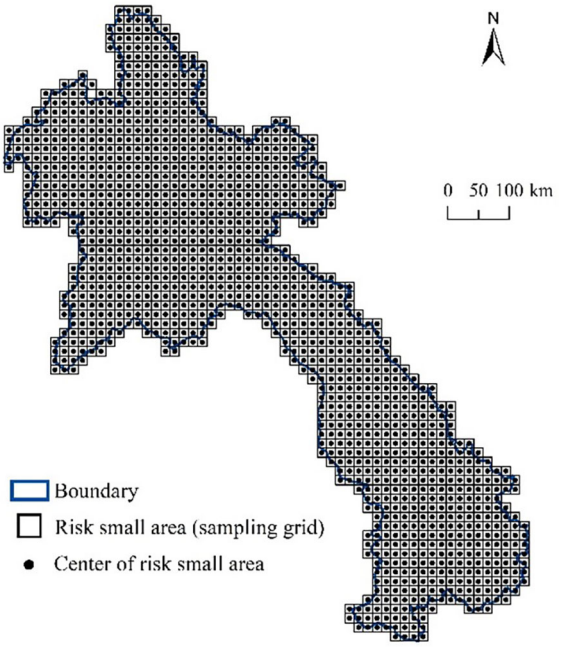


FIGURE 2
Division of Laos into 15 × 15 km grids for calculation of the ecological hazard area.

hazard indices quantitatively describe the effects of natural and anthropogenic influences on ecosystems (Bayliss et al., 2012). The present study reduced information redundancy among individual indicators by using the relative areas of landscape components, the landscape vulnerability index, and the landscape interference index. Fragstats 4.2 software was used to construct the landscape ecological hazard evaluation model using the moving window method and index calculation, which was taken to represent the ecological hazard of the center point of each plot (Li et al., 2017). The index was calculated as follows (Equation 1):

$$\mathrm{ERI}_i = \sum_{i=1}^{N} \frac{A_{ki}}{A_k} R_i, \quad (1)$$

where ERI_i is the index of landscape ecological hazard in the i^{th} sample unit; A_{ki} is the areal extent of landscape type i in the k^{th} sample unit; A_k is the areal extent of the k^{th} sample unit; R_i is the index of landscape loss.

2.6 Exploratory analysis of spatial data

Analysis of spatial data is used to find the rules under which spatial data is distributed and its spatial heterogeneity (Anselin, 1996). The present study applied exploratory spatial data analysis to analyze the ecological hazard of Laos and to identify associated patterns. The present study used global and local spatial

TABLE 2 Methods used to calculate the landscape pattern indices.

Sequence number	Exponential	Symbol	Formula	Formula description	Ecological meaning
(1)	Landscape fragmentation	*Ci*	$C_i=n_i/A_i$	n_i is patch number of class i, A_i is the total area of class i	This index expresses the degree of fragmentation of the entire landscape or a particular landscape type at a given time and given nature. Under natural or human interference, the landscape tends to be complex, homogeneous, and continuous from a single, homogeneous, and continuous whole. In the process of heterogeneous and discontinuous patch mosaics, the larger the value, the lower the internal stability of the landscape unit, and the lower the stability of the corresponding landscape ecosystem.
(2)	Landscape isolation	*Ni*	$N_i = \frac{1}{2}\sqrt{\frac{n_i}{A}} \times \frac{A}{A_i}$	A is the total area of the entire landscape	This index expresses the separation degree of individual distribution of different elements or patches in a particular landscape type. The greater the separation degree, the more scattered the landscape is in geographical distribution, the more complex the landscape distribution, and the higher the degree of fragmentation.
(3)	Landscape dominance	*Di*	$D_i = \frac{(Q_i + M_i)}{4} + \frac{L_i}{2}$ $Q_i = \frac{n_i}{N_i}$ $M_i = \frac{\delta_i}{\delta}$ $L_i = \frac{A_i}{A}$	δ_i is the sample number of patch i, δ is the total number of samples	The value of this index is used to measure the importance of patches in the landscape, and its size directly reflects the impact of patches on the formation and change of landscape patterns.
(4)	Landscape disturbance	*Si*	$S_i = aC_i + bN_i + cD_i$	a, b and c are weights of indices C_i, N_i and $D_i$respectively. where a = 0.5, b = 0.3, c = 0.2, and $a + b + c$ =1	This index is used to reflect the extent to which ecosystems represented by different landscapes are disturbed (mainly by human activities).
(5)	Landscape vulnerability	*Fi*	Obtained by artificial assignment and normalization		It indicates the vulnerability of the ecosystem represented by the landscape type when it is disturbed by the outside world, and this value is related to its stage in the natural succession process of the landscape. In general, ecosystems in the primary succession stage, with simple food chain structures and low biodiversity indices, are relatively fragile.
(6)	Landscape loss	*Ri*	$R_i = S_i \times F_i$		This index reflects the degree of loss of natural attributes of ecosystems represented by different landscape types when they are disturbed by nature and humans.

autocorrelation to characterize the ecological hazard of Laos. The geodetector was used to identify factors driving the spatial and temporal variation in ecological hazard in Laos.

2.6.1 Ecological hazard assessment

The ecological hazard index has random and structural spatial variation. The present study analyzed the spatial heterogeneity of the index of landscape ecological hazard. This was achieved by fitting the ecological hazard index to a semi-variogram in ArcGIS to obtain the best-fitting model (Hengl, 2009). The present study then applied Kriging interpolation to interpolate the three datasets (Oliver and Webster, 1990), following which spatial distributions of the hazard to the landscape ecology were generated. The present study applied the transfer matrix model to analyze landscape ecological hazard over different periods. The grading standards used in the present study to evaluate results were consistent with those used in previous studies (Li et al., 2020; Gong et al., 2021; Zhang et al., 2022). Landscape ecological hazard in 2020 was separated into five grades using the natural discontinuity point method: (1) minimal ecological hazard (ERI< 0.0709); (2) low ecological hazard (0.0709 ≤ ERI< 0.1157); (3) moderate ecological hazard (0.1157 ≤ ERI< 0.1773); (4) high ecological hazard (0.1773< ERI ≤ 0.2782); (5) extreme ecological hazard (ERI ≥ 0.2782). Data for the remaining two periods were categorized in the same way that for the 2020s.

2.6.2 Analysis of autocorrelation in the spatial dimension

Autocorrelation in the spatial dimension can be used to identify spatial interdependence between two or more variables (Koenig, 1999). The present study applied spatial autocorrelation to analyze the distributions of variables and correlations between variables (Anselin, 2003). Identification of spatial correlation between hazard attributes of adjacent areas was conducted by applying spatial autocorrelation. GeoDa and ArcGIS software were used to apply the global spatial autocorrelation index Moran's I and the local spatial autocorrelation index LISA to evaluate spatial differences in ecological hazard. Moran's I have a value of −1 and 1 and can be

used to reflect similarities in a particular attribute between adjacent units. Moran's I > 0,< 0, and = 0 indicate a positive, negative, and no correlation, respectively, and aggregated, dispersed, and random research units, respectively (Diniz-Filho et al., 2003). Moran's I was calculated as follows (Equation 2):

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}(x_i-\overline{x})(x_j-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}}, \quad (2)$$

where x_i and x_j represent variable x at adjacent points, x represents the mean of the variables, ω_{ij} indicates the adjacent weight, and n represents total hazard points.

Moran's I can only indicate the distribution of the index of ecological hazard. Therefore, the current study integrated the LISA index to allow further exploration of the aggregation of the local hazard index and to identify abnormal spatial characteristics. The LISA index can be divided into four types: (1) High-High; (2) High-Low; (3) Low-High; (4) Low-Low. The LISA index was used to represent different spatial clustering distributions and was calculated as follows (Equation 3):

$$I_i = \frac{x_i-\overline{x}}{S^2}\sum_{j\neq is}^{n'}\omega_{ij}(x_j-\overline{x}) \quad (3)$$

where n' represents the size of the sample and S^2 represents statistical variance.

2.6.3 Analysis of the geodetector attribution

A geodetector is a statistical method used to identify drivers of geographic spatial characteristics by identifying heterogeneity in spatial stratification of events. Heterogeneity in the spatial stratification represents variance in a region that exceeds the within-layer sum of the variance. The geodetector used in the present study incorporated risk factor, ecological, and interaction detectors (Zhu et al., 2020). The present study selected the factor detector for use in the geodetector through reference to relevant previous research results (Huang et al., 2020; Liang et al., 2022; Li et al., 2022; Xu et al., 2023). The present study identified 14 socioeconomic and regional accessibility factors affecting the landscape ecological hazard in Laos, including the digital elevation model (DEM), slope, fractional vegetation cover, annual average temperature, annual precipitation, organic matter content, population density, kilometer grid GDP, night light, distance from the road, distance to water, distance from the town center, and distance to parks and reserves.

3 Results

3.1 Temporal changes in patterns of landscape

The present study calculated the landscape pattern index of each landscape type in the study area in 2000, 2010, and 2020 using the statistical analysis function in Fragstats 4.2 and Excel 2010 (Table 3). Interactions between natural and anthropogenic influences result in changes in the area and number of patches of each landscape, resulting in changes to the hazard index. As shown in Table 3, forest, cultivated land, and grassland represented the main landscape types in Laos. There was a decreasing trend in forest area over the last two decades, whereas there were increasing trends in cultivated land and grassland. The area of water bodies first increased from 2000 to 2010 and then decreased from 2010 to 2020. While artificial surfaces had the smallest area, there was a significant increase in artificial surfaces in Laos between 2010 and 2020, with that in 2020 exceeding that in 2000 and 2010 by factors of 6.18 and 4.65, respectively. The number of grassland patches significantly exceeded those of other land covers, resulting in this land cover obtaining the most extensive fragmentation index. The large fragmentation index of grassland resulted in a small random scattered distribution. Forest showed the second-largest number of patches and random distribution, with decreasing patch numbers over time.

Many factors affect landscape ecological hazard, including the vulnerability, disturbance, type, and land use structure of landscapes. Ecological hazard can be characterized by landscape loss. Artificial surfaces showed the most significant loss index, followed by grassland, wetland, water bodies, and cultivated land, while forest had the lowest loss index. Despite the high loss index of artificial surfaces, this land cover had minimal influence on landscape ecological hazard due to its limited landscape area. The most extensive landscape loss index of artificial surfaces could be attributed to its highest vulnerability among the assessed landcover types. The landscape loss index also varied among the different periods. There were continual decreases in landscape loss of artificial surfaces and water bodies. The change in artificial landscape surfaces is most significant, changing from 0.2473 in 2000 to 0.1136 in 2020, representing an increase factor of 2.18 over 20 years. While this decreasing trend reduced the influence of artificial surfaces on landscape ecological hazard, its influence remained significant. The landscape loss indices of other landscape types generally remained unchanged.

3.2 Spatiotemporal evolution of landscape ecological hazard

3.2.1 Distribution of landscape ecological hazard level

Laos showed low, moderate, and high ecological hazard from 2000 to 2020, with the cumulative areas of these three categories accounting for 79.19%–91.06% of the total area (Figure 3). The distributions of landscape ecological hazard in Laos were similar among the three-time points. Ecological hazard in the study area decreased from north to south. Areas of high ecological hazard were mainly distributed around areas of extreme ecological hazard, whereas areas of low ecological hazard were mainly adjacent to those of minimal ecological hazard. The complex geographical environments in the northwest, central east, and south of the study area showed clear vertical zonal characteristics. Various landscape patches showed a staggered distribution, such as forest land, resulting in high separation, serious fragmentation, and low connectivity. Interactions among different landscapes were

TABLE 3 Indices of patterns of landscape in Laos from 2000 to 2020.

Landscape type	Time	Area/ km^2	Number	Fragmentation index	Abruption index	Predominance index	Obstruction index	Fragility index	Dominance index
Cultivated land	2000	23016.3200	10940	0.0048	0.1170	0.2268	0.0828	0.16	0.0133
	2010	25712.9000	10437	0.0041	0.1022	0.2371	0.0801	0.16	0.0128
	2020	28030.5100	12657	0.0045	0.1032	0.2424	0.0817	0.16	0.0131
Forest	2000	217551.7375	176586	0.0081	0.0497	0.6356	0.1461	0.02	0.0029
	2010	214094.3575	181997	0.0085	0.0513	0.6307	0.1458	0.02	0.0029
	2020	210427.8975	178951	0.0085	0.0517	0.6249	0.1447	0.02	0.0029
Grassland	2000	21438.1700	1123986	0.5243	1.2737	0.4816	0.7406	0.11	0.0815
	2010	21617.7500	1134656	0.5249	1.2675	0.4809	0.7389	0.11	0.0813
	2020	21850.9600	1104741	0.5056	1.2373	0.4807	0.7201	0.11	0.0792
Wetland	2000	163.3483	269	0.0165	2.5860	0.0128	0.7866	0.07	0.0551
	2010	208.5183	894	0.0429	3.6884	0.0161	1.1312	0.07	0.0792
	2020	201.2846	938	0.0466	3.9138	0.0172	1.2009	0.07	0.0841
Water bodies	2000	2518.6820	27150	0.1078	1.6849	0.1706	0.5935	0.04	0.0237
	2010	2295.1350	21250	0.0926	1.6337	0.1528	0.5670	0.04	0.0227
	2020	2743.4990	20202	0.0736	1.3326	0.1626	0.4691	0.04	0.0188
Artificial Surfaces	2000	134.2784	444	0.0331	4.0415	0.0381	1.2366	0.2	0.2473
	2010	178.2611	568	0.0319	3.4390	0.0444	1.0565	0.2	0.2113
	2020	829.6782	3180	0.0383	1.7483	0.1216	0.5680	0.2	0.1136

10.3389/fevo.2023.1276239

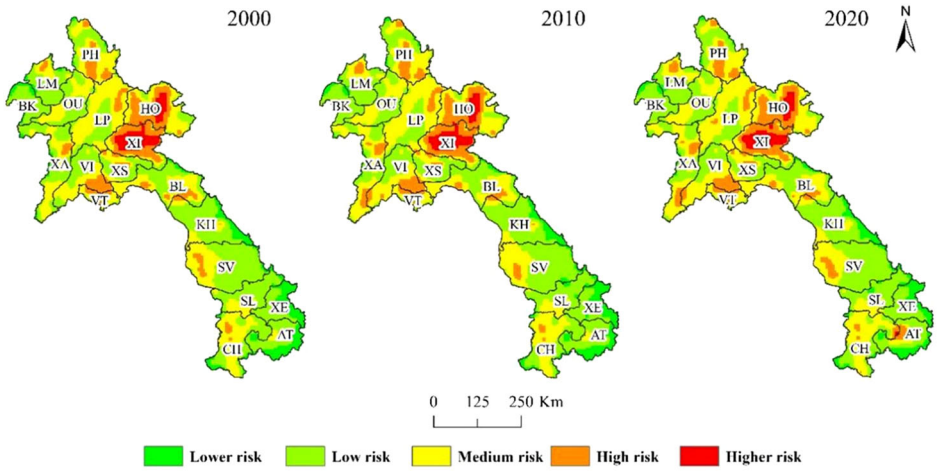


FIGURE 3
Spatial distribution of landscape ecological hazard falling into various categories in Laos in 2000–2020. PH, Phongsali; LM, Louang Namtha; BK, Bokeo; OU, Oudomsai; LP, Louangphrabang; HO, Houaphan; XA, Xaignabouri; VI, Viangchan; XI, Xiangkhoang; XS, Xaisomboun; VT, Vientiane; BL, Bolikhamxai; KH, Khammouan; SL, Saravan; SV, Savannakhet; XE, Xekong; CH, Champasak; AT, Attapu.

obstructed, producing elevated ecological hazard. In addition, the region implements relatively redundant agricultural production methods. Long-term unsustainable agricultural production activities have resulted in frequent transformations between landscape components (Castella et al., 2013) and an overall decline in ecological environmental quality, thereby increasing the ecological hazard of the area. Areas of higher ecological hazard were concentrated in the Xiangkhoang Plateau in northeastern Laos, HO, and XI provinces. The dominant types of landscape were forest, grassland, and cultivated land. These landcover types showed scattered and fragmented distributions, forming a clear vertical band spectrum. There were clear fragmentation characteristics of landscape patches. This fragmentation has destroyed the integrity and stability of the ecosystem, leading to high ecological hazard. Areas of high ecological hazard were concentrated in the transition from higher to moderate ecological hazard. These areas were widely distributed in the PH, LM, LP, HO, XI, XA, VI, VT, BL, SV, and CH provinces. Change in the distribution of high ecological hazard involved extending areas to surrounding areas, during which areas of minimal ecological hazard transition to high ecological hazard. This change was most obvious in the XA province in west Laos and CH and AT provinces in the south. Most high ecological hazard areas were in the mountainous and plateau areas characterized by high terrain. The distribution of areas of moderate ecological hazard and associated temporal change were similar to those of areas of high ecological hazard, with areas of moderate ecological hazard located adjacent to areas of high ecological hazard. Areas of moderate ecological hazard consisted mainly of cultivated land and artificial surfaces. The region experienced considerable disturbance from anthropogenic activities due to the expansion of cultivated land and urban areas during the middle and late stages. However, the stability of agricultural land and construction land remained strong, with these land cover types resistant to changes in the environment and anthropogenic endeavors. Areas of low ecological hazard were mainly in BK province, north of the study area, and in KH and SV provinces in central Laos. Low ecological hazard dominated SL, XE, and AT in the south, with forest and grassland being the main land cover types. The establishment of national parks and nature reserves resulted in increases in the region's stability, thereby increasing the region's resilience to environmental changes and anthropogenic endeavors. The natural landscape of the study area is well-preserved and is not susceptible to human interference, resulting in low landscape loss.

3.2.2 Transformations in land cover type associated with landscape ecological hazard level

There were increases in landscape ecological hazard in Laos from 2000 to 2020. The areas of high/higher, moderate, and low/minimal ecological hazard increased by 1,947.8.1 km^2, increased by 6,513.9 km^2, and decreased by 8,461.8 km^2, respectively (Table 4). The Sankey map of landscape ecological hazard in Laos over the 2-decade study period (Figures 4, 5) shows that the area of extreme ecological hazard increased from 2.86% to 3.02% (407.9 km^2). Areas of minimal ecological hazard mainly transitioned to low and moderate ecological hazard. There was less transition in areas of high and extreme ecological hazard, with changes in ecological hazard in these areas having minimal impact on overall ecological hazard in Laos. Processes contributing to the transition of areas to moderate ecological hazard were more complex, indicating the need for increased ecological stability. There needs to be an increased focus on maintaining ecosystem stability in the region to avoid the intensification of landscape fragmentation and a transition to a higher level of ecological hazard. Future effects of human activities and climate change will inevitably increase interference in and the destruction of the natural landscape in the study area, resulting in a significantly expanded area of medium ecological hazard and a concurrent decrease in the area of lower ecological hazard. Therefore, it is expected that the ecological state of the landscape in Laos will decline in the future. Urban expansion results in fragmentation of

10.3389/fevo.2023.1276239

TABLE 4 Relative area of landscape ecological hazard categories in Laos for 2000–2020 (km^2).

Ecological hazard class	Area		
	2000	2010	2020
I	18091.76	17947.25	16111.05
II	121664.60	117621.62	115167.05
III	87874.67	91482.90	94438.32
IV	29003.34	29645.50	31353.69
V	7974.86	7838.98	7567.01

I: lower hazard, II: low hazard, III: medium hazard, IV: high hazard, V: higher hazard.

landscape patches along the urban fringe, which in turn leads to changes in landscape structure and function.

3.2.3 Spatial characteristics of landscape ecological hazard

As shown in Figure 6, the present study conducted global autocorrelation based on the three periods of landscape ecological hazard in the study area. The Moran's I values of the three periods were 0.651, 0.642, and 0.623, respectively. The significant Z and P values within the confidence interval suggested significant positive correlations in the landscape ecological hazard index in the study area between the three periods. The distribution of Moran's I index in each quadrant showed that landscape ecological hazard among the three periods were spatially clustered and affected each other. There was a decreasing trend in the global Moran's I, indicating a weakening in spatial autocorrelation and a decrease in spatial convergence.

Further analysis of local autocorrelation within the ecological hazard index of Laos from 2000 to 2020 (Figure 7) showed "high-high" and "low-low" aggregations. The distributions of aggregations remained relatively concentrated and stable, consistent with the distribution of Moran's I and the spatial distribution of ecological hazard. There were few "low-high" and "high-low" aggregations, which were scattered around high- and low-value aggregations. These aggregations gradually homogenized with adjacent units. "Low-low" aggregations were mainly in the south and southeast regions with lower human impact. The dominant land cover categories in these areas were forest, grassland, and cultivated land. "High-high" aggregations were concentrated in the XI and HO provinces on the Chuankuang Plateau, with these areas

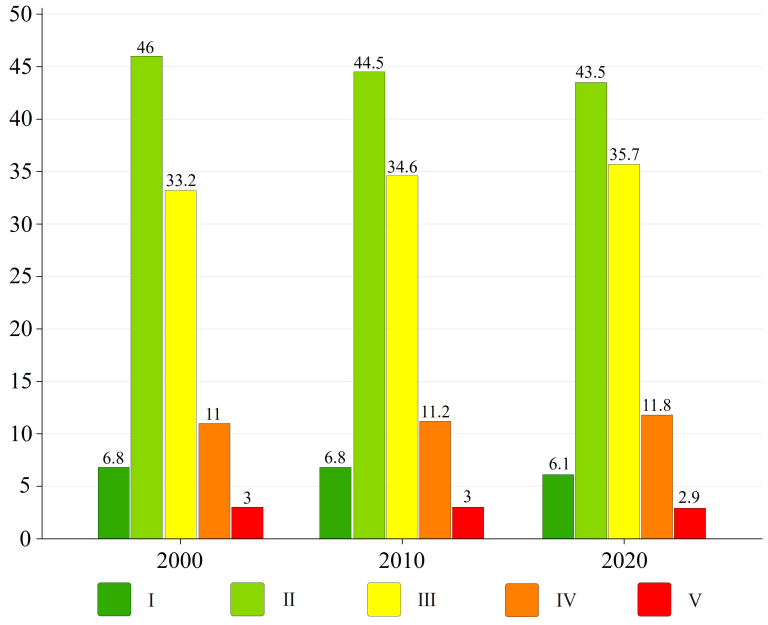


FIGURE 4
Proportion of landscape ecological hazard classes in Laos, 2000–2020 (%).

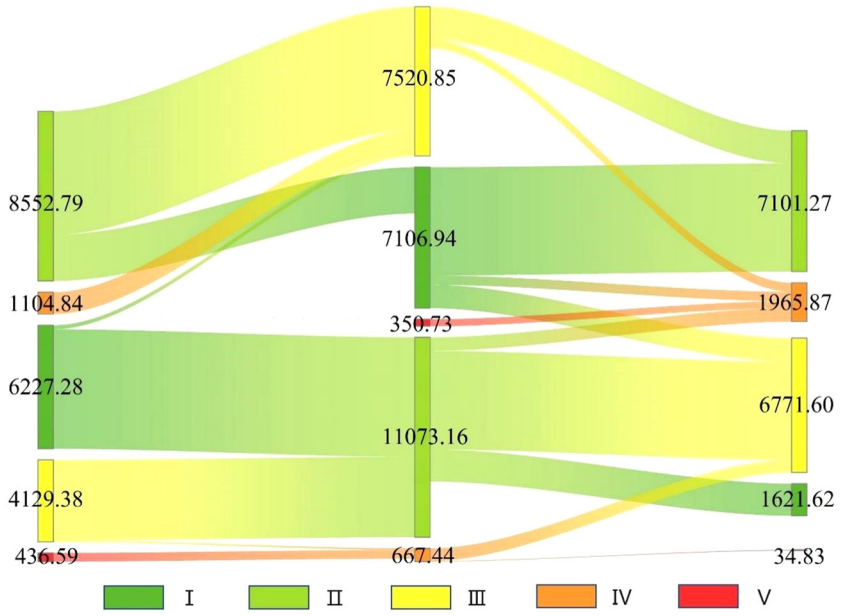


FIGURE 5
Sankey diagram showing the transfer of landscape ecological hazard in Laos between 2000–2020 (km^2). I: lower hazard, II: low hazard, III: medium hazard, IV: high hazard, V: higher hazard.

characterized by high landscape fragmentation and isolation, high altitude, considerable topographic relief, high deforestation, and severe deterioration of the ecological environment. Some "high-high" aggregations were concentrated in urban agglomerations characterized by dense population and high socioeconomic development. These areas included the intersection between PH province in northern Laos and VI and VT provinces near the central capital. Small aggregations also occurred in the urban areas of XA, BL, SV, and CH provinces. The main land cover types in these areas included water bodies, cultivated land, and artificial surfaces. Landscape connectivity was reduced in low urban agglomerations by human disturbance. Anthropogenic endeavors disrupted the natural evolution process in areas of high ecological vulnerability along the Mekong River, resulting in fragmented habitat and high ecological hazard. There were increasing trends in the "low-low" and "high-high" aggregations over the 2-decade study period. This result indicated an increase in polarization in local ecological hazard in Laos.

3.2.4 Factors influencing landscape ecological hazard

3.2.4.1 Factors influencing overall landscape ecological hazard

The study selected 14 natural, social, economic, and regional factors influencing landscape ecological hazard. Table 5 illustrates each factor's relative contribution to landscape ecological hazard. The results showed that each factor significantly affected landscape ecological hazard. However, the effects of anthropogenic influences dominated the spatiotemporal evolution of landscape ecological hazard at the whole study area scale. The relative contributions of

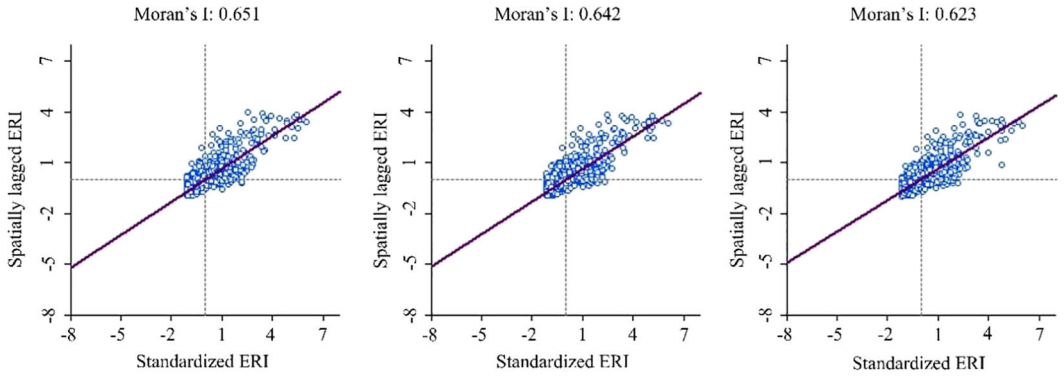


FIGURE 6
Scatter plot of standardize ecological hazard index (ERI) vs spatially lagged ERI and the associated Moran's I in Laos from 2000 to 2020.

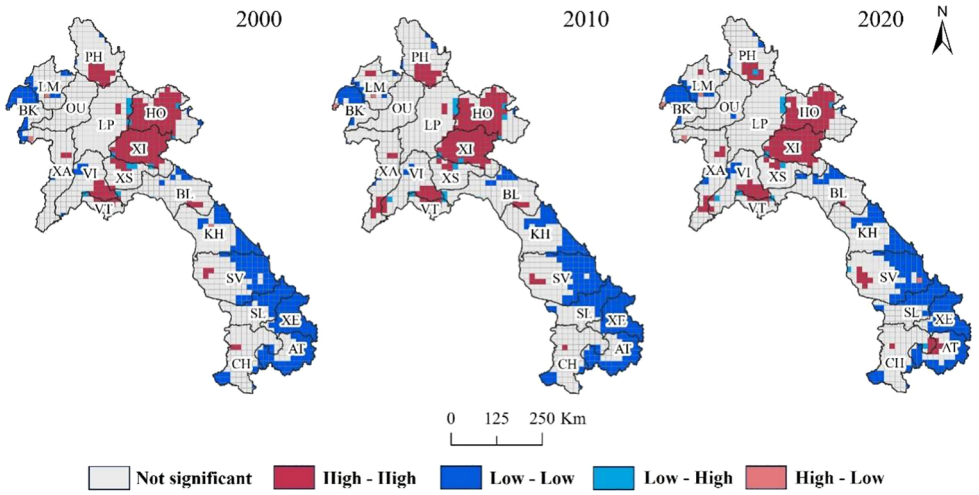


FIGURE 7
Map of local autocorrelation in ecological hazard in Laos from 2000 to 2020. PH, Phongsali; LM, Louang Namtha; BK, Bokeo; OU, Oudomsai; LP, Louangphrabang; HO, Houaphan; XA, Xaignabouri; VI, Viangchan; XI, Xiangkhoang; XS, Xaisomboun; VT, Vientiane; BL, Bolikhamxai; KH, Khammouan; SL, Saravan; SV, Savannakhet; XE, Xekong; CH, Champasak; AT, Attapu.

gross domestic product (GDP), distance from the town center, and distance from a road were close to or exceeded 10% over all three periods. The effect of the distance from a road on landscape ecological hazard showed a rising trend over the 2-decade study period. In contrast, the influences of night light and population density showed downward trends. The influences of other anthropogenic influences fluctuated, with the highest variation in GDP attributable to poverty alleviation policies implemented in Laos in 2010 (Phimphanthavong, 2013). Natural influences had little impact on overall landscape ecological hazard in Laos from 2000 to 2020 due to forest dominating the landscape (79.5%), followed by cultivated land (10.6%) and grassland (8.3%). Among natural influences, annual precipitation, annual average temperature, soil organic matter content, vegetation coverage, distance from protected areas, and distance from water had significant impacts on ecological hazard of Laos of between 5% and 10%. The relatively high influence of annual precipitation on the ecological hazard in 2010 could be attributed to extensive flooding on the Vientiane Plain and in Zhanbasai (Baiyinbaoligao et al., 2020). There were increasing trends in the impacts of annual average temperature, vegetation coverage, distance from protected areas, and distance from water on ecological hazard. This result could be attributed to alterations in land use due to anthropogenic endeavors (Sunderlin, 2006; Lestrelin et al., 2012). Elevation and slope showed minimal impacts on ecological hazard in Laos, mainly due to the area's inaccessibility with significant elevation differences and a high slope preventing agricultural development and human settlement (Wang et al., 2019). However, the impact of natural influences on ecological hazard in Laos will continue to increase with increasing societal demands and the intensification of anthropogenic endeavors. While natural and anthropogenic influences affected the spatiotemporal changes in landscape ecological hazard in the study area, anthropogenic endeavors were the main drivers of deterioration in landscape ecological hazard.

3.2.4.2 Influencing factors of local landscape ecological hazard

The large scale of the study area (Kapustka et al., 2001) prevents the comprehensive analysis of factors influencing landscape ecological hazard. Therefore, the present study analyzed local characteristics of landscape ecological hazard. The minimum landscape ecological hazard index in 2020 (0.12) was used as a threshold in the analysis, in which the change in the ecological hazard index between 2000 and 2020 of > 0.12,< −0.12, and between −0.12 and 0.12 indicated an increase, decrease, and no change in landscape ecological hazard, respectively. Ten areas that showed increased landscape ecological hazard was categorized as critical areas of ecological hazard (Figure 8). The factors influencing ecological hazard in these ten areas (a–j) were then identified. The results showed that interacting natural influences were drivers of ecological hazard, with human socioeconomic factors also playing important roles. As shown in Table 6, GDP had the highest impacts on ecological hazard in b, e, and g. The deterioration in ecological hazard in b was at the junction of BK and XA provinces in the economic zone of the Golden Triangle. Commercial mining of gems and gold constitutes significant local economic activities (Kyophilavong, 2009). The area of deterioration in e spanned XA and VI provinces. These areas have rich wood and lignite resources and host important crop production areas (Pathammavong et al., 2017). The area of deterioration in g was in central and western SV province. SV province contains the second largest city in Laos, after Vientiane, and adjacent to Thailand. This region hosts intense economic activities, particularly near the Mekong River and the local bus terminal (Fujita and Phanvilay, 2008). "Distance from the town center" had the highest effect on ecological hazard in northern CH province. This area constitutes an important political and economic center in Laos with a rich historical and natural heritage. Consequently, this area has a prosperous tourism industry. "Distance

TABLE 5 Factors influencing landscape ecological hazard (%) in Laos between 2000 and 2020.

Influencing factor	Factor	2000	2010	2020
Natural factor	Elevation	3.83	2.21	2.58
	Slope	4.10	2.37	3.29
	Annual precipitation	6.65	9.42	5.58
	Annual average temperature	8.85	8.09	9.50
	Vegetation coverage	8.32	8.93	10.40
	Soil organic matter content	5.66	4.46	5.72
Social and economic factor	Night light	1.42	0.91	0.84
	Population density	4.28	3.48	3.42
	Kilometer grid GDP	12.62	10.15	12.24
Regional accessibility factor	Distance from town center	12.71	10.41	11.19
	Distance from road	9.75	9.82	11.58
	Distance from protected area	4.29	5.17	8.74
	Distance from water	10.75	11.45	12.80

from road" had the most impact on ecological hazard in b, c, and i. Areas b and c are experiencing rapid economic development, which has in turn led to investment in road infrastructure. The area of extreme ecological hazard in i hosts a vital highway that connects Laos with Thailand, Cambodia, and Vietnam. In particular, the opening of the Laos-Japan Bridge across the Mekong River that connects the Polofen Plateau, Thailand, and Si Phan Don has resulted in increased trade between Laos and Thailand and a flourishing Talat Sao Heung market near the bridge (Pathammavong et al., 2017). "Distance from a protected area" had the most enormous impact on ecological hazard in g. This area hosts significant natural resources and is adjacent to protected areas, including the Xe Bang Nouan National Biodiversity Reserve in the south, the Dong Phou Vieng National Reserve in the southeast, and the Phou Xang He National Reserve in the north. "Distance from water" had the most considerable impact on ecological hazard in c in north-central XA province. This is the only area in Laos to the west of the Mekong River and is a flat floodplain. This area hosts the largest lake in Laos (Pokhrel et al., 2018). "Night light" and "population density" showed minimal impacts on ecological hazard. Among natural influences, "elevation" had significant impacts on the ecological hazard in c, i, and j; "slope" in c and h; "annual precipitation" in e and j; "annual average temperature" in b, c and j; "vegetation coverage" in c and j; "soil organic matter content" in f and j. The mountainous terrain restricted the rapid expansion of urbanization in b and c. These areas had scattered construction land and relatively concentrated anthropogenic endeavors, resulting in high landscape disturbance. Anthropogenic influences had significant impacts on landscape ecological hazard. The proximity of natural and artificial landscape patches formed a barrier to urban expansion in g. Increased ecological hazard in j was driven by both natural and anthropogenic influences. The natural influences included "elevation", "annual precipitation", "annual average temperature", and "vegetation coverage". Anthropogenic influences included ""distance from the town center" and "distance from the road". In summary, at a local scale, landscape ecological hazard was affected by both human and natural influences, although the former dominated.

4 Discussion

4.1 The impact of alterations in land use/cover on landscape ecological hazard

Alterations in land use/cover are essential drivers of global environmental change and impact ecosystem services and biodiversity, thereby contributing to landscape ecological hazard (Wang et al., 2021). Previous studies (Lin et al., 2018; Zhang et al., 2018; Hoque et al., 2020) have shown that alterations in land use/cover directly drive ecological hazard. Therefore, assessment of the impacts of potential alterations in land use/cover on the ecological environment can assist in the optimization of land use for the management of ecological hazard. The prediction of alterations in land use/cover under

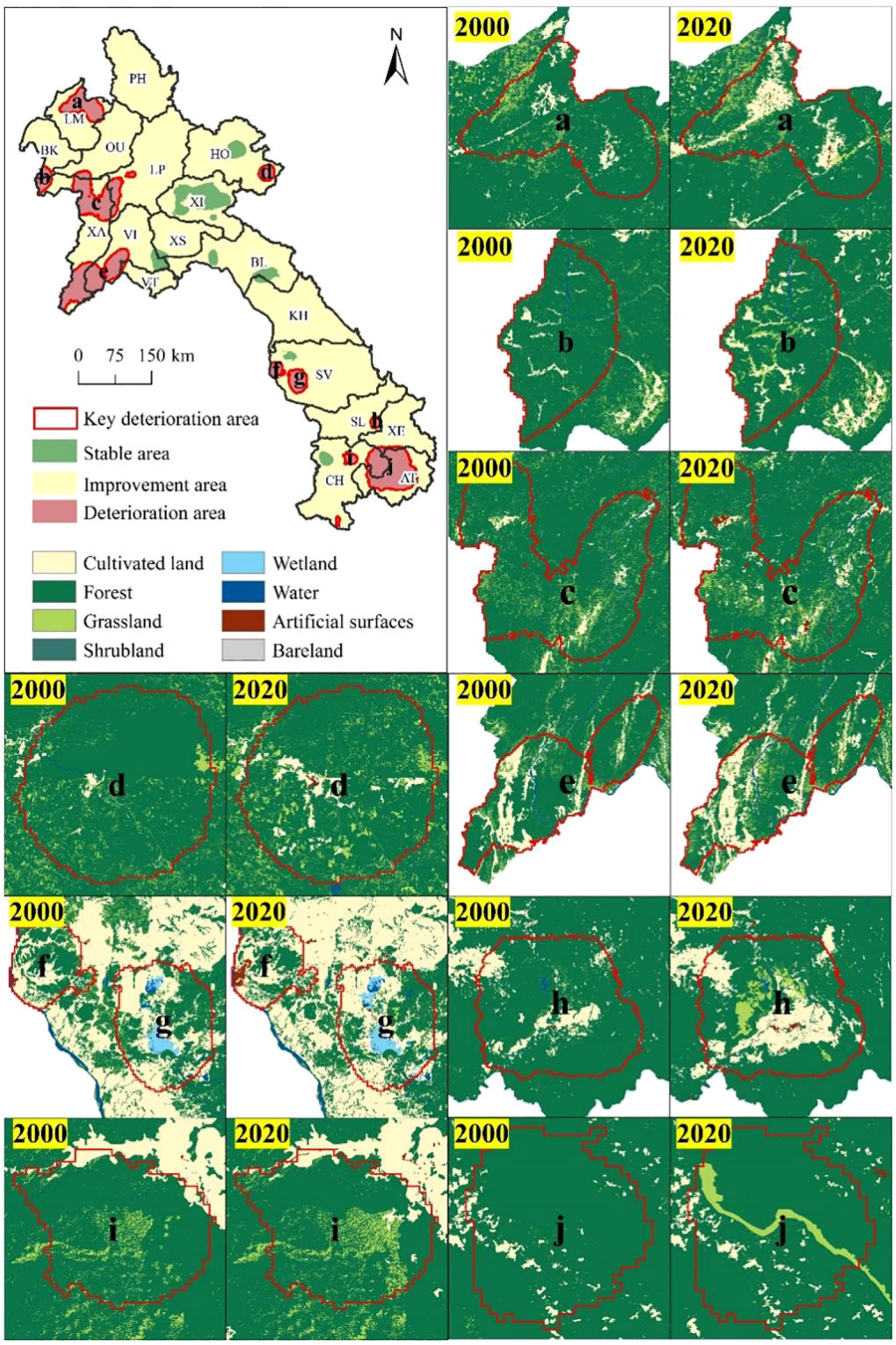


FIGURE 8
Changes in landscape ecological hazard in ecological conservation area in Laos and changes in land cover in deteriorating areas between 2000–2020. PH, Phongsali; LM, Louang Namtha; BK, Bokeo; OU, Oudomsai; LP, Louangphrabang; HO, Houaphan; XA, Xaignabouri; VI, Viangchan; XI, Xiangkhoang; XS, Xaisomboun; VT, Vientiane; BL, Bolikhamxai; KH, Khammouan; SL, Saravan; SV, Savannakhet; XE, Xekong; CH, Champasak; AT, Attapu. Ten areas (a–j) that showed increased landscape ecological hazard were categorized as critical areas of ecological hazard.

various scenarios can help formulate management strategies to minimize ecological hazard. The most comprehensive evaluation of landscape indices and ecological hazard is based on land use/cover (Li et al., 2020). Alterations in land use/cover driven by intense socioeconomic activities can change landscape patterns and processes, impacting ecological hazard. A fundamental dynamic relationship exists between internal drivers and alterations in land use/cover.

4.2 Changes in landscape patterns and scale effects

Changes in landscape patterns are driven by changes to ecological processes under natural or anthropogenic influences (Chen et al., 2013). The evolution of landscape patterns typically shows spatial heterogeneity and scale dependence (Kapustka et al., 2001). An in-depth study of the correlation between patterns of

10.3389/fevo.2023.1276239

TABLE 6 Factors affecting landscape ecological hazard (%) in Laos between 2000 and 2020.

Landscape ecological hazard deterioration area	Elevation	Slope	Annual precipitation	Annual average temperature	Vegetation coverage	Soil organic matter content	Night light	Population density	Kilometer grid GDP	Distance from town center	Distance from road	Distance from protected area	Distance from water
a	12.33	9.54	15.40	8.73	8.26	0.21	0.52	1.11	0.46	21.22	5.71	23.55	0.00
b	15.32	7.95	5.73	26.60	16.66	0.02	0.02	0.02	41.41	22.02	82.22	8.93	13.88
c	49.19	40.38	2.96	65.69	48.42	2.96	2.96	1.01	2.96	11.63	61.94	21.62	78.75
d	12.44	1.49	0.43	18.44	3.67	0.43	0.07	0.34	0.43	7.03	39.37	30.54	0.00
e	1.25	3.78	30.91	6.49	12.48	0.19	0.68	0.36	31.23	11.35	49.74	21.50	30.58
f	0.83	0.08	0.83	1.02	6.68	45.29	0.15	1.10	5.42	0.88	2.43	0.00	12.67
g	0.50	2.63	0.50	5.17	3.36	10.46	0.29	0.71	40.82	0.91	36.46	66.08	22.08
h	10.38	13.26	15.74	7.58	9.62	7.67	0.30	2.77	23.86	12.84	14.58	21.01	34.78
i	19.15	6.57	5.45	0.09	6.76	1.15	1.21	3.87	15.15	15.07	17.32	5.05	28.31
j	26.10	7.57	31.05	29.87	21.59	32.91	0.00	4.66	0.00	33.48	76.64	7.07	0.00

landscape and ecological process must consider the scale effect. This is because there is a need to focus on spatial structure and ecological process within landscape ecological hazard assessment, and the accuracy of the assessment depends on the choice of scale (Chen et al., 2022). Scale effects within patterns of landscape have remained an essential topic within landscape ecology research. Past related studies have focused on scale rules governing changes in landscape patterns and the construction of scale analysis methods. The study of the impact of scale in landscape ecology typically requires the consideration of two essential aspects, namely spatial resolution and extent (Suter, 1990), with the former relating to the spatial resolution of land use data and the latter to the size of the research units (Hunsaker et al., 1990). Recent studies on scale effects on landscape patterns have focused on spatial resolution, while there have been relatively few studies on the impact of spatial extent. There have been even fewer studies that have researched both aspects (Bayliss et al., 2012), and the few existing studies have primarily been conducted at watershed and urban landscape scales (Hope, 2006). The selection of an appropriate scale is vital for studying the scale effect, as the selection of a scale that is too large or too small can result in essential landscape details being ignored and too much focus on local processes (Peng et al., 2014). Quantitative studies of landscape pattern should fully consider factors affecting the landscape pattern index and the characteristics of landscape data. This is because the responses of different landscape pattern indices to scale effect differ as they are impacted by specific factors. Addressing the above complexity requires a multi-scale comprehensive evaluation focusing on an in-depth exploration of the scale effects of different influencing factors.

4.3 Mechanisms driving landscape ecological hazard in Laos

There has always been a focus on analyzing driving mechanisms of landscape ecological hazard within landscape ecology research (Qu et al., 2021). This analysis can help understand the evolution of landscape patterns for achieving sustainable development (Banks-Leite et al., 2022). However, the complexity of ecosystems prevents clear identification of direct drivers of the evolution of landscape patterns. Past studies have mainly applied correlation analysis to identify factors influencing landscape ecological hazard (Suter, 1990). Many studies have focused on the drivers of landscape ecological hazard conducted at different scales (Lin et al., 2021; Qu et al., 2021; Ai et al., 2022). These studies have shown that 60% of landscape ecological hazard can be attributed to human socioeconomic activities, whereas the remaining 40% is related to climate change. The results of the present study showed that the evolution of landscape ecological hazard in Laos over the last 20 years has been driven by both natural and anthropogenic influences. At a whole study area level, natural influences had the highest impact on the evolution of landscape ecological hazard, whereas anthropogenic influences had the most substantial local impacts. While past studies have resulted in a basic understanding of factors affecting landscape ecological hazard in Laos, the key factors driving landscape ecological hazard differ among regions. The present study

identified factors affecting local landscape ecological hazard and the underlying driving mechanisms.

4.4 Management of landscape ecological hazard prevention in Laos

The results of the present study showed that changes in the structure and quality of landscape forest, grassland, and cultivated land were critical to the ecological stability of Laos. The areas of different landscape ecological hazard face different challenges and pressures due to differences in natural and socioeconomic conditions. Therefore, these areas should adopt risk prevention and environment governance strategies suited to local conditions: (1) Areas of high and extreme ecological hazard should strengthen the conservation and restoration of forest, grassland, and cultivated land. These efforts should include biodiversity conservation, the return to forest and grassland of farmland, and comprehensive management of soil erosion and desertification; (2) Transformations occurring in areas of moderate ecological hazard are more complex, and these areas have weak ecological stability. Within these areas, there is a need to increase the connections between forest, grassland, and various landscape types to facilitate afforestation and grass cover. These efforts can help reduce the rate at which areas of moderate ecological hazard transfer to areas of high ecological hazard; (3) There are extensive forest areas in regions of low and minimal ecological hazard. Regardless, these areas face challenges of a simple forest structure and degradation of soil, water, and other ecological functions. There is a need to increase the resilience of the ecological environment in these regions by regulating the intensity of development and continuously improving the ecological environment.

5 Conclusion

The present study focused on Laos and constructed a landscape ecological hazard evaluation model utilizing a landscape pattern index. This model was used to analyze the spatiotemporal evolution of landscape ecological hazard and used geodetector to measure the degrees to which natural factors, social and economic factors, and regional accessibility factors explained localized and overall landscape ecological hazard in the study area. The factors influencing landscape ecological hazard were explored from global and local perspectives. The results of the present study are summarized below.

(1) Natural and anthropogenic stresses were drivers of changes to the area and number of patches of each landscape type, which in turn resulted in changes in the corresponding landscape type index. Landscape patch number reflected the heterogeneity of the landscape and showed a specific positive correlation with the fragmentation index. During the study period, the main land cover types in Laos were forest, cultivated land, and grassland. Changes in landscape types were higher between 2000 and 2020, with decreases in the area of forest (7123.84 km^2) and increases in the areas of other landscape types categorized as artificial surfaces. Areas of landscape types categorized as cultivated land and artificial surfaces increased significantly, which resulting in a transition from small, dispersed patches of landscape types to concentrated, contiguous, larger patches of landscape. Concurrently, the spatial aggregation of the landscape gradually increased, whereas wetland showed the highest abruption and fragmentation indices. Elevation showed an inverse relationship to spatial aggregation of the landscape. The other types of landscape patterns in the study area remained mostly stable.

(2) Areas of high ecological hazard were concentrated in central and northeastern Laos, whereas areas of low ecological hazard were concentrated in the northwest and southeast. There was an overall increase in average landscape ecological hazard in Laos over the study period, corresponding to a deterioration in the overall ecological status. The areas of high and extreme ecological hazard increased, whereas areas of low and minimal ecological hazard decreased. Transformations occurring in areas of moderate ecological hazard were relatively complex, thereby necessitating the need to focus on avoiding further deterioration due to fragmentation of landscape patches. The development of forestry and agriculture in the study area, needs to consider maintaining the original landscape as much as possible. In addition, there should be a focus on reducing the spatial scales of forestry and agriculture through intensive farming methods, which can increase the industrialization of agriculture and forestry while facilitating improved conservation of the local landscape environment.

(3) There were significant positive spatial correlations in the landscape ecological hazard index with relative stability in high- and low-value aggregation areas. There were significant differences between the spatial distributions of ecological pressure and ecological response, and ecological hazard in some areas was affected by multiple factors. Under the current landscape structure, there is a need to increaser ecological management in critical areas by optimizing resource allocation, conserving the natural environment, and ensuring sustainable development in a stable direction. The drivers of landscape aggregation in areas lower economic value and lower agglomeration in Laos remain poorly understood due to limited knowledge, indicating the need for further studies.

(4) The factors driving changes to land use patterns and landscape ecological hazard in Laos are interdependent and interact. At a national scale, natural factors had little impact on landscape ecological hazard in Laos, whereas human factors played a dominant role. The evolution of landscape ecological hazard in Laos is a function of the joint influences of natural and human factors. The present study restricted factor detection analysis to the application of a geodetector for identifying the contributions of influencing factors. However, the interactions between factors were not considered. Future research should further study the individual and interactive impacts of factors on landscape ecological hazard. Concurrently, future studies should focus on mechanisms of interaction between landscape ecological hazard and ecological processes and the ecological consequences, thereby increasing the practical significance of such studies.

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

JM: Conceptualization, Data curation, Methodology, Software, Validation, Writing – original draft. VK: Conceptualization, Validation, Writing – review & editing. JW: Conceptualization, Methodology, Validation, Writing – review & editing. JZ: Investigation, Software, Visualization, Writing – review & editing. WL: Data curation, Formal Analysis, Visualization, Writing – review & editing. XZ: Formal Analysis, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Details of all funding sources should be provided, including grant numbers if applicable. Please ensure to add all necessary funding information, as after publication this is no longer possible. This research was funded by the Multi-Government International Science and Technology Innovation Cooperation Key Project of the National Key Research and Development Program of China for the "Environmental monitoring and assessment of land use/land cover change impact on ecological security using geospatial technologies" (2018YFE0184300), the National Natural Science Foundation of China (41961060), and the China Scholarship Council (202008090261).

Acknowledgments

Thanks to the Academic Editor, and the Editor. In addition, we also want to thank the Program for Innovative Research Team (in Science and Technology) in the University of Yunnan Province [grant number IRTSTYN].

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

Ai, J., Yu, K., Zeng, Z., Yang, L., Liu, Y., and Liu, J. (2022). Assessing the dynamic landscape ecological risk and its driving forces in an island city based on optimal spatial scales: Haitan Island, China. *Ecol. Indic.* 137, 108771. doi: 10.1016/j.ecolind.2022.108771

Anselin, L. (1996). *Interactive Techniques and Exploratory Spatial Data Analysis* (West Virginia University: Regional Research Institute Working Papers).

Anselin, L. (2003). *An introduction to spatial autocorrelation analysis with GeoDa* (Champagne-Urbana, Illinois: Spatial Analysis Laboratory, University of Illinois). Available at: https://www.semanticscholar.org/paper/Interactive-Techniques-and-Exploratory-Spatial-Data-Anselin/1ff35696e7a12a09cf95abbb2778717c4e23cb73.

Bai, H., and Weng, L. (2023). Ecological security pattern construction and zoning along the China-Laos Railway based on the potential-connectedness-resilience framework. *Ecol. Indic.* 146, 109773. doi: 10.1016/j.ecolind.2022.109773

Baiyinbaoligao, Liu, H., Chen, X., and Mu, X. (2020). "Overview of the Mekong River Basin," Flood Prevention and Drought Relief in Mekong River Basin. Singapore: Springer. 1–25.

Banks-Leite, C., Betts, M. G., Ewers, R. M., Orme, C. D. L., and Pigot, A. L. (2022). The macroecology of landscape ecology. *Trends Ecol. Evol.* 37, 480–487. doi: 10.1016/j.tree.2022.01.005

Bartell, S. M. (2006). Biomarkers, bioindicators, and ecological risk assessment—A brief review and evaluation. *Environ. Bioindicators.* 1, 60–73. doi: 10.1080/15555270591004920

Bayliss, P., van Dam, R. A., and Bartolo, R. E. (2012). Quantitative ecological risk assessment of the Magela creek floodplain in Kakadu National Park, Australia: comparing point source risks from the ranger uranium mine to diffuse landscape-scale risks. *Hum. Ecol. Risk Assess.* 18, 115–151. doi: 10.1080/10807039.2012.632290

Castella, J. C., Lestrelin, G., Hett, C., Bourgoin, J., Fitriana, Y. R., Heinimann, A., et al. (2013). Effects of landscape segregation on livelihood vulnerability: Moving from extensive shifting cultivation to rotational agriculture and natural forests in northern Laos. *Hum. Ecol.* 41, 63–76. doi: 10.1007/s10745-012-9538-8

Chapman, P. M., Fairbrother, A., and Brown, D. (1998). A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ. Toxicol. Chem.* 17, 99–108. doi: 10.1002/etc.5620170112

Chen, S., Chen, B., and Fath, B. D. (2013). Ecological risk assessment on the system scale: A review of state-of-the-art models and future perspectives. *Ecol. Modell.* 250, 25–33. doi: 10.1016/j.ecolmodel.2012.10.015

Chen, Y., Wang, J., Kurbanov, E., Thomas, A., Sha, J., Jiao, Y., et al. (2022). Ecological security assessment at different spatial scales in central Yunnan Province, China. *PloS One* 17, e0270267. doi: 10.1371/journal.pone.0270267

Diniz-Filho, J. A. F., Bini, L. M., and Hawkins, B. A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Glob Ecol. Biogeogr.* 12, 53–64. doi: 10.1046/j.1466-822X.2003.00322.x

Fujita, Y., and Phanvilay, K. (2008). Land and forest allocation in Lao People's Democratic Republic: comparison of case studies from community-based natural resource management research. *Soc. Nat. Resour.* 21, 120–133. doi: 10.1080/08941920701681490

Gao, L., Tao, F., Liu, R., Wang, Z., Leng, H., and Zhou, T. (2022). Multi-scenario simulation and ecological risk analysis of land use based on the PLUS model: A case study of Nanjing. *Sustain Cities Soc* 85, 104055. doi: 10.1016/j.scs.2022.104055

Gong, J., Cao, E., Xie, Y., Xu, C., Li, H., and Yan, L. (2021). Integrating ecosystem services and landscape ecological risk into adaptive management: Insights from a western mountain-basin area, China. *J. Environ. Manage.* 281, 111817. doi: 10.1016/j.jenvman.2020.111817

Hengl, T. (2009). "A practical guide to geostatistical mapping," in *Extended edition of the EUR 22904 EN Scientific and Technical Research series report published by 10 Office for Official Publications of the European Communities*, *2nd edn*(Luxembourg: Geoderma, Elsevier BV), 293. Available at: https://www.semanticscholar.org/paper/A-Practical-Guide-to-Geostatistical-Mapping-Hengl/756c4f3e378febc3978210934e2e3fb9c67221b9.

Hope, B. K. (2006). An examination of ecological risk assessment and management practices. *Environ. Int.* 32, 983–995. doi: 10.1016/j.envint.2006.06.005

Hoque, M. Z., Cui, S., Islam, I., Xu, L., and Tang, J. (2020). Future impact of land use/land cover changes on ecosystem services in the lower Meghna river estuary, Bangladesh. *Sustainability* 12, 2112. doi: 10.3390/su12052112

Huang, Q., Zhu, X., Liu, C., Wu, W., Liu, F., and Zhang, X. (2020). Spatial-temporal evolution and determinants of the belt and road initiative: A maximum entropy gravity model approach. *Chin. Geogr. Sci.* 30, 839–854. doi: 10.1007/s11769-020-1144-x

Hunsaker, C. T., Graham, R. L., Suter, G. W.II, O'Neill, R. V., Barnthouse, L. W., and Gardner, R. H. (1990). Assessing ecological risk on a regional scale. *Environ. Manage.* 14, 325–332. doi: 10.1007/BF02394200

Kapustka, L. A., Galbraith, H., Luxon, M., and Yocum, J. (2001). Using landscape ecology to focus ecological risk assessment and guide risk management decision-making. *Toxicol. Ind. Health* 17, 236–246. doi: 10.1191/0748233701th121oa

Koenig, W. D. (1999). Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.* 14, 22–26. doi: 10.1016/S0169-5347(98)01533-X

Kyophilavong, P. (2009). "Mining sector in Laos," in *Major Industries and Business Chance in CLMV Countries BRC Report* (Thailand: Bangkok Research Center (BRC), vol. 2. , 69–100. Available at: https://www.researchgate.net/publication/281693388_Mining_Sector_in_Laos.

Lestrelin, G., Castella, J. C., and Bourgoin, J. (2012). Territorialising sustainable development: The politics of land-use planning in Laos. *J. Contemp Asia.* 42, 581–602. doi: 10.1080/00472336.2012.706745

Li, J., Pu, R., Gong, H., Luo, X., Ye, M., and Feng, B. (2017). Evolution characteristics of landscape ecological risk patterns in coastal zones in Zhejiang province, China. *Sustainability* 9, 584. doi: 10.3390/su9040584

Li, W., Wang, Y., Xie, S., Sun, R., and Cheng, X. (2020). Impacts of landscape multifunctionality change on landscape ecological risk in a megacity, China: A case study of Beijing. *Ecol. Indic.* 117, 106681. doi: 10.1016/j.ecolind.2020.106681

Li, J., Wang, J., Zhang, J., Liu, C., He, S., and Liu, L. (2022). Growing-season vegetation coverage patterns and driving factors in the China-Myanmar Economic Corridor based on Google Earth Engine and geographic detector. *Ecol. Indic.* 136, 108620. doi: 10.1016/j.ecolind.2022.108620

Liang, Y., Li, S., Zeng, J., and Wu, T. (2022). Examining the impact of multidimensional accessibility on regional poverty in Laos. *Appl. Geogr.* 148, 102789. doi: 10.1016/j.apgeog.2022.102789

Lin, D., Liu, F., Zhang, J., Hao, H., and Zhang, Q. (2021). Research progress on ecological risk assessment based on multifunctional landscape. *J. Resour Ecol.* 12, 260–267. doi: 10.5814/j.issn.1674-764x.2021.02.012

Lin, X., Xu, M., Cao, C. P., Singh, R., Chen, W., and Ju, H. (2018). Land-use/land-cover changes and their influence on the ecosystem in Chengdu City, China during the period of 1992–2018. *Sustainability* 10, 3580. doi: 10.3390/su10103580

Liu, L., Li, J., Wang, J., Liu, F., Cole, J., Sha, J., et al. (2022). The establishment of an eco-environmental evaluation model for southwest China and eastern South Africa based on the DPSIR framework. *Ecol. Indic.* 145, 109687. doi: 10.1016/j.ecolind.2022.109687

Liu, Y. G., Wang, N. L., Wang, L. G., Zhao, Y. Q., and Wu, X. B. (2013). Application of Gis in regional ecological risk assessment of water resources. *Environ. Eng. Manag J.* 12, 1465–1474. doi: 10.30638/eemj.2013.180

Malekmohammadi, B., and Rahimi Blouchi, L. (2014). Ecological risk assessment of wetland ecosystems using Multi Criteria Decision Making and Geographic Information System. *Ecol. Indic.* 41, 133–144. doi: 10.1016/j.ecolind.2014.01.038

Munns, J. W. R., Rea, A. W., Suter, II, G.W., Martin, L., Blake-Hedges, L., Crk, T., et al. (2016). Ecosystem services as assessment endpoints for ecological risk assessment. *Integr. Environ. Assess. Manage.* 12, 522–528. doi: 10.1002/ieam.1707

Oliver, M. A., and Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. *Int. J. Geogr. Inf Syst.* 4, 313–332. doi: 10.1080/02693799008941549

Pathammavong, B., Kenney-Lazar, M., and Sayaraj, E. V. (2017). Financing the 450 Year Road: Land expropriation and politics A'll the way down'in Vientiane, Laos. *Dev. Change.* 48, 1417–1438. doi: 10.1111/dech.12339

Paustenbach, D. J. (2015). *Human and Ecological Risk Assessment: Theory and Practice (Wiley Classics Library)* (London: John Wiley & Sons). Available at: https://www.tandfonline.com/doi/abs/10.1080/713610025.

Peng, J., Liu, Y., Pan, Y., Zhao, Z., Song, Z., and Wang, Y. (2014). Study on the correlation between ecological risk due to natural disaster and landscape pattern-process: review and prospect. *Adv. Earth Sci.* 29, 1186. doi: 10.11867/j.issn.1001-8166.2014.10.1186

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., and Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* 20, 503–510. doi: 10.1016/j.tree.2005.05.011

Phimphanthavong, H. (2013). The determinants of sustainable development in Laos. *Int. J. Acad. Res. Manage. (IJARM).* 3, 2296–1747.

Pokhrel, Y., Burbano, M., Roush, J., Kang, H., Sridhar, V., and Hyndman, D. W. (2018). A review of the integrated effects of changing climate, land use, and dams on Mekong River hydrology. *Water* 10, 266. doi: 10.3390/w10030266

Qu, Y., Zong, H., Su, D., Ping, Z., and Guan, M. (2021). Land use change and its impact on landscape ecological risk in typical areas of the Yellow River Basin in China. *Int. J. Environ. Res. Public Health* 18, 11301. doi: 10.3390/ijerph182111301

Ran, P., Hu, S., Frazier, A. E., Qu, S., Yu, D., and Tong, L. (2022). Exploring changes in landscape ecological risk in the Yangtze River Economic Belt from a spatiotemporal perspective. *Ecol. Indic.* 137, 108744. doi: 10.1016/j.ecolind.2022.108744

Su, S., Xiao, R., Jiang, Z., and Zhang, Y. (2012). Characterizing landscape pattern and ecosystem service value changes for urbanization impacts at an eco-regional scale. *Appl. Geogr.* 34, 295–305. doi: 10.1016/j.apgeog.2011.12.001

Sunderlin, W. D. (2006). Poverty alleviation through community forestry in Cambodia, Laos, and Vietnam: An assessment of the potential.. *For. Policy Econ.* 8, 386–396.

Suter, II, G.W. (1990). Endpoints for regional ecological risk assessments. *Environ. Manage.* 14, 9–23. doi: 10.1007/BF02394015

Suter, II, G.W. (2001). Applicability of indicator monitoring to ecological risk assessment. *Ecol. Indic.* 1, 101–112. doi: 10.1016/S1470-160X(01)00011-5

Wang, H., Liu, X., Zhao, C., Chang, Y., Liu, Y., Zang, F., et al. (2021). Spatial-temporal pattern analysis of landscape ecological risk assessment based on land use/land cover change in Baishuijiang National Nature Reserve in Gansu Province, China. *Ecol Indic.* 124, 107454.

Wang, J., Sui, L., Yang, X., Wang, Z., Ge, D., Kang, J., et al. (2019). Economic globalization impacts on the ecological environment of inland developing countries: A case study of Laos from the perspective of the land use/cover change. *Sustainability* 11, 3940. doi: 10.3390/su11143940

Xu, X., Peng, Y., and Qin, W. (2023). Simulation, prediction and driving factor analysis of ecological risk in Savan District, Laos. *Front. Environ. Sci.* 10, 1058792. doi: 10.3389/fenvs.2022.1058792

Yanes, A., Botero, C. M., Arrizabalaga, M., and Vásquez, J. G. (2019). Methodological proposal for ecological risk assessment of the coastal zone of Antioquia, Colombia. *Ecol. Eng.* 130, 242–251. doi: 10.1016/j.ecoleng.2017.12.010

Zhang, W., Chang, W. J., Zhu, Z. C., and Hui, Z. (2020). Landscape ecological risk assessment of Chinese coastal cities based on land use change. *Appl. Geogr.* 117, 102174. doi: 10.1016/j.apgeog.2020.102174

Zhang, D., Jing, P., Sun, P., Ren, H., and Ai, Z. (2022). The non-significant correlation between landscape ecological risk and ecosystem services in Xi'an Metropolitan Area, China. *Ecol. Indic.* 141, 109118. doi: 10.1016/j.ecolind.2022.109118

Zhang, F., Yushanjiang, A., and Wang, D. (2018). Ecological risk assessment due to land use/cover changes (LUCC) in Jinghe County, Xinjiang, China from 1990 to 2014 based on landscape patterns and spatial statistics. *Environ. Earth Sci.* 77, 491. doi: 10.1007/s12665-018-7676-z

Zhang, S., Zhong, Q., Cheng, D., Xu, C., Chang, Y., Lin, Y., et al. (2022). Landscape ecological risk projection based on the PLUS model under the localized shared socioeconomic pathways in the Fujian Delta region. *Ecol. Indic.* 136, 108642. doi: 10.1016/j.ecolind.2022.108642

Zhu, L., Meng, J., and Zhu, L. (2020). Applying Geodetector to disentangle the contributions of natural and anthropogenic factors to NDVI variations in the middle reaches of the Heihe River Basin. *Ecol. Indic.* 117, 106545. doi: 10.1016/j.ecolind.2020.106545

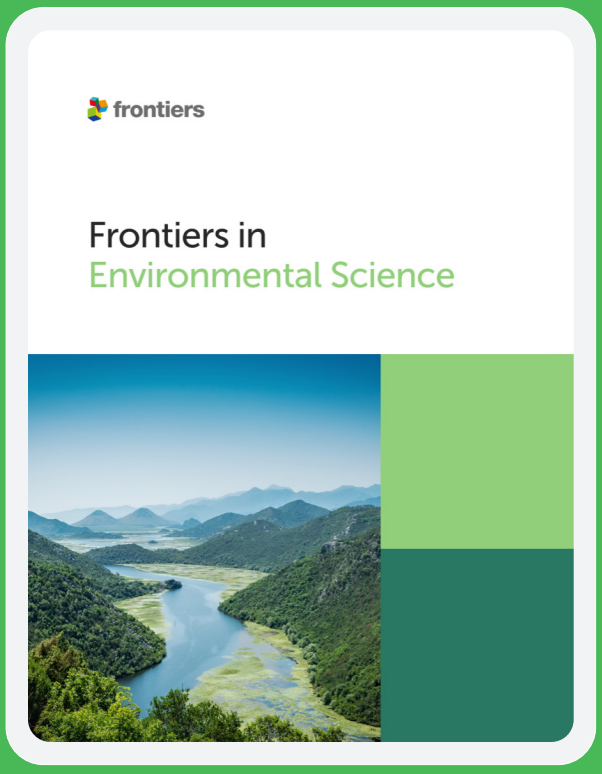
Frontiers in Environmental Science

Explores the anthropogenic impact on our natural world

An innovative journal that advances knowledge of the natural world and its intersections with human society. It supports the formulation of policies that lead to a more inhabitable and sustainable world.

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