frontiers | Research Topics

Advances in autonomous ships (AS) for ocean observation

Edited by
Xinyu Zhang, Defeng Wu, Yanzhen Gu, Gorka Bidegain and Yanlong Chen

Coordinated by
Chengbo Wang

Published in
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-5584-2
DOI 10.3389/978-2-8325-5584-2

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

Advances in autonomous ships (AS) for ocean observation

Topic editors
Xinyu Zhang — Dalian Maritime University, China
Defeng Wu — Jimei University, China
Yanzhen Gu — Zhejiang University, China
Gorka Bidegain — University of the Basque Country, Spain
Yanlong Chen — National Marine Environmental Monitoring Center, China

Topic coordinator
Chengbo Wang — University of Science and Technology of China, China

Citation
Zhang, X., Wu, D., Gu, Y., Bidegain, G., Chen, Y., Wang, C., eds. (2024). *Advances in autonomous ships (AS) for ocean observation*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5584-2

Table of contents

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 09 October 2024
DOI 10.3389/fmars.2024.1498084

OPEN ACCESS

EDITED AND REVIEWED BY
Hervé Claustre,
Centre National de la Recherche Scientifique (CNRS), France

*CORRESPONDENCE
Yanlong Chen
ylchen_dl@163.com
Chengbo Wang
wangcb_dlmu@163.com

RECEIVED 18 September 2024
ACCEPTED 27 September 2024
PUBLISHED 09 October 2024

CITATION
Zhang X, Chen Y, Bidegain G, Wu D, Gu Y and Wang C (2024) Editorial: Advances in autonomous ships (AS) for ocean observation. *Front. Mar. Sci.* 11:1498084. doi: 10.3389/fmars.2024.1498084

Editorial: Advances in autonomous ships (AS) for ocean observation

Xinyu Zhang[1], Yanlong Chen[2]*, Gorka Bidegain[3,4], Defeng Wu[5], Yanzhen Gu[6] and Chengbo Wang[7]*

[1]College of Navigation, Dalian Maritime University, Dalian, China, [2]Marine Remote Sensing Technology Team, National Marine Environmental Monitoring Center, Dalian, China, [3]Department of Applied Mathematics, Faculty of Engineering of Gipuzkoa, University of the Basque Country (UPV/EHU), Donostia, Gipuzkoa, Spain, [4]Research Centre for Experimental Marine Biology and Biotechnology, Plentzia Marine Station, University of the Basque Country (PiE-UPV/EHU), Plentzia, Bizkaia, Spain, [5]School of Marine Engineering, Jimei University, Xiamen, China, [6]Ocean College, Zhejiang University, Hangzhou, China, [7]Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei, China

KEYWORDS

autonomous ships, ocean observation, task decision-making, path planning, control, data analysis

Editorial on the Research Topic
Advances in autonomous ships (AS) for ocean observation

Introduction

Ocean observation is the basis for understanding and studying marine science. In recent years, the application of autonomous ships (AS), including Unmanned Surface Vessels (USVs), Autonomous Underwater Vehicles (AUVs), and Remotely Operated Vehicles (ROVs), in ocean observation has gained significant traction due to their capability to perform maritime autonomous tasks of oceans efficiently and safely in challenging marine environments. Compared with traditional technical means, the unique technical capability of ASs in marine environment observation is the ability to maneuver on demand under the influence of complex marine environments. Therefore, giving full play to its controllable maneuverability and realizing its perception, task decision-making, path planning, control, and perception data analysis is the key to its application. Equipped with advanced sensors and instruments, these vessels can gather critical ocean data over large areas and long durations, providing invaluable insights for marine scientists.

This editorial aims to highlight the latest advancements in AS technology and their implications for ocean science, particularly the integration of Artificial Intelligence (AI) and Machine Learning (ML). These innovations have the potential to greatly enhance the efficiency and accuracy of ocean observation, transforming the field of marine science.

Contributing articles and main conclusions

This Research Topic comprises eleven high-quality papers, each contributing to many different aspects of autonomous ships (AS) for ocean observation. In the realm of enhanced data collection techniques, Berild et al. sampled river plume fronts in three-dimensional space using AUVs. This model addresses critical challenges in coastal environments impacted by climate change and human activities. In another study, AUVs equipped with interferometric side-scan sonar were used to monitor aquaculture setups in high-energy shallow water environments (Peck et al.). Lei et al. developed a novel calibration method for the Simulating Waves Nearshore wave model, incorporating the white-capping dissipation term. Validated across diverse global locations, including the South China Sea, Gulf of Mexico, and Mediterranean Sea, this method demonstrates broad applicability in wave modeling. For the detection of small marine targets, Cheng et al. proposed an enhanced method based on the YOLOv7 model to detect small targets in SSS images, and introduced a global attention mechanism to focus on global information and extract target features. Experimental results show that this method can be applied to autonomous target detection in USVs and AUVs, thereby enhancing the autonomous operation capability of unmanned autonomous ocean observation platforms. The development of hydrodynamic simulation tools for ROVs has led to better understanding of the forces acting on these vehicles during operation (Zhang et al.). Such simulations are instrumental in improving the design and maneuverability of underwater vehicles, which is essential for complex tasks such as monitoring volcanic activities around active volcanoes (Tada et al.). In complex ocean environments, multiple ASs are required to collaborate to complete observation tasks. Kang et al. demonstrated the potential to improve the efficiency of maritime operations through collaborative ocean observation research by communicating heterogeneous USVs. Furthermore, adaptive terminal sliding mode control schemes have been developed to maintain the formation of USVs and ROVs even under deceptive attacks (Zhang et al.). In terms of innovative imaging technologies for marine science, to address the challenges posed by adverse weather conditions, such as rain, and haze, a prompt-based learning method was proposed for maritime image restoration by He et al. This method enhances the quality of maritime images, which is essential for navigation, fishing, and search and rescue operations. Additionally, hybrid dynamic transformers have been developed for underwater image super-resolution (He et al.), significantly improving the clarity and detail of underwater imagery. In the aspect of maritime and ocean observation understanding and decision support, Li et al. introduced a framework utilizing knowledge graph technology to analyze maritime data. By integrating Automatic Identification System data with spatial information from port facilities, they created semantic connections among ships, berths, and waterways. This approach enhances ship identification and berth allocation, improving decision-making for intelligent maritime systems.

In summary, these collective efforts underscore a comprehensive approach to advancing maritime research and technology. By leveraging the capabilities of Autonomous Ships (ASs) and integrating sophisticated modeling, autonomous systems, image processing, and data analysis techniques, researchers are addressing complex challenges in marine science. These advancements not only enhance our ability to monitor and understand marine environments more effectively but also improve the efficiency and safety of oceanographic research. The integration of AI and ML within AS technology exemplifies how innovation is transforming ocean observation, offering valuable insights into oceanic systems and facilitating better management of marine resources.

Author contributions

XZ: Writing – original draft, Writing – review & editing. YC: Writing – original draft, Writing – review & editing. GB: Writing – original draft, Writing – review & editing. DW: Writing – original draft, Writing – review & editing. YG: Writing – original draft, Writing – review & editing. CW: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by National Natural Science Foundation of China under Grant No. 52371359.

Acknowledgments

We are grateful to all authors and reviewers for their hard work on this Research Topic, on behalf of the Guest Associate Editors. We anticipate that this will stimulate more research into advances in autonomous ships (AS) for ocean observation.

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

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 30 November 2023
DOI 10.3389/fmars.2023.1298727

OPEN ACCESS

EDITED BY
Daniele Hauser,
UMR8190 Laboratoire Atmosphères, Milieux, Observations Spatiales (LATMOS), France

REVIEWED BY
Aifeng Tao,
Hohai University, China
José Pinho,
University of Minho, Portugal

*CORRESPONDENCE
Yanzhen Gu
guyanzhen@zju.edu.cn
Fangguo Zhai
gfzhai@ouc.edu.cn

RECEIVED 22 September 2023
ACCEPTED 13 November 2023
PUBLISHED 30 November 2023

CITATION
Lei Z, Wu W, Gu Y, Zhai F and Li P (2023) A general method to determine the optimal whitecapping dissipation coefficient in the SWAN model.
Front. Mar. Sci. 10:1298727.
doi: 10.3389/fmars.2023.1298727

A general method to determine the optimal whitecapping dissipation coefficient in the SWAN model

Zhifeng Lei[1], Wenfan Wu[2], Yanzhen Gu[1,3,4,5]*, Fangguo Zhai[2]* and Peiliang Li[1,3,5]

[1]Institute of Physical Oceanography and Remote Sensing, Ocean College, Zhejiang University, Zhoushan, China, [2]College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao, China, [3]Hainan Institute, Zhejiang University, Sanya, China, [4]Laoshan Laboratory, Qingdao, China, [5]Hainan Observation and Research Station of Ecological Environment and Fishery Resource in Yazhou Bay, Sanya, China

Whitecapping dissipation is a critical term in affecting the accuracy of wave height modeling. However, the whitecapping dissipation coefficient (C_{ds}), as a primary factor influencing whitecapping, is commonly determined through trial and error in various studies. In this study, we present a general method for calibrating the Simulating Waves Nearshore (SWAN) wave model using the whitecapping dissipation term, demonstrated through a detailed study in the South China Sea (SCS). Theoretical analysis reveals that the optimal C_{ds} value shows a one-to-one correspondence with the applied wind field. Expectedly, under high-quality wind field conditions, the optimal C_{ds} values tend to fall within a narrow range, regardless of the model domain or time span. Numerical experiments executed in the SCS further consolidated this inference, encompassing two common wind input schemes (ST6 and YAN) and three distinct whitecapping dissipation schemes (KOMEN, JANSSEN, and WST). Based on the experimental results, we have identified an optimal C_{ds} range for each whitecapping dissipation scheme. C_{ds} values within the optimal range consistently outperformed the default C_{ds} in the SWAN model. Subsequent experiments verified the method's applicability to the Gulf of Mexico and the Mediterranean Sea. The findings suggest that this research holds substantial promise for practical applications on a global scale.

KEYWORDS

SWAN, whitecapping dissipation coefficient, wind errors, ERA5, SARAL

1 Introduction

In recent decades, the third-generation wave models that can solve the spectral action balance equation without assuming *a priori* spectral shape (The WAMDI Group, 1988; Booij et al., 1999) have been widely developed and applied worldwide (Cavaleri et al., 2020; Shao et al., 2023). Among the terms in the wave action equation,

whitecapping, responsible for energy dissipation in deep water, remains one of the least understood physical aspects (Rogers et al., 2002; Cavaleri et al., 2019). Several whitecapping expressions had been proposed (Hasselmann, 1974; Komen et al., 1984; Janssen, 1992; Banner et al., 2000; Alves and Banner, 2003; Van der Westhuysen et al., 2007). Early whitecapping expressions were adjusted based on closing the energy balance of waves in fully developed conditions, as exemplified in the work of Komen et al. (1984) (hereinafter referred to as the KOMEN expression or KOMEN). Alves and Banner (2003) refined the KOMEN expression based on observations, proposing an expression predominantly reliant on the ratio of azimuthal-integrated spectral saturation to the saturation spectrum. Van der Westhuysen et al. (2007) proposed a novel whitecapping expression based on Alves and Banner (2003) but removed the dependence on mean spectra, increasing its suitability for nearshore applications. The Simulating Waves Nearshore (SWAN) model, one of the prominent representatives of third-generation wave models, provides 6 whitecapping expressions with 19 adjustable parameters (The SWAN team, 2021a).

Whitecapping plays a fundamental role in achieving the correct energy balance and significantly influences the accuracy of wave models (Roland and Ardhuin, 2014; Cavaleri et al., 2019). Therefore, the selection and tuning of the whitecapping scheme are crucial (Sun et al., 2022). The whitecapping dissipation coefficient (C_{ds}) is a key parameter that integrally controls the whitecapping dissipation, which is not dependent on the wave steepness or the wave number (Sun et al., 2022). Among all the parameters in the whitecapping schemes, C_{ds} is usually used as a tuning parameter in the calibration process (Cavaleri et al., 2018; Cavaleri et al., 2020). Extensive research indicates that suitable schemes and optimal parameters may vary by region or forcing wind field (Shao et al., 2023). For instance, in the Bohai Sea, the KOMEN expression effectively represents the wind-wave characteristics with the C_{ds} of 2.2E-5 (Lv et al., 2014). Appendini et al. (2013) and Amarouche et al. (2019) improved the model performance using the expression proposed by Janssen (1992) (hereinafter referred to as the JANSSEN expression or JANSSEN) but with entirely different optimal C_{ds} in the Mediterranean Sea (MS). Appendini et al. (2013) achieved the optimal combination of C_{ds} at 1.5 and δ (the coefficient determining the dependence of whitecapping on the wave number) at 0.7, whereas Amarouche et al. (2019) found the optimal C_{ds} to be 1.0. Off the west coast of Norway, the expression proposed by Van der Westhuysen et al. (2007) provided the best performance with mixed swell-wind sea conditions (Van der Westhuysen et al., 2007; Christakos et al., 2021) found that default settings of the whitecapping dissipation scheme commonly led to overestimation of the peak frequency and underestimation of the energy level of the spectral peak during high wind speed conditions. Consequently, wave parameters such as significant wave height (SWH) and mean wave period may be underestimated (Elkut et al., 2021; Umesh and Behera, 2021). Although calibrating the model using C_{ds} lacks a valid physical basis, as the simulated whitecapping dissipation may not accurately reflect realistic conditions, this approach is widely acknowledged for improving practical wave simulations (Wu et al., 2021; Bujak et al., 2023).

Most wave modeling studies typically involve conducting multiple experiments over a certain range of C_{ds} and determining the optimal C_{ds} based on simulation results (e.g., Akpinar and Ponce de León, 2016; Kutupoğlu et al., 2018; Bingölbali et al., 2019; Sun et al., 2022). Wu et al. (2021) proposed a novel C_{ds} calibration method that requires at least two experiments to determine the optimal C_{ds}. However, this method relies on a fitting formula based on experimental results, inevitably introducing fitting errors. Although both methods can obtain the optimal C_{ds}, they require significant time and computational resources. Therefore, finding a general method to determine the optimal C_{ds} value is necessary. Overall, the accuracy of wave model results, particularly in SWH, is strongly influenced by the forcing wind field and source term parameterization (Cavaleri and Bertotti, 1997; Zhai et al., 2021). The wind field provides positive energy flux to the wave model, while the dissipation term contributes to negative energy flux (Babanin et al., 2010). Thus, there exists a potential relationship between the wind field and C_{ds}. Based on this concept, it is theoretically feasible to calibrate the wave model.

This study aims to propose a general method for determining the optimal C_{ds} to improve the efficiency of wave simulation. The remaining parts of this paper are structured as follows. Section 2 describes the study area and bathymetry data. Section 3 details the primary data and methods, including the basic principle of SWAN, observations, and error metrics. Section 4 presents the theoretical basis of this work. In Section 5, numerical experiments are conducted to explore the characteristics of the optimal C_{ds}. Section 6 discusses the applicability of the conclusions we have obtained to different regions. Finally, Section 7 provides the conclusion.

2 Study area and bathymetry

The study area of this work encompasses the South China Sea (SCS; 104°E–124°E, 0°–25°N), as delineated by the solid black line box in Figure 1. The SCS is a typical semi-enclosed marine region, connected to the Pacific Ocean and the Indian Ocean through narrow straits or channels (Su et al., 2017; Ou et al., 2018). It features intricate topography, characterized by three distinct elements: the continental shelf that connects to the land, the continental slope at the outer edge of the continental shelf, and the central basin. As depicted in Figure 1, this region is marked by significant variations in water depth, with a maximum depth exceeding 5500 meters and an average depth of approximately 1200 meters. The general pattern is one of shallow waters in the north and south and deeper waters in the central area (Zhang et al., 2020). Given its strategic importance in shipping and trade routes, along with its abundant reserves of oil, natural gas, and fisheries, the SCS holds significant economic and geopolitical value (Wang, 2021).

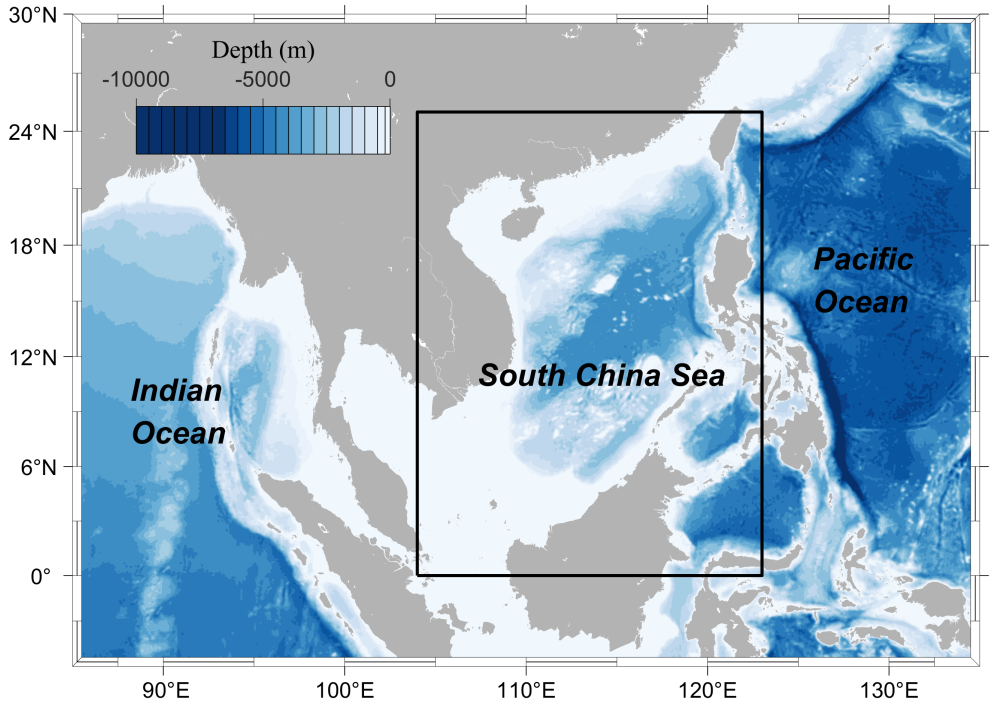


FIGURE 1
Bathymetry map of the study area. The black solid line frame denotes the SCS.

The bathymetry data used in this study was interpolated from the General Bathymetric Chart of the Oceans (GEBCO[1]) dataset. The GEBCO is a global terrain model that provides elevation data on a 15-arc-second interval grid with a resolution of almost 0.46 km (Weatherall et al., 2021). Due to its high resolution, GEBCO accurately depicts near-shore and deep-sea terrain, making it widely used in wave simulation (Akpinar et al., 2016; Kutupoğlu et al., 2018; Beyramzade et al., 2019; Sun et al., 2022).

3 Data and methods

3.1 SWAN model description

The third-generation wave model SWAN is developed at the Delft University of Technology (The SWAN team, 2021b). SWAN is well known for its implicit schemes and iteration techniques, which make the model performance more robust and economic, especially in shallow shelf seas. As a third-generation wind-wave model, SWAN computes the rate of change of wave action density (N) as follows

$$\frac{\partial N}{\partial t} + \frac{\partial c_x N}{\partial x} + \frac{\partial c_y N}{\partial y} + \frac{\partial c_\sigma N}{\partial \sigma} + \frac{\partial c_\theta N}{\partial \theta} = \frac{S_{tot}}{\sigma} \tag{1}$$

The first term on the left-hand-side represents the rate of change of N with time, and the second and third terms represent the propagation of waves in geographic space. c_x and c_y are the wave propagation velocities in the zonal and meridional directions respectively. The fourth term represents the shifting of the radian frequency (σ) due to variations in depth and mean currents. The fifth term demotes the depth-induced and current-induced refraction in wave propagation direction (θ). The right side of this equation, S_{tot} is the superposition of all sink and source terms:

$$S_{tot} = S_{in} + S_{ds,w} + S_{ds,b} + S_{ds,br} + S_{nl3} + S_{nl4} \tag{2}$$

These six terms denote, respectively, wave growth by the wind, wave decay due to whitecapping, bottom friction and depth-induced wave breaking, nonlinear transfer of wave energy through three-wave and four-wave interactions. Detailed descriptions of these terms can be found in the SWAN scientific and technical documentation (The SWAN team, 2021b).

In SWAN, the whitecapping expressions are based on a pulse-based model (Hasselmann, 1974), and remodified by the WAMDI Group (1988):

$$S_{ds,w}(\sigma,\theta) = -\Gamma\tilde{\sigma}\frac{k}{\tilde{k}}E(\sigma,\theta) \tag{3}$$

in which $\tilde{\sigma}$ and $\tilde{k}$ represent the mean frequency and mean wave number respectively, and k is the wave number. Γ is a coefficient related to the wave steepness and has been adapted by Günther et al. (1992) from Janssen (1992):

$$\Gamma = C_{ds}\left((1-\delta) + \delta\frac{k}{\tilde{k}}\right)\left(\frac{\tilde{s}}{\tilde{s}_{PM}}\right)^p \tag{4}$$

where $\tilde{s}$ is the overall wave steepness, and C_{ds}, δ and p are tunable coefficients. $\tilde{s}_{PM} = \sqrt{3.02 \times 10^{-3}}$ is the value of $\tilde{s}$ for the Pierson-Moskowitz spectrum (Pierson and Moskowitz, 1964). C_{ds} has two different choices in SWAN, namely KOMEN and JANSSEN. The default value of C_{ds} is 2.36E-5 for KOMEN but 4.5 for JANSSEN.

1 https://download.gebco.net/.

Van der Westhuysen et al. (2007) divided the dissipation mode into breaking and non-breaking waves, which were active in different parts of the spectrum:

$$S_{ds,w}(\sigma,\theta) = f_{br}(\sigma)S_{ds,break} + [1 - f_{br}(\sigma)]S_{ds,non-break} \quad (5)$$

$$S_{ds,break}(\sigma,\theta) = -C_{ds}\left(\frac{B(k)}{B_r}\right)^{\frac{p}{2}}(tanh(kh))^{\frac{2-p_0}{4}}\sqrt{gk}E(\sigma,\theta) \quad (6)$$

where $S_{ds,break}$ and $S_{ds,\mathrm{non}-break}$ are the contribution by breaking and non-breaking waves, respectively. $B(k) = \int_0^{2\pi} c_g k^3 E(\sigma,\theta)d\theta$ is the azimuthal-integrated spectral saturation, B_r is a threshold saturation level. B_r and C_{ds} are both tunable parameters and the default settings in SWAN are B_r = 1.75E-3 and C_{ds} = 5.0E-5. However, the theory of non-breaking low-frequency waves is not yet mature, so $S_{ds,\mathrm{non}-break}$ is usually replaced by KOMEN.

The scheme of Van der Westhuysen et al. (2007) is always used in conjunction with the wind input scheme of Yan (1987) (hereinafter referred to as YAN), and the expression is given as

$$S_{in,YAN}(\sigma,\theta) = \left\{\left[C_1\left(\frac{U_\star}{c_{ph}}\right)^2 + C_2\left(\frac{U_\star}{c_{ph}}\right) + C_3\right]cos(\theta-\theta_w) + C_4\right\}\sigma E(\sigma,\theta) \quad (7)$$

where C_1 = 4.0E-2, C_2 = 5.52E-3, C_3 = 5.2E-5, C_4 = −3.02E-4 are constants (The SWAN team, 2021b), $U_\star$ and c_{ph} are the friction velocity and phase speed respectively. In SWAN, the whitecapping scheme of Van der Westhuysen et al. (2007) and YAN are usually treated as a stand-alone scheme, hereinafter referred to as WST.

Since the implementation of the "ST6" source term package (Rogers et al., 2012; The SWAN team, 2021b) (hereinafter referred to as ST6), its good performance at different spatial scales and weather conditions has made it widely used (Liu et al., 2019). According to Zieger et al. (2015) and Rogers et al. (2012), the wind input expression of ST6 is given as

$$S_{in,ST6}(\sigma,\theta) = \frac{\rho_a}{\rho_w}\sigma\{2.8 - [1 + \tan h(10\sqrt{B_n}W - 11)]\}\sqrt{B_n}WE(\sigma,\theta) \quad (8)$$

$$W = W_1(\sigma,\theta) - a_0 W_2(\sigma,\theta) \quad (9)$$

where B_n is the spectral saturation. W_1 and W_2 represent the positive and adverse wind inputs respectively and their magnitudes are dependent on the friction velocity $U_\star$. a_0 is the wind scaling coefficient.

3.2 Model setup

In this study, we used the hindcast model SWAN Cycle III version 41.31AB[2]. The SWAN model is operated in the third generation and non-stationary mode with a spatial resolution of 0.25° × 0.25°. A time step of 30 minutes is adopted, and each time step is iterated up to a maximum of 5 times. The JONSWAP (Joint North Sea Wave Project) spectrum is divided into 72 directions and frequency bins between 0.04 Hz and 1.0 Hz. The JONSWAP spectrum is used for the bottom friction with C_b (the bottom friction coefficient) setting to 0.038 (Hasselmann et al., 1973). The study period spans from 2017 to 2021. To investigate the effect of different whitecapping schemes, we evaluate three schemes: KOMEN, JANSSEN, and WST. For KOMEN and JANSSEN, we employ the ST6 with the wind drag formula developed by Hwang (2011) and the wind scaling coefficient a_0 set to 28. For WST, we still use YAN as the wind input scheme. Due to computational resource limitations, we mainly conduct numerical experiments using these two wind input schemes, ST6 and YAN. The detailed settings of all experiments are provided in Table 1.

3.3 Atmospheric forcing data and observations

Forcing wind fields significantly affect the accuracy of wave models (Kutupoğlu et al., 2018; Yang et al., 2022). In this study, we utilized 10-m wind speeds (U_{10}) from three high-quality wind products to drive the wave model, namely the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis (ERA5[3]), the Cross-Calibrated Multi-Platform Version 2.0 (CCMP[4]), and the National Centers for Environmental Prediction (NCEP) Final Reanalysis Data (FNL[5]).

The latest atmospheric reanalysis data from ECMWF is ERA5, which supersedes ERA-Interim since September 2019 (Jiang et al., 2022). ERA5 offers a finer spatial grid, higher temporal resolution, and more vertical levels compared to ERA-Interim (Hersbach et al., 2020). The dataset used in this study has a horizontal resolution of 0.25° × 0.25° and a temporal resolution of 1 hour. Previous studies have demonstrated the exceptional performance of ERA5 in our study area (Zhang et al., 2020; Feng et al., 2022; Yang et al., 2022; Zhai et al., 2023). Therefore, ERA5 was selected as the primary forcing wind field to drive the model.

CCMP is a Level-3 ocean vector wind analysis product that provides high-quality global wind field data with a six-hour temporal resolution from 1988 to the present and a spatial resolution of 0.25° × 0.25° (Wentz, 2015; Mears et al., 2019; Wu et al., 2022). Experimental validation conducted by Atlas et al. (2011) demonstrated a significant improvement in the accuracy of CCMP data compared to wind field measurements from individual satellite platforms, rendering it well-suited for oceanic and atmospheric research.

FNL is a global reanalysis product with a six-hour temporal resolution spanning from 1999 to the present and a spatial resolution of 1.00° × 1.00° (Appendini et al., 2013; Chen et al., 2020). It employs an advanced data assimilation system and assimilates observation data from various sources. The product is founded upon the Global Data Assimilation System (GDAS) and is

2 https://swanmodel.sourceforge.io/.

3 https://cds.climate.copernicus.eu/.

4 https://data.remss.com/ccmp/v02.0/.

5 https://rda.ucar.edu/datasets/ds083.2/.

TABLE 1 Parameter settings of SWAN model.

Model physics	Parameterization scheme	Parameters	Values
Wind input	ST6	*HWANG*	–
		a_0	28
	YAN	–	–
Triad wave–wave interactions	LTA	*Ur*	0.01
Quadruplet wave interactions	DIA	λ	0.25
		Cnl4	3E+7
Bottom friction	JONSWAP	C_b	0.038
Depth-induced wave breaking	CONSTANT	α	1.0
		γ	0.73
Whitecapping	KOMEN	*cds2*	–
	JANSSEN	*cds1*	–
	WST	*cds2*	–

prepared operationally every 6 hours using the identical model and assimilation scheme as the NCEP operational Global Forecast System (GFS).

The Satellite with ARgos and ALtiKa (SARAL[6]) project is a joint mission operated by the Indian Space Agency (ISRO) and the French Space Agency (CNES), designed for ocean observations (Verron et al., 2015). AltiKa, SARAL's primary payload, is the first spaceborne altimeter operating at the Ka-band frequency (35.75 GHz). The higher frequency leads to a smaller footprint (8 km diameter) and so a better spatial resolution (Verron et al., 2021). Since March 2013, SARAL has been providing along-track data for various physical oceanographic parameters on a global scale, including sea surface wind speed, SWH, and sea surface height (Verron et al., 2015). Recent studies have affirmed SARAL's high accuracy, data quality, and availability (Sepulveda et al., 2015; Sharma et al., 2022). SARAL offers a range of processed data products at various levels. In this study, we utilized a delayed-mode version, specifically the Nadir altimeter Geophysical Data Record (GDR).We performed interpolation on the wind-forcing data and simulated SWH data through temporal (cubic spline) and spatial (nearest-neighbor) methods to align them with the altimeter data. To ensure the reliability of our validation results, we excluded the altimeter data that was more than 5 kilometers away from the nearest grid points. The variations in the quantity of valid altimeter data under different conditions are detailed in Table 2.

3.4 Error metrics

To accurately quantify the model performance, we employed two commonly used error metrics, including the index of agreement (*d*) proposed by Willmott (1982), and *Slope*. The specific formula for the *d index* is presented below

$$d = 1 - \frac{\sum_1^n (S_i - O_i)^2}{\sum_1^n (|S_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (10)$$

where O_i is the observed value, $\bar{O}$ is the mean value of the observed data, S_i is the value of the wind products or model outputs, and n is the sample size. The *d index* displays the differences between simulated and observed means and variances, which reflect sensitivity to outliers in the observation data and insensitivity to additional and proportional variances between simulated and observed values (Zheng et al., 2023). Moreover, the *d index* is a standardized metric, with values ranging between 0 and 1, where values closer to 1 indicate higher consistency between two datasets (Willmott, 1982). The *Slope* provides an indication of the direction of errors and is calculated as the linear regression coefficient in the regression model y = cx. A *Slope* value greater (or less) than 1 signifies that S_i tends to be larger (or smaller) than O_i. This paper will primarily use the *d index* to measure the consistency between S_i and O_i, while the *Slope* will be used as a secondary measure to assess the direction of errors.

4 Theoretical basis

4.1 The wind errors and C_{ds}

With the wind input scheme determined, the forcing wind field, the only input variable, directly determines the magnitude of the input energy. However, errors in the wind field can affect the simulated SWH. Positive errors in the wind field lead to larger simulated SWH, while negative errors lead to smaller simulated SWH (Wu et al., 2020). Taking physical quantities in real environment as reference, positive errors in the forcing wind field require greater dissipation energy than in reality to maintain energy balance, whereas negative errors require less. It can be inferred that C_{ds}, which is the primary factor influencing dissipation energy, may exhibit a compensatory relationship with wind errors. In order to

6 ftp://ftp-access.aviso.altimetry.fr.

TABLE 2 Amount of valid altimeter data under different conditions.

Gribed data	ERA5	CCMP	FNL	Simulated SWH
Amount of valid data	68,827	67,929	16,803	68,181

explore this relationship, sensitivity experiments were conducted. Figure 2A illustrates simulations where the wind field was scaled by factors of 0.8, 0.9, 1, 1.1, and 1.2, with a constant C_{ds} value of 0.22E-5. Likewise, in Figure 2B, simulations were performed using the same wind field but varying C_{ds} values. A comparison between Figures 2A, B reveals that, when C_{ds} is held constant, the simulated SWH gradually increases as the wind field increases. Conversely, when the wind field remains constant, the simulated SWH gradually decreases with increasing C_{ds}. Hence, when the wind errors are determined, the corresponding optimal value of C_{ds} can also be determined.

4.2 The errors in the wind fields

Currently, a wide range of wind field products are available, and with advancements in observation and assimilation techniques, these products consistently demonstrate high quality (Wu et al., 2020; Wu et al., 2022). However, the utilization of diverse assimilation data and methods in different wind field products results in variations in their errors. Figure 3 illustrates the interannual variations in the errors of three wind field products: ERA5, FNL, and CCMP. Analysis of the *Slope* values reveals a consistent underestimation of actual wind fields by ERA5 over 2017–2021, corroborating findings from previous studies (Shi et al., 2021; Son et al., 2023; Zhai et al., 2023). Furthermore, in 2019 and 2020, all three wind field products exhibited varying degrees of underestimation. When examining the trend of the *d index*, CCMP consistently displayed the highest quality with a *d index* of approximately 0.92, while FNL exhibited the poorest quality with a *d index* of around 0.89. In 2020, there was a fluctuation in the quality of these three wind field products compared to the norm, with a *d index* of approximately 0.86. Overall, the *d index* of the three high-quality wind field products ranges from 0.86 to 0.93, and the *Slope* values are also near 1. This indicates that the errors in each wind field product are relatively stable.

Based on the findings in Section 4, it is hypothesized that when using high-quality wind field products, the optimal C_{ds} values will exhibit minimal variability. Therefore, conducting a series of numerical experiments to validate this hypothesis is essential.

5 Numerical experiments

In this section, we will conduct an extensive sensitivity analysis to validate whether the hypothesis proposed in the previous chapter still holds under different seasons, years, and wind field types. If confirmed, we will also determine the specific range of the optimal C_{ds}, which can greatly enhance the accuracy of wave simulation.

According to the conclusions in Section 4, there is a monotonic relationship between C_{ds} and simulated SWH. In other words, as C_{ds} increases, there will inevitably be an optimal simulation effect at a certain value. To determine this optimal simulation accuracy, we search for the maximum value of the *d index*. Guided by this principle, we initiate our experiments using one-tenth of the default C_{ds} as the starting point and apply a step size of 0.1E-5 or 0.1.

5.1 Sensitivity to different seasons

Firstly, we examine the seasonal characteristics of the optimal C_{ds} values under three distinct whitecapping dissipation schemes

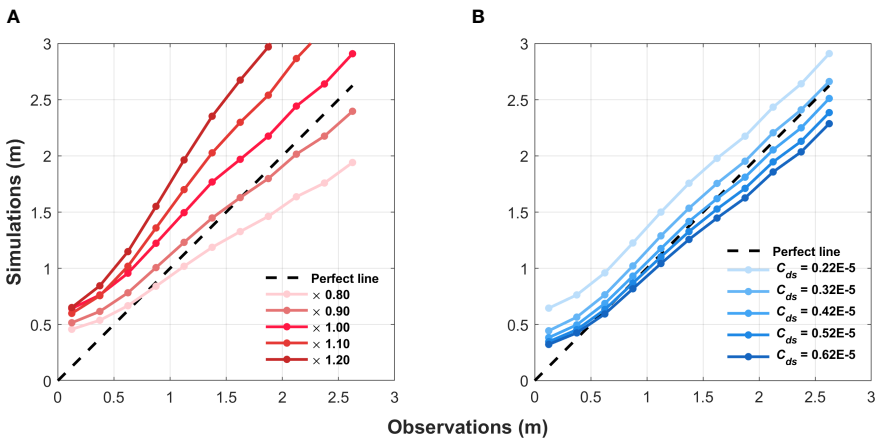


FIGURE 2
Bin-averaged scatterplots of simulated SWH versus the observations under **(A)** variable wind fields and **(B)** variable C_{ds}. In panel **(A)**, ERA5 U_{10} is scaled at 0.8, 0.9, 1.1, and 1.2 times with a fixed C_{ds} value of 0.22E-5. In panel **(B)**, C_{ds} values are varied at 0.22E-5, 0.32E-5, 0.42E-5, 0.52E-5, and 0.62E-5, while ERA5 U_{10} remains unchanged.

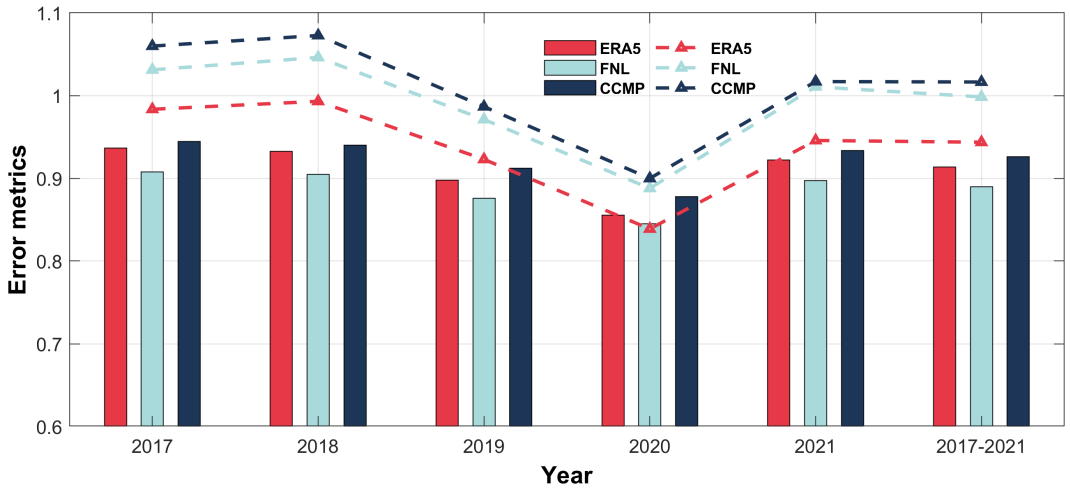


FIGURE 3
Time series analysis of wind field quality. Red, sky blue, and dark blue are ERA5, FNL and CCMP, respectively. The bars represent the *d index*, while the dashed line indicates the *Slope* value. The final point on the X-axis represents the overall error for a five-year period (2017–2021).

using ERA5 as the forcing wind field. Figure 4 illustrates the *d index* represented by purple bars, the *Slope* depicted by a black curve, and the maximum value of the *d index* indicated by a green dashed line, corresponding to the position of the optimal C_{ds} values. The comparison of the *d index* for each season reveals that the simulation performance in spring is notably inferior to that of other seasons, even when considering the optimal C_{ds} values, as the *d index* remains below 0.90. Analyzing the simulation results of the three schemes across different seasons shows that WST only achieves a *d index* above 0.90 in winter, suggesting that the overall simulation performance of KOMEN and JANSSEN surpasses that of WST.

Regarding the distribution of the optimal C_{ds} values, KOMEN remains fixed at 0.52E-5, while JANSSEN ranges from [0.66, 0.76], and WST fluctuates within the range of [0.72E-5, 1.12E-5]. It is important to note that reducing the step size of C_{ds} and increasing the number of experiments may yield more accurate values for the optimal C_{ds}, but this is expected to have only a minor impact on the experimental results. Based on the experimental outcomes, it is evident that the three whitecapping dissipation schemes exhibit seasonal fluctuations in the optimal C_{ds} values. However, the magnitude of these fluctuations is minimal.

Based on the observed variations in the *Slope*, it is evident that an increase in C_{ds} corresponds to a gradual decrease in the *Slope*, implying a systematic decline in the simulated SWH. This finding reinforces the conclusion established in Section 4. Additionally, it is worth mentioning that at the optimal C_{ds} values, all of the *Slope* consistently fall below 1, indicating an underestimation of the simulated SWH relative to the observed values. Notably, WST exhibits the most pronounced underestimation among the evaluated schemes.

5.2 Sensitivity to different years

To determine the optimal C_{ds} values for each year, we expanded the range of C_{ds} based on the experiments in Section 5.1. Figure 5 illustrates the interannual characteristics of the optimal C_{ds} values for the three whitecapping dissipation schemes. The optimal C_{ds} range is [0.42E-5, 0.52E-5] for KOMEN, [0.46, 0.86] for JANSSEN, and [0.72E-5, 1.12E-5] for WST. Comparing the seasonal characteristics of the optimal C_{ds} values, JANSSEN exhibits a slightly larger fluctuation, while the other two schemes remain relatively consistent. Overall, we can conclude that the interannual variability of the optimal C_{ds} values is also very small.

Regarding the *d index*, the simulation performance in 2020 exhibits the poorest results, with the *d index* falling below 0.90 for all three schemes. When examining the *d index* of the optimal C_{ds} values, the simulation performance of KOMEN and JANSSEN is superior to that of WST, even in the year with the worst simulation performance in 2020.

As highlighted in Section 4.2, the quality of the wind field in 2020 displayed fluctuations, with its *Slope* significantly lower than that of other years. This discrepancy indicates a severe underestimation of the actual wind field by the wind field product in 2020. Referring to Figure 5, it is evident that the optimal C_{ds} value for 2020 is the smallest among the five years. This finding further strengthens the confirmation of the compensatory relationship between the wind errors and C_{ds}, as elucidated in Section 4.

To address the potential impact of short-term disruptions on the overall quality of the wind field, we considered the period from 2017 to 2021 as a unit, as depicted in Figures 5P–R. Comparing the optimal C_{ds} values for the five years to that of 2021 reveals a complete equivalence between them. In Figure 3, the *Slope* and *d*

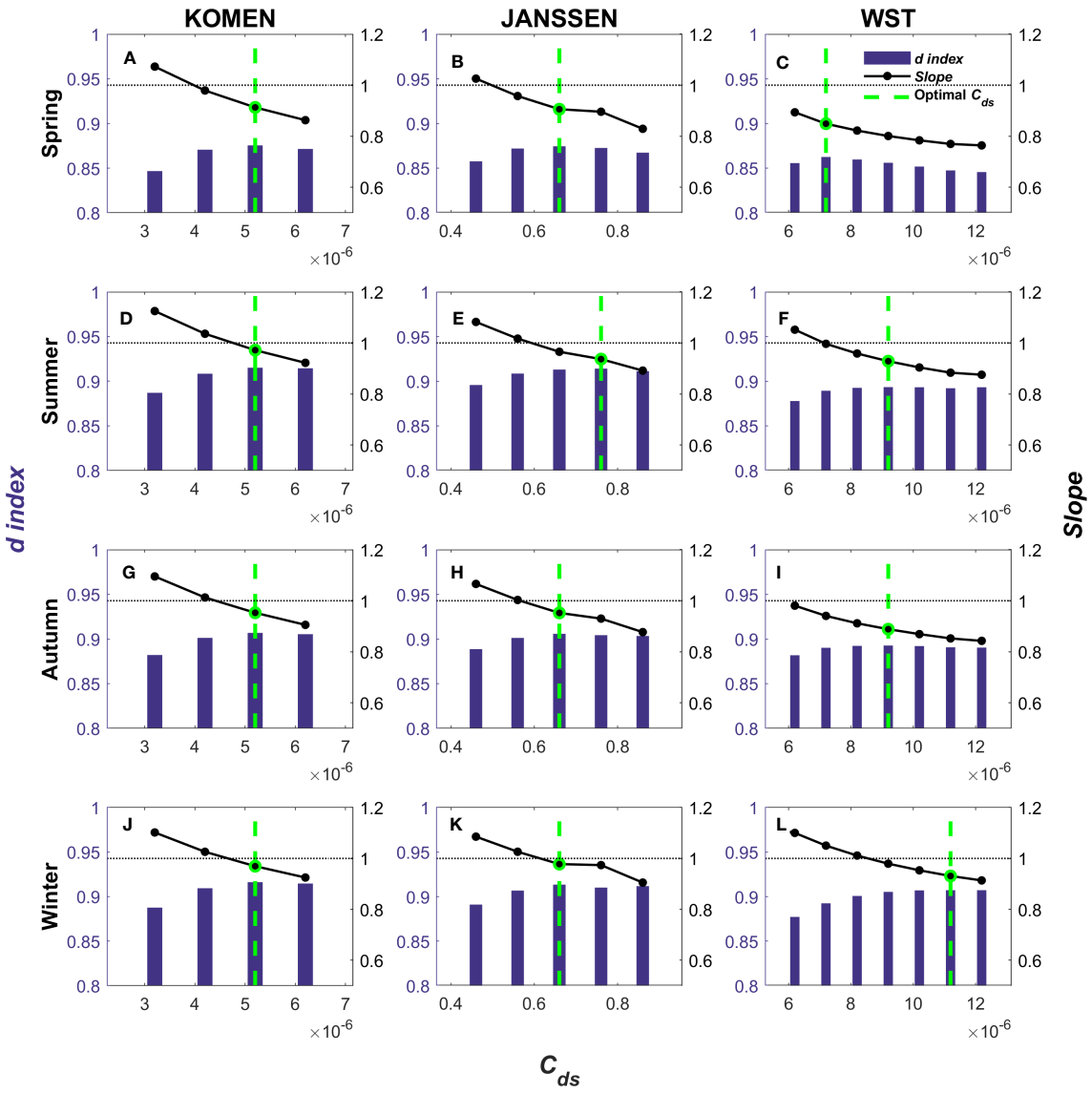


FIGURE 4
The distributions of *d index* and *Slope* for spring **(A–C)**, summer **(D–F)**, autumn **(G–I)**, and winter **(J–L)**. From left to right, **(A, D, G, J)** are KOMEN; **(B, E, H, K)** are JANSSEN; **(C, F, I, L)** are WST. The purple bars show the *d index* for each experiment, corresponding to the left Y-axis. The solid black line shows the *Slope* distribution corresponding to the right Y-axis. The green dotted line indicates the experiment corresponding to the maximum value of the *d index*, that is, the experiment corresponding to the optimal C_{ds}. The black dashed line represents the *Slope* equal to 1.

index of ERA5 for the five-year timeframe closely approximate those of 2021, providing further evidence of the direct relationship between wind errors and the optimal C_{ds} values. Furthermore, by comparing the *d index* in close proximity to the optimal C_{ds} values, we observe minimal differences among them. This finding implies that the optimal value attained for the five-year period consistently yields favorable simulation performance annually. In essence, the longer the simulation timeframe, the more accurately the main characteristics of the optimal C_{ds} can be reflected. Therefore, a simulation duration of five years will be maintained for subsequent validation experiments.

5.3 Sensitivity to different wind fields

The preceding section focused on the optimal C_{ds} characteristics when using ERA5 as the forcing wind field. In order to assess the generalizability of the optimal C_{ds} features, we conducted experiments using CCMP and FNL as alternative forcing wind fields.

As shown in Figure 3, the *Slope* reaches the highest when the CCMP is applied, while it is the lowest for the ERA5 winds. Figure 6 provides an illustration of the optimal C_{ds} values, showing that CCMP generally has the highest C_{ds} values among the three wind

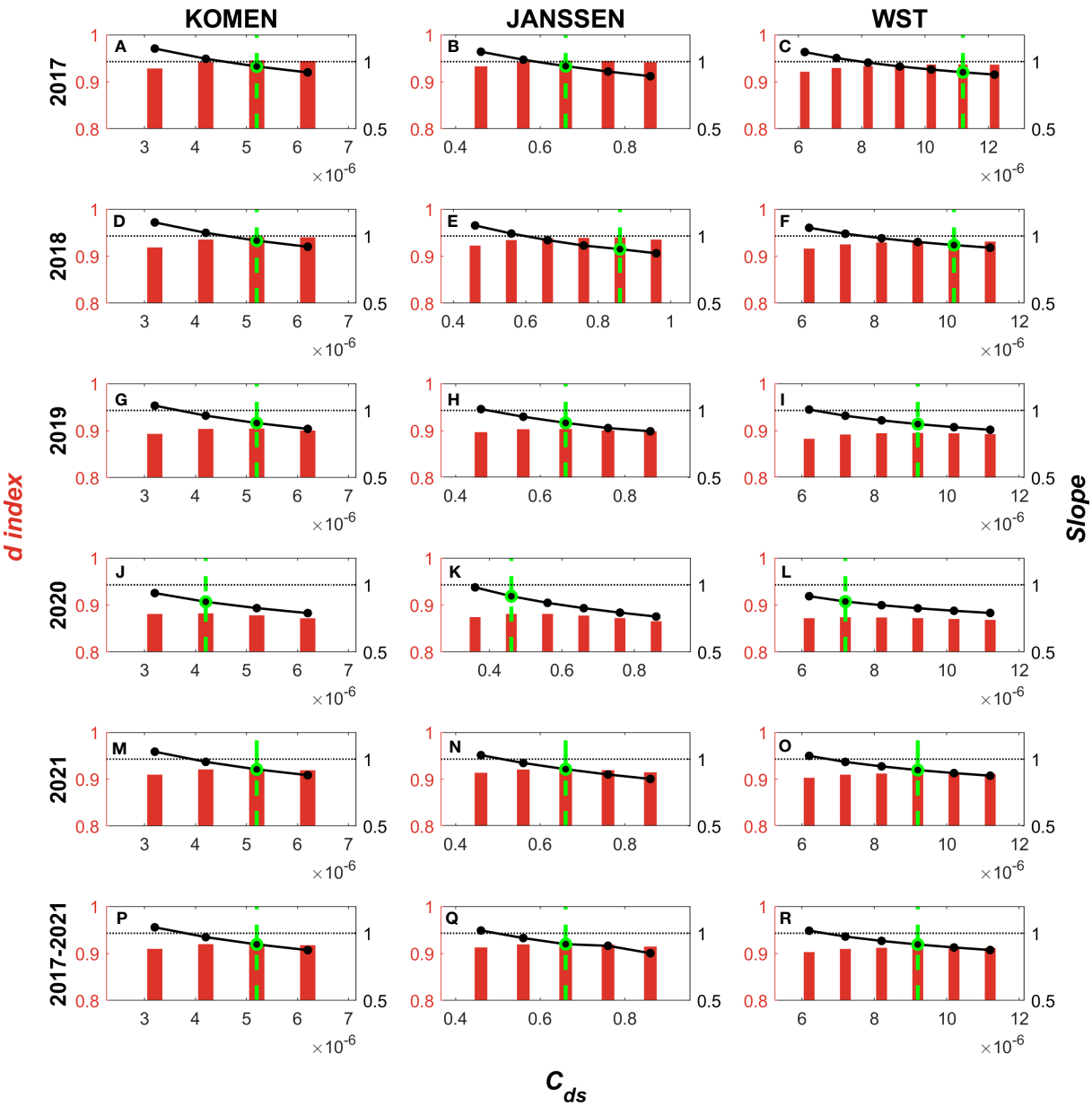


FIGURE 5
The distributions of *d index* and *Slope* for 2017 **(A–C)**, 2018 **(D–F)**, 2019 **(G–I)**, 2020 **(J–L)**, 2021 **(M–O)** and the five years **(P–R)**. From left to right, **(A, D, G, M, J, P)** are KOMEN; **(B, E, H, K, N, Q)** are JANSSEN; **(C, F, I, L, O, R)** are WST. The crimson bars show the *d index* for each experiment, corresponding to the left Y-axis. The other settings are the same as in Figure 4.

field products, while ERA5 exhibits the lowest values. This finding provides further confirmation of the compensatory relationship between wind field errors and C_{ds}. Remarkably, for KOMEN, it is observed that the optimal C_{ds} values are identical across all three wind field products, at 0.52E-5. The optimal C_{ds} range for JANSSEN is [0.66, 0.76], while for WST, it fluctuates within the range of [0.92E-5, 1.12E-5]. Similar to the seasonal and interannual characteristics, the fluctuation range of the optimal C_{ds} among different wind field products is also minimal.

Figure 7 presents the differences in simulation performance between the optimal C_{ds} values and the default C_{ds} values across all experiments. The optimal C_{ds} values are represented by the black lines in the figure, corresponding to the right y-axis. The maximum value on the right y-axis represents the default C_{ds} for each scheme. It is evident that the range of the optimal C_{ds} values for all three schemes is significantly narrower compared to the default values of the model. Specifically, for the individual whitecapping dissipation schemes, the optimal C_{ds} range is [0.42E-5, 0.52E-5] for KOMEN, [0.46, 0.86] for JANSSEN, and [0.72E-5, 1.12E-5] for WST. From the experiments conducted above, we can conclude that the variations in wind field errors across different time scales and types indeed result in fluctuations in the optimal C_{ds} values. However, due to the overall stability of wind field errors, the optimal C_{ds} values also fluctuate within a small range. These

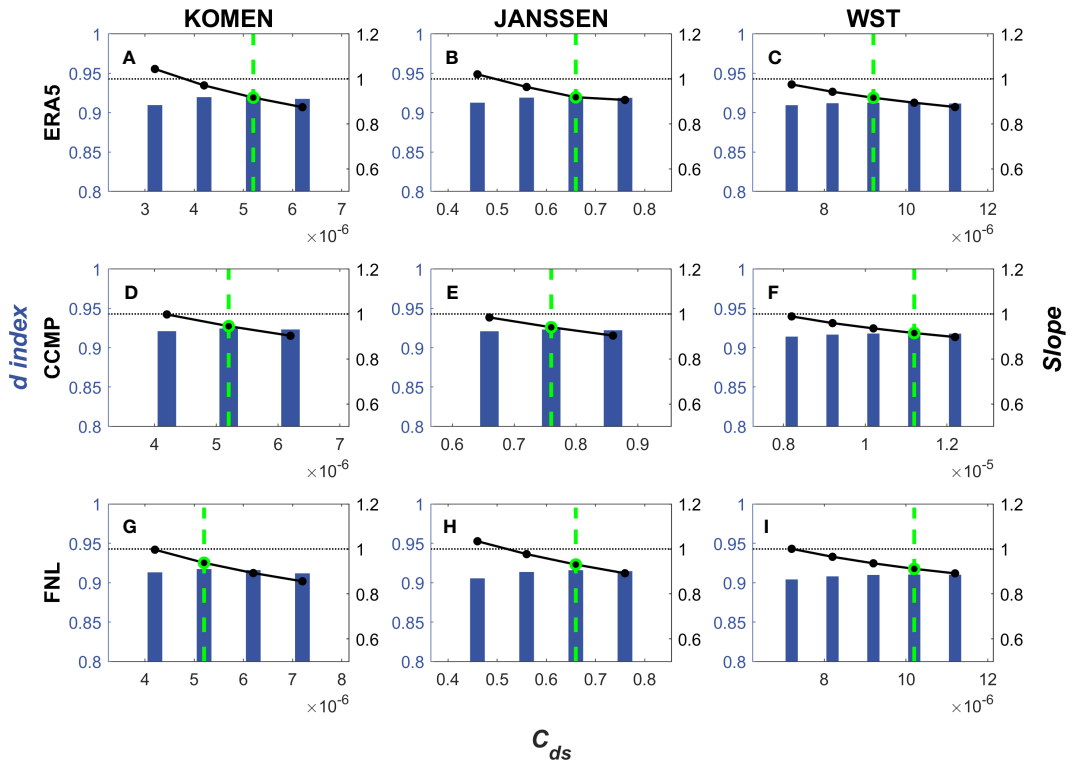


FIGURE 6
The distribution of *d index* and *Slope* of ERA5 **(A–C)**, CCMP **(D–F)**, and FNL **(G–I)**. From left to right, **(A, D, G)** are KOMEN; **(B, E, H)** are JANSSEN; **(C, F, I)** are WST. The blue bars show the *d index* for each experiment, corresponding to the left Y-axis. The other settings are the same as in Figure 4.

results demonstrate the reliability of the theoretical basis proposed in Section 4.

The red and dark blue bars in Figure 7 represent the *d index* of the optimal C_{ds} and default C_{ds}, respectively, corresponding to the left y-axis of the figure. It can be observed that the *d index* of the optimal C_{ds} values for all three schemes is consistently around 0.90, indicating significantly improved simulation performance compared to the default values. The most notable enhancement is observed for JANSSEN with the optimal C_{ds} values, while the improvement is minimal for WST. This finding suggests that the default value of WST exhibits the best simulation performance among the three schemes. From the previous experimental results, we have discovered that C_{ds} values near the optimal C_{ds} also yield excellent simulation performance. In other words, selecting any C_{ds} value within the proposed optimal value range would result in improved simulation performance than the default value of the model.

6 Discussion

This study aims to propose a general method for determining the optimal C_{ds} to enhance the accuracy of ocean wave simulation. However, the primary focus of this paper is on the SCS. Therefore, to determine the wider applicability of the optimal C_{ds} intervals, two additional regions, namely the Gulf of Mexico (GoM; 98°W–78°W, 17°N–31°N) and the MS (8°W–37°E, 28°N–45°N), were selected as study areas. The GoM and the MS share similarities with the SCS, featuring complex topography and drastic variations in water depth. They also serve as vital maritime routes and regions rich in natural resources (Huerta and Harry, 2012; Appendini et al., 2013; Elkut et al., 2021; Beyramzadeh and Siadatmousavi, 2022). Therefore, choosing these areas as subjects of study holds heightened practical relevance. The time range, model settings, and observational data remained consistent with this study. Similarly, we filtered the raw data by excluding data points that were more than 5 kilometers away from the grid points. As a result, there were 44,756 valid data for the GoM and 84,272 valid data for the MS. To effectively utilize computational resources, we only used ERA5 to drive the model.

Figure 8A displays the bathymetry map for the GoM and the MS. For the GoM (Figures 8B, D, F), optimal C_{ds} values are determined as 0.42E-5 for KOMEN, 0.46 for JANSSEN, and 0.82E-5 for WST, respectively. Similarly, for the MS (Figures 8C, E, G), the optimal C_{ds} values for the respective schemes are 0.52E-5, 0.66, and 1.02E-5. All the optimal C_{ds} values for both regions fall within the proposed range, demonstrating the robust applicability of the proposed viewpoint across different regions.

This study primarily focuses on two wind input schemes, ST6 and YAN, in SWAN. However, SWAN offers multiple other wind input schemes to choose from. Different wind input schemes can

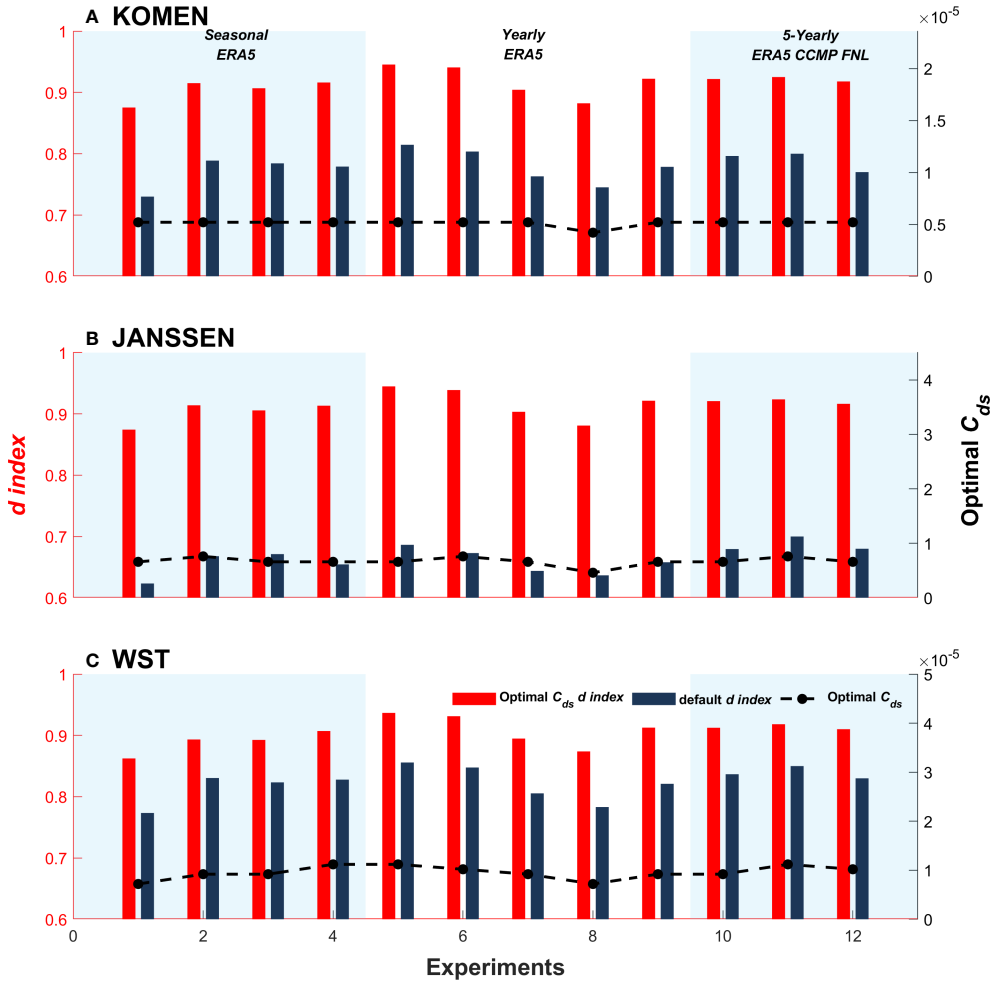


FIGURE 7
The difference in model performance between the optimal C_{ds} and the default C_{ds} for KOMEN **(A)**, JANSSEN **(B)**, and WST **(C)**. The X-axis represents the experiment number: 1) 1–4 are seasonal results driven by ERA5; 2) 5–9 are interannual results driven by ERA5; 3) 10–12 are 5-year simulation results driven by ERA5, CCMP, and FNL. The red and dark blue bars represent the *d index* for the optimal C_{ds} and the default C_{ds}, corresponding to the left Y-axis. The dash black line shows the optimal C_{ds} values distribution for all experiments, corresponding to the right Y-axis.

result in variations in the energy input to the wave model (Wang and Huang, 2004; Adcock and Taylor, 2018), potentially leading to variations in the optimal C_{ds} range. Nevertheless, the optimal C_{ds} values will always fluctuates within a very small range.

When utilizing the optimal C_{ds} range, two considerations must be taken into account. Firstly, our conclusions are based on a large volume of observational data. In cases where the validation dataset is insufficient or the simulation duration is too short, such as during typhoon events, the limited number of samples can lead to potential errors. Therefore, it's essential to exercise caution when applying the optimal C_{ds} range under these circumstances. Additionally, for nearshore wave simulations, where whitecapping dissipation no longer dominates, the contributions of bottom friction and depth-induced breaking to the dissipation process become prominent (Xu et al., 2013; Peng et al., 2023). Moreover, errors in nearshore wind fields can increase. Therefore, adjusting a single parameter alone may not yield satisfactory simulation results. In such cases, the utilization of the optimal C_{ds} range also needs to be approached with caution. Typhoon and nearshore wave simulations hold significant research value, and these aspects will be the main focus of our future investigations.

7 Conclusion

Calibrating wave models is of paramount importance for simulating SWH, with C_{ds} often serving as a primary calibration parameter. Nevertheless, determining the optimal C_{ds} value is a challenging task. This study, through theoretical analysis and numerical experiments, provides a robust optimal C_{ds} range. Within this range, the accuracy of simulated SWH for any C_{ds} value is better than the model's default C_{ds}.

Specifically, we begin by revealing a direct relationship between wind errors and the optimal C_{ds} through sensitivity experiments. Through a comprehensive evaluation of high-quality wind field products using satellite observational data, we have discovered that the errors of these wind field products are stable. This mechanism

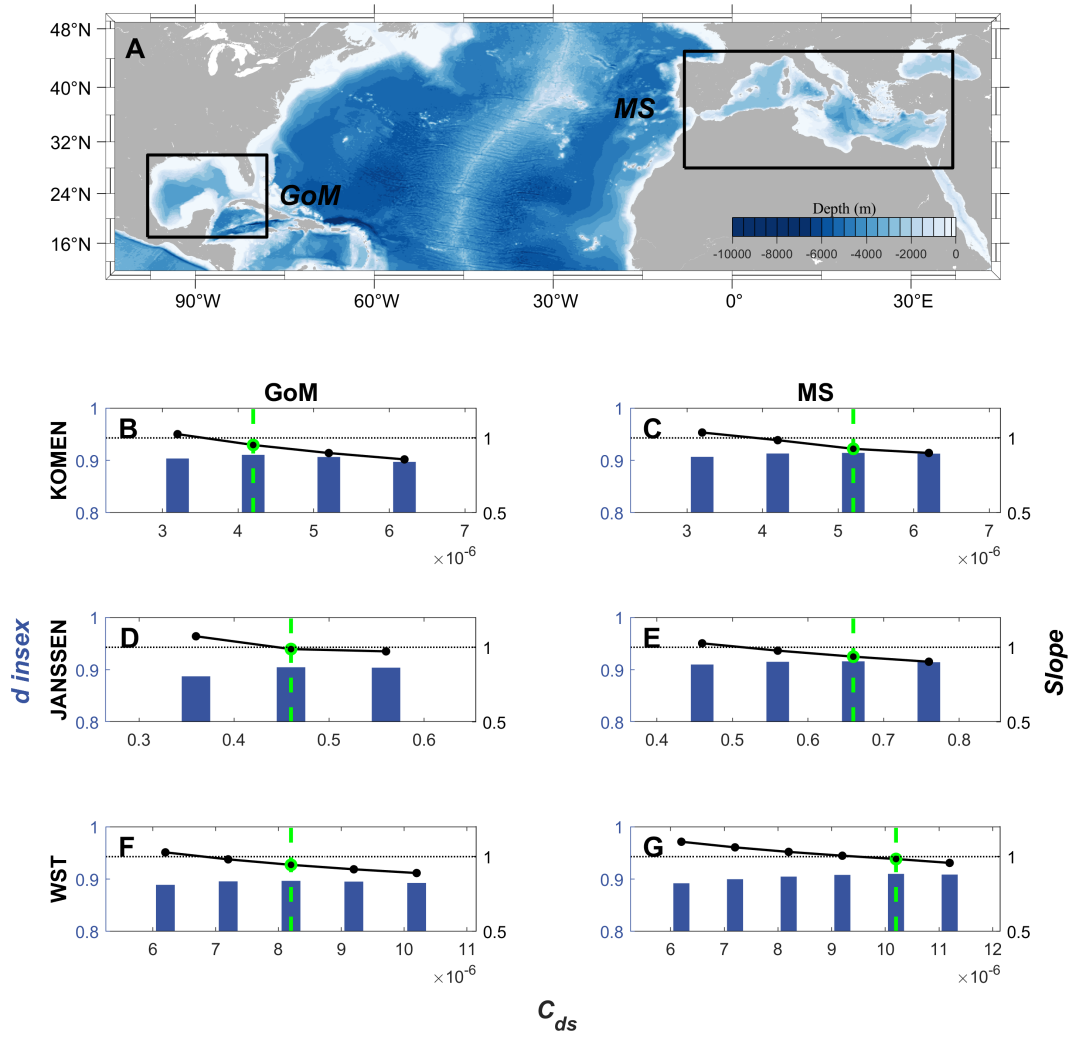


FIGURE 8
The distribution of *d index* and *Slope* in the GoM **(B, D, F)** and MS **(C, E, G)**. From top to bottom, **(A)** is a schematic of the simulated area; **(B, C)** are KOMEN; **(D, E)** are JANSSEN; **(F, G)** are WST. The other settings are the same as in Figure 6.

suggests that, under high-quality wind field conditions, the optimal C_{ds} values will exhibit small fluctuations.

To verify this conjecture, we conducted a five-year wave simulation in the SCS using SWAN, covering various scenarios with different time scales and wind field types. Employing two wind input schemes (ST6 and YAN) and three whitecapping dissipation schemes (KOMEN, JANSSEN, and WST), we observed that the optimal C_{ds} values for all three whitecapping dissipation schemes fluctuated within a narrow range. Specifically, the optimal C_{ds} range was [0.42E-5, 0.52E-5] for KOMEN, [0.46, 0.86] for JANSSEN, and [0.72E-5, 1.12E-5] for WST. These results demonstrate the applicability of our proposed viewpoint across different time scales and wind field types. Furthermore, compared to the model's default C_{ds}, we found that any C_{ds} within the optimal range had better simulation performance.

To investigate the applicability of the optimal C_{ds} characteristics in different regions, we conducted similar experiments in the GoM and the MS using ERA5. The findings demonstrated that the optimal C_{ds} values in these regions aligned with the suggested range, affirming the universality of this approach across a global scale.

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

ZL: Formal Analysis, Methodology, Validation, Writing – original draft. WW: Data curation, Investigation, Visualization, Writing – review & editing. YG: Resources, Software, Writing – review & editing. FZ: Conceptualization, Methodology, Writing – review & editing. PL: Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Major program of Laoshan Laboratory for Marine Science and Technology (Grant Number LSKJ202202903), the Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City (Grant Number 2021CXLH0020), the Scientific and technological projects of Zhoushan (Grant Numbers 2022C01004 and 2022C81010), the National Science Foundation of China (Grant Number 42176016) and the Shandong Provincial Natural Science Foundation (Grant Number ZR2020MD059).

Acknowledgments

We acknowledge the development team of the SWAN wave model at the Delft University of Technology. We would also like to express our gratitude to the Hainan Observation and Research Station of Ecological Environment and Fishery Resource in Yazhou Bay for their invaluable support in providing data for this 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

Adcock, T. A. A., and Taylor, P. H. (2018). Ocean Wave Non-Linearity and Wind Input in Directional Seas: Energy Input During Wave-Group Focussing. *American Society of Mechanical Engineers*. V07BT06A052. doi: 10.1115/OMAE2018-77998

Akpinar, A., Bingolbali, B., and Van Vledder, G. (2016). Wind and wave characteristics in the Black Sea based on the SWAN wave model forced with the CFSR winds. *OCEAN Eng.* 126, 276–298. doi: 10.1016/j.oceaneng.2016.09.026

Akpinar, A., and Ponce de León, S. (2016). An assessment of the wind re-analyses in the modelling of an extreme sea state in the Black Sea. *Dyn. Atmospheres Oceans* 73, 61–75. doi: 10.1016/j.dynatmoce.2015.12.002

Alves, J. H. G. M., and Banner, M. L. (2003). Performance of a saturation-based dissipation-rate source term in modeling the fetch-limited evolution of wind waves. *J. Phys. Oceanogr.* 33, 1274–1298. doi: 10.1175/1520-0485(2003)033<1274:POASDS>2.0.CO;2

Amarouche, K., Akpınar, A., Bachari, N. E. I., Çakmak, R. E., and Houma, F. (2019). Evaluation of a high-resolution wave hindcast model SWAN for the West Mediterranean basin. *Appl. Ocean Res.* 84, 225–241. doi: 10.1016/j.apor.2019.01.014

Appendini, C. M., Torres-Freyermuth, A., Oropeza, F., Salles, P., López, J., and Mendoza, E. T. (2013). Wave modeling performance in the Gulf of Mexico and Western Caribbean: Wind reanalyses assessment. *Appl. Ocean Res.* 39, 20–30. doi: 10.1016/j.apor.2012.09.004

Atlas, R., Hoffman, R. N., Ardizzone, J., Leidner, S. M., Jusem, J. C., Smith, D. K., et al. (2011). A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bull. Am. Meteorol. Soc* 92, 157–174. doi: 10.1175/2010BAMS2946.1

Babanin, A. V., Tsagareli, K. N., Young, I. R., and Walker, D. J. (2010). Numerical investigation of spectral evolution of wind waves. Part II: dissipation term and evolution tests. *J. Phys. Oceanogr.* 40, 667–683. doi: 10.1175/2009JPO4370.1

Banner, M. L., Babanin, A. V., and Young, I. R. (2000). Breaking probability for dominant waves on the sea surface. *J. Phys. Oceanogr.* 30, 3145–3160. doi: 10.1175/1520-0485(2000)030<3145:BPFDWO>2.0.CO;2

Beyramzade, M., Siadatmousavi, S. M., and Majidy Nik, M. (2019). Skill assessment of SWAN model in the red sea using different wind data. *Reg. Stud. Mar. Sci.* 30, 100714. doi: 10.1016/j.rsma.2019.100714

Beyramzadeh, M., and Siadatmousavi, S. M. (2022). Skill assessment of different quadruplet wave-wave interaction formulations in the WAVEWATCH-III model with application to the Gulf of Mexico. *Appl. Ocean Res.* 127, 103316. doi: 10.1016/j.apor.2022.103316

Bingölbali, B., Akpınar, A., Jafali, H., and Vledder, G. P. V. (2019). Downscaling of wave climate in the western Black Sea. *Ocean Eng.* 172, 31–45. doi: 10.1016/j.oceaneng.2018.11.042

Booij, N., Ris, R. C., and Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *J. Geophys. Res. Oceans* 104, 7649–7666. doi: 10.1029/98JC02622

Bujak, D., Lončar, G., Carević, D., and Kulić, T. (2023). The feasibility of the ERA5 forced numerical wave model in fetch-limited basins. *J. Mar. Sci. Eng.* 11, 59. doi: 10.3390/jmse11010059

Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J.-R., Breivik, Ø., et al. (2018). Wave modelling in coastal and inner seas. *Prog. Oceanogr.* 167, 164–233. doi: 10.1016/j.pocean.2018.03.010

Cavaleri, L., Barbariol, F., and Benetazzo, A. (2020). Wind–wave modeling: where we are, where to go. *J. Mar. Sci. Eng.* 8, 260. doi: 10.3390/jmse8040260

Cavaleri, L., Barbariol, F., Benetazzo, A., and Waseda, T. (2019). Ocean wave physics and modeling: the message from the 2019 WISE meeting. *Bull. Am. Meteorol. Soc* 100, ES297–ES300. doi: 10.1175/BAMS-D-19-0195.1

Cavaleri, L., and Bertotti, L. (1997). In search of the correct wind and wave fields in a minor basin. *Mon. Weather Rev.* 125, 1964–1975. doi: 10.1175/1520-0493(1997)125<1964:ISOTCW>2.0.CO;2

Chen, C., Sasa, K., Ohsawa, T., Kashiwagi, M., Prpić-Oršić, J., and Mizojiri, T. (2020). Comparative assessment of NCEP and ECMWF global datasets and numerical approaches on rough sea ship navigation based on numerical simulation and shipboard measurements. *Appl. Ocean Res.* 101, 102219. doi: 10.1016/j.apor.2020.102219

Christakos, K., Björkqvist, J.-V., Tuomi, L., Furevik, B. R., and Breivik, Ø. (2021). Modelling wave growth in narrow fetch geometries: The white-capping and wind input formulations. *Ocean Model.* 157, 101730. doi: 10.1016/j.ocemod.2020.101730

Elkut, A. E., Taha, M. T., Abu Zed, A. B. E., Eid, F. M., and Abdallah, A. M. (2021). Wind-wave hindcast using modified ECMWF ERA-Interim wind field in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* 252, 107267. doi: 10.1016/j.ecss.2021.107267

Feng, Z., Hu, P., Li, S., and Mo, D. (2022). Prediction of significant wave height in offshore China based on the machine learning method. *J. Mar. Sci. Eng.* 10, 836. doi: 10.3390/jmse10060836

Günther, H., Hasselmann, S., and Janssen, P. A. E. M. (1992). The WAM Model cycle 4. *Deutsches Klima Rechenzentrum*. doi: 10.2312/WDCC/DKRZ_Report_No04

Hasselmann, K. (1974). On the spectral dissipation of ocean waves due to white capping. *Bound.-Layer Meteorol.* 6, 107–127. doi: 10.1007/BF00232479

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., et al. (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut Hydrogr Z* 8, 1–95.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc* 146, 1999–2049. doi: 10.1002/qj.3803

Huerta, A. D., and Harry, D. L. (2012). Wilson cycles, tectonic inheritance, and rifting of the North American Gulf of Mexico continental margin. *Geosphere* 8, 374–385. doi: 10.1130/GES00725.1

Hwang, P. A. (2011). A note on the ocean surface roughness spectrum. *J. Atmospheric Ocean. Technol.* 28, 436–443. doi: 10.1175/2010JTECHO812.1

Janssen, P. A. E. M. (1992). "Consequences of the effect of surface gravity waves on the mean air flow," in *Breaking waves*. Eds. M. L. Banner and R. H. J. Grimshaw (Berlin, Heidelberg: Springer Berlin Heidelberg), 193–198. doi: 10.1007/978-3-642-84847-6_19

Jiang, Y., Rong, Z., Li, P., Qin, T., Yu, X., Chi, Y., et al. (2022). Modeling waves over the Changjiang River Estuary using a high-resolution unstructured SWAN model. *Ocean Model.* 173, 102007. doi: 10.1016/j.ocemod.2022.102007

Komen, G. J., Hasselmann, S., and Hasselmann, K. (1984). On the existence of a fully developed wind-sea spectrum. *J. Phys. Oceanogr.* 14, 1271–1285. doi: 10.1175/1520-0485(1984)014<1271:OTEOAF>2.0.CO;2

Kutupoğlu, V., Çakmak, R. E., Akpınar, A., and van Vledder, G. P. (2018). Setup and evaluation of a SWAN wind wave model for the Sea of Marmara. *Ocean Eng.* 165, 450–464. doi: 10.1016/j.oceaneng.2018.07.053

Liu, Q., Rogers, W. E., Babanin, A. V., Young, I. R., Romero, L., Zieger, S., et al. (2019). Observation-based source terms in the third-generation wave model WAVEWATCH III: updates and verification. *J. Phys. Oceanogr.* 49, 489–517. doi: 10.1175/JPO-D-18-0137.1

Lv, X., Yuan, D., Ma, X., and Tao, J. (2014). Wave characteristics analysis in Bohai Sea based on ECMWF wind field. *Ocean Eng.* 91, 159–171. doi: 10.1016/j.oceaneng.2014.09.010

Mears, C. A., Scott, J., Wentz, F. J., Ricciardulli, L., Leidner, S. M., Hoffman, R., et al. (2019). A near-real-time version of the cross-calibrated multiplatform (CCMP) ocean surface wind velocity data set. *J. Geophys. Res. Oceans* 124, 6997–7010. doi: 10.1029/2019JC015367

Ou, Y., Zhai, F., and Li, P. (2018). Interannual wave climate variability in the Taiwan Strait and its relationship to ENSO events. *J. Oceanol. Limnol.* 36, 2110–2129. doi: 10.1007/s00343-019-7301-3

Peng, J., Mao, M., and Xia, M. (2023). Dynamics of wave generation and dissipation processes during cold wave events in the Bohai Sea. *Estuar. Coast. Shelf Sci.* 280, 108161. doi: 10.1016/j.ecss.2022.108161

Pierson, W. J., and Moskowitz, L. (1964). A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.* 69, 5181–5190. doi: 10.1029/JZ069i024p05181

Rogers, W. E., Babanin, A. V., and Wang, D. W. (2012). Observation-consistent input and whitecapping dissipation in a model for wind-generated surface waves: description and simple calculations. *J. Atmospheric Ocean. Technol.* 29, 1329–1346. doi: 10.1175/JTECH-D-11-00092.1

Rogers, W. E., Kaihatu, J. M., Petit, H. A. H., Booij, N., and Holthuijsen, L. H. (2002). Diffusion reduction in an arbitrary scale third generation wind wave model. *Ocean Eng.* 29, 1357–1390. doi: 10.1016/S0029-8018(01)00080-4

Roland, A., and Ardhuin, F. (2014). On the developments of spectral wave models: numerics and parameterizations for the coastal ocean. *Ocean Dyn.* 64, 833–846. doi: 10.1007/s10236-014-0711-z

Sepulveda, H. H., Queffeulou, P., and Ardhuin, F. (2015). Assessment of SARAL/altiKa wave height measurements relative to buoy, jason-2, and cryosat-2 data. *Mar. Geod.* 38, 449–465. doi: 10.1080/01490419.2014.1000470

Shao, Z., Liang, B., Sun, W., Mao, R., and Lee, D. (2023). Whitecapping term analysis of extreme wind wave modelling considering spectral characteristics and water depth. *Cont. Shelf Res.* 254, 104909. doi: 10.1016/j.csr.2022.104909

Sharma, R., Chaudhary, A., Seemanth, M., Bhowmick, S. A., Agarwal, N., Verron, J., et al. (2022). SARAL/AltiKa data analysis for oceanographic research: Impact of drifting and post star sensor anomaly phases. *Adv. Space Res.* 69, 2349–2361. doi: 10.1016/j.asr.2021.12.008

Shi, H., Cao, X., Li, Q., Li, D., Sun, J., You, Z., et al. (2021). Evaluating the accuracy of ERA5 wave reanalysis in the water around China. *J. Ocean Univ. China* 20, 1–9. doi: 10.1007/s11802-021-4496-7

Son, D., Jun, K., Kwon, J.-I., Yoo, J., and Park, S.-H. (2023). Improvement of wave predictions in marginal seas around Korea through correction of simulated sea winds. *Appl. Ocean Res.* 130, 103433. doi: 10.1016/j.apor.2022.103433

Su, H., Wei, C., Jiang, S., Li, P., and Zhai, F. (2017). Revisiting the seasonal wave height variability in the South China Sea with merged satellite altimetry observations. *Acta Oceanol. Sin.* 36, 38–50. doi: 10.1007/s13131-017-1073-4

Sun, W., Liang, B., Shao, Z., and Wang, Z. (2022). Analysis of Komen scheme in the SWAN model for the whitecapping dissipation during the tropical cyclone. *Ocean Eng.* 266, 113060. doi: 10.1016/j.oceaneng.2022.113060

The SWAN team. (2021a). SWAN: User Manual (SWAN Cycle III version 41.31AB). *Delft Univ. Technol.* Available at: https://swanmodel.sourceforge.io/.

The SWAN team. (2021b). SWAN: Scientific and technical documentation (SWAN Cycle III version 41.31AB). *Delft Univ. Technol.* Available at: https://swanmodel.sourceforge.io/.

The WAMDI Group (1988). The WAM model—A third generation ocean wave prediction model. *J. Phys. Oceanogr.* 18, 1775–1810. doi: 10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2

Umesh, P. A., and Behera, M. R. (2021). On the improvements in nearshore wave height predictions using nested SWAN-SWASH modelling in the eastern coastal waters of India. *Ocean Eng.* 236, 109550. doi: 10.1016/j.oceaneng.2021.109550

Van der Westhuysen, A. J., Zijlema, M., and Battjes, J. A. (2007). Nonlinear saturation-based whitecapping dissipation in SWAN for deep and shallow water. *Coast. Eng.* 54, 151–170. doi: 10.1016/j.coastaleng.2006.08.006

Verron, J., Bonnefond, P., Andersen, O., Ardhuin, F., Bergé-Nguyen, M., Bhowmick, S., et al. (2021). The SARAL/AltiKa mission: A step forward to the future of altimetry. *Adv. Space Res.* 68, 808–828. doi: 10.1016/j.asr.2020.01.030

Verron, J., Sengenes, P., Lambin, J., Noubel, J., Steunou, N., Guillot, A., et al. (2015). The SARAL/altiKa altimetry satellite mission. *Mar. Geod.* 38, 2–21. doi: 10.1080/01490419.2014.1000471

Wang, H. (2021) *Analysis of spatial-temporal characteristics of waves and prediction of significant wave height in the South China Sea*. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1021879989.nh.

Wang, W., and Huang, R. X. (2004). Wind energy input to the surface waves. *J. Phys. Oceanogr.* 34, 1276–1280. doi: 10.1175/1520-0485(2004)034<1276:WEITTS>2.0.CO;2

Weatherall, P., Tozer, B., Arndt, J. E., Bazhenova, E., Bringensparr, C., Castro, C., et al. (2021). The GEBCO_2021 Grid - a continuous terrain model of the global oceans and land. *NERC EDS Br. Oceanogr. Data Cent. NOC.* doi: 10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f

Wentz, F. J. (2015). A 17-yr climate record of environmental parameters derived from the tropical rainfall measuring mission (TRMM) microwave imager. *J. Clim.* 28, 6882–6902. doi: 10.1175/JCLI-D-15-0155.1

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc* 63, 1309–1313. doi: 10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Wu, W., Li, P., Zhai, F., Gu, Y., and Liu, Z. (2020). Evaluation of different wind resources in simulating wave height for the Bohai, Yellow, and East China Seas (BYES) with SWAN model. *Cont. Shelf Res.* 207, 104217. doi: 10.1016/j.csr.2020.104217

Wu, W., Liu, Z., Zhai, F., Li, P., Gu, Y., and Wu, K. (2021). A quantitative method to calibrate the SWAN wave model based on the whitecapping dissipation term. *Appl. Ocean Res.* 114, 102785. doi: 10.1016/j.apor.2021.102785

Wu, S., Liu, J., Zhang, G., Han, B., Wu, R., and Chen, D. (2022). Evaluation of NCEP-CFSv2, ERA5, and CCMP wind datasets against buoy observations over Zhejiang nearshore waters. *Ocean Eng.* 259, 111832. doi: 10.1016/j.oceaneng.2022.111832

Xu, F., Perrie, W., and Solomon, S. (2013). Shallow water dissipation processes for wind waves off the mackenzie delta. *Atmosphere-Ocean* 51, 296–308. doi: 10.1080/07055900.2013.794123

Yan, L. (1987). An improved wind input source term for third generation ocean wave modelling. *NASA Tech. ReportsN88-*, 1–24.

Yang, Z., Lin, Y., and Dong, S. (2022). Weather window and efficiency assessment of offshore wind power construction in China adjacent seas using the calibrated SWAN model. *Ocean Eng.* 259, 111933. doi: 10.1016/j.oceaneng.2022.111933

Zhai, R., Huang, C., Yang, W., Tang, L., and Zhang, W. (2023). Applicability evaluation of ERA5 wind and wave reanalysis data in the South China Sea. *J. Oceanol. Limnol.* 41, 495–517. doi: 10.1007/s00343-022-2047-8

Zhai, F., Wu, W., Gu, Y., Li, P., and Liu, Z. (2021). Dynamics of the seasonal wave height variability in the South China Sea. *Int. J. Climatol.* 41, 934–951. doi: 10.1002/joc.6707

Zhang, X., Cheng, L., Zhang, F., Wu, J., Li, S., Liu, J., et al. (2020). Evaluation of multi-source forcing datasets for drift trajectory prediction using Lagrangian models in the South China Sea. *Appl. Ocean Res.* 104, 102395. doi: 10.1016/j.apor.2020.102395

Zheng, Z., Ali, M., Jamei, M., Xiang, Y., Abdulla, S., Yaseen, Z. M., et al. (2023). Multivariate data decomposition based deep learning approach to forecast one-day ahead significant wave height for ocean energy generation. *Renew. Sustain. Energy Rev.* 185, 113645. doi: 10.1016/j.rser.2023.113645

Zieger, S., Babanin, A. V., Erick Rogers, W., and Young, I. R. (2015). Observation-based source terms in the third-generation wave model WAVEWATCH. *Ocean Model.* 96, 2–25. doi: 10.1016/j.ocemod.2015.07.014

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 January 2024
DOI 10.3389/fmars.2023.1319719

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Shiqiang Yan,
City University of London, United Kingdom
Jiucai Jin,
Ministry of Natural Resources, China

*CORRESPONDENCE
Martin Outzen Berild
martin.o.berild@ntnu.no

RECEIVED 11 October 2023
ACCEPTED 21 December 2023
PUBLISHED 17 January 2024

CITATION
Berild MO, Ge Y, Eidsvik J, Fuglstad G-A and Ellingsen I (2024) Efficient 3D real-time adaptive AUV sampling of a river plume front. *Front. Mar. Sci.* 10:1319719. doi: 10.3389/fmars.2023.1319719

Efficient 3D real-time adaptive AUV sampling of a river plume front

Martin Outzen Berild[1]*, Yaolin Ge[1], Jo Eidsvik[1], Geir-Arne Fuglstad[1] and Ingrid Ellingsen[2]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway, [2]Fisheries and New Biomarine Industry, SINTEF Ocean, Trondheim, Norway

The coastal environment faces multiple challenges due to climate change and human activities. Sustainable marine resource management necessitates knowledge, and development of efficient ocean sampling approaches is increasingly important for understanding the ocean processes. Currents, winds, and freshwater runoff make ocean variables such as salinity very heterogeneous, and standard statistical models can be unreasonable for describing such complex environments. We employ a class of Gaussian Markov random fields that learns complex spatial dependencies and variability from numerical ocean model data. The suggested model further benefits from fast computations using sparse matrices, and this facilitates real-time model updating and adaptive sampling routines on an autonomous underwater vehicle. To justify our approach, we compare its performance in a simulation experiment with a similar approach using a more standard statistical model. We show that our suggested modeling framework outperforms the current state of the art for modeling such spatial fields. Then, the approach is tested in a field experiment using two autonomous underwater vehicles for characterizing the three-dimensional fresh-/saltwater front in the sea outside Trondheim, Norway. One vehicle is running an adaptive path planning algorithm while the other runs a preprogrammed path. The objective of adaptive sampling is to reduce the variance of the excursion set to classify freshwater and more saline fjord water masses. Results show that the adaptive strategy conducts effective sampling of the frontal region of the river plume.

KEYWORDS

adaptive sampling, ocean modeling, autonomous underwater vehicle, Gaussian random field, stochastic partial differential equations, surrogate model

1 Introduction

Human activities and pollution are heavily impacting the world's oceans (Halpern et al., 2008). Anthropogenic climate change and local intrusion from industries can lead to fundamentally altered ocean ecosystems, challenging species distributions, loss in biodiversity, incidence of disease, and more (Hoegh-Guldberg and Bruno, 2010; Doney et al., 2012). The changes in ecosystem structure further influence important services such as carbon sequestration, oxygen production, and nutrient food chains. In order to achieve a more sustainable utilization of marine resources and services, we need to enhance our insight. Developing smart technologies for efficient monitoring of the ocean can provide information that enables us to identify adverse effects and guide development of countermeasures, and it can hence be vital in saving or maintaining local ecosystems. Commonly used ocean observation technologies are buoys, drifters, satellites, unmanned surface vehicles, Argo floats, underwater gliders, cabled seafloor observatories, autonomous underwater vehicles (AUVs), hadal landers, or some coupled system of these technologies (see, e.g., Lin and Yang (2020) for an overview). Ocean monitoring systems are advancing from simple and static single sensors systems to dynamic and multisensor systems that can cover a large spectrum of temporal and spatial scales. With the drive in artificial intelligence and robotic systems, there is also a development toward intelligent sampling systems where observations of various kinds are gathered and processed where and when it is considered valuable.

With the improved affordability and functionality of AUVs, the research literature has seen many advances lately; Zhang et al. (2012; 2013) used deterministic algorithms to map coastal temperature upwellings; Das et al. (2015) demonstrated AUV mission planning for informative plankton sampling; Fossum et al. (2018) monitored large temperature gradients by adaptively choosing surveys paths that substantially reduce the uncertainty in the statistical temperature model; Fossum et al. (2019) conducted a 3D AUV survey for chlorophyll-a mapping; Mo-Bjørkelund et al. (2020) employed hexagonal grids for equilateral survey paths to adaptively explore large temperature gradients; Foss et al. (2022) used a 2D spatiotemporal model onboard an AUV to supervise mining waste seafill; Fonseca et al. (2023) compared satellite imagery and adaptive AUV sampling results for predicting algal blooms. These examples from recent research activity have advanced the field of ocean monitoring with AUVs by going from planar (sea-surface) fields to volumetric fields, in the combination of various data sources, or by presenting a novel algorithm for adaptive exploration.

Considering the vastness of our ocean, it is extremely difficult to obtain sufficient data to cover the full range of scale and resolution desired. Instead, one must rely on a combination of different data sources and sophisticated modeling tools. To fill in the gaps effectively, one can further proactively plan targeted and high-precision sampling campaigns that will improve predictions and support decision-making. At its core, these tasks relate to statistical methods that can combine various data sources for prediction and for evaluating data sampling designs to optimize further data-gathering efforts.

In this work, we combine the fields of oceanography, statistics, and robotics to effectively monitor freshwater frontal regions of river outlets in three dimensions (north, east, depth) using AUVs. Specifically, we conduct sampling in the Nidelva River running into the fjord outside Trondheim, Norway (see Figure 1). The freshwater coming from the river is mixing slowly with the more saline fjord waters, which can cause a sharp gradient between the different water masses.

At our availability, we have output from a complex numerical ocean model Slagstad and McClimans (2005), henceforth referred

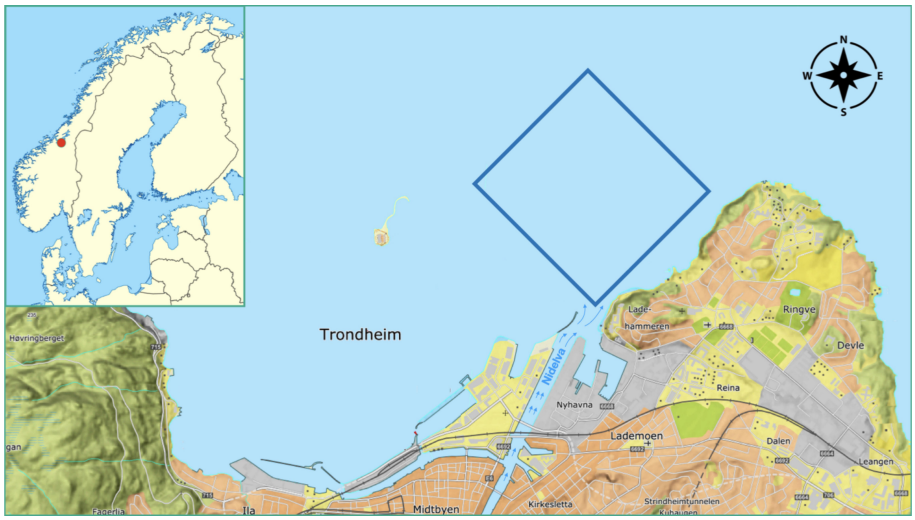


FIGURE 1
Map of the operational area in the fjord outside Trondheim, Norway. The location of Trondheim is indicated by the red circle on the map of Scandinavia in the top left corner. We have 3D numerical ocean model data at a high lateral and depth resolution in these coastal waters. The blue square just north of the Nidelva River outlet indicates the boundaries of the autonomous underwater vehicle mission in a map view. The operation domain extends from the sea surface to 5-m depth.

to as SINMOD. Along with many other physical oceanography variables, SINMOD outputs salinity at every grid node in a dense spatial (3D) and temporal grid. Even though this model carries much physical insight, the salinity output can be systematically biased, and we will calibrate and update the salinity by deploying an AUV. In this way, the SINMOD data are used to form a prior model for the salinity trends and variations at the time of the AUV deployment. We fit a Gaussian process prior as a surrogate model to the numerical ocean model SINMOD. This surrogate model has the advantage that it can be updated onboard the AUV, and it can hence assimilate *in-situ* data efficiently. Moreover, this surrogate model enables fast evaluation of various AUV sampling designs in real time whereas it is maneuvering in the water. Fossum et al. (2021) used similar methods to fit a surrogate model from SINMOD, but only in 2D space. Ge et al. (2023) used a 3D Gaussian surrogate prior model, but only for a small-size grid and assuming a much simpler spatial dependency structure.

This paper brings together many elements, and the novelty lies in a more realistic description of spatial correlations with a complex model learned from SINMOD data (Berild and Fuglstad, 2023). The approach is made computationally feasible through modern techniques using a 3D Gaussian random fields with a Markov property, and this enables adaptive AUV sampling based on the new surrogate model in a large-size 3D waypoint graph used during AUV deployments. Additionally, we

- fit the more realistic 3D statistical model to 3D numerical ocean data, and develop a fast algorithm for updating this model onboard an AUV during field deployment,
- develop methods for adaptive path-planning in the context of 3D space with the more realistic model onboard the AUV,
- show through a simulation study, based on SINMOD, that the more realistic statistical model allows an AUV to sample and map the ocean domain better than with a standard statistical model,
- run two AUVs simultaneously in the ocean and show that the combination of an intelligent adaptive survey design and the more realistic model outperforms a standard pre-scripted AUV sampling plan.

In Section 2, we describe the numerical ocean model and its statistical surrogate model. In Section 3, we present the data assimilation part and our approach for adaptive AUV sampling designs. In Section 4, we study properties of the suggested methods in a simulation study. In Section 5, we show results of deployments with one adaptive AUV mission and one preprogrammed mission. In Section 6, we conclude and point to future work.

2 Prior model for salinity

Consider a three-dimensional ocean domain $\mathcal{D} \subseteq \mathbb{R}^3$, where $x(s)$ represents the salinity field at a specific location $s = (\text{latitude}, \text{longitude}, \text{depth})^T \in \mathcal{D}$. The salinity in this ocean domain exhibits both spatial and temporal variations. However, we focus on short-term AUV deployments and simplify our analysis by excluding temporal effects.

2.1 Numerical ocean model

An approximation of the salinity field is achieved using the complex numerical ocean model SINMOD, developed by SINTEF ocean (Slagstad and McClimans, 2005). SINMOD is a three-dimensional model based on the primitive equations, solved using finite difference methods on a regular grid with horizontal cell sizes of 20 km × 20 km, which are nested in several steps down to 32m × 32m for the bay outside Trondheim. The model employs varying vertical resolutions, allowing for higher resolution near the dynamic surface and more uniform resolution in deeper waters. Atmospheric forces (obtained from forecasts available at https://www.met.no), freshwater outflows (data from HBV model (Beldring et al., 2003) provided by the Norwegian Water Resources and Energy Directorate (NVE)), and tides (https://www.tpxo.net/) drive the model. SINMOD offers numerical simulations of multiple ocean variables, including temperature and currents as well as salinity. It is a multipurpose tool that has been used for instance in the prediction of Arctic ocean primary production by leveraging physical–biological coupling (Slagstad et al., 2015; Vernet et al., 2021), in quantifying the effects of the aquaculture structures for large-scale cages by specifying and incorporating drag parameters in SINMOD (Broch et al., 2020), and coupled with the particle dispersion of waste from fish farming (Broch et al., 2017), oil production (Nepstad et al., 2022), or mine tailings (Berget et al., 2018; Nepstad et al., 2020; Berget et al., 2023). For a more comprehensive explanation of the SINMOD methodology, readers are directed to Slagstad and McClimans (2005).

In the current paper, we are only using the salinity outputs from SINMOD. Figure 2 shows an example of SINMOD salinity data and an excursion set (salinity ≤25.4 g/kg) separating water masses into freshwater/saltwater in the fjord outside Trondheim. We notice that the river plume has lower salinity than the surrounding brackish water. There are very low salinity levels (around 5 g/kg–10 g/kg) in the river outlet, whereas the salinity increases further out in the fjord (around 31 g/kg). The temporal resolution of the SINMOD numerical model used for these simulations is 10 min. Salinity is measured in grams salinity per kilogram water (g/kg), which is dimensionless and equal to ‰ and sometimes referred to as the practical salinity unit (PSU).

2.2 Surrogate model with spatially varying anisotropy

In-situ salinity observations made with an AUV are assumed to be more accurate than the forecast provided by SINMOD. However, an AUV measurement only characterizes the salinity at the specific location where the measurement was taken, whereas a model like SINMOD or similar is required to extrapolate variables in space and time. For onboard computing, SINMOD is however too computationally intensive, and it is challenging to assimilate AUV

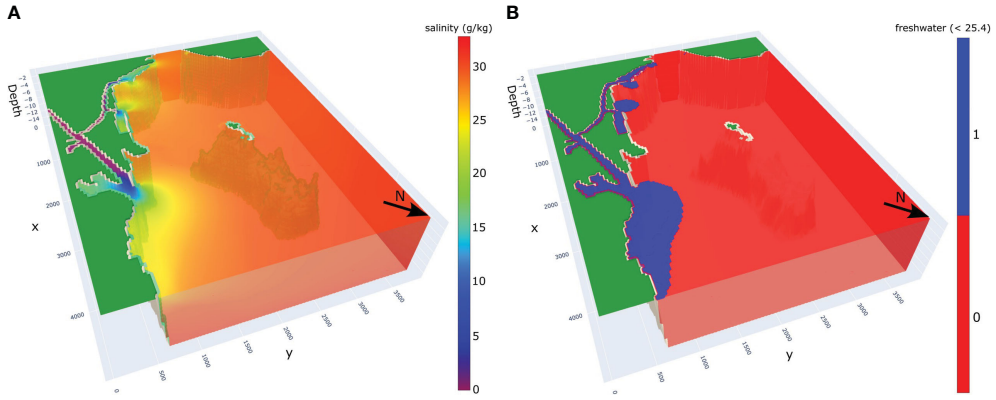


FIGURE 2
Salinity simulation **(A)** and corresponding freshwater excursion **(B)** from the numerical ocean model SINMOD for 08/09/2022.

observations in real time with a full-fledged numerical ocean model. Instead, a surrogate model can be trained from the numerical model. It forms an approximate representation of the underlying physical model and is highly applicable for different tasks that require fast updating. Statistical models with spatial effects have shown very suitable for such a task (Gramacy, 2020), and we employ a particular statistical surrogate model for the numerical ocean model salinity data here.

We use the spatial statistical model presented in Berild and Fuglstad (2023), where the 3D salinity field is modeled as a Gaussian Markov random field (GMRF) that allows sparse matrix computations and realistic modeling *via* spatial variability in the directional dependencies and the variance components. This model is an extension of Lindgren et al. (2011) and Fuglstad et al. (2015).

Assuming that the 3D discretization of the domain $\mathcal{D} \subseteq \mathbb{R}^3$ consists of $n_1 \times n_2 \times n_3$ grid cells, the salinity field, $x(\boldsymbol{s})$, is represented by a vector of concatenated field values of size $n = n_1 n_2 n_3$. In the application, we have $n_1 = 50$, $n_2 = 45$, $n_3 = 6$ with 32×32 m^2 lateral resolution and 1-m depth resolution. The vector x of salinity values is modeled by a Gaussian distribution, i.e.,

$$\boldsymbol{x} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma), \qquad \Sigma = \mathbf{Q}^{-1}. \tag{1}$$

Here, the $\sim$ symbol means "distributed according to," and $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ refers to the n-variate Gaussian (or normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, where its inverse, namely, the precision matrix, is denoted $\mathbf{Q}$.

There is much flexibility in choosing the mean vector and covariance matrix in Equation (1), and the Gaussian distribution can hence form quite realistic surrogate models. The mean vector $\boldsymbol{\mu}$ of the salinity field captures the spatial trends of the field, which in our case entails fresher water near the river gradually getting more saline going out in the fjord. To form a realistic covariance structure, the idea of Berild and Fuglstad (2023) is to form a random process for $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{\mu}$ *via* differential operators and Gaussian noise forming a stochastic partial differential equation (SPDE) as

$$(\kappa^2(s) - \nabla \cdot \mathbf{H}(s) \nabla)u(s) = \mathcal{W}(s). \tag{2}$$

Here, $\boldsymbol{s}$ is a location in the domain of interest $\mathcal{D} \subseteq \mathbb{R}^3$, $u(\boldsymbol{s})$ is the spatially varying deviation from the trend, and and $\mathcal{W}(s)$ is a Gaussian white noise process (with zero mean and statistically independent values), where as $\kappa(\mathbf{s}) = \kappa(\boldsymbol{s}; \boldsymbol{\theta}) > 0$ and $\mathbf{H}(\boldsymbol{s}) = \mathbf{H}(\boldsymbol{s}; \boldsymbol{\theta}) > 0$ as differentiable are model components controlled by parameters θ that regulate the variability and dependency within the process.

Equation (2) is solved locally for the zero-mean random field $u(\boldsymbol{s})$ using numerical integration and differentiation on a discretization of the domain of interest $\mathcal{D}$. The solution is $\boldsymbol{u} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}^{-1})$, where the precision matrix $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta})$ inherits the sparsity of the differential operators in Equation (2), and it describes the Markov structure in the GMRF model. This structure is very important for our purposes because it enables fast matrix factorization and matrix-vector computations. Hence, the GMRF formulation means that we can update the model onboard the AUV. It is also used in the sampling design evaluations. Without this sparsity, the Gaussian surrogate model could not scale up the magnitude of the ocean mass in 3D (Berild and Fuglstad, 2023).

A detailed description of the model is provided in the Supplementary Material.

2.3 Parameter estimation for salinity field

In order to estimate the parameters and components of the statistical GMRF model for salinity, we utilize numerical ocean model data from SINMOD as the training dataset. These data are denoted as $y(\boldsymbol{s}_i, t_j)$, where $\boldsymbol{s}_i \in \mathcal{D}$ represents the location of cell $i \in [1,\ldots,n]$ at time t_j for SINMOD realization $j = 1,\ldots,T$. The surrogate data model is then

$$y(\boldsymbol{s}_i) \sim \mathcal{N}(x(\boldsymbol{s}_i), \sigma_S^2), \qquad i = 1, \ldots, n,$$

where σ_S^2 is an unstructured noise variance of the SINMOD dataset.

We estimate a location-dependent mean $\mu(\boldsymbol{s}_i)$ of the GMRF using the empirical average across all replicates t_j as:

$$\hat{\mu}(\boldsymbol{s}_i) = \frac{1}{T}\sum_{j=1}^{T} y(\boldsymbol{s}_i, t_j), \quad \forall \boldsymbol{s}_i \in \mathcal{D}. \tag{3}$$

We compute an estimate $\hat{\boldsymbol{\theta}}$ of the covariance parameters of the GMRF by maximizing the likelihood function $L(\boldsymbol{\theta})$, given residual data from an autoregressive model fitted to the SINMOD data [see Supplementary Material and (Berild and Fuglstad, 2023)]. In Sections 4 and 5, the covariance parameters of the models are estimated using a dataset from the SINMOD model, which spans 144 timesteps over the course of 1 day with a temporal resolution of 10 min. This dataset includes observations for all locations within the spatial field at each timestep. Optimizing the likelihood of this rather sophisticated covariance model is not straightforward, but it gets less difficult with more data and this also improves the accuracy of estimates. Berild and Fuglstad (2023) suggest that at least 10 timesteps of the whole field should be used to find reasonable parameter values for such a flexible model.

Figure 3 shows the prior mean (Equation (3)), the prior variance of the n-variate Gaussian distribution $\boldsymbol{x}$, and the corresponding spatial correlation of the marked location. The mean salinity clearly increases going north in the fjord, away from the river outlet. The salinity variance is larger near the river. For the correlation, we notice non-circular contours indicative of anisotropy. Here, the correlation appears to be stronger in the directions where salinity is expected to be similar to that of the reference location.

3 Adaptive AUV sampling

We now delve into our approach for adaptive AUV exploration. One part of this involves continuous updates of the GMRF surrogate model through onboard data assimilation of the *in-situ* AUV salinity data. Another part is the strategic planning of the next AUV sampling locations.

3.1 Conditioning to AUV data

Assume that the AUV gathers *in-situ* data at m locations or design points $\boldsymbol{d} = \boldsymbol{d}_1, \ldots, \boldsymbol{d}_m$, where $\boldsymbol{d}_j \in \mathcal{D}$. In practice, these locations form an AUV design (a trajectory). Data $y(\boldsymbol{d}_j), j = 1, \ldots, m$, are noisy measurements of the salinity $x(\boldsymbol{d}_j)$ at the location $\boldsymbol{d}_j$ where they are made. We organize the data in a length-m vector $\boldsymbol{y}$, and we allocate these observations to the correct grid locations by using a size $m \times n$ selection matrix $\mathbf{A}$. This matrix has a single 1 entry in each row, and otherwise only 0 entries. With this structure, it selects the m indices in the length-n vector x of discretized salinity field variables in Equation (1). The measurement model is then

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N_m(\mathbf{0}, \sigma^2_{\text{auv}}\mathbf{I}_m).$$

Here, the variance σ^2_{auv} of the independent additive noise terms aggregates the AUV positioning error and measurement noise. This variance parameter is specified from existing AUV data.

The conditional model for salinity $\boldsymbol{x}$, given measurements $\boldsymbol{y}$, is Gaussian distributed with an updated precision matrix

$$\mathbf{Q}_{\text{C}} = \mathbf{Q} + \mathbf{A}^{\text{T}}\mathbf{A}/\sigma^2_{\text{auv}}, \tag{4}$$

and conditional mean

$$\boldsymbol{\mu}_{\text{C}} = \boldsymbol{\mu} + \mathbf{Q}_{\text{C}}^{-1}\mathbf{A}^{\text{T}}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu})/\sigma^2_{\text{auv}}. \tag{5}$$

With the sparse precision matrices, the updating in Equations (4, 5) can be computed very fast.

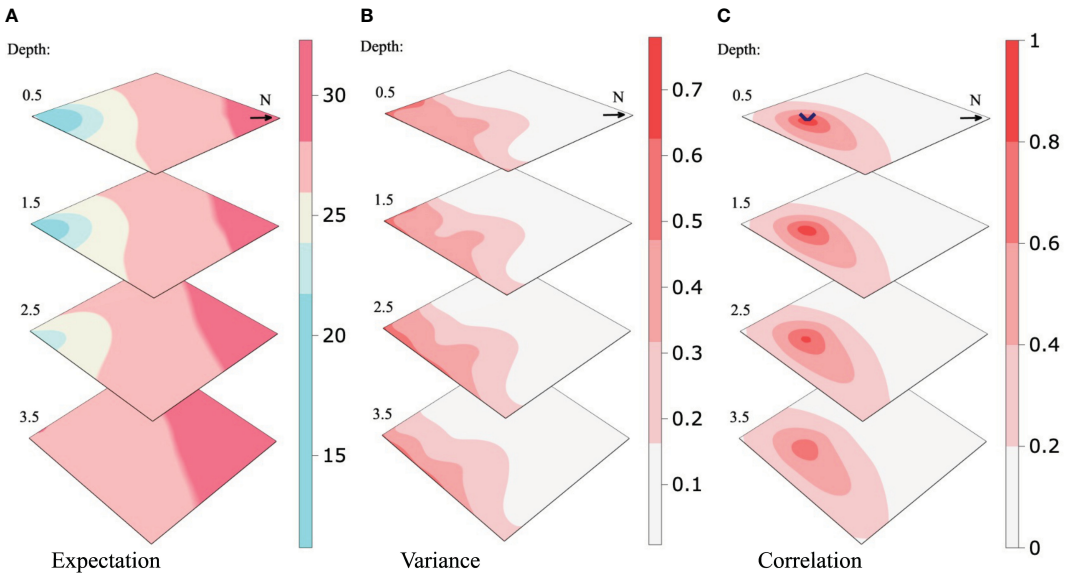


FIGURE 3
Prior expectation **(A)**, the variance of the process model **(B)**, and the spatial correlation of the location highlighted **(C)**. The N-arrow is the cardinal north.

Given a series of observations collected with the AUV along a straight line from the river plume and straight north, we calculate the conditional precision matrix and mean using Equations (4) and (5) of the model estimated in Section 2.3. Using the precision, we calculate the inverse diagonal (the conditional variance of the field), and from this the correlation about a location in space. We demonstrate the effect of data conditioning using a visualization of the conditional expectation, conditional variance, and conditional correlation given a series of updates, which are shown in Figures 4-6, respectively. Figure 4 indicates that the river water is going further north than anticipated in the prior mean. In Figure 5, we see that the variance is reduced where the AUV has visited, and as a consequence, the correlation range shown in Figure 6 gets lower. Dense data sampling tends to reduce the spatial correlation.

3.2 Excursion sets and plume mapping criterion

One goal of the AUV sampling is to improve the characterization of the plume front defined in our case as the zone separating fresh river waters and more saline fjord waters. Following Fossum et al. (2021) and Ge et al. (2023), we use the uncertainty in the random set of excursions below a salinity threshold ℓ to distinguish river and fjord water. The excursion set is defined by

$$\mathrm{ES} = \{\boldsymbol{s} \in \mathcal{D} : x(\boldsymbol{s}) < \ell\}.$$

The associated excursion probability (EP) and the Bernoulli variance (BV) is

$$\mathrm{EP}(\boldsymbol{s}) = P(x(\boldsymbol{s}) < \ell), \quad \mathrm{BV}(\boldsymbol{s}) = \mathrm{EP}(\boldsymbol{s})[1 - \mathrm{EP}(\boldsymbol{s})] \qquad s \in \mathcal{D}.$$

The BV is near 0 at locations where the EP is near 0 or 1, whereas it is at its maximum value 0.25 at locations, which have EP equal to 0.5.

When AUV salinity data $\boldsymbol{y}$ are available, we get a conditional GMRF and conditional EPs and BVs. Effective AUV sampling designs get salinity data that can pull these EPs closer to 0 or 1, and in doing so, one reduces the uncertainty of the river plume front. Conditional on salinity data $\boldsymbol{y} = \boldsymbol{y}_d$ according to design $\boldsymbol{d} \in \mathcal{D}$. The conditional EPs and BVs are

$$P(x(\boldsymbol{s}) < \ell|\boldsymbol{y}_{\boldsymbol{d}}), \quad P(x(\boldsymbol{s}) < \ell|\boldsymbol{y}_{\boldsymbol{d}})[1 - P(x(\boldsymbol{s}) < \ell|\boldsymbol{y}_{\boldsymbol{d}})].$$

Design plans must be made before the data $\boldsymbol{y}_{\boldsymbol{d}}$ is revealed, and we take the expectation over the data when calculating the most effective design. Focusing on improved spatial mapping of the river plume front, it is natural to integrate the objective criterion over all locations in the domain. The expected integrated Bernoulli variance (EIBV) of a design $\boldsymbol{d}$ is then defined by

$$\mathrm{EIBV}_d(\boldsymbol{\mu}, \mathbf{Q}) = \int_{\mathcal{D}} E_{y_d}\{P(x(s) < \ell|\boldsymbol{y}_{\boldsymbol{d}})[1 - P(x(s) < \ell|\boldsymbol{y}_{\boldsymbol{d}})]\}d\boldsymbol{s}. \quad (6)$$

For the GMRF surrogate model specified by mean $\boldsymbol{\mu}$ and precision $\mathbf{Q}$, the EIBV for a design $\boldsymbol{d}$ has a closed form involving sums of bivariate cumulative distribution functions Φ_2 for the Gaussian distribution. In this expression, the design is here involved *via* a one-entry structure of the selection matrix $\mathbf{A} = \mathbf{A}_{\boldsymbol{d}}$. The closed-form solution facilitates very fast computations of multiple sampling designs. The complete derivations of the closed forms are in the Supplementary Material. See also Fossum et al. (2021) and Ge et al. (2023). In our approach with the sparse GMRF model, we use Monte Carlo sampling from the conditional model to approximate the variance reduction components that are required in the EIBV (see Supplementary Material).

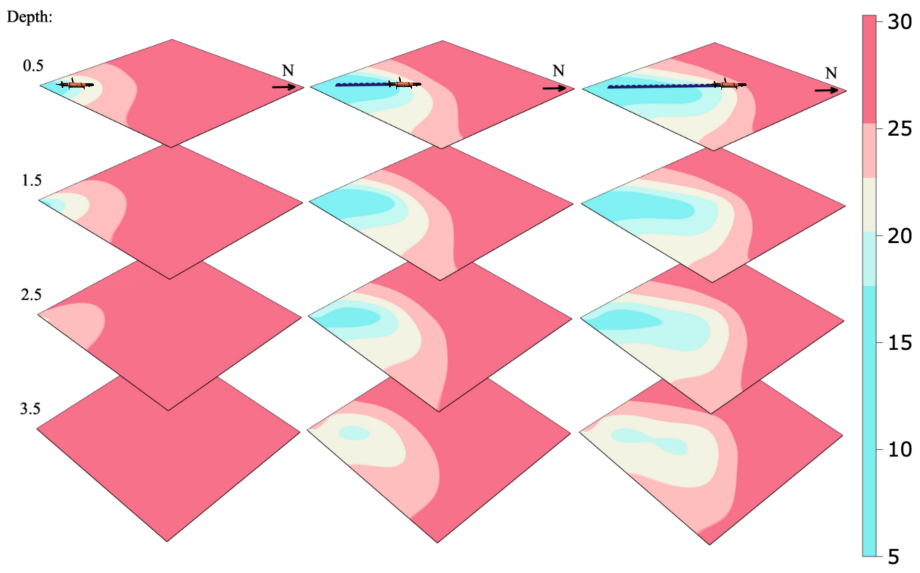


FIGURE 4
Conditional expectation given AUV measurements along a fixed transect path at a 0.5-m depth.

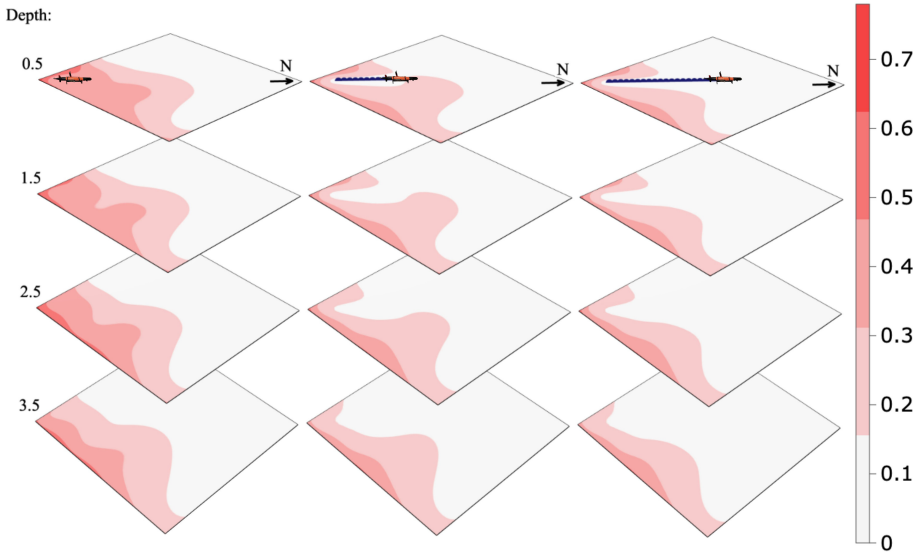


FIGURE 5
Conditional variance in the process model given AUV measurements along a fixed transect path at a 0.5-m depth.

3.3 Adaptive AUV sampling algorithm

The AUV cannot navigate to all possible design locations. Rather, its continued path is constrained by the current location and the possible maneuvers it can perform. We let $\mathcal{P} \subset \mathcal{D}$ denote the possible designs the AUV can choose from, defined by directions (straight, left, right, up, down) from the current AUV location. The chosen design is the one that minimizes the EIBV in Equation (6). This means that

$$\mathbf{d}^* = \text{argmax}_{\mathbf{d}\in\mathcal{P}} \text{EIBV}_d(\boldsymbol{\mu}, \mathbf{Q}) . \quad (7)$$

During the AUV operation, this kind of design choice is done at many time points, and with an updated model that is conditional on all the data gathered up to this point. In this way, we utilize the benefits of robotic intelligence to navigate the uncertain ocean plume zone.

We outline a myopic adaptive sampling algorithm in the 3D domain. This is a sequential selection of waypoints or grid nodes where the AUV data are sampled and the model updated. The myopic approach represents a heuristic optimization strategy for the AUV operation that does not anticipate potential data or navigation choices beyond the current time. It makes the optimal choice based on the expected values at the current time alone.

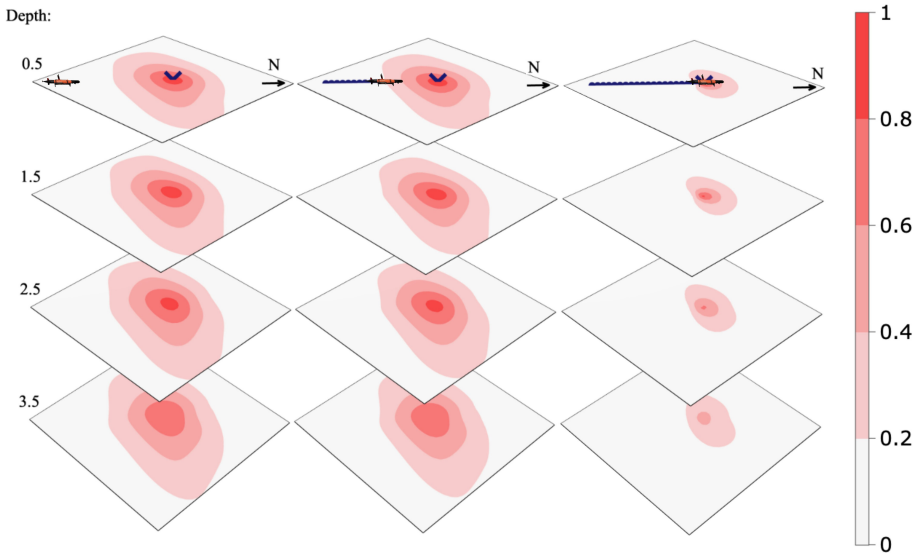


FIGURE 6
Conditional correlation of the marked point given AUV measurements along a fixed transect path at a 0.5-m depth.

Figure 7A shows the idea of adaptive sampling in a sketch with a cycle of tasks where one leads to the next. Here, the AUV senses the salinity, updates its onboard model, and plans where to navigate to, and then it continues on the next cycle.

Hence, at the planning stage, the computer onboard the AUV solves Equation (7) to navigate in promising 3D directions (Figure 7B). To compensate for the time it takes to do the computation, and to make the system near real time, asynchronous parallel computing is applied to compensate for the excessive computing time onboard.

Algorithm 1 shows the main steps of this adaptive AUV sampling approach. In this algorithm, we use t to indicate subsequent stages of AUV sampling. At stage t, the updated mean in the onboard surrogate model is denoted $\boldsymbol{\mu}_{C,t}$ and the updated precision is $\mathbf{Q}_{C,t}$. The selection matrix $\mathbf{A}_t = \mathbf{A}(\boldsymbol{d}_t)$ is formed based on the most promising design $\boldsymbol{d}_t$ at each stage. This design $\boldsymbol{d}_t$ is chosen among several possible designs $\mathcal{P}_t \subset \mathcal{D}$ that vary depending on where the AUV is at the current stage and the operational navigation opportunities it has according to the grid. In our implementation, the AUV can continue from its current location to go straight ahead, or turn left, right, up, or down. It cannot return back to its previous grid location (Figure 7B). There are natural exceptions at the grid boundary.

Initialization: Prior model $\boldsymbol{\mu}_{C,0} = \boldsymbol{\mu}$, $\mathbf{Q}_{C,0} = \mathbf{Q}$. Set start location $\boldsymbol{d}_0$. Set $t = 1$.
while True do
  Plan:
    Choose design that reduces EIBV the most

$$\boldsymbol{d}_t = \arg\min \mathrm{EIBV}_{\boldsymbol{d}}(\boldsymbol{\mu}_{C,t-1}, \mathbf{Q}_{C,t-1})$$

$$d \in P_t$$

  Form selection matrix $\mathbf{A}_t = \mathbf{A}(\boldsymbol{d}_t)$.
  Act/sense:
    Move according to design $\boldsymbol{d}_t$ collecting measurements $\boldsymbol{y}_t$
  Model Update:
    With the collected measurements, update the GMRF

$$\mathbf{Q}_{C,t} = \mathbf{Q}_{C,t-1} + \mathbf{A}_t^T \mathbf{A}_t / \sigma^2_{\text{auv}}$$

$$\boldsymbol{\mu}_{C,t} = \boldsymbol{\mu}_{C,t-1} + \mathbf{Q}^{-1}_{C,t} \mathbf{A}_t^T (\boldsymbol{y}_t - \mathbf{A}_t \boldsymbol{\mu}_{C,t-1}) / \sigma^2_{\text{auv}}$$

    Set $t = t + 1$.
end while

Algorithm 1. Myopic EIBV minimizing sampling with a GMRF surrogate model.

4 Simulation study

In this section, we conduct a simulated experiment to evaluate the performance of our approach for monitoring the three-dimensional freshwater plume of the Nidelva River in Trondheim, Norway. The operational area is outlined in Figure 1. Specifically, we will compare the effectiveness of the suggested complex GRF model and a more standard model. The complex model is discretized with a resolution of 32m x 32m square cells in the horizontal plane and the standard model with a hexagonal grid with a lateral neighbor distance of 120 m. Both models have 1-m-depth increments ranging from 0.5 m to 5.5 m, resulting in a total of $n = 50 \times 45 \times 6$ spatial location for the complex model and 1,098 for the standard model. This is in line with the capabilities of the AUVs' onboard computer.

Initially, both models are estimated on the SINMOD data within the operational area in order to form a prior field. The standard model is specified using a standard variogram analysis, resulting in a Matérn covariance with a lateral correlation range of 550 m, a vertical range of 2 m, a prior marginal variance of 1, and a nugget effect of 0.4 (see Section 2.4 of Cressie (1993) for a description of this spatial data analysis method). The parameters of the complex model are estimated through the approach described in Section 2.3, and detailed in the Supplementary Material and Berild and Fuglstad (2023). Both models use the empirical average across all timesteps (replicates) of the SINMOD data, Equation (3), as its prior expectation.

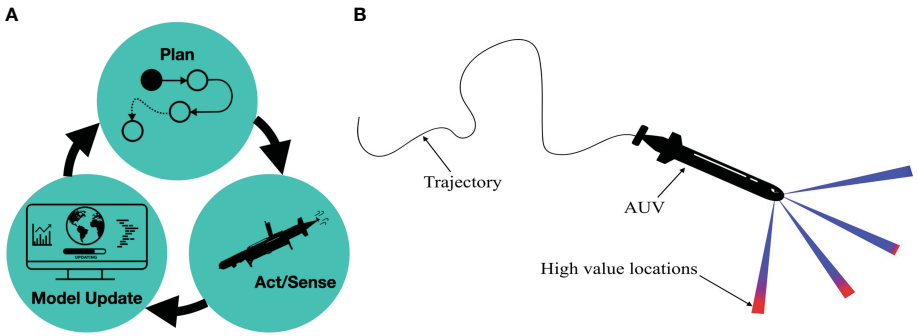


FIGURE 7
Illustration of the adaptive sampling mechanism. Visualization of the adaptive sampling design **(A)**. The AUV evaluates the potential high-value locations to determine the next visiting waypoint **(B)**. Red colors represent more interesting next waypoints.

In order to obtain performance statistics, we ran $L = 100$ simulated field experiments where the AUV is equipped with either one of the models estimated above and tasked with monitoring the salinity field according to Algorithm 1. The AUV is in this simulation environment exploring a SINMOD dataset from the 09/11/2022 with an assumed additional Gaussian noise term with standard deviation 0.12. This noise represents positional error and measurement error in a real experimental setting. Moreover, the AUV is set to travel at 1 m/s and each simulated field experiment is run for $T = 25$ sequential steps of Algorithm 1, i.e., visiting 25 spatial locations, where the starting location is kept the same for each run.

Within the lth simulated experiment and after visiting the tth location, the following three metrics are calculated: integrated Bernoulli variance (IBV), root mean squared error (RMSE), and classification error (CE). Let x_l be the ground truth (SINMOD data) in the lth experiment. Then, we calculate the metrics as

$$\mathrm{IBV}_{l,t} = \sum_{i=1}^{n} \mathrm{EP}_{l,t}(\boldsymbol{s}_i)1 - \mathrm{EP}_{l,t}(\boldsymbol{s}_i)],$$

$$\mathrm{RMSE}_{l,t} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[x_l(\boldsymbol{s}_i) - \mu_{C,l,t}(\boldsymbol{s}_i)]^2},$$

$$\mathrm{CE}_{l,t} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(\mathrm{E}_l(\boldsymbol{s}_i) \neq \hat{\mathrm{E}}_{l,t}(\boldsymbol{s}_i)), \qquad \mathrm{E}_l(\boldsymbol{s}_i) = \mathbb{I}(x_l(\boldsymbol{s}_i) \leq \ell),$$

$$\hat{\mathrm{E}}_{l,t}(\boldsymbol{s}_i) = \mathbb{I}(\mu_{C,l,t}(\boldsymbol{s}_i) \leq \ell),$$

where $\mathbb{I}$ is the indicator function, $t \in [0,T]$, where $T = 25$ indicates the sequential step, and $l = 1, \ldots, L$ with $L = 100$ replicate field experiment. The summary statistics of these metrics from the L replicated experiments are shown in Figure 8. The solid lines are the average across all L replicates at time t for each metric, e.g.,

$$\widehat{\mathrm{IBV}}_t = \frac{1}{L}\sum_{l=1}^{L}\mathrm{IBV}_{l,t},$$

and similarly the error bars show the empirical standard deviation at time t across all L replicates as

$$\mathbb{SD}(\mathrm{IBV}_t) = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(\mathrm{IBV}_{l,t} - \widehat{\mathrm{IBV}}_t)^2}.$$

Each display has one of the metrics on the second axis and time stages on the first axis. For the IBV and CE criteria, the percentage reduction compared with the starting value is shown since the models are constructed differently and therefore will also differ prior to the mission, as can be seen in the middle RMSE display.

The IBV reduction (Figure 8A) indicates the ability of the AUV to capture the river plume boundary. A lower IBV means that the AUV is better at sampling the frontal salinity region separating river and fjord water masses. In this spatial example, the IBV has a tendency of going down, even though it could increase at some stages (because data pull probabilities closer to 0.5). The complex GMRF model clearly achieves lower IBV than the simpler model. After some stages, the curve for the complex GMRF model declines rapidly, indicating that the AUV is efficient at exploring the boundary. This means that incorporating a more realistic covariance structure helps the AUV choose the best designs and it tends to move in the right direction.

The RMSE plot (Figure 8B) reflects the similarity between the ground truth and the updated field. The ground truth is here the same as what the AUV is sampling, i.e., the SINMOD dataset from 09/11/2022, but without the added noise term. A lower RMSE means that the AUV is gathering data that helps in predicting the salinity field. Again, the complex model is performing much better than the simpler one. For CE (Figure 8C), a lower value means that the updated model is good at classifying the excursion set associated with the ground truth. The complex model has CE results that are declining faster than the simpler model. The complex model performs better than the standard model due to its versatile capability and flexibility. However, we do realize that training such models often requires expert knowledge and it can be a laborious process to fine-tune the parameters for such a complex model.

In all displays of Figure 8, we observe larger metric variability for the complex GMRF model. The underlying reason for this is the Monte Carlo variance in the EIBV calculation for the GMRF model (see Section 3.2). With the relatively small sampling size, the Monte Carlo error is still not negligible and, influenced by this estimate, the

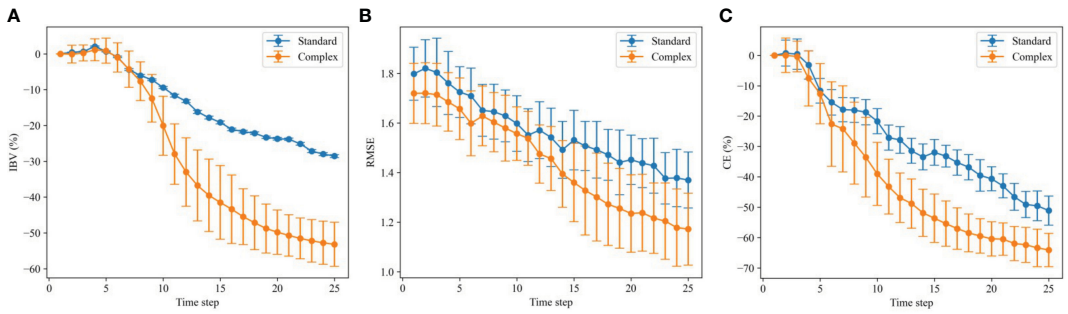


FIGURE 8
Variation in integrated Bernoulli variance **(A)**, root mean square error **(B)**, and classification error **(C)** over the 100 replicate runs with the standard model (blue) and the complex model (orange). The solid lines show the averages, and the vertical error bars show the empirical standard deviations of the respective metrics.

directional sampling decision made by the AUV exhibits more small-scale variability than that of the standard model, which has a closed-form variance expression. Over many replicates, the variability in metrics then gets larger for the GMRF model, especially for the IBV, which relates directly to the AUV sampling decision criterion.

5 Results of Nidelva mission

The field experiment was executed in the Nidelva River plume outside Trondheim, Norway, on the 08/09/2022. The duration of this field deployment spanned 1.5 h. Figure 1 shows the operational area.

5.1 Experimental setup

For this experiment, two AUVs are deployed. This is intended to not only increase the amount of data collected but also enable us to compare the performance of our embedded system under similar conditions. One of the AUVs was programmed with the adaptive sampling algorithm, whereas the other was running with a preprogrammed path plan onboard.

LAUV (Light Autonomous Underwater Vehicle) Harald and LAUV Roald (Figure 9) from the Applied Underwater Robotics Laboratory at NTNU were employed for this mission. LAUV Roald was programmed to carry out the adaptive experiment, and LAUV Harald was programmed to conduct the predesigned plan. To measure the salinity in the water, LAUVs Harald and Roald use CTD sensors, or conductivity, temperature, and depth sensors. Harald uses a SeaBird SBE 49 FastCAT and Roald an AML OEM SV Xchange. Despite being from different manufacturers, the specifications from the suppliers indicate that they should have the same level of precision and accuracy. All the essential scripts were integrated onboard on the backseat NVIDIA Jetson TX2 CPU. For hardware and software in the loop testing and the actual deployment, we relied on the framework developed by Mo-Bjørkelund et al. (2020). The onboard implementation of Algorithm requires Robot Operating Systems (ROS) (Quigley, 2009) and a software bridge to the LAUV, running DUNE (DUNE: Unified Navigation Environment Pinto et al. (2013)) embedded and communicating over the Inter Module Communication (IMC) message protocol (LSTS, 2022).

The software bridge between ROS and IMC was adapted from the Swedish Maritime Robotics Centers implementation of a ROS-IMC bridge (Bhat et al., 2020) (https://github.com/smarc-project/imc_ros_bridge) to include messages going from ROS to the vehicle. In addition, a wrapper for the vehicle IMC messages was used, facilitating interaction between the adaptive software and the vehicle. The communication bridge and framework between ROS and IMC use the same back-seat interface as Pinto et al. (2018), with IMC messages being transmitted over Transmission Control Protocol (TCP) (Cerf and Kahn, 1974) between the main CPU and the auxiliary CPU in the AUV. The adaptive code was running on the auxiliary CPU in order to preserve the integrity of the main CPU. For illustration, a flowchart containing the main software components is presented in Figure 10.

Before conducting the principal deployment, we gained understanding of the sea conditions *via* a preliminary survey. We first launched the pre-survey adaptive mission with a reasonable threshold based on our belief field and then updated the threshold to be 25.4 g/kg after observing the updated salinity field from the pre-survey run.

5.2 Field operation

The AUVs started moving from their starting location around 12:50 a.m. We received the “Mission Complete” message from the AUVs around 14:15 p.m., which marks the end of the operation.

In Figure 11, the results of the AUV conducting adaptive sampling are displayed. Here, we plot the AUV path (black) and the updated posterior mean salinity field over time steps. The AUV

FIGURE 9
LAUV Roald is on expedition with an adaptive sampling algorithm onboard. The AUV is around 2 m long and runs at around 1 m/s. It is doing a 3D sampling mission at depths ranging from 0.5 m to 5.5 m.

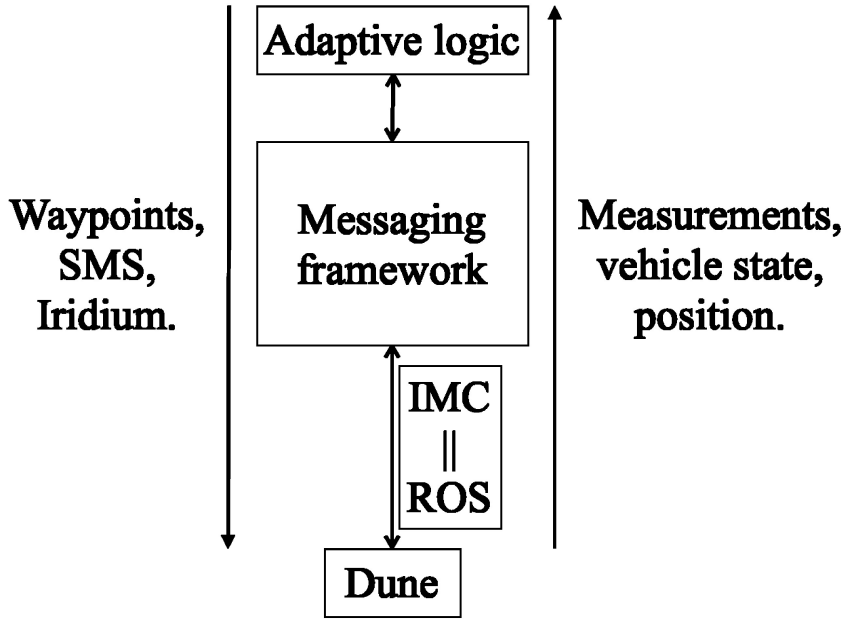


FIGURE 10
The diagram of the software component in the adaptive sampling system. The main CPU of the AUV is running DUNE (Pinto et al., 2013), whereas IMC (LSTS, 2022) messages are sent through TCP (Cerf and Kahn, 1974) to a secondary CPU, where the adaptive code and ROS (Quigley, 2009) are executed.

began near the river mouth and gradually moved toward the frontal region, occasionally diving to the deeper layers. In total, the AUV traveled approximately 9.734 km, with a coverage of 6.9% of the field at 0.5 m, 6.3% at 1.5 m, 1.6% at 2.5 m, and 0.1% at 3.5m. Both the AUVs were set to travel at 1.5 m/s, but the speed varies widely because of conditions in the ocean. During the mission, the plume expanded due to the tidal effect, so the AUV attempted to follow the front more closely. Interestingly, the AUV did not dive deeper than 2.5 m. This can be attributed to the fact that the water becomes more homogeneous and saline when it is too deep, and the river plume tends to stay close to the upper layers. Also, this can be an effect of the model learning from observation closer to the surface. It should be noted that the trajectory presented in Figure 11 exhibits a seemingly disjointed appearance. This phenomenon arises from the methodology employed, wherein observed measurements are allocated to the nearest center of the grid cell corresponding to their actual spatial location. Furthermore, this disjointedness is exacerbated by the movement of the AUV across various-depth layers, which introduces discontinuities in the trajectories within each layer.

In Figure 12, the salinity prediction results of the AUV's pre-planned sampling are displayed. The path was designed to maximize the sampling coverage and consequently reduce the variance of the field. The AUV was programmed to move along the path with a consistent YoYo pattern. This pattern involves the AUV moving between 0.5 m and 5.5 m repeatedly. The

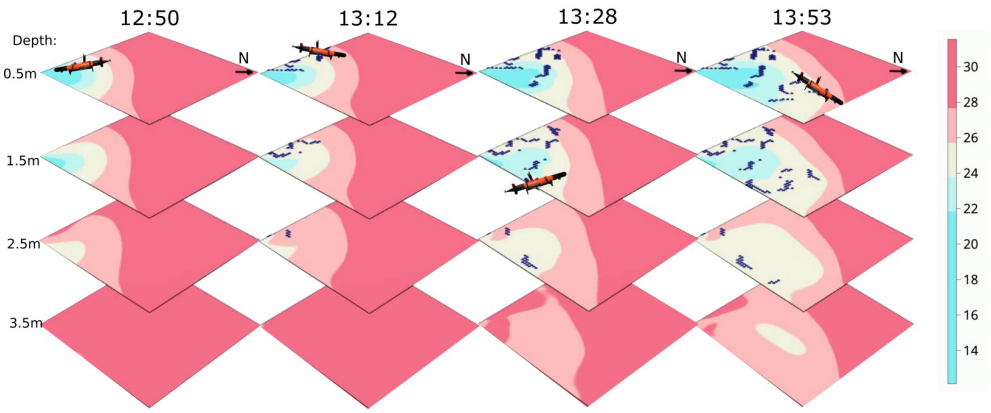


FIGURE 11
Salinity prediction during the adaptive sampling mission 08/09/2022. The AUV (black) began close to the river mouth and gradually moved toward the frontal region and dived to deeper layers occasionally.

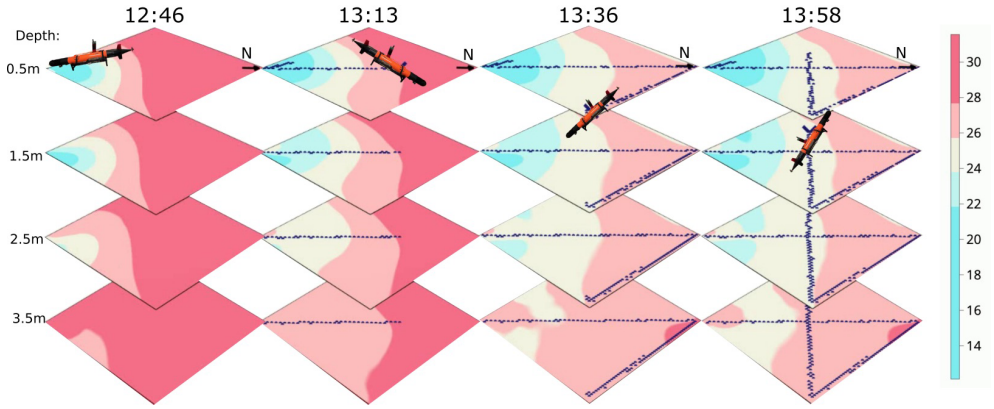


FIGURE 12
Salinity prediction during the fixed path mission on 08/09/2022. The AUV (black) aims to cover the spatial domain.

preprogrammed path approach ensures a more systematic and exhaustive coverage of the volume, providing a broader perspective but lacking the pinpoint accuracy on such a large and rapidly changing ocean volume. The path the AUV traveled along was approximately 9.346 km with a coverage of 7.9% at 0.5 m, 8.3% at 1.5 m, 9.1% at 2.5 m, 8.6% at 3.5 m, 7.9% at 4.5 m, and 7.2% at 5.5 m.

Given the unpredictable nature of the location of the freshwater front, it is virtually impossible to preplan precise sampling paths. The shifts and movements of the plume demand a real-time responsive approach like adaptive sampling. This is also evident when comparing Figures 11, 12. On the other hand, if a broad overview of the ocean volume is the goal, then a preplanned design likely is useful to ensure a systematic coverage of the region, leaving minimal gaps in the data collection. Also, note that diving 1 m is significantly more time efficient than moving 32 m in the horizontal plane for the AUV, this can be viewed by the coverage in each layer by two missions. The fixed path mission has good coverage within each layer whereas the adaptive mission mostly considers the top two layers. Because of this, it could be interesting in future work to consider adaptive sampling in only the horizontal plane and to always move in a YoYo pattern.

5.3 SINMOD and AUV data comparison

Even though salinity is only one state variable in SINMOD, it is useful and interesting to compare the AUV salinity measurements with the predictions made by SINMOD, as it will give information on the overall performance of the hydrodynamic model. Using all the data collecting by both AUVs, we compare the location-specific observations made by the AUVs with the associated SINMOD predictions for this day.

As illustrated in Figure 13, there is a clear inclination of SINMOD to overpredict salinity values. This trend is evident as a majority of the AUV measurements are situated below the zero error line (dotted line). In shallow water regions, both SINMOD and the AUV measurements exhibited high salinity variability, which is reasonable considering the freshwater influx from the river and local disturbances. For deeper waters, the salinity is higher for both models and more concentrated, with a bias around 3.5 g/kg between the SINMOD predictions and AUV measurements. The highest measured salinity value by the AUV was 28.0 g/kg, whereas the highest value from SINMOD was 31.5 g/kg, further confirming that the numerical model overestimates salinity both for the water in the river plume and in the brackish layer in the fjord.

While this discrepancy between SINMOD and the actual salinity field is evident, this will not impact the learning of the covariance structure, which captures the spatial correlation and variability within the data and is independent of any systematic bias. That said, the observed overestimation in SINMOD does set a prior expectation in our model that is skewed slightly high and will initially impact the adaptive sampling algorithm.

6 Conclusions

In this study, we have presented an approach for effective 3D (north, east, depth) sampling of salinity in a river plume front, employing a realistic and flexible spatial covariance model running onboard an AUV in real time. Results of a deployment in the Trondheim fjord demonstrate that prior inputs from the SINMOD numerical ocean model are effectively calibrated with the *in-situ* AUV measurements. In a mission focusing on mapping the frontal region, the AUV adapts naturally to the updated situation and zig-zags near the plume front to improve its spatial characteristics. Moreover, it is evident that the adaptive approach holds a distinct advantage over the preplanned method when it comes to accurately monitoring dynamic zones like the river plume front.

In Section 2.3, we estimate our surrogate model with the prior mean and covariance structure, which in the visualization both appear reasonable and are well-suited for the domain of interest. Furthermore, as detailed in Section 4, the surrogate model consistently outperformed a standard benchmark model across several key performance metrics. Lastly, in Section 5.2, we present the results of the field mission. Figure 11 depicts the AUV's

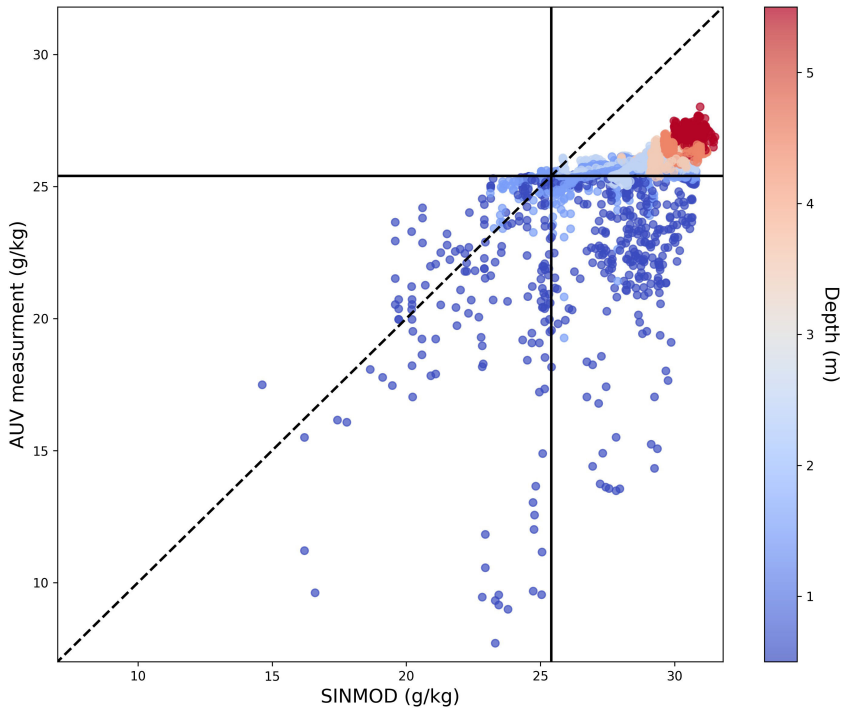


FIGURE 13
Comparison of salinity values estimated from the numerical ocean model (x-axis) and salinity measurements collected with the AUV on the 8th of September 2022 (y-axis) where the color shows the corresponding depths. The dotted line shows the zero error line.

predictions throughout the adaptive sampling mission, which we interpret as indicative of reasonable adaptive sampling behavior under the given conditions. The surrogate model, while promising in these results, does necessitate further refinement to fully realize its potential in this context. Firstly, refining its parameterization slightly to simplify the likelihood surface can potentially improve the optimization process significantly. Furthermore, estimating the covariance structure to innovations constructed from the SINMOD data, as described in Berild and Fuglstad (2023), is not guaranteed to be accurate in removing the temporal effect in the data, thus making it challenging to ascertain if the final structure is only capturing the spatial effect.

The prior models used in this work included 3D space with no temporal variation. A natural extension is to include temporal variation in the prior, which could be done in a Gaussian framework assuming known advection and diffusion (Foss et al., 2022). However, more research is required to develop realistic space–time models for frontal regions, such as that associated with river plumes, while maintaining the computational efficiency required to conduct expansive field surveying as considered in this study. Lastly, our exploration was confined to a near-sighted myopic sampling scheme. Future avenues might explore more sophisticated strategies (Bai et al., 2021), using longer sampling horizons where one can look ahead and anticipate the information gained by traversing longer distances with the AUV while also accounting for operational constraints.

In closing, it is important to highlight that while our study is centered on separating ocean masses of low (freshwater plume) and high (brackish water) salinity concentrations, we believe that this approach transfers well to other applications in physical or biological oceanography, such as polar melting water, high chlorophyll concentrations, oxygen or carbon content, or pollution detection.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.17605/OSF.IO/RHDKM.

Author contributions

MB: Methodology, Software, Writing – original draft, Data curation, Formal analysis. YG: Methodology, Software, Writing – original draft, Data curation, Formal analysis. JE: Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. G-AF: Conceptualization, Methodology, Supervision, Writing – review & editing. IE: Data curation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is

funded by the Norwegian Research Council through the MASCOT project 305445.

Acknowledgments

We acknowledge the aid and support of AURlab (https://www.ntnu.edu/aur-lab) in lending us the AUVs and help in running multiple field mission.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1319719/full#supplementary-material

References

Bai, S., Shan, T., Chen, F., Liu, L., and Englot, B. (2021). Information-driven path planning. *Curr. Robotics Rep.* 2, 177–188. doi: 10.1007/s43154-021-00045-6

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., and Voksø, A. (2003). Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrology Earth System Sci.* 7, 304–316. doi: 10.5194/hess-7-304-2003

Berget, G. E., Eidsvik, J., Alver, M. O., and Johansen, T. A. (2023). Dynamic stochastic modeling for adaptive sampling of environmental variables using an auv. *Autonomous Robots* 47, 483–502.

Berget, G. E., Fossum, T. O., Johansen, T. A., Eidsvik, J., and Rajan, K. (2018). Adaptive sampling of ocean processes using an auv with a gaussian proxy model. *IFAC-PapersOnLine* 51, 238–243.

Berild, M. O., and Fuglstad, G.-A. (2023). Spatially varying anisotropy for gaussian random fields in three-dimensional space. *Spatial Stat* 55, 100750.

Bhat, S., Torroba, I., Özkahraman, Ö., Bore, N., Sprague, C. I., Xie, Y., et al. (2020). "A cyber-physical system for hydrobatic auvs: system integration and field demonstration," in *2020 IEEE/OES Autonomous Underwater Vehicles Symposium (AUV)*. 1–8 (St. Johns, NL, Canada: IEEE).

Broch, O. J., Daae, R. L., Ellingsen, I. H., Nepstad, R., Bendiksen, E.Å., Reed, J. L., et al. (2017). Spatiotemporal dispersal and deposition of fish farm wastes: a model study from central Norway. *Front. Mar. Sci.* 4, 199.

Broch, O. J., Klebert, P., Michelsen, F. A., and Alver, M. O. (2020). Multiscale modelling of cage effects on the transport of effluents from open aquaculture systems. *PloS One* 15, e0228502.

Cerf, V., and Kahn, R. (1974). A protocol for packet network intercommunication. *IEEE Trans. Commun.* 22, 637–648. doi: 10.1109/TCOM.1974.1092259

Cressie, N. (1993). *Statistics for spatial data* (New York: John Wiley & Sons).

Das, J., Py, F., Harvey, J. B., Ryan, J. P., Gellene, A., Graham, R., et al. (2015). Data-driven robotic sampling for marine ecosystem monitoring. *Int. J. Robotics Res.* 34, 1435–1452.

Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., et al. (2012). Climate change impacts on marine ecosystems. *Annu. Rev. Mar. Sci.* 4, 11–37. doi: 10.1146/annurev-marine-041911-111611

Fonseca, J., Bhat, S., Lock, M., Stenius, I., and Johansson, K. H. (2023). *Adaptive sampling of algal blooms using autonomous underwater vehicle and satellite imagery: Experimental validation in the baltic sea, arXiv preprint arXiv:2305.00774.*

Foss, K. H., Berget, G. E., and Eidsvik, J. (2022). Using an autonomous underwater vehicle with onboard stochastic advection-diffusion models to map excursion sets of environmental variables. *Environmetrics* 33, e2702.

Fossum, T. O., Eidsvik, J., Ellingsen, I., Alver, M. O., Fragoso, G. M., Johnsen, G., et al. (2018). Information-driven robotic sampling in the coastal ocean. *J. Field Robotics* 35, 1101–1121. doi: 10.1002/rob.21805

Fossum, T. O., Fragoso, G. M., Davies, E. J., Ullgren, J. E., Mendes, R., Johnsen, G., et al. (2019). Toward adaptive robotic sampling of phytoplankton in the coastal ocean. *Sci. Robotics* 4, eaav3041. doi: 10.1126/scirobotics.aav3041

Fossum, T. O., Travelletti, C., Eidsvik, J., Ginsbourger, D., and Rajan, K. (2021). Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling. *Ann. Appl. Stat* 15, 597–618. doi: 10.1214/21-AOAS1451

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015). Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sin.* 25, 115–133.

Ge, Y., Eidsvik, J., and Mo-Bjørkelund, T. (2023). 3d adaptive auv sampling for classification of water masses. *IEEE J. Ocean Eng.* 48, 626–639.

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences* (New York: CRC press).

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., et al. (2008). A global map of human impact on marine ecosystems. *Science* 319, 948–952. doi: 10.1126/science.1149345

Hoegh-Guldberg, O., and Bruno, J. F. (2010). The impact of climate change on the world's marine ecosystems. *Science* 328, 1523–1528. doi: 10.1126/science.1189930

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Society: Ser. B (Statistical Methodology)* 73, 423–498.

Lin, M., and Yang, C. (2020). Ocean observation technologies: A review. *Chin. J. Mechanical Eng.* 33, 32. doi: 10.1186/s10033-020-00449-z

LSTS (2022) *Inter module communication protocal*. Available at: https://lsts.pt/docs/imc/master (Accessed Accessed: 2022-11-01). Dataset.

Mo-Bjørkelund, T., Fossum, T. O., Norgren, P., and Ludvigsen, M. (2020). "Hexagonal grid graph as a basis for adaptive sampling of ocean gradients using AUVs," in *Global Oceans 2020*. 1–5 (Singapore: U.S. Gulf Coast), ISBN: . doi: 10.1109/IEEECONF38699.2020.9389324

Nepstad, R., Liste, M., Alver, M. O., Nordam, T., Davies, E., and Glette, T. (2020). High-resolution numerical modelling of a marine mine tailings discharge in western Norway. *Regional Stud. Mar. Sci.* 39, 101404.

Nepstad, R., Nordam, T., Ellingsen, I. H., Eisenhauer, L., Litzler, E., and Kotzakoulakis, K. (2022). Impact of flow field resolution on produced water transport in lagrangian and eulerian models. *Mar. pollut. Bull.* 182, 113928.

Pinto, J., Dias, P. S., Martins, R., Fortuna, J., Marques, E., and Sousa, J. (2013). "The lsts toolchain for networked vehicle systems," in *2013 MTS/IEEE OCEANS - Bergen*. 1–9. doi: 10.1109/OCEANS-Bergen.2013.6608148

Pinto, J., Mendes, R., da Silva, J. C. B., Dias, J. M., and de Sousa, J. B. (2018). "Multiple autonomous vehicles applied to plume detection and tracking," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*. 1–6. doi: 10.1109/OCEANSKOBE.2018.8558802

Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., and Leibs, J. (2009). ROS: an open-source Robot Operating System. *ICRA workshop on open source software*. 3, 5. Kobe, Japan.

Slagstad, D., and McClimans, T. A. (2005). Modeling the ecosystem dynamics of the Barents sea including the marginal ice zone: I. Physical and chemical oceanography. *J. Mar. Syst.* 58, 1–18. doi: 10.1016/j.jmarsys.2005.05.005

Slagstad, D., Wassmann, P. F., and Ellingsen, I. (2015). Physical constrains and productivity in the future arctic ocean. *Front. Mar. Sci.* 2, 85.

Vernet, M., Ellingsen, I., Marchese, C., Bélanger, S., Cape, M., Slagstad, D., et al. (2021). Spatial variability in rates of net primary production (npp) and onset of the spring bloom in Greenland shelf waters. *Prog. Oceanography* 198, 102655.

Zhang, Y., Bellingham, J. G., Ryan, J. P., Kieft, B., and Stanway, M. J. (2013). "Two-dimensional mapping and tracking of a coastal upwelling front by an autonomous underwater vehicle," in *2013 OCEANS - San Diego*. 1–4, ISBN: . doi: 10.23919/OCEANS.2013.6741231

Zhang, Y., Ryan, J. P., Bellingham, J. G., Harvey, J. B. J., and McEwen, R. S. (2012). Autonomous detection and sampling of water types and fronts in a coastal upwelling system by an autonomous underwater vehicle. *Limnology Oceanog: Methods* 10, 934–951. doi: 10.4319/lom.2012.10.934

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 08 February 2024
DOI 10.3389/fmars.2024.1320361

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Yong Ma,
Wuhan University of Technology, China
Nitin Agarwala,
National Maritime Foundation, India
Zhang Guichen,
Shanghai Maritime University, China
Guoqing Zhang,
Dalian Maritime University, China

*CORRESPONDENCE
Yang Liu
godkissyou@126.com

RECEIVED 12 October 2023
ACCEPTED 16 January 2024
PUBLISHED 08 February 2024

CITATION
Zhang Q, Zhang S, Liu Y, Zhang Y and Hu Y (2024) Adaptive terminal sliding mode control for USV-ROVs formation under deceptive attacks. *Front. Mar. Sci.* 11:1320361. doi: 10.3389/fmars.2024.1320361

Adaptive terminal sliding mode control for USV-ROVs formation under deceptive attacks

Qiang Zhang, Sihang Zhang, Yang Liu*, Yan Zhang and Yancai Hu

School of Navigation and Shipping, Shandong Jiaotong University, Weihai, Shandong, China

This work investigates the cooperative formation control problem of unmanned surface vehicle-remotely operated vehicles (USV-ROVs) subject to uncertainties under deceptive attacks. In the control design, with the utilization of the desired formation as well as the geometric position between USV and ROVs, a geometric transformation approach is developed and a geometric constraint relationship of governing formation positions is derived. Under the terminal sliding mode control (TSMC) design framework, a novel terminal sliding surface is crafted to circumvent the singularity issue. To further bolster robustness, using the sliding mode damper concept, a variable damping reaching law is devised. To refrain from the effectiveness of attacks and uncertainties, the adaptive technique is integrated into the TSMC framework. To save the communication resources, an event-triggering mechanism is established between the distributed controller and ROVs. Then, an event-triggered adaptive finite-time cooperative formation control scheme is developed for the USV-ROVs. The Lyapunov theory analysis shows that the cooperative formation control issue of USV-ROVs is realized and the deceptive attack can be suppressed efficaciously. The simulation, comparison, and quantitative analysis demonstrate the relative effectiveness and superiority of the developed scheme.

KEYWORDS

cooperative formation, USV, ROVs, deceptive attack, sliding mode control

1 Introduction

In recent years, the field of marine engineering has witnessed a significant shift towards the exploration of collaborative multi-intelligent body formation control. This research direction involves the coordination and cooperation of multiple intelligences, which has the advantages of improving survivability, reducing operation cost, and expanding operation range compared with the traditional operation of a single intelligence (Xu et al., 2022). Intelligent aircraft and underwater operation equipment such as unmanned surface vehicles (USV), remotely operated vehicles (ROV), and autonomous underwater vehicles (AUV) play an important role in related engineering scenarios (Li et al., 2023b). Therefore,

control issues become critical in order for multi-intelligence bodies to achieve safe cooperative formation navigation and to perform smoothly according to pre-designed missions. In USVs in unmanned control mode, the network becomes an important medium for executing tasks, and in open network links between sensors, controllers, and actuators, the exchanged data are vulnerable to attacks because of protocol vulnerabilities, misconfigurations, and other situations. In recent years, a number of maritime security incidents related to network attacks have been triggered around the world, raising serious security concerns (Swaszek et al., 2013). Hence, addressing the issue of network attacks is pivotal for achieving secure formation navigation of USV-ROVs and the successful execution of assigned missions.

Network attack methods are generally categorized into Denial of Service (DoS) attacks, replay attacks, and deceptive injection attacks (Chen et al., 2019). DoS attacks are crafted to overwhelm a system's resources, rendering it incapable of processing legitimate service requests. Replay attacks and deceptive injection attacks can be collectively categorized as deceptive attacks, centered around manipulating and corrupting transmitted data within the network. Replay attacks involve maliciously resending the same data repeatedly, constituting a specialized form of deception attack. In deceptive injection attacks, attackers inject spoofed commands into sensors that receive ship data, subsequently integrated into the control system. In contrast, deceptive injection attacks are more harmful to USV-ROVs formation control because of their stealthy nature (Ding et al., 2016). Maritime security mechanisms for network control systems (NCS) have primarily focused on IT-level designs (Li et al., 2022), such as setting up firewalls, to enhance security control and detect potential threats at the computer system level. However, these measures alone are insufficient for regulating the control of formation devices when they come under deceptive attacks (Sandberg et al., 2015). To address this issue, security control strategies need to be developed at the control level to ensure the safety and integrity of marine operations, in conjunction with security software. Researchers have explored various control techniques, such as sliding mode control (Shi et al., 2018a), observer-based control (Jin, 2018), and adaptive control (Zhang and Zhang, 2015), due to their robustness in dealing with system uncertainties and external disturbances. These techniques constitute a resilient compensation mechanism (Wu et al., 2019) to mitigate the impact of attacks on the system. Sliding mode control has attracted many scholars to study it because it relies less on the accuracy of the system model and has less influence of system parameter variations and external perturbations compared to other control techniques. W Wang (Wang et al., 2021) investigated the problem of formation control for uncertain systems under deceptive attacks, incorporating adaptive methods to improve control efficiency, performance, and system resilience. T Yin (Yin et al., 2022) investigated the problem of event-based intermittent formation control under deceptive attacks by introducing an averaging method in the event triggering mechanism to reduce the burden on the network bandwidth. DW Zhang (Zhang and Liu, 2023) proposed a predictive sliding mode control method to improve the closed-loop system robustness of the formation by introducing a sliding variable to defend against random deceptive attacks. In the field of cyber security control, the issue of chattering associated with sliding mode control is often overlooked (Saihi et al., 2019). Additionally, some researchers have employed the theory of uncertain system state estimation to model attack damage control commands as virtual unknown states. They transformed these commands into unknown state estimation problems, established adaptive resilience policies using continuous incentive mechanisms, and designed event-triggered filters to counter deceptive attacks (Liu et al., 2020). Other studies have proposed novel observers (Huang and Dong, 2019) to address attack problems from a modeling perspective by algebraically modifying the attack abstraction model of the device. However, the effectiveness of resilient compensation and the associated constraints still require further exploration and validation due to the sudden and unpredictable nature of attacks.

In addition to the aforementioned challenges, formation control differs from single-vessel trajectory tracking control as it requires maintaining a specific formation. Common formation control methods include the leading-following method, virtual structure method, behavior-based method, and artificial potential field method; among these, the leading-following method has garnered extensive attention due to its simplicity and practical applicability. Essentially, this method involves designating an individual within the formation as a navigator and the remaining members as followers; a relationship is then established between the navigator and the followers to achieve effective formation control. H Xu (Xu et al., 2023) et al. proposed a leading-follower formation control method combined with a fixed-time perturbation observer, which solved the problem of USVs formation control in the presence of unknown perturbations. L Ding (Ding and Guo, 2012) et al. aimed at the synergy problem between USVs, combined the leading-follower method and backstepping method together to design a new formation control law. Based on the virtual leader strategy, M Fu (Fu et al., 2018) et al. designed a new control law by simplifying the backstepping control technique through coordinate transformation, and proved the stability of the whole formation system. Based on graph topology theory, Li Y (Yun and Ying, 2016) et al. introduced a hybrid leading-follower method to design control strategies to control the formation sequentially according to the motion patterns in different phases, and Lin et al. used the pilot-follower method to divide the ships in the formation into virtual leader and follower, and introduced an observer to eliminate the virtual leader and follower. Lin A (Lin et al., 2018) et al. used the leading-follower method to categorize the ships in the formation into a virtual leader and followers, introduced observers to eliminate unknown disturbances, and reduced the complexity of designing formation controllers by transforming the formation control into the design of controllers for three subsystems. The above studies conducted intelligent control research on different tasks from the perspective of equipment coordination, but most of them focused on power localization and did not bring out the unmanned advantages that USVs themselves possess, and the motion control of USV-ROVs in formation coordination was not further studied.

During navigation, USV-ROVs are affected by a variety of uncertainties in addition to network attacks, such as external

perturbations caused by wind, waves, and currents, and internal uncertainties such as ship dynamic uncertainties and model errors (Li et al., 2023a). USV-ROV formation systems are characterized by offshore operations, and the time-varying disturbances caused by sea winds and waves to the USV-ROV are the primary control challenges from the external environment (Skjetne et al., 2005). These components can affect the control performance of the whole formation control system and reduce the robustness of the system. In order to deal with these problems, some effective uncertainty estimation and reconstruction methods have been presented, such as sliding mode, observer, and parameter adaptive techniques, among which terminal sliding mode control has been widely used by many scholars due to its strong anti-interference ability. Yang G (Yang and Chen, 2020) et al. proposed a power convergence law, which can adaptively select parameters to improve the convergence speed of the convergence law. Kang Z (Kang et al., 2020) et al. proposed a double power convergence law, which adaptively combined the terminal sliding mode with the double power convergence law, which can better accomplish the tracking task while weakening the jitter phenomenon. Tian Y (Tian et al., 2020) et al. designed a fast convergence fixed-time non-singular terminal sliding mode method for a class of second-order non-linear systems with matching uncertainty, which avoids the singularity phenomenon, improves the fast response capability of the system, and suppresses the jitter phenomenon of the sliding mode control. Sun R (Sun et al., 2018) et al. combined a neural network based neural network-based formation switching strategy and terminal sliding mode control technology to design the controller, which effectively improves the control accuracy and reduces the real-time computational load of the controller. Many scholars have achieved good results in the field of network attack formation and terminal sliding mode control, but since USV-ROVs are a kind of multi-input non-linear system, the design is usually more complicated, and the convergence law has a crucial role in the rapidity and stability of the controller. Based on the aforementioned approach, the primary objective of this paper is to design a novel terminal sliding mode control to address the formation problem under deceptive attacks. The main contributions of this paper are as follows.

1. In this paper, mathematical abstraction modeling of stealthy attacks and perturbations caused by internal and external factors including actuator matching perturbations is carried out in part 2. Aiming at the stealthy nature of a deceptive attack, a sliding mode control scheme is adopted from the perspective of control system robustness. In part 3.1, considering the singularity and trajectory tracking time requirements, a new terminal sliding mode surface design is carried out to complete the convergence phase and sliding phase in a limited time to improve the control efficiency.
2. In part 3.2, based on the dynamic characteristics of the sliding mode convergence phase, a dynamic dampener is introduced, and a novel variable damping sliding mode convergence law is developed. In this approach, the convergence speed at the junction of convergence and arrival phases have been reduced, thereby minimizing sliding mode chattering. Moreover, it increases dynamic speed when moving away from the sliding mode surface to improve response efficiency, leading to better response and compensation effects in sliding mode control.
3. To tackle the challenges related to saturation filtering, a non-linear saturation fitting function has been devised in part 3.3, which is coupled with an event-triggering mechanism. This innovative approach serves to decrease the utilization of communication resources and effectively mitigates communication congestion triggered by deceptive attacks.

Notations: $\mathrm{R}^{m\times n}$ is $m \times n$-dimensional Euclidean space, $\sin(\cdot)$ is sinusoidal function, $\cos(\cdot)$ is cosine function, $\|\cdot\|$ is Euclidean parameter, $\mathrm{sign}(\cdot)$ is symbolic function, $\tanh(\cdot)$ is hyperbolic tangent function, $\mathrm{sat}(\cdot)$ is saturation function, $\mathrm{diag}(\cdot)$ is diagonal matrix, and $\min\{\cdot\}$ and $\max\{\cdot\}$ are minimum and maximum values, respectively.

2 Problem description

2.1 Mathematical model

The investigation of USV-ROVs formation control in this paper is based on the three degrees of freedom planar motion synergy problem (Min et al., 2020; Fan et al., 2021; Chen et al., 2023), wherein the kinematics of the two devices are described within the generalized coordinate system.

The kinematics and dynamics equations of USV (Zhu and Du, 2018) are described as (Equations 1) and (2).

$$\dot{q}_{usv} = J_{usv}(\theta_{usv})u_{usv} \tag{1}$$

$$\dot{u}_{usv} = \tilde{B}_{\mathrm{usv}}\tau_{usv} - \widehat{F}_{rov}(u_{usv}) + \bar{\tau}_{usvd} \tag{2}$$

where $J_{usv}(\theta_{usv}) = [\cos(\theta_{usv}), -\sin(\theta_{usv}), 0; \sin(\theta_{usv}), \cos(\theta_{usv}), 0; 0, 0, 1]$ is the transformation matrices for dynamics and kinematics in the coordinate system; $\dot{q}_{usv} = [\dot{x}_{usv}, \dot{y}_{usv}, \dot{\theta}_{usv}]^{\mathrm{T}}$ is the vector of first-order derivatives of the position coordinate vector. $\widehat{F}_{usv}(u_{usv}) = M_{usv}^{-1}C_{usv}(u_{usv})u_{usv} + M_{usv}^{-1}D_{usv}u_{usv}$ is uncertain. $\tau_{usv} = [\tau_{usv_1}, \tau_{usv_2}, \tau_{usv_3}]^{\mathrm{T}}$ is the ship system input. $\bar{\tau}_{usvd} = M_{usv}^{-1}\bar{\tau}_{dusvf}$ is the external perturbation of the converted system, with low-frequency accretion characteristics. $\tilde{B}_{\mathrm{usv}} = M_{usv}^{-1}$ is a matrix of system control input conversion coefficients. M_{usv} is the ship type relationship and hydrodynamic additional relationship matrix, $C_{usv}(u_{usv})$ is the Coriolis centripetal matrix, and D_{usv} is the hydrodynamic damping coefficient matrix.

A tracked differential ROV with incomplete constraint characteristics is selected for underwater navigation according to the requirements of ocean engineering. The kinematics and dynamics equations of ROV (Wu and Wang, 2020) are described as (Equations 3) and (4).

$$\dot{q}_{rov} = J_{rov}(\theta_{rov})u_{rov} \tag{3}$$

$$\dot{u}_{rov} = \tilde{B}_{rov}\tau_{rov} - \hat{F}_{rov}(\dot{q}_{rov}) - \bar{\tau}_{rovd} \tag{4}$$

where $J_{rov}(\theta_{rov}) = [\cos(\theta_{rov}),\ d_{rov}\sin(\theta_{rov});\ \sin(\theta_{rov}),\ -d_{rov}\cos(\theta_{rov});\ 0,\ 1]$ is the ROVs transformation matrix; $\dot{q}_{rov} = [\dot{x}_{rov}, \dot{y}_{rov}, \dot{\theta}_{rov}]^{\mathrm{T}}$ is the first derivative vector of the position coordinate vector. The robot system input $\tau_{rov} = [\tau_{rov1},\ \tau_{rov2}]^{\mathrm{T}}$ is the left and right wheel drive torque vector. $\bar{\tau}_{rovd} = \bar{M}_{rov}^{-1}S_{rov}^{\mathrm{T}}(q_{rov})\bar{\tau}_{rovdf}$ is the external perturbation of the converted system, with low-frequency accretion characteristics. $\hat{F}_{rov}(\dot{q}_{rov}) = \bar{M}_{rov}^{-1}S_{rov}^{\mathrm{T}}(q(t))\bar{F}_{rov}(\dot{q}(t))$ is uncertain. $\bar{M}^{-1} = (mr)^{-1}[J,\ 0;\ 0,\ m]^{T}$. $\bar{F}_{rov}(\dot{q}_{rov})$ is the friction matrix; $A_{rov}^{\mathrm{T}}(q_{rov}) = [\sin(\theta_{rov}),\ -\cos(\theta_{rov}),\ -d_{rov}]^{\mathrm{T}}$ is a non-complete constraint matrix; λ_{rov} is the Lagrange multiplier. $\tilde{B}_{rov} = \left[\frac{1}{m_{rov}r_{rov}},\ \frac{1}{m_{rov}r_{rov}};\ \frac{b_{rov}}{J_{rov}r_{rov}},\ \frac{-b_{rov}}{J_{rov}r_{rov}}\right]^{\mathrm{T}}$ is the matrix of system input conversion coefficients, J_{rov} is the moment of inertia; m_{rov} is the quality of ROV; b_{rov} is the ROV width; and r_{rov} is the driving wheel radius.

2.2 Attack models

The USV-ROV formation system, employing a centralized control scheme and formation strategy for command transmission, faces security risks in its semi-open wireless network environment. Deceptive attacks are a common threat, involving IP hijacking and the injection of false data, bypassing firewalls and compromising the control of USV-ROV formations. Additionally, these attacks can result in offline damage to wireline switching equipment, impacting overall formation effectiveness. Existing network control systems often neglect post-attack damage and lack adequate countermeasures. Therefore, it is crucial to design security control strategies specifically at the control level to comprehensively address these threats.

Remark 1: Element $i = \bar{l} \in \{a_0,\ a_1,\ a_2, \cdots a_{\bar{L}}\}$ in the USV identification, $i = a \in \{a_1,\ a_2, \cdots a_N\}$ represents the leader USV and the virtual leader sequence, and $\hat{L} = \bar{L} + 1$. The USV smart device with device identification sequence $i \in \{a_{\bar{L}+1}, \cdots a_N\}$ is a follower of the ROV device of $i = b \in \{b_1,\ b_2, \cdots b_M\}$, $G = N + M - \hat{L}$. Then $g = \{a_{\bar{L}+1}, \cdots a_N,\ b_1,\ b_2, \cdots b_M\}$ is the set of follower identification.

A stealth deceptive attack uses a hidden identity to access the IP and injects bounded interval false data that mimics normal data to the control command data; the deceptive data injection logic is shown in Figure 1.

From a control instruction perspective, this attack can be abstracted as an unknown additive signal (Ito and Takanami, 1997). Consequently, the actuator input signal modeled by the stealth deceptive attack is (Equation 5).

$$\bar{\tau}_i^a = \tau_i^* + \zeta_i^\beta \tag{5}$$

where $\bar{\tau}_i^a(k)$ is a control command sent by the controller to be transmitted to the system inputs in the physical perception layer after a stealthy deceptive attack at the receiver side of the network layer. ζ_i^β is bounded unknown data. τ_i^* is the NCS output.

The success rate of the attack can be represented using the Bernoulli series, and the final form of the false data is (Equation 6).

$$\zeta_i^\beta = \sigma_i^\beta \mu_i^\beta \tag{6}$$

where σ_i^β is the attack rate concomitant matrix with Bernoulli distribution properties and μ_i^β is the false data value. An attacker injects bounded random data in order to reduce the probability of being captured by a security program; $\bar{\xi}_i^\beta \geq \left\|\xi_i^\beta\right\| \geq 0$, $\left\|\sigma_i^\beta\right\| = 0$ is the tampering probability when the attack is unsuccessful or not performed during the three interactions.

2.3 Formation model

During the actual operational process, each device exhibits a saturation characteristic in its execution performance. To represent this physical attribute, a saturation function is introduced. The control input is (Equation 7).

$$\tau_{si} = sat\ (\bar{\tau}_i^a) = \begin{cases} \bar{\tau}_i^a & \|\bar{\tau}_i^a\| \leq M_\tau \\ M_\tau & \bar{\tau}_i^a > M_\tau \end{cases} \tag{7}$$

where $M_\tau = \tau_M$ is the constant absolute value of the physical saturation of the actuator and $sat\ (\cdot)$ is the saturation function.

The global motion model of USV and ROV considering attack, matching, external disturbance, and uncertainty is (Equation 8).

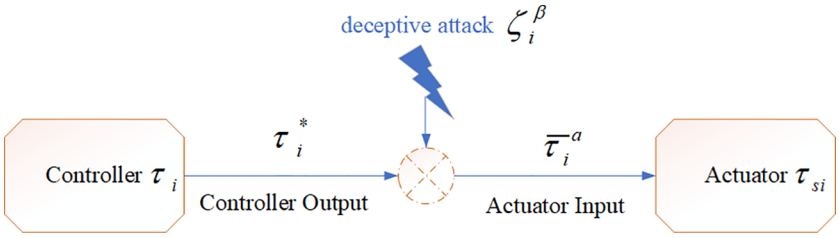


FIGURE 1
Stealth deceptive attack data injection.

$$\begin{cases} \dot{q}_i = S_i(q_i)u_i \\ \dot{u}_i = \tilde{B}_i sat\left(\tau_i^{\star} + \zeta_i^{\beta}\right) - \widehat{F}_i(\dot{q}) - \tau_{id} \end{cases} \tag{8}$$

where i is the model physical item of equipment, $i = a$ is the USV equipment identification sequence, and $i = b$ is the ROV equipment identification sequence.

$\Phi_{lg}^{D} = [L_{lg}^{D};\ \psi_{lg}^{D};\ \psi_{gl}^{D}]$ is the matrix of geometric features of the desired formation of leaders and followers in the global generalized coordinate system, $\Phi_{lg} = [L_{lg};\ \psi_{lg};\ \psi_{gl}]$ is the real formation feature matrix, $i \in g$ is the follower device, and $i \in l$ is the leader device. Taking the geometric position of a single ship and a single robot as an example, the formation relationship of the formation system is shown in Figure 2.

Assumption 1: The motion state of the formation system represents the actual configuration of the formation once each local equipment achieves a specific geometric arrangement within the formation.

The essence of synergistic control of formation geometry and formation motion is that the control system allows the current formation geometry feature Φ_{lg} described as

$$\lim_{t\to\infty}(\Phi_{lg}^{D} - \Phi_{lg}) = 0 \tag{9}$$

(Equation 9) illustrates if the collaborative goal can be achieved is dependent on the transformation form of the dynamic characteristics of each device. The relationship between the dynamic positions of each device and the dynamic transformation characteristics is (Equation 10).

$$\dot{q}_l = \dot{q}_g + G_{\dot{q}lg} = \begin{bmatrix} \dot{x}_l \\ \dot{y}_l \\ w_l \end{bmatrix} = \begin{bmatrix} \dot{x}_g \\ \dot{y}_g \\ w_g \end{bmatrix} + \begin{bmatrix} G_{\dot{x}lg} \\ G_{\dot{y}lg} \\ G_{\dot{\theta}lg} \end{bmatrix} \tag{10}$$

where $G_{\dot{x}lg} = d_g \sin(\theta_g)w_g - \dot{L}_{lg}\cos(\theta_l + \psi_{lg}) + L_{lg}(w_l + \dot{\psi}_{lg})\sin(\theta_l + \psi_{lg})$ is the motion conversion item of the X-axis; $G_{\dot{y}lg} = d_g\cos(\theta_g)w_g - \dot{L}_{lg}\sin(\theta_l + \psi_{lg}) - L_{lg}(w_l + \dot{\psi}_{lg})\cos(\theta_l + \psi_{lg})$ is the dynamic component conversion item in the direction of the Y-axis; and $G_{\dot{\theta}lg} = \dot{\psi}_{lg} + \dot{\psi}_{gl} - d_g\cos(\psi_{gl})w_g$ is the motion conversion item of the speed.

After collation, it becomes evident that the formation motion difference model that requires control within the formation scheme is (Equation 11). Subsequently, this formation relationship matrix will be utilized for formation control system design with a view to achieving the target expectation $\lim_{t\to\infty}\eta_i^E = 0$ to accomplish the formation control task.

$$\eta_i^E = \eta_i^D - \eta_i^s = \begin{bmatrix} \eta_l^d \\ \eta_g^d \\ \bar{\eta}_l^{\beta} \\ \bar{\eta}_g^{\beta} \end{bmatrix} - \begin{bmatrix} \eta_l \\ \eta_g \\ \bar{\eta}_l \\ \bar{\eta}_g \end{bmatrix} \tag{11}$$

where η_l^d is the expected position vector of the leader and η_l is the position vector of the leader. According to assumption 1, they are all virtual leaders, the formation differences are split according to kinematics and dynamics to give dynamic errors $\dot{\eta}_{ei} = \begin{bmatrix} \dot{\eta}_l^d - \dot{\eta}_l \\ \dot{\eta}_g^d - \dot{\eta}_g \end{bmatrix}$

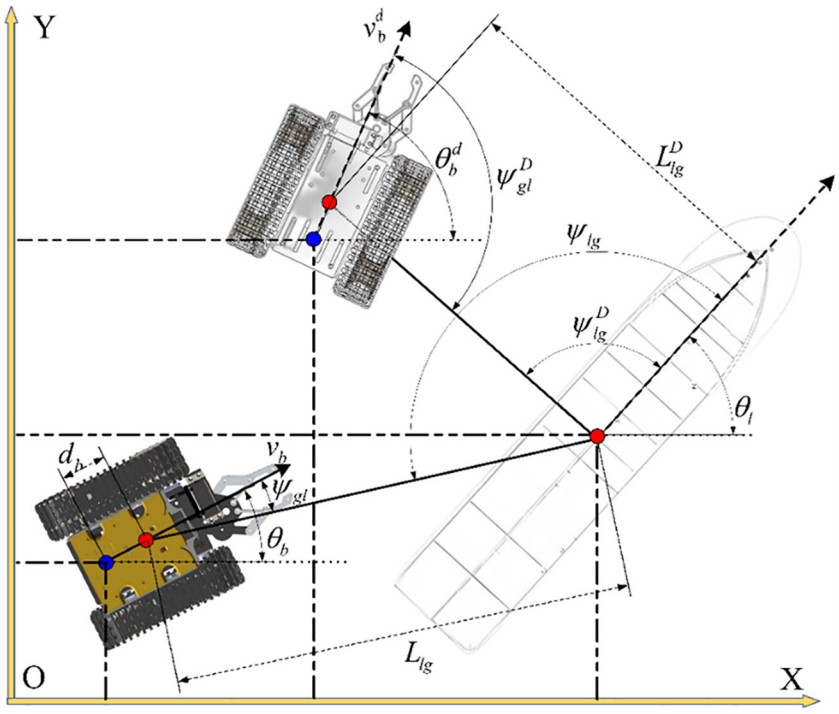


FIGURE 2
Geometric relationship of USV-ROV formation system.

10.3389/fmars.2024.1320361

and $\dot{\eta}_{ei} = \begin{bmatrix} \dot{\eta}_l^{\beta} - \dot{\eta}_l \\ \dot{\eta}_g^{\beta} - \dot{\eta}_g \end{bmatrix}$.

To facilitate the heterogeneous formation control, the error is divided by the equipment labeling, the first-order derivation of (Equation 11) is carried out and brought into the model (10) deformation can be shown that the dynamic formation cooperative model is (Equation 12).

$$\begin{cases} \dot{E}_{qi} = (S_i^d(q_i^d)u_i^d - G_{qlg}^d) - S_i(q_i)u_i \\ \dot{E}_{ui} = \dot{u}_i^{\beta} - \tilde{B}_i\tau_{si} - \widehat{F}_i(\dot{q}) - \overline{\tau}_{id} \end{cases} \tag{12}$$

where u_l^{β} is the virtual movement expectation of the leader, u_g^{β} is the virtual movement expectation of the follower, and $i = l$ and $i = g$ are the speed states of the leader and the followers, respectively.

If every device, under the influence of two synergy programs, attains the expected state simultaneously, the formation's motion state can be maintained. The geometric synergy program successfully guides the controlled system to achieve the desired geometric characteristics, signifying the completion of the formation restoration process. Moreover, if the speed of each device, governed by the motion synergy program, tends to converge, it indicates that the formation system has achieved the anticipated motion state.

3 Controller design and stability analysis

3.1 Design and analysis of non-singular terminal sliding mode surface

To address the singularity issue of the terminal sliding mode surface, following the traditional terminal sliding mode surface design concept (Min et al., 2020), the non-singular terminal sliding mode surface can be expressed as

$$S_{ui} = a_{i1}E_{ui} + a_{i2}\int \|E_{ui}\|^{\frac{1}{4}} sign[E_{ui}]d\gamma_t + a_{i3}\|E_{ui}\|^{\frac{9}{4}} sign[E_{ui}] \tag{13}$$

where a_{i1}, a_{i2}, and a_{i3} are normal constant diagonal matrices.

Deriving (Equation 13), the result shown in (Equation 14) can be obtained. It is evident that there are no singular points in $\dot{S}_u$, signifying that the non-singular sliding mode surface will not encounter issues related to singularities.

$$\dot{S}_{ui} = (a_{i1} + \frac{9a_{i3}}{4}\|E_{ui}\|^{\frac{5}{4}} sign[E_{ui}])\dot{E}_{ui} + a_{i2}\|E_{ui}\|^{\frac{1}{4}} sign[E_{ui}] \tag{14}$$

Analysis of (Equation 13) shows that the design meets the sliding mode dynamic standard Equation 15) (Shi et al., 2018b).

$$\begin{cases} S_{ui} > 0, \quad E_{ui} > 0 \\ S_{ui} = 0, \quad E_{ui} = 0 \\ S_{ui} < 0, \quad E_{ui} < 0 \end{cases} \tag{15}$$

To verify that the terminal sliding mode (13) can converge in a finite time, assuming that when $\dot{S}_{ui} = \frac{dS_{ui}}{dt} = 0$, transform (14), and the definite integral is calculated on both sides in the convergence time $T_{su\star}$ interval, and according to $-m_{ue}(T_{su\star} - 0) \geq -\int_0^{T_{su\star}} \frac{9a_{i3}}{4}\|E_{ui}\|^{\frac{5}{4}}dt \geq -M_{ue}(T_{su\star} - 0)$, it can be described as

$$-T_{su\star} \geq a_{i2}^{-1}\int_{E_{ui}(0)}^{E_{ui}(T_{su\star})} a_{i1}\|E_{ui}\|^{\frac{1}{4}}d(\|E_{ui}\|) \geq -a_{i2}^{-1}a_{i1}\|E_{ui}(0)\|^{\frac{3}{4}} \tag{16}$$

where M_{ue} and m_{ue} are the maximum and minimum values of $\frac{9a_{i3}}{4}\|E_{ui}\|^{\frac{5}{4}}$, respectively.

It can be seen that the terminal sliding mode surface (13) can converge in finite time $T_{su\star} \leq a_{i2}^{-1}a_{i1}\|E_{ui}(0)\|^{\frac{3}{4}}$ when $E_{ui}(0) \neq 0$. It is evident that the terminal sliding mode surface can achieve convergence during the sliding phase within a finite time.

3.2 Dynamically adjusted approach rate design

In pursuit of achieving global finite-time convergence of the sliding mode while optimizing chattering and enhancing robust response capability, a dynamic damper is designed based on the dynamic information of the sliding mode can be described in (Equation 17). Additionally, based on the traditional reaching law, a dynamic adjustment reaching law can be described in (Equation 18).

$$D_i(S_{ui}) = (2 + S_{ui}^2)\ln(2 + S_{ui}^2) - 2\ln(2) \tag{17}$$

$$\dot{S}_{Di} = -\varepsilon_{ui}D(S_{ui})sign[S_{ui}] \tag{18}$$

where $D_i(S_{ui}) = diag\{D_{i1}(S_{ui}) \cdots D_{ik_i}(S_{ui})\}$, ε_{ui} is a diagonal matrix with positive definite coefficients. When $E_{ui} \to 0$, dynamically adjust the approach law $\dot{S}_{Di} \to 0$, which can prove that it will not cause over-switching of the control approach and weaken the chattering. According to (Equation 18), it can be seen that the closer to the sliding mode surface, the smaller the reaching law, and the farther away from the sliding mode surface, the larger the reaching law value.

Lemma 1 (Chen et al., 2023). If a continuous non-linear system, such as $\dot{x}_\epsilon = F_\epsilon(x_\epsilon)$, is controllable, then for $x_\varepsilon \not\equiv 0$, the Lyapunov function $V_\epsilon \in C^2$ exists and is positive definite. If $\dot{V}_\varepsilon(x_\varepsilon) + k_{\chi 1}V_\varepsilon(x_\varepsilon) + k_{\chi 2}V_\varepsilon^o(x_\varepsilon) - \Delta_M \leq 0$, $\Delta_M > 0$, $k_{\chi 1} > 0$, $k_{\chi 1} > 0$, and $0 < o < 1$, then it can be stabilized in finite time, and $T_\epsilon \leq k_{\chi 2}^{-1}(k_{\chi 1} - k_{\chi 1}o)^{-1}(k_{\chi 2}V_\varepsilon^{1-\kappa}(x_\varepsilon(t_0)) + k_{\chi 1})$ is the finite time horizon.

Establishing Lyapunov positive definite function $V_{SD\star} = 0.5(S_{ui}^{\mathrm{T}}S_{ui})^2$ and deriving it, the equation can be gotten:

$$\begin{aligned} \dot{V}_{SD\star} &= -S_{ui}^T S_{ui} S_{ui}^T(\varepsilon_{ui}D_{zi}(S_{ui})sign[S_{ui}] + \kappa_{ui}S_{ui}) \\ &\leq -S_{ui}^T S_{ui}[\varepsilon_u \|S_{ui}^T\|(1 + c_{b\star}\|S_{ui}\| - 0.5c_{b\star}^2) - 2\ln(2)\varepsilon_u S_{ui}^T] \\ &\leq -\varepsilon_{uD}V_{SD\star} - \varepsilon_{u\star}\dot{V}_{SD\star}^{\frac{3}{4}} \end{aligned} \tag{19}$$

where $c_{b\star} \geq 0.25$, $\varepsilon_{u\star} = \min\{c_{b\star} - 2\ln(2)\varepsilon_u\}$, $\varepsilon_{uD} = \min\{\varepsilon_u - 0.5\varepsilon_u c_{b\star}^2\}$, $\varepsilon_u - 0.5\varepsilon_u c_{b\star}^2 > 0$. (Equation 19) satisfies Lemma 1; it can be seen that the finite time approach can be realized under the action of this reaching law.

3.3 Event trigger mechanism design

Taking into consideration the control physical saturation characteristics of each device, the event trigger condition can be described as

$$\begin{cases} \tau_i^* = \tau_{ik}^*, \forall t \in [t_k, t_{k+1}) \\ \tau_{ik}^* = \inf\left\{ \tau_i^*,\ \|\xi_{i\tau}\| .. \partial_{ik} \left\| \tau_i^* \right\| + \ell_{ik} \right\} \end{cases} \tag{20}$$

where $\xi_{i\tau} = \tau_i^* - \tau_{ik}^*$; ∂_{ik} is the trigger condition value, $1 > \partial_{ik} > 0$, $\|\ell_{ik}\| \geq 0$. Triggered at the time t_k, $k \in R^+$, at which point the control input is updated. If the calculation at the control end cannot satisfy (Equation 20), the control command transmission will not be performed, and the USV-ROVs will execute the control command in $[t_k, t_{k+1})$ under the action of the zero-order keeper. $\tau_i^* = L_{id} \tanh(\tau_i) + E_{id}$ is the controller output after smooth saturation fitting, L_{id} is the control threshold parameter for saturated non-linear filtering, and E_{id} is the filtering error.

3.4 Formation controller design

The dynamic adjustment terminal sliding mode formation controller will be designed in two steps, and the control process is depicted in Figure 3.

To simplify the algorithm design and address the issue of heterogeneous equipment with varying orders, derived after introducing auxiliary variables in (12), the new error surface is (Equation 21).

$$\dot{E}_{Qi} = \begin{bmatrix} \dot{E}_{Qa} \\ \dot{E}_{Sb} \end{bmatrix} = \begin{bmatrix} Q_{\lambda a} E_{qa} - Q_{qa} \\ E_{qb} \end{bmatrix} \tag{21}$$

where E_{Sb} and E_{Qa} are the ROV auxiliary sliding mode surfaces, $E_{Sb} = E_{qb}$, $E_{Qa} = Q_{\lambda a} E_{qa} - Q_{qa}$, $Q_{\lambda a} = [1, 0, 0; 0, 1, 0; 0, 0, 0]$, $Q_{qa} = [d - d\cos(\theta_e), -d\sin(\theta_e) - \theta_e, (1 - \cos(\theta_e))^{0.5}]^{\mathrm{T}}$.

Based on the system model (Equation 8) and the coordination model (Equation 16), the problem of geometric kinematics coordination for the formation is reformulated as the coordination and stabilization problem of the relative pose coordinates of each equipment. The virtual inversion finite-time kinematics controller is (Equation 22).

$$u_{Di} = W_{Qi} E_{Qi} + \bar{W}_{Ei} \tag{22}$$

where $W_{Qi} = [W_{Qa}, 0; 0, W_{Qb}]$, $\bar{W}_{Ei} = [W_{Ea}, W_{Eb}]^{\mathrm{T}}$. Among them, $W_{Qa} = [k_{a2}, 0, 1; 0, \partial_a k_{a1}, 0]$ and $W_{Qb} = diag(\alpha_{sbx}, \alpha_{sby}, \alpha_{sb\theta})$ are positive parameter matrices of the kinematic controller, $W_{Eb} = D_{zb}(E_{Sb}) sign[E_{Sb}] + \eta_b^d$ and $W_{Ea} = [-1 - v_{ar} E_{Qax} - w_{ar}\theta_{ae} + v_{ar}; \beta_a v_{ar} \sin(\theta_{ae}) + w_{ar}]$ are auxiliary quantities. β_a, ∂_a, α_{sbx}, α_{sby}, and $\alpha_{sb\theta}$ are all normal constants, and $Z_{bEq} = D_{zb}(E_{Sb})$ is a dynamic regulator.

Lemma 2 (Gao et al., 2016). The Radial Basis Function Neyral Network (RBFNN) is cited because of the non-linear characteristics of system uncertainty. If there exists m-dimensional compact set $\Xi^{m_\kappa} \subseteq \mathrm{R}^{m_\kappa} \rightarrow \mathrm{R}$ and there is an unknown non-linear function $f_\kappa(Q)$ with initial value 0 defined on Ξ^{m_κ}, then the RBFNN approximator is used to fit the dynamic values of $f_\kappa(Q_\kappa) = W_\kappa^{*\mathrm{T}} Z_\kappa(Q_\kappa) + e_{Z\kappa}(Q_\kappa)$, $\forall Q_\kappa \in \Xi^{m_\kappa}$, where $e_{Z\kappa}(Q_\kappa)$ is the bounded RBF fitting error, which is defined on the compact set Ξ^{m_κ}. $|e_Z(Q)| \leq \bar{e}_Z$, and $\bar{e}_Z$ is the maximum nuclear distance. To improve the non-linear local approximation ability, the Gaussian function $Z_\kappa(Q_\kappa) = \exp((Q_\kappa - \kappa)^{\mathrm{T}}(Q_\kappa - \kappa)/-l^2)$ is selected as the smooth kernel function. The κ is approaching the center column distance vector. The l is varying constant value. W_κ^* is the order m-dimensional weight row vector of optimally fitted, as $W_\kappa^* = \arg\left(\min_{\hat{W}_\kappa^*} \left\{ \sup_{Q_\kappa \in \Xi^{m_\kappa}} \left| \hat{W}_\kappa^{*\mathrm{T}} Z_\kappa(Q_\kappa) - f_\kappa(Q_\kappa) \right| \right\} \right)$, where $\hat{W}_\kappa^*$ is the minimum estimate of W_κ^*.

Using the RBF neural network in Lemma 2 to estimate the dynamic uncertainty of the model, the RBF neural network approximator $\hat{F}_i(\eta_i) = W_i^{*T} Z_{Wi}(\eta_i) + e_{Zi}(\eta_i)$ is constructed and brought into (Equation 12), and in (Equations 21) and (22), once the outer-loop model is organized based on the classification of

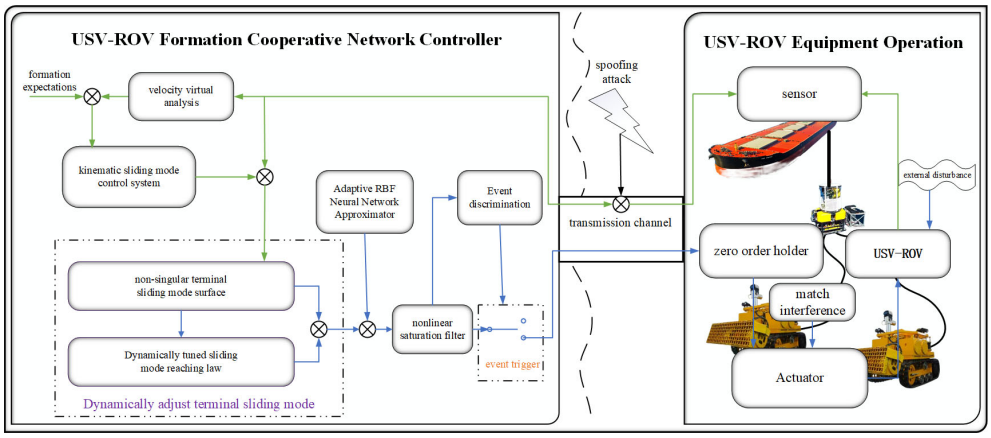


FIGURE 3
Schematic diagram of USV-ROV dynamic adjustment terminal skid formation control process.

formation equipment, using the design concept of the back-stepping virtual control law, the error dynamic Equation 23 can be described as

$$\dot{E}_{ui} = \dot{\bar{\eta}}_i^{\beta} - (\tilde{B}_i \tau_{si} - W_i^{*T} Z_{Wi}(\eta_i) - e_{Zi}(\eta_i) - \bar{\tau}_{id}) \tag{23}$$

where $e_{Zi}(\eta_i)$ is the approximation error of device i and W_i^{*T} is the neural network weight matrix. $Z_{Wi}(\eta_i)$ is the hyperbolic tangent kernel function for feature extraction.

Utilize terminal sliding mode surfaces and dynamically regulate convergence laws to exploit sliding mode robust response with two-stage global finite time convergence capability for passive resilient aggression tolerant formation control of deceptive attacks.

$$\dot{S}_{Di} = -\epsilon_{ui} D_{zi}(S_{ui}) sign[S_{ui}] + \kappa_{ui} S_{ui} \tag{24}$$

where $D_{zi}(S_{ui}) = (\varsigma_{ui} + S_{ui}^2)\ln(\varsigma_{ui} + S_{ui}^2) - \ln(\varsigma_{ui})$ is the sliding mode reaching law dynamic regulator.

Based on the model (8) and the dynamic adjustment approach law (Equation 24), the inner loop formation controller and the adaptive law are described in (Equations 25) and (26), respectively.

$$\begin{aligned}\tau_i = (\tilde{B}_i A_{i2})^{-1}[&\varepsilon_{ui} sign[S_{ui}] \ln[1 + \|S_{ui}\|^2](1 + \ln[1 + \|S_{ui}\|^2]^{\|S_{ui}\|^2 - 1}) \\ &+ \hat{\varphi}_{\tau i} G_{ti} \tanh(D_{ti} S_{ui}) + G_{Hi} \hat{\varphi}_{\tau i} S_{ui} H_{di}^4 + A_{i2} \dot{\bar{\eta}}_i^{\beta}]\end{aligned} \tag{25}$$

$$\dot{\hat{\varphi}}_{\tau i} = G_{ti} S_u \tanh(D_{ti} S_{ui}) - b_{\tau i} \hat{\varphi}_{\tau i} + G_{Hi} \|S_{ui}\|^2 \tanh(\|S_{ui}\|^2) H_{di}^4 \tag{26}$$

where G_t, D_{ti}, G_{Hi}, and $b_{\tau i}$ are bounded normal constants. $A_{i1} = a_{i2} sign(E_{ui}) \|E_{ui}\|^{\frac{1}{4}}$, $A_{i2} = a_{i1} + \frac{9a_{i3}}{4} \|E_{ui}\|^{\frac{5}{4}} sign(E_{ui})$. $\dot{\hat{\varphi}}_{\tau i}$ is an adaptive RBF disturbance compensator, $\varphi_\tau(t) = \max\left\{\|A_{i2}\tilde{B}_i \tau_M\|, \left\|\tilde{B}_i e_{\tau\star}\right\|, \left\|A_{i2}\zeta_i^{\beta}(1 + e_{Zi})\right\|, \|A_{i2} W^*\|, \|A_{i2}\bar{\tau}_d(t)\|, A_{i1}\right\}$, $H_{di} = \left\|\tilde{B}\tau_{\max}\right\| + 5 + \|Z\| + \|u_e\|$, $\|\varphi_{\tau i}\| \le \bar{\varphi}_{\tau i}$.

3.5 Stability analysis

The stability verification of the kinematics controller and the dynamics controller will be conducted separately.

A. Kinematics stability proof

Based on the formation system, the Lyapunov direct method is utilized to design $V_O = \frac{1}{2} E_{Qi}^{\mathrm{T}} E_{Qi}$, and deriving it, one can get

$$\dot{V}_O = \dot{V}_{O1} + \dot{V}_{O2} = E_{Qa}^{\mathrm{T}} \frac{dE_{Qa}}{dt} + E_{Qi}^{\mathrm{T}} \frac{dE_{Qi}}{dt} \tag{27}$$

Substituting (Equations 22) and (24) into (Equation 27) results in, when combined with Young's inequality, the equation

$$\begin{aligned}\dot{V}_O &\le -L_{Od1} V_{O1} - L_{Od2} V_{O1}^{0.5} + L_{Od3} - \varepsilon_{zb1}^{*} V_{O2} - \varepsilon_{zb2}^{*} V_{O2}^{0.5} + \varepsilon_{zb3} \\ &\le -L_{Os1} V_O - L_{Os2} V_O^{0.5} + L_{Os3}\end{aligned} \tag{28}$$

where $L_{Od1} = \min\{l_2, l_4, l_6\}$, $L_{Od1} = \min\{l_1, l_3, l_5\}$, $L_{Od3} = 6$, $l_1 = k_{a2}^{0.5}$, $l_2 = k_{a2}$, $l_3 = \sqrt{\partial_a k_{a1}}$, $l_4 = \partial_a k_{a1}$, and $l_5 = l_6 = 1$. $\varepsilon_{zb1}^{*} = \min\{\varepsilon_{zb}\}$, $\varepsilon_{zb2}^{*} = \min\{\epsilon_{zb}(c_{sb2} - c_{sb1} - \varsigma_{ub}\ln(\varsigma_{ub}))\}$, $c_{sb2} > c_{sb1} \ge 0.25$, and $c_{sb2} - c_{sb1} \ge \varsigma_{ub}\ln(\varsigma_{ub}) \ge 0$. $L_{Os1} = \min\left\{L_{Od1}, \varepsilon_{zb1}^{*}\right\}$, $L_{Os2} = \min\left\{L_{Od2}, \varepsilon_{zb2}^{*}\right\}$, and $L_{Os3} = \max\{\varepsilon_{zb3}, L_{Od3}\}$. (Equation 28) can prove that the dual-loop controller can stabilize the global motion of the formation system in a finite time.

B. Proof of dynamic stability of USV-ROV system

Based on (Equations 8), (14), and (26), design Lyapunov function $V_u = \frac{1}{2} S_{ui}^{\mathrm{T}} S_{ui} + \frac{1}{2} \tilde{\varphi}_{\tau i}^{\mathrm{T}} \tilde{\varphi}_{\tau i}$, where $\tilde{\varphi}_{\tau i} = \varphi_{\tau i} - \hat{\varphi}_{\tau i}$. Deriving V_u, using the norm inequality and the minimum learning parameter reduction method, combined with Young's inequality, the Equation 29 can be gotten:

$$\begin{aligned}\dot{V}_u \le\ & 0.5\varepsilon_{ui}\|S_{ui}\|^2 - \varepsilon_{iu}(c_1 - c_2)\|S_{ui}\| - (0.25 b_{\tau i} - 0.5)\tilde{\varphi}_{\tau i}^T \tilde{\varphi}_{\tau i} - 0.25\|\tilde{\varphi}_{\tau i}\| \\ &+ \varepsilon_{ui} c_2^2 + 0.2785 D_{ti}^{-1} G_{ti}\, \bar{\varphi}_{\tau i} + 0.5\tilde{\varphi}_{\tau i}^T \varphi_{\tau i} + 0.5 b_{\tau i} \varphi_{\tau i}^T \varphi_{\tau i} + 1/16 b_{\tau i}\end{aligned} \tag{29}$$

where $c_2 > c_1$ and $b_{\tau i} > 2$.

Lemma 3 (Liu et al., 2020). For $x_\varepsilon, y_\varepsilon \in \mathrm{R}^{n_\varepsilon}$, any $(x_\varepsilon, y_{\varepsilon 1}), (x_\varepsilon, y_{\varepsilon 2}) \in \Omega$, there exists a constant $L_\epsilon \ge 0$ such that $\|f_\varepsilon(x_\varepsilon, y_1) - f_\varepsilon(x_\varepsilon, y_{\varepsilon 2})\| \le L_\varepsilon \|(x_\varepsilon, y_{\varepsilon 1}) - (x_\varepsilon, y_{\varepsilon 2})\|$ exists, then $f_\varepsilon(\cdot)$ is said to satisfy the Lipschitz condition on the tight set Ω.

Based on (Equation 27), the Lyapunov global function $V_{D\star} = V_O + V_u$ is established, and after derivation and simplification, the equation can be gotten:

$$\begin{aligned}\dot{V}_{D\star} &\le -\lambda_1(\|S_{ui}\| + \|\tilde{\varphi}_{ui}\| + \left\|E_{qb}\right\|) - \lambda_2(\|S_{ui}\|^2 + \|\tilde{\varphi}_{ui}\|^2 + E_{qb}^{\mathrm{T}} E_{qb}) + \xi \\ &\le -\lambda_1 V_{D\star}^{0.5} - \lambda_2 V_{D\star} + \xi\end{aligned} \tag{30}$$

where $\lambda_1 = \min\{\varepsilon_{ui}(c_1 - c_2), 0.25, L_{Os2}\}$, $\lambda_2 = \min\{0.5 G_s \varepsilon_u, (0.25 b_\tau - 0.5), L_{OS1}\}$. The boundary term is $\xi = \max\left\{2 + G_s \varepsilon_u c_2^2 + 0.2785 D_t^{-1} G_t\, \bar{\varphi}_\tau(t) + \xi_1\right\}$. It can be seen that (Equation 30) can converge in a finite time.

Based on the convergence and boundedness of the closed-loop system, when $t \to T_u$, it can be seen that there is boundedness $\left\|V_{D\star}\right\|, , \frac{\xi}{\lambda_1^2}$ in the system. $T_{su\star}$ is the bounded time for convergence, $T_{su\star}, , \frac{2}{\lambda_1}\ln\left[\frac{\lambda_1 V^{0.5}(0) + \lambda_2}{\lambda_2}\right]$.

C. Event trigger proof of Zeno-free phenomenon

Since $\|t_{k+1} - t_k\| > t_{u\star}$, $t_{u\star} \in \mathrm{N}^+$, $\left\|\tau_i^{*}\right\| \le \tau_M$ can be known according to (Equation 20), thus $\|\xi_{i\tau}\| = \left\|\tau_i^{*} - \tau_{ik}^{*}\right\| \le \tau_\star$, where $\tau_\star > 0$. Then, integrating the trigger error $\xi_{i\tau}$ with t, it can be gotten that

$$\begin{aligned}\frac{d\|\xi_{i\tau}\|}{dt} &= \mathrm{sign}[\dot{\xi}_{i\tau}]\dot{\xi}_{i\tau} \\ &\leqslant \frac{d\|L_{id}\tanh(\tau_i) + E_{id}\|}{dt}\end{aligned} \tag{31}$$

According to Equation 31 and the controller output time correlation, and the input is limited by saturation τ_M, $(t_{k+1} - t_k) t_{u\star}^{-1} \ldots t_{u\star}^{-1} dt \ldots (\tau_M t_{u\star})^{-1} d\xi_{i\tau} \ldots (\lim_{t \to t_{k+1}} \xi_{i\tau}(t) - \xi_{i\tau}(t_k))(\tau_M t_{u\star})^{-1} \ldots (\partial ik\left\|\tau_i^{*}\right\| + \ell_{ik})(\tau_M t_{u\star})^{-1}$ can be obtained. From this we can know that $\inf\left\{t_{u\star}\right\} = (\partial_{ik}\left\|\tau_i^{*}\right\| + \ell_{ik})/(\tau_M t_{u\star})$, so the trigger condition (20) can avoid the proof of Zeno's phenomenon.

Hence, the stability of the entire formation control system has been demonstrated.

4 Simulation analysis

For the scientific rigor of the simulation, the ROV in the USV-ROV formation system uses the submarine operation ROV in the literature (Chen et al., 2023); the mass is 15kg, the width is 0.5m, the driving wheel radius is 0.1m, the distance between the center of mass and the center of gravity is 0.05m, the driving input is $[15\mathrm{N}\cdot\mathrm{m},\ 15\mathrm{N}\cdot\mathrm{m}]^{\mathrm{T}}$, and the moment of inertia is $J_{rov} = 5\mathrm{kg}\cdot\mathrm{m}^2$. The USV chooses the 1 : 70 ship CyberShip II (Zhang et al., 2022) of the Norwegian University of Science and Technology as the simulation object, which has double symmetrical propellers, double tail rudders, and single side thrusters; the mass is 23.8kg the length of the hull is 1.255m, the width is 0.29m, and the drive input is $[5\mathrm{N},\ 5\mathrm{N},\ 2.5\mathrm{N}\cdot\mathrm{m}]^{\mathrm{T}}$. At the same time, the parameters are $d_{11} = 12\mathrm{kg/s}$, $d_{22} = 17\mathrm{kg/s}$, $d_{33} = 0.5\mathrm{kg/s}$, $d_{23} = d_{32} = 0.2\mathrm{kg/s}$, $m_{11} = 25.5\mathrm{kg}$, $m_{22} = 33.8\mathrm{kg}$, $m_{33} = 2.76\mathrm{kg}$, and $m_{23} = m_{32} = 1.0948\mathrm{kg}$.

The stealthy deception attack probability is $\sigma_i^{\beta} = 0.7$, and $\mu_i^{\beta} \in [-2,\ 2]$ is the attack random tampering amount. $\tau_{da} = [\tau_{dav},\ \tau_{dar},\ \tau_{daw}]$ are wind, wave, and current disturbances (Zhang et al., 2022), $\tau_{daw} = 0.25\cos(t) + 0.25N_w$, $\tau_{dav} = 0.2\sin(0.5t + \pi/4) + N_w$, and $\tau_{dar} = 0.18\sin(0.5t) + 0.1\cos(t) + N_w$; $\tau_{db} = [1 + N_w,\ 1.5\sin(t) + 1.5\cos(t),\ \sin(1.5t) + \cos(1.5t)]$ is the amount of external disturbance (Chen et al., 2023) encountered by the ROV during its movement, and N_w is Gaussian white noise with a power of one. The total simulation time is 200s, and the step size of the simulation is 0.01s. The amount and number of USV three-channel attack tampering is shown in Figure 4.

To verify the control performance of the control algorithm in different tasks, the curvature-transformed trajectory that synthesizes the broken line and curve characteristics is selected as the reference trajectory of the virtual leader $q_{a_0} = [x_r,\ y_r,\ \theta_r]^{\mathrm{T}}$; to facilitate the verification of the USV-ROV formation control effect, one USV is selected as the leader, and two ROVs are selected as follows: $q_{a_{1s}}(0) = [4,\ 8,\ 1.8^{-1}\pi]$ is the USV initial position, and $q_{b_1}(0) = [3,\ 7,\ 1.8^{-1}\pi]^{\mathrm{T}}$ and $q_{b_2}(0) = [6,\ 5,\ 1.8^{-1}\pi]^{\mathrm{T}}$ are ROV initial positions. To verify the superiority of the algorithm in this paper, $S_{ui2} = a_{i1}E_{ui} + a_{i2}\int E_{ui}d\gamma_t + a_{i3}\dot{E}_{ui}$ is introduced, and the comparison controller is designed in combination with the exponential reaching law, and Equation 32 can be obtained

$$\tau_{i2} = (\tilde{B}_i a_{i1})^{-1}[\varepsilon_{ui}sign[S_{ui2}] + \hat{\varphi}_{\tau i}G_{ti}\tanh(D_{ti}S_{ui2}) + G_{Hi}\hat{\varphi}_{\tau i}S_{ui2}H_{di}^4 + a_{i1}\dot{\bar{\eta}}_i^{\beta}] \tag{32}$$

The parameters for the formation kinematics and dynamics controller can be described in Table 1.

To quantitatively assess the effectiveness of the algorithm presented in this paper, the following metrics are employed: mean integral absolute control (MIAC) is used to compute input energy, mean integral square error (MISE) is utilized to measure error control accuracy, and mean integral total variation (MITV) is designed to assess incoming communication transmission frequencies. The

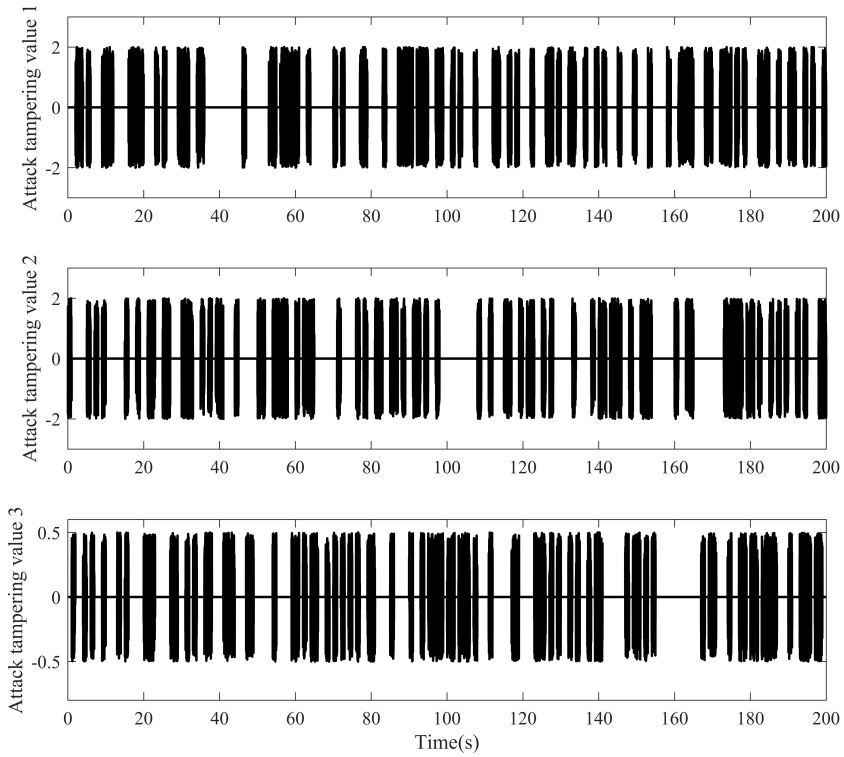


FIGURE 4
Deceptive attack onset time and amount of tampering.

results are summarized in Tables 2 and 3, providing quantitative insights. $\mathrm{MIAC} = (t_f - t_0)^{-1}\int_{t_0}^{t_f}\|\tau_{si}\|d\mu$, $\mathrm{MITV} = (t_f - t_0)^{-1}\int_{t_0}^{t_f}\|\tau_{si}(\mu+1) - \tau_{si}(\mu)\|d\mu$, $\mathrm{MISE} = (t_f - t_0)^{-1}\int_{t_0}^{t_f}\|E_D\|^2 d\mu$.

Figure 5 illustrates the formation recovery and hold control achieved under deceptive attacks using the controller (25) proposed in this paper. Position error comparisons in Figures 6–8 reveal that the algorithm presented in this study exhibits higher control accuracy than controller (32). Under the control of controller (25), the formation can attain the desired position in approximately 18 seconds, demonstrating superior tracking efficiency compared to controller (32). Quantitative values of MISE in Tables 2, 3 further confirm the enhanced control accuracy of this paper's algorithm, showcasing improvements of 12.82 %, 8.92 %, and 28.00 %, respectively, for position error compared to controller (32). Additionally, during the attack phase of 50s-150s in Figures 6–8, the error fluctuation of controller (25) is smaller than that of controller (32), indicating better stability for the proposed algorithm in this paper.

Analyzing Figures 9–11 reveals that, in comparison to controller (32), controller (25) is more effective in reducing jitter. The sliding phase converges faster to the expectation $S_{ui} = 0$, and during the sliding mode switching in the attack phase, controller (25) experiences less jitter under the influence of dynamic convergence law (24), achieving faster convergence and meeting design expectations. Taking 140-160s in Figures 10–12 as an illustration, the proposed method exhibits smaller jitter fluctuations and shorter convergence recovery response times compared to controller (32).

Figures 12, 13 illustrate that, in contrast to controller (32), controller (25) integrates the adaptive parameter estimator (26) designed by the RBF neural network with the influence of the minimum learning parameter. This integration establishes a virtual model for deceptive attacks and external disturbances, allowing for effective compensation of input anomalies. Controller (25) exhibits resistance against the continuous excitation of state changes, thereby countering the effects of deceptive attacks and disturbances.

TABLE 1 Formation controller parameters.

Parameters	Value	Parameters	Value
α_{sax}	$diag$ (0.1, 0.01, 10)	k_{b1}	0.4
α_{say}	$diag$ (0.1, 0.2, 5.5)	k_{b2}	1.15
$\alpha_{sa\theta}$	$diag$ (0.1, 0.1, 100.2)	∂_b	0.5
a_{i1}	$diag$ (100, 185, 150, 1.5, 90)	G_{Hir}	$diag\,(5\times10^{-5},0,0)$
a_{i2}	$diag$ (100, 0.01, 0.5, 0.8, 150)	G_{Hiv}	$diag\,(10^{-11},10^{-5},10^{-5})$
a_{i3}	$diag$ (0.65, 15.5, 55, 150, 0.01)	G_{Hiw}	$diag\,(1,2\times10^{-5},2\times10^{-5})$
D_{tir}	$diag$ (25, 0, 0)	G_{tir}	$diag\,(0.005,0,0)$
D_{tiv}	$diag$ (10, 10, 78)	G_{tiv}	$diag\,(10,15\times10^{-7},15\times10^{-7})$
D_{tiw}	$diag\,(100,78,78)$	G_{tiw}	$diag\,(0.05,10^{-5},10^{-5})$
ϵ_{ri}	$diag$ (0.05, 0, 0)	$b_{\tau iv}$	$diag\,(180,195,195)$
ϵ_{wi}	$diag$ (0.0115, 0.025, 0.025)	$b_{\tau ir}$	$diag\,(130,0,0)$
ϵ_{vi}	$diag$ (0.05, 0.015, 0.015)	$b_{\tau iw}$	$diag\,(80,140,140)$

TABLE 2 Quantitative analysis of control effect of controller (25).

Evaluation criteria	MIAC	MISE	MITV	Transfer times
equipment a_1	[0.8, 1.19, 2.66]	[3.62, 3.32, 4.60]	[7.8, 6.19, 6.01]	[7980, 8610]
equipment b_1	[0.9, 1.69]	[3.35, 2.82, 2.02]	[5.9, 6.31]	[8433, 9106]
equipment b_2	[1.1, 1.09]	[3.10, 2.93, 2.66]	[5.1, 5.90]	[6621, 7638]

TABLE 3 Quantitative analysis of control effect of controller (32).

Evaluation criteria	MIAC	MISE	MITV	Transfer times
equipment a_1	[1.8, 2.37, 4.06]	[4.12, 4.01, 4.71]	[8.61, 5.90, 8.21]	20000
equipment b_1	[1.20, 3.71]	[4.27, 3.46, 2.52]	[5.85, 7.01]	20000
equipment b_2	[0.85, 2.52]	[2.87, 2.73, 2.81]	[6.43, 6.40]	20000

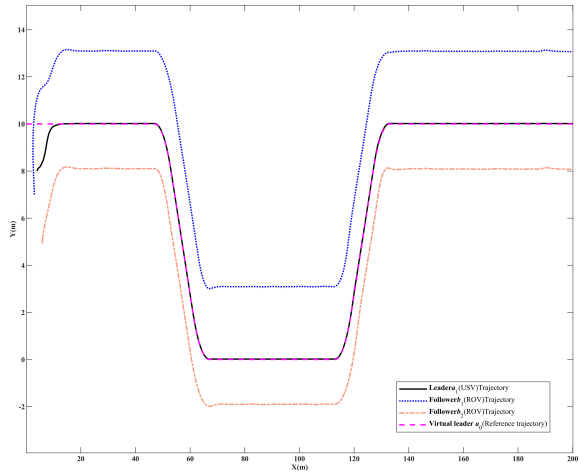


FIGURE 5
USV-ROV formation collaborative tracking movement.

Examining Figures 14–16 and the MIAC in Tables 2 and 3, let us consider the time interval of 50-100s. In comparison to controller (32), the input of the controller (25) proposed in this paper, despite being exposed to deceptive attacks, exhibits a smaller jitter range, fewer high-frequency spikes, conserves more control resources, and demonstrates superior control dynamics.

Observing the MITV and transfer times in Tables 2 and 3, it is evident that controller (25), in contrast to controller (32), employs an event triggering mechanism with saturation characteristics to govern the controller output. This approach conserves approximately 50 % of control resources, enhancing control robustness within the rated power limit. Simultaneously, it improves transfer regulation performance, achieving superior control efficacy under deceptive attacks.

5 Conclusion

This paper addresses the control of USV-ROVs formation systems in the presence of deceptive data injection. A controller is designed considering dynamic uncertainties, deceptive attacks, unknown time-varying environmental disturbances, equipment matching disturbances, and input saturation constraints. First, a formation control scheme is designed to successfully realize the desired operational formation through geometric transformations. Then, a new terminal sliding mode surface is designed to solve the singularity

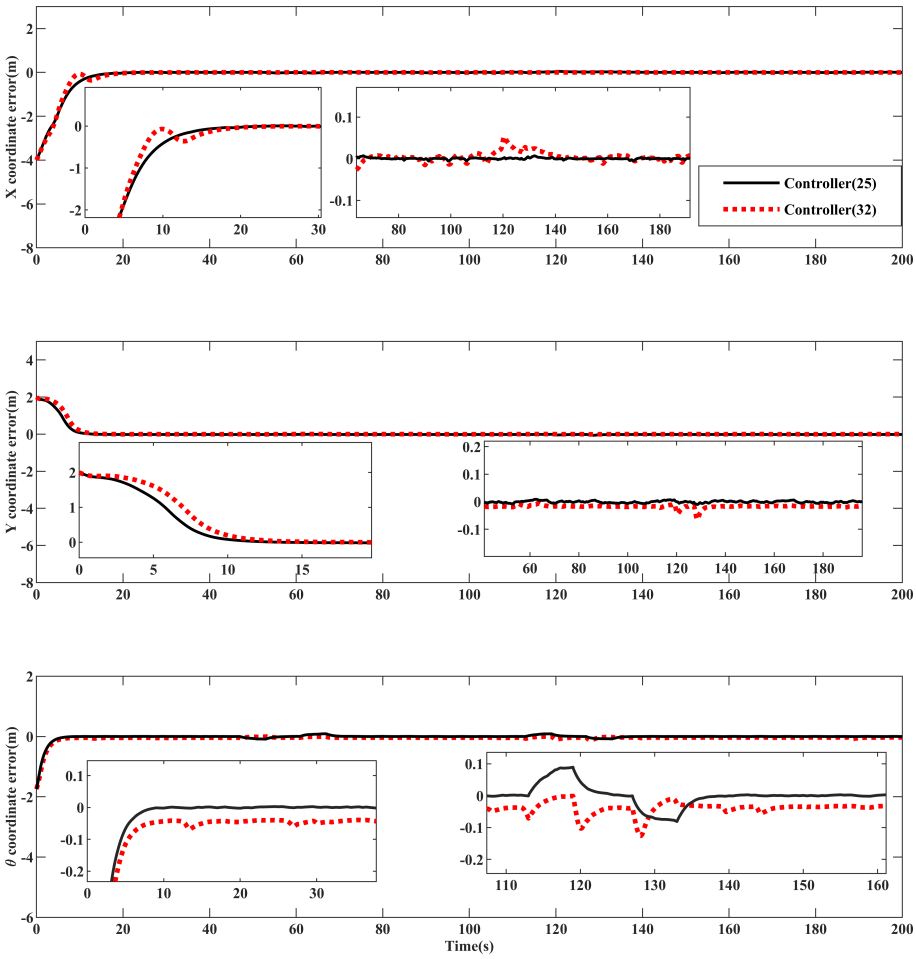


FIGURE 6
Comparison of position error convergence of USV equipment a_1.

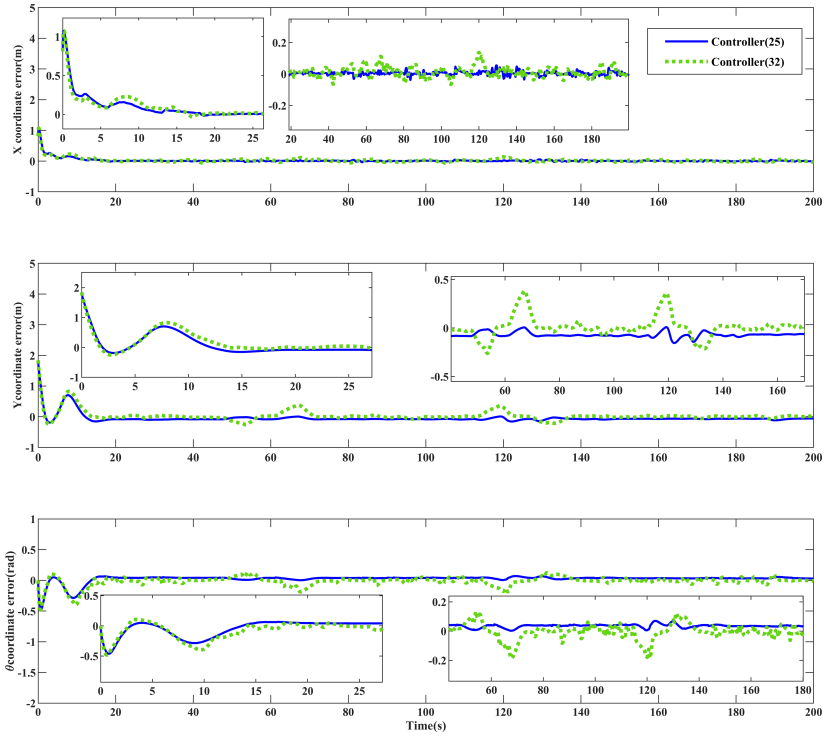


FIGURE 7
Convergence comparison of position error of ROV equipment b_1.

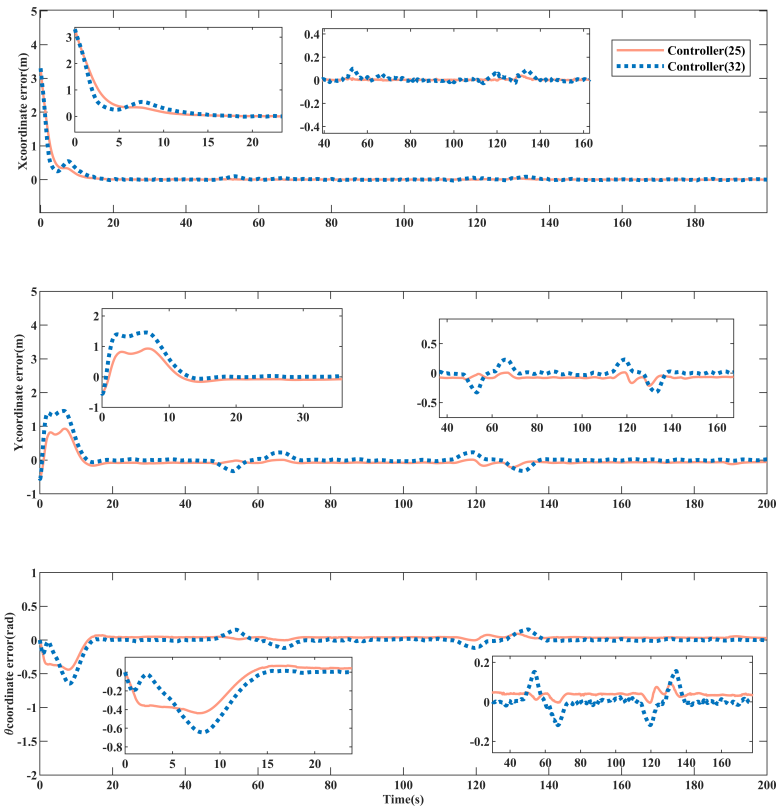


FIGURE 8
Convergence comparison of position error of ROV equipment b_2.

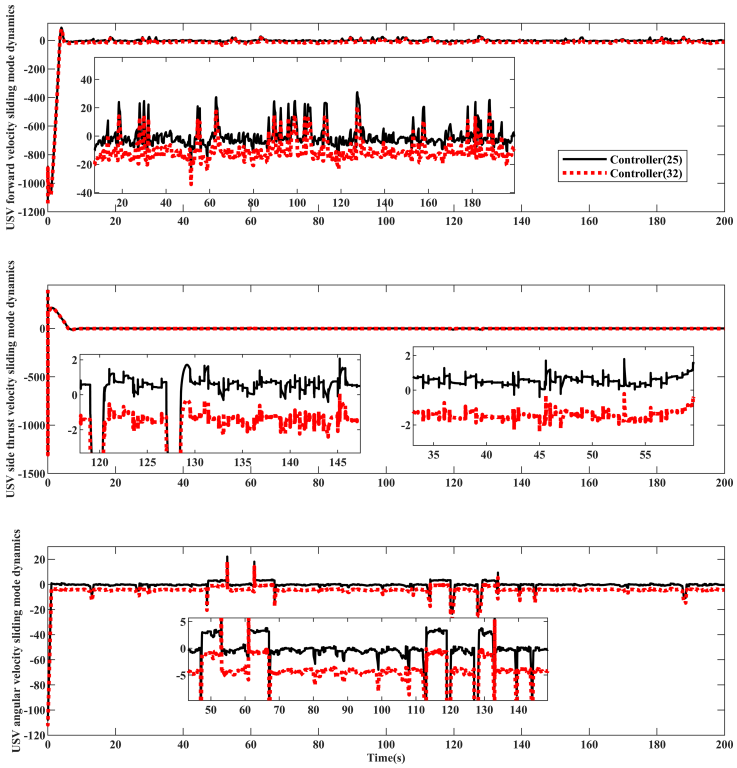


FIGURE 9
Dynamic comparison of sliding mode of USV equipment a_1.

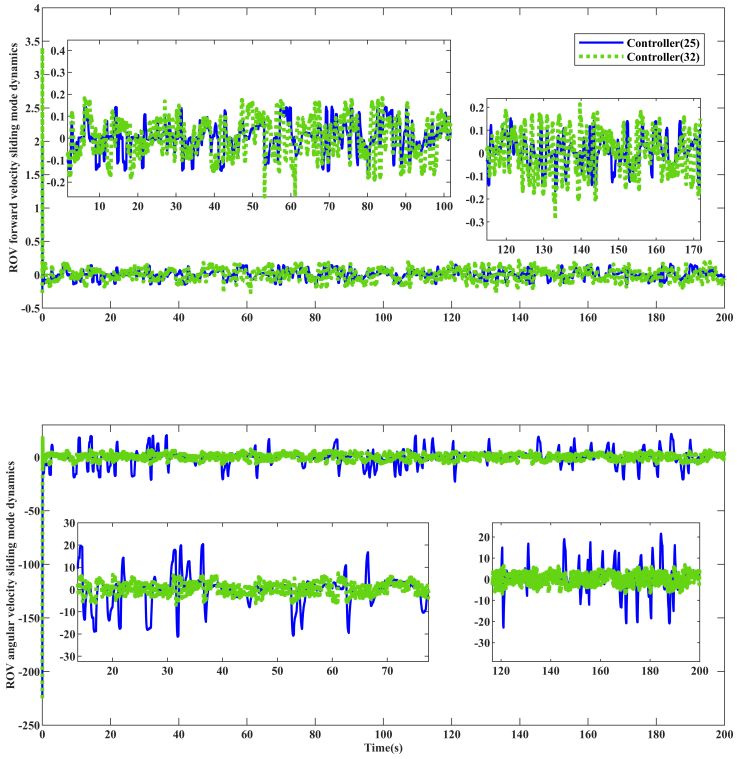


FIGURE 10
Dynamic comparison of sliding mode of ROV equipment b_1.

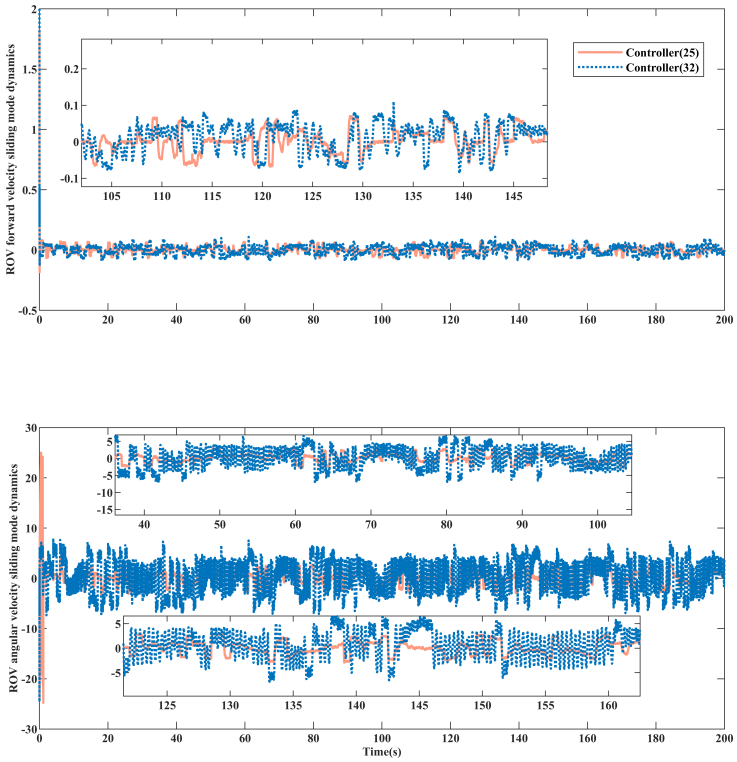


FIGURE 11
Dynamic comparison of sliding mode of ROV equipment b_2.

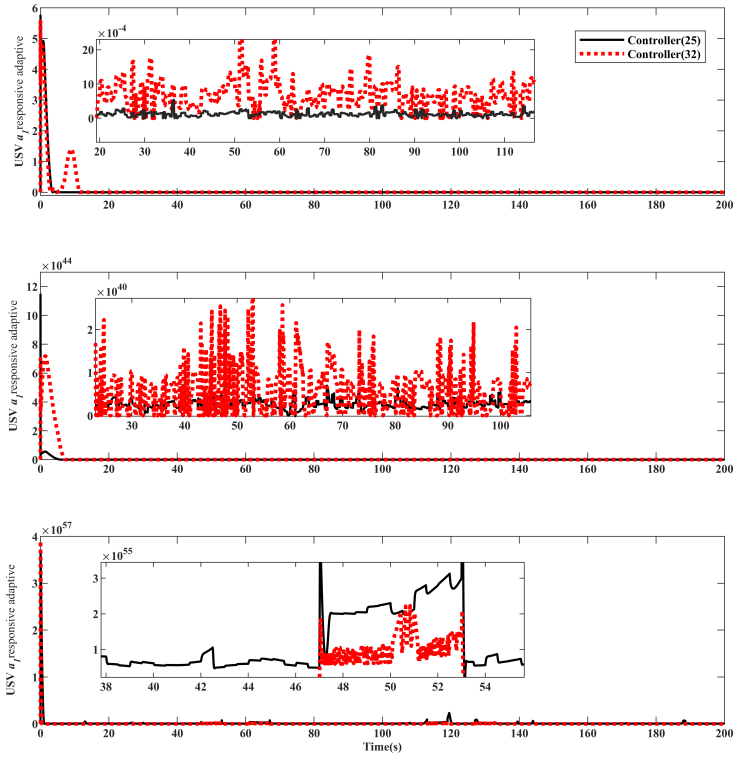


FIGURE 12
Comparison of adaptive approximation response of USV equipment a_1.

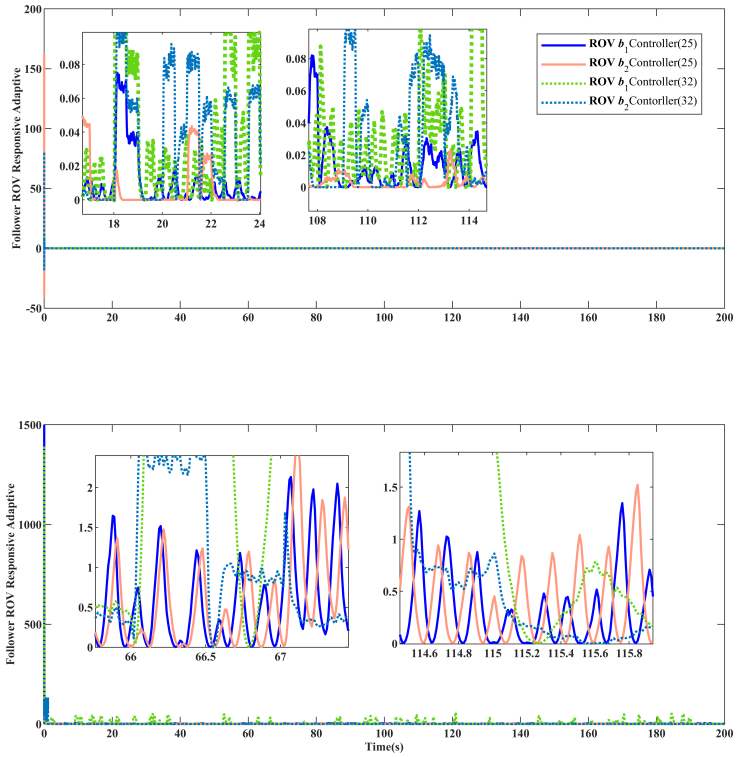


FIGURE 13
Comparison of adaptive approach response of ROV equipment.

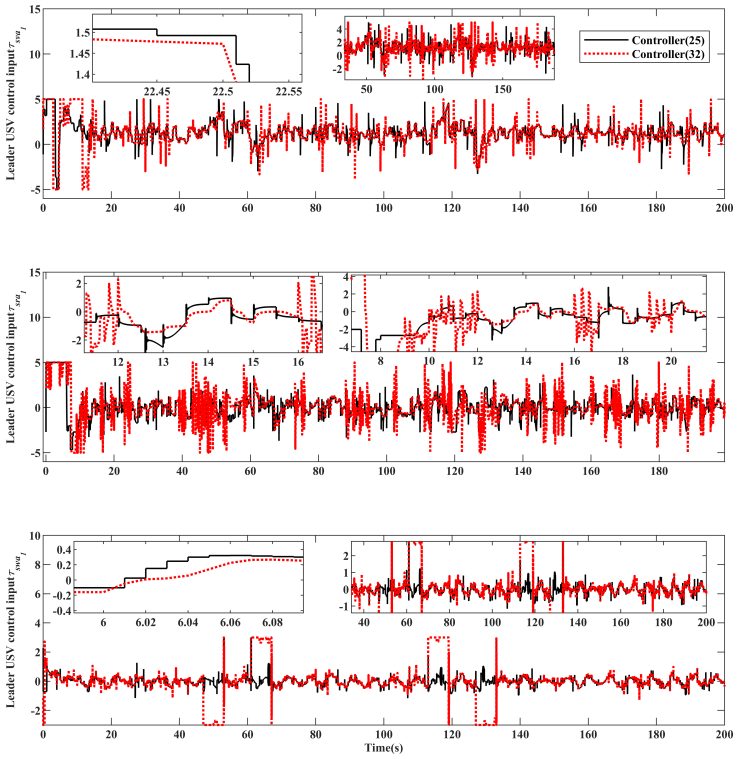


FIGURE 14
USV equipment a_1 system input comparison.

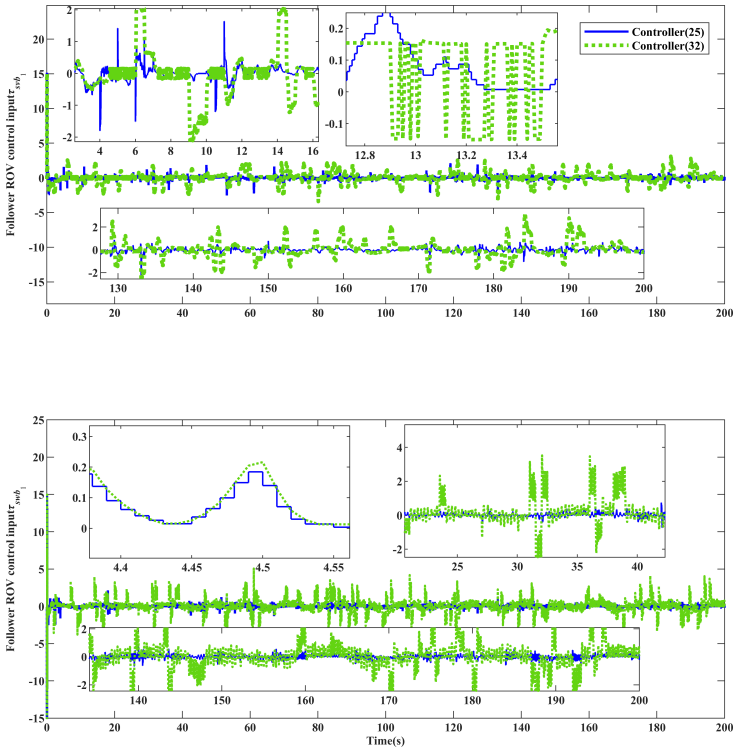


FIGURE 15
ROV equipment b_1 system input comparison.

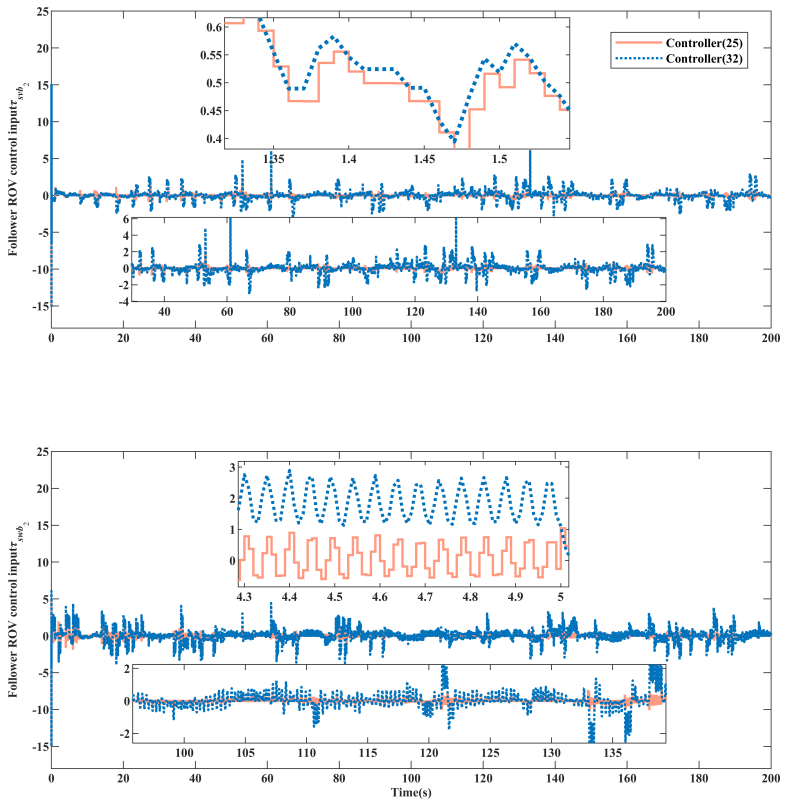


FIGURE 16
ROV equipment b_2 system input comparison.

problem of the terminal sliding mode surface while achieving global finite time convergence. After that, a new dynamic convergence law is designed to speed up the convergence response, optimize the sliding mode jitter problem, and ensure robustness under deceptive attacks. Finally, a new controller is designed using an event-triggered mechanism, RBF neural networks, and adaptive techniques to avoid the effects of deceptive attacks and uncertainty, save control resources, and enhance robustness. To verify the control stability of the algorithm under the influence of deceptive attacks, experiments were conducted using single USV and dual ROVs formation configurations. These results were then compared with the performance of a conventional sliding mode control algorithm. The experiments confirm the feasibility and superiority of the method proposed in this paper and ultimately demonstrate the effectiveness of the robust adaptive compensation mechanism against deceptive attacks.

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

QZ: Writing – review & editing. SZ: Writing – original draft. YL: Writing – review & editing. YZ: Writing – review & editing. YH: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The research carried out in this paper was mainly supported by the Shandong Provincial Natural Science Foundation grant (No. ZR2022ME087) and partly supported by the Shandong Provincial Science and Technology Department of Science and Technology Small and Medium-sized Enterprises Innovation Capacity Enhancement Project (No. 2021TSGC1394), as well as being partly supported by the project (No. 2022KJ210) in the Shandong province colleges and universities "Youth Innovation Team Plan".

Acknowledgments

The authors would like to express their sincere gratitude to Qiyu Chen and Gaoyang Guo for their valuable comments during the writing of the paper, and to the reviewers for their positive and constructive comments.

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

Chen, Q., Zhang, Q., Hu, Y., Liu, Y., and Wu, H. (2023). Euclidean distance damping–based adaptive sliding mode fault-tolerant event-triggered trajectory-tracking control. *Proc. Institution Mechanical Engineers Part I: J. Syst. Control Eng.* 237 (3), 551–567. doi: 10.1177/09596518221126584

Chen, W., Ding, D., Dong, H., and Wei, G. (2019). Distributed resilient filtering for power systems subject to denial-of-service attacks. *IEEE Trans. Systems Man Cybernetics: Syst.* 49 (8), 1688–1697. doi: 10.1109/TSMC.2019.2905253

Ding, L., and Guo, G. (2012). Formation control for ship fleet based on backstepping. *Control Decision* 27 (2), 299–303.

Ding, D., Wang, Z., Ho, D. W., and Wei, G. (2016). Observer-based event-triggering consensus control for multiagent systems with lossy sensors and cyber-attacks. *IEEE Trans. cybernetics* 47 (8), 1936–1947. doi: 10.1109/TCYB.2016.2582802

Fan, Y., Liu, B., Wang, G., and Mu, D. (2021). Adaptive fast non-singular terminal sliding mode path following control for an underactuated unmanned surface vehicle with uncertainties and unknown disturbances. *Sensors* 21 (22), 7454. doi: 10.3390/s21227454

Fu, M. Y., Wang, D. S., and Wang, C. L. (2018). Formation control for water-jet USV based on bio-inspired method. *China Ocean Engineering. 32*, 117–122. doi: 10.1007/s13344-018-0013-1

Gao, H., Song, Y., and Wen, C. (2016). Backstepping design of adaptive neural fault-tolerant control for MIMO nonlinear systems. *IEEE Trans. Neural Networks Learn. Syst.* 28 (11), 2605–2613. doi: 10.1109/TNNLS.2016.2599009

Huang, X., and Dong, J. (2019). Reliable control of cyber-physical systems under sensor and actuator attacks: An identifier-critic based integral sliding-mode control approach. *Neurocomputing* 361, 229–242. doi: 10.1016/j.neucom.2019.06.069

Ito, T., and Takanami, I. (1997). "On fault injection approaches for fault tolerance of feedforward neural networks," in *Proceedings sixth asian test symposium (ATS'97)*. 88–93 (IEEE).

Jin, X. (2018). Fault-tolerant iterative learning control for mobile robots non-repetitive trajectory tracking with output constraints. *Automatica* 94, 63–71. doi: 10.1016/j.automatica.2018.04.011

Kang, Z., Yu, H., and Li, C. (2020). Variable-parameter double-power reaching law sliding mode control method. *Automatika* 61 (3), 345–351. doi: 10.1080/00051144.2020.1757965

Li, X., Wen, C., Wang, J., Xing, L., and Li, X. (2022). Jamming-resilient synchronization of networked lagrangian systems with quantized sampling data. *IEEE Trans. Ind. Inf.* 18 (12), 8724–8734. doi: 10.1109/TII.2022.3148395

Li, J., Zhang, G., Shan, Q., and Zhang, W. (2023a). A novel cooperative design for USV-UAV systems: 3-D mapping guidance and adaptive fuzzy control. *IEEE Trans. Control Network Systems.* 10 (2), 564–574. doi: 10.1109/TCNS.2022.3220705

Li, J., Zhang, G., Zhang, W., Shan, Q., and Zhang, W. (2023b). Cooperative path following control of USV-UAVs considering low design complexity and command transmission requirements. *IEEE Trans. Intelligent Vehicles*. doi: 10.1109/TIV.2023.3317336

Lin, A., Jiang, D., and Zeng, J. (2018). Underactuated ship formation control with input saturation. *Acta Autom. Sin.* 33, 1496–1504.

Liu, J., Wang, Y., Cao, J., Yue, D., and Xie, X. (2020). Secure adaptive-event-triggered filter design with input constraint and hybrid cyber attack. *IEEE Trans. Cybernetics* 51 (8), 4000–4010. doi: 10.1109/TCYB.2020.3003752

Min, H., Xu, S., and Zhang, Z. (2020). Adaptive finite-time stabilization of stochastic nonlinear systems subject to full-state constraints and input saturation. *IEEE Trans. Automatic Control* 66 (3), 1306–1313. doi: 10.1109/TAC.2020.2990173

Saihi, L., Ferroudji, F., Berbaoui, B., Bakou, Y., Koussa, K., Meguellati, F., et al. (2019). "Hybrid control based on sliding mode fuzzy of DFIG power associated WECS," in *AIP conference proceedings*, vol. 2123, No. 1). (AIP Publishing).

Sandberg, H., Amin, S., and Johansson, K. H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Syst. Magazine* 35 (1), 20–23. doi: 10.1109/MCS.2014.2364708

Shi, J., Zhou, D., Yang, Y., and Sun, J. (2018a). Fault tolerant multivehicle formation control framework with applications in multiquadrotor systems. *Sci. China Inf. Sci.* 61, 1–3. doi: 10.1007/s11432-018-9486-3

Shi, J., Zhou, D., Yang, Y., and Sun, J. (2018b). Fault tolerant multivehicle formation control framework with applications in multiquadrotor systems. *Sci. China Inf. Sci.* 61, 1–3. doi: 10.1007/s11432-018-9486-3

Skjetne, R., Fossen, T. I., and Kokotović, P. V. (2005). Adaptive maneuvering, with experiments, for a model ship in a marine control laboratory. *Automatica* 41 (2), 289–298. doi: 10.1016/j.automatica.2004.10.006

Sun, R., Wang, J., Zhang, D., and Shao, X. (2018). Neural network-based sliding mode control for atmospheric-actuated spacecraft formation using switching strategy. *Adv. Space Res.* 61 (3), 914–926. doi: 10.1016/j.asr.2017.11.011

Swaszek, P. F., Hartnett, R. J., Kempe, M. V., and Johnson, G. W. (2013). "Analysis of a simple, multi-receiver GPS spoof detector," in *Proceedings of the 2013 international technical meeting of the institute of navigation*, 884–892.

Tian, Y., Cai, Y., and Deng, Y. (2020). A fast nonsingular terminal sliding mode control method for nonlinear systems with fixed-time stability guarantees. *IEEE Access* 8, 60444–60454. doi: 10.1109/ACCESS.2020.2980044

Wang, W., Han, Z., Liu, K., and Lü, J. (2021). Distributed adaptive resilient formation control of uncertain nonholonomic mobile robots under deception attacks. *IEEE Trans. Circuits Syst. I: Regular Papers* 68 (9), 3822–3835. doi: 10.1109/TCSI.2021.3096937

Wu, Y., and Wang, Y. (2020). Asymptotic tracking control of uncertain nonholonomic wheeled mobile robot with actuator saturation and external disturbances. *Neural Computing Appl.* 32 (12), 8735–8745. doi: 10.1007/s00521-019-04373-9

Wu, C., Wu, L., Liu, J., and Jiang, Z. P. (2019). Active defense-based resilient sliding mode control under denial-of-service attacks. *IEEE Trans. Inf. Forensics Secur.* 15, 237–249. doi: 10.1109/TIFS.2019.2917373

Xu, H., Cui, G., Ma, Q., Li, Z., and Hao, W. (2023). Fixed-time disturbance observer-based distributed formation control for multiple QUAVs. *IEEE Trans. Circuits Syst. II: Express Briefs.* doi: 10.1109/TCSII.2022.3233438

Xu, X. P., Yan, X. T., Yang, W. Y., An, K., Huang, W., and Wang, Y. (2022). Algorithms and applications of intelligent swarm cooperative control: A comprehensive survey. *Prog. Aerospace Sci.* 135, 100869. doi: 10.1016/j.paerosci.2022.100869

Yang, G. Y., and Chen, S. Y. (2020). Piecewise fast multi-power reaching law: Basis for sliding mode control algorithm. *Measurement Control* 53 (9-10), 1929–1942. doi: 10.1177/0020294020964246

Yin, T., Gu, Z., and Park, J. H. (2022). Event-based intermittent formation control of multi-UAV systems under deception attacks. *IEEE Trans. Neural Networks Learn. Systems.* doi: 10.1109/TNNLS.2022.3227101

Yun, L., and Ying, J. (2016). Combination of leader-follower method and potential function about ship formation control. *Control Theory Appl. 33*, 1259–1264.

Zhang, G., and Zhang, X. (2015). A novel DVS guidance principle and robust adaptive path-following control for underactuated ships using low frequency gain-learning. *ISA Trans.* 56, 75–85. doi: 10.1016/j.isatra.2014.12.002

Zhang, G., Liu, S., and Zhang, X. (2022). Adaptive distributed fault-tolerant control for underactuated surface vehicles with bridge-to-bridge event-triggered mechanism. *Ocean Eng.* 262, 112205. doi: 10.1016/j.oceaneng.2022.112205

Zhang, D. W., and Liu, G. P. (2023). Predictive sliding-mode control for networked high-order fully actuated multiagents under random deception attacks. *IEEE Trans. Systems Man Cybernetics: Systems.* doi: 10.1109/TSMC.2023.3311438

Zhu, G., and Du, J. (2018). Global robust adaptive trajectory tracking control for surface ships under input saturation. *IEEE J. Oceanic Eng.* 45 (2), 442–450. doi: 10.1109/JOE.2018.2877895

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 12 March 2024
DOI 10.3389/fmars.2024.1348883

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Lanyong Zhang,
Harbin Engineering University, China
Tingkai Chen,
Dalian Maritime University, China
Boguslaw Cyganek,
AGH University of Science and Technology, Poland

*CORRESPONDENCE
Feihu Zhang
feihu.zhang@nwpu.edu.cn

RECEIVED 03 December 2023
ACCEPTED 20 February 2024
PUBLISHED 12 March 2024

CITATION
Cheng C, Wang C, Yang D, Wen X, Liu W and Zhang F (2024) Underwater small target detection based on dynamic convolution and attention mechanism.
Front. Mar. Sci. 11:1348883.
doi: 10.3389/fmars.2024.1348883

Underwater small target detection based on dynamic convolution and attention mechanism

Chensheng Cheng, Can Wang, Dianyu Yang, Xin Wen, Weidong Liu and Feihu Zhang*

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

In ocean observation missions, unmanned autonomous ocean observation platforms play a crucial role, with precise target detection technology serving as a key support for the autonomous operation of unmanned platforms. Among various underwater sensing devices, side-scan sonar (SSS) has become a primary tool for wide-area underwater detection due to its extensive detection range. However, current research on target detection with SSS primarily focuses on large targets such as sunken ships and aircraft, lacking investigations into small targets. In this study, we collected data on underwater small targets using an unmanned boat equipped with SSS and proposed an enhancement method based on the YOLOv7 model for detecting small targets in SSS images. First, to obtain more accurate initial anchor boxes, we replaced the original k-means algorithm with the k-means++ algorithm. Next, we replaced ordinary convolution blocks in the backbone network with Omni-dimensional Dynamic Convolution (ODConv) to enhance the feature extraction capability for small targets. Subsequently, we inserted a Global Attention Mechanism (GAM) into the neck network to focus on global information and extract target features, effectively addressing the issue of sparse target features in SSS images. Finally, we mitigated the harmful gradients produced by low-quality annotated data by adopting Wise-IoU (WIoU) to improve the detection accuracy of small targets in SSS images. Through validation on the test set, the proposed method showed a significant improvement compared to the original YOLOv7, with increases of 5.05% and 2.51% in *mAP*@0.5 and *mAP*@0.5: 0.95 indicators, respectively. The proposed method demonstrated excellent performance in detecting small targets in SSS images and can be applied to the detection of underwater mines and small equipment, providing effective support for underwater small target detection tasks.

KEYWORDS

side-scan sonar, underwater target detection, YOLOv7, K-Means++, ODConv, GAM, WIoU

1 Introduction

Due to the distinctive attributes of the underwater environment, optical imaging techniques face substantial limitations when deployed underwater. Conversely, sound waves experience minimal attenuation in water, rendering side-scan sonar (SSS) a prevalent tool for underwater target detection.

Sonar target detection methods can be categorized into traditional techniques and Convolutional Neural Network (CNN)-based approaches. Conventional sonar image detection methods predominantly employ pixel-based (Chen et al., 2014), feature-based (Mukherjee et al., 2011), and echo-based (Raghuvanshi et al., 2014) strategies. These methods utilize manually crafted filters founded on pixel value characteristics, grayscale thresholds, or *a priori* information about the targets for detection. However, underwater settings are intricate, and sonar echoes contend with self-noise, reverberation noise, and environmental noise. Consequently, sonar images exhibit low resolution, blurred edge details, and significant speckle noise, complicating the identification of dependable pixel traits and grayscale thresholds. Furthermore, owing to the diminutive illuminated regions and ambiguous target features in acoustic images, even for the same target, discrepancies in the sonar's position, depth, and angle can lead to variations in the morphological attributes of the target within sonar images. Hence, existing conventional algorithms encounter notable constraints in terms of technical feasibility, time requirements, and applicability when confronted with intricate sonar target detection scenarios. A pressing necessity exists for a detection algorithm that remains robust against fluctuations in target morphology in sonar images, mitigates erroneous detections and omissions induced by background noise interference, and exhibits commendable generalization capabilities.

In comparison to traditional methodologies, deep learning approaches rooted in CNN offer substantial advantages due to their capacity to autonomously acquire and extract deep-level features from images. The learned feature parameters often outperform manually devised counterparts, resulting in significantly heightened detection accuracy when applied to large datasets, as compared to traditional methods. Presently, CNN-based object detection methodologies within the domain of optical image processing have attained a mature stage of development. Researchers have progressively extended the application of these technologies to various inspection tasks, such as steel defect detection (Yang et al., 2021; Zhao et al., 2021), medical image analysis (Bhattacharya et al., 2021; Jia et al., 2022), marine life detection (Chen et al., 2021; Wang et al., 2023c), radar image interpretation (Hou et al., 2021; Zhang et al., 2021a), agricultural product inspection (Soeb et al., 2023; Yang et al., 2023), and more. Significant achievements have been made in each of these fields. Moreover, CNN-based methods can also be employed for image enhancement to improve the quality of blurry images and enhance the recognition of regions of interest (Chen et al., 2023; Wang et al., 2023b), thereby enhancing the effectiveness of target detection. Therefore, investigating how to apply CNN-based object detection methods more efficiently to the field of underwater acoustic image target detection is a highly worthwhile research endeavor. Furthermore, this research can contribute to addressing the challenges associated with underwater acoustic image target detection difficulties.

As of now, employing deep learning techniques for target detection in SSS images still faces several challenges (Le et al., 2020; Neupane and Seok, 2020; Hożyń, 2021). Firstly, current target detection networks typically rely on anchor box initializations derived from extensive optical datasets, which may not necessarily be suitable for our unique SSS dataset. Consequently, there is a need to re-cluster and generate anchor box initializations customized to specific dataset. Secondly, factors such as sound wave propagation loss, refraction, and scattering often result in acquired sonar images exhibiting characteristics such as low contrast, strong speckle noise, and blurry target edges. In comparison to conventional camera images, sonar images significantly differ in terms of texture diversity, color saturation, and feature resolution. Hence, it is imperative to enhance the feature extraction capability of the backbone network and apply appropriate attention mechanisms to target features in sonar images, aiming to improve detection accuracy. Lastly, due to the formidable challenges associated with collecting SSS image data, obtaining a sufficient quantity of thoroughly comprehensive and high-quality image data for network training is challenging. This necessitates making the most of all available data, including some lower-quality data, to maximize the average detection accuracy.

In response to these challenges, this paper takes full consideration of the unique characteristics of the SSS dataset. Four improvements are made to the YOLOv7 network to enhance its detection performance for small targets in SSS images. The effectiveness of the proposed improvements is validated through multiple experiments. The main contributions of this paper are as follows:

1) We replaced the k-means algorithm with k-means++ to recluster the annotated bounding boxes in the SSS dataset, thereby obtaining initial anchor boxes that are more suitable for the sizes of small targets in the dataset.
2) We replaced the static convolutional blocks in the backbone network with Omni-dimensional Dynamic Convolution (ODConv), considering the multi-dimensional information of convolutional kernels. This substitution enhances the feature extraction capability of the network without significantly increasing the number of parameters.
3) In the neck network, five global attention mechanism (GAM) modules are introduced, taking into account global information and enhancing the capability to extract target features. This addresses the challenge of feature sparsity commonly found in SSS images.
4) In the loss function section, we introduced Wise-IoU (WIoU) to address the issue of poor quality in SSS data. Such an improvement can alleviate the adverse impact of low-quality data on gradients, leading to higher data utilization and, consequently, an improvement in the detection accuracy of the trained model.

The remaining sections of this paper are structured as follows. Section 2 elaborates on related research concerning underwater acoustic target detection. In Section 3, we detail the methodology

adopted in this study. The experimental procedure and outcome presentation are outlined in Section 4. Finally, Section 5 provides a summary of this paper and offers prospects for future work.

2 Related work

Extensive research has been undertaken in the domain of underwater acoustic image target detection (Lee et al., 2018; Zhang et al., 2021b; Kim et al., 2022; Tang et al., 2023). These endeavors encompass the design of specialized functional modules tailored to data characteristics or the adaptation and enhancement of networks originally well-suited for optical data to underwater acoustic data.

(Jin et al., 2019) devised EchoNet, a deep neural network architecture that leverages transfer learning to detect sizable objects like airplanes and submerged vessels in forward-looking sonar images (Fan et al., 2021). introduced a 32-layer residual network to replace ResNet50/101 in MASK-RCNN, streamlining the network's parameter count while upholding object detection accuracy. They also adopted the Adagard optimizer in place of SGD and evaluated the detection accuracy of the network model through cross-training with a collection of 2500 sonar images (Singh and Valdenegro-Toro, 2021). conducted a comparison of diverse target segmentation networks, including LinkNet, DeepLabV3, PSPNet, and UNet, based on an extensive dataset of over 1800 forward-looking sonar images. Their investigation revealed that a UNet network employing ResNet34 as the backbone, tailored for their sonar dataset, achieved the most favorable outcomes. This network was subsequently applied to the detection and segmentation of marine debris (Xiao et al., 2021). addressed shadow information in acoustic images by introducing a shadow capture module capable of capturing and utilizing shadow data within the feature map. This module, compatible with CNN models, incurred a modest parameter increase and displayed portability. The incorporation of shadow features improved detection accuracy (Wang et al., 2021). proposed AGFE-Net, a novel sonar image target detection algorithm. This algorithm extended the receptive field of convolutional kernels through multi-scale receptive field feature extraction blocks and self-attention mechanisms, thus acquiring multi-scale feature information from sonar images and enhancing feature correlations. Employing a bidirectional feature pyramid network and an adaptive feature fusion block enabled the acquisition of deep semantic features, suppression of background noise interference, and precise prediction box selection through an adaptive non-maximum suppression algorithm, ultimately enhancing target localization accuracy. To address the issue of suboptimal transfer learning results due to significant domain gaps between optical and sonar images (Li et al., 2023a), introduced a transfer learning method for sonar image classification and object detection known as the Texture Feature Removal Network. They considered texture features in images as domain-specific features and mitigated domain gaps by discarding these domain-specific features, facilitating a more seamless knowledge transfer process. This innovative approach aims to bridge the gap between optical and sonar image analysis, enhancing the effectiveness of transfer learning techniques.

Due to the YOLO series of networks' excellent detection performance and ease of deployment, they have found wide application in the field of underwater target detection. Additionally, researchers have made numerous enhancements to the YOLO series networks, making them even more suitable for object detection in underwater acoustic images. In order to address the limitations in detection performance and low detection accuracy resulting from multi-scale image inputs (Li et al., 2023b), proposed an underwater target detection neural network based on the YOLOv3 algorithm, enhanced with spatial pyramid pooling. The improved neural network demonstrated promising results in the detection of underwater targets, including shipwrecks, schools of fish, and seafloor topography (Li et al., 2021). introduced an enhanced RBF-SE-YOLOv5 network that reallocates channel information weights to enhance effective information extraction. This enhancement entailed refining the backbone network of the original model and integrating it with RBFNet, thus improving the network's receptive field, feature representation, and capacity to learn vital information. The study demonstrated that amplifying perception information in high receptive fields and integrating multi-scale information augments the efficacy of vital feature extraction. The proposed algorithm notably enhances effective feature extraction, comprehensively captures global information, and mitigates prediction errors and issues of low credibility. Addressing the deficiency in detecting small targets in underwater sonar images (Wang et al., 2022), harnessed the YOLOv5 framework for marine debris detection. They introduced a multi-branch shuttle network into YOLOv5s and replaced YOLOv5s' neck network with BiFPN to augment detection performance. The study also analyzed the impact of uneven target data distribution and network scale on model performance, thereby furnishing reference solutions for ensuring accuracy and speed in target detection (Zhang et al., 2022a), grounded in the YOLOv5 framework, employed the IOU value between initial anchor boxes and target boxes instead of YOLOv5's Euclidean distance as the clustering criterion. This refinement brought the initial anchor boxes closer to true values, enhancing network convergence speed. Additionally, they introduced coordinate information by appending pixel coordinates of the image as extra channels to the feature map and performing convolution operations, consequently amplifying the accuracy of the detection module's localization regression (Li et al., 2023c). proposed MA-YOLOv7, a YOLOv7-based network that incorporates multi-scale information fusion and attention mechanisms for target detection and filtering in images. They also introduced a target localization method to determine target positions (latitude and longitude).

However, current research primarily revolves around employing SSS to detect large targets such as airplanes and sunken ships, or using forward-looking sonar to detect small targets at close range. There remains a significant dearth of research focused on utilizing SSS for wide-ranging detection of small underwater targets. This paper constructs a small target SSS dataset based on data collected during experiments and conducts a comprehensive study on small target detection methods in SSS. The

primary objective is to facilitate the advancement of the field of small target detection in SSS.

3 Improved methods

YOLOv7, introduced in 2022, stands as a one-stage object detection network (Wang et al., 2023a). It demonstrates outstanding proficiency in both detection speed and accuracy compared to other detection algorithms. In this study, we improved the YOLOv7 model and, through multiple experimental validations, identified four effective improvement points, as illustrated in Figure 1. We applied these enhancements to small target detection in SSS images, achieving notable improvements in detection performance compared to the original YOLOv7, as evidenced by significant enhancements in detection metrics.

3.1 K-means++

To enhance both efficiency and accuracy in detection, this study employs the k-means++ (Arthur and Vassilvitskii, 2007) technique to supplant the k-means approach, initially employed in YOLOv7, for clustering anchor boxes within the dataset. In the conventional k-means method, the first phase involves the random generation of n cluster centers from the data samples. Subsequently, the Euclidean distance between each sample and the cluster centers is computed, and the sample is assigned to the cluster center exhibiting the smallest Euclidean distance. In the subsequent phase, the cluster centers are reevaluated, and samples are reclassified. This iterative process is repeated until the cluster centers reach stability.

The k-means++ method represents an enhancement over the conventional k-means approach. Unlike generating all cluster centers randomly in a single instance, k-means++ generates one cluster center at a time. It calculates the Euclidean distance $D(x)$ between all samples and the cluster center, subsequently deriving the probability of each sample being chosen as the next cluster center through the Equation 1.

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (1)$$

Subsequently, the next cluster center is chosen via the roulette wheel selection method. This sequence of steps is reiterated until n cluster centers are generated. After this stage, the ensuing process resembles that of the conventional k-means algorithm: the cluster centers are updated, samples are reclassified, and these steps are iterated until the cluster centers achieve stability. While the k-means++ algorithm invests more time in selecting initial cluster centers, once these initial centers are established, the convergence speed accelerates, yielding cluster centers that hold greater representativeness. This approach mitigates the challenge of becoming trapped in local optima.

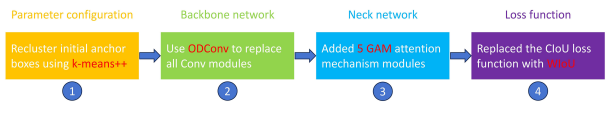


FIGURE 1
Improvements made to YOLOv7.

3.2 Omni-dimensional dynamic convolution

In current neural networks, the majority typically employ static convolutional kernels. However, recent research on dynamic convolutions suggests calculating relevant weights based on the input and linearly combining n convolutional kernels according to these weights. This makes the convolution operation dependent on the input, leading to a significant improvement in neural network accuracy. The experimental results from (Li et al., 2022) demonstrate that the use of ODConv enhances the detection performance for small targets. Therefore, in this study, all convolutional operations in the YOLOv7 backbone network are replaced with ODConv to enhance the detection performance of the network.

The core innovation of ODConv lies in its multi-dimensional dynamic attention mechanism. Traditional dynamic convolution typically achieves dynamism only in the dimension of the number of convolutional kernels, by weighting and combining multiple kernels to adapt to different input features. ODConv extends this concept further by dynamically adjusting not only the number of convolutional kernels but also three other dimensions: spatial size, input channel number, and output channel number. This means that ODConv can adapt more finely to the features of input data, thereby improving the effectiveness of feature extraction.

Additionally, ODConv employs a parallel strategy to simultaneously learn attention across different dimensions. This strategy allows the network to efficiently process features on each dimension while ensuring complementarity and synergy among the dimensions. This is particularly beneficial for handling complex features in SSS images. The network structure is illustrated in Figure 2.

The output after ODConv can be expressed using the Equation 2.

$$y = (d_1 \odot c_1 \odot b_1 \odot a_1 \odot W_1 + \ldots + d_n \odot c_n \odot b_n \odot a_n \odot W_n) * x \quad (2)$$

where a represents the attention parameter for the spatial dimensions of the convolutional kernel, b represents the attention parameter for the input channel dimensions, c represents the attention parameter for the output channel dimensions, d represents the attention parameter for the convolutional kernel W.

3.3 Global attention mechanism

The incorporation of attention mechanisms within neural networks draws inspiration from human visual attention, enhancing feature extraction by assigning distinct weights to various channels within neural network feature layers. This strategy enables the model to concentrate on pertinent

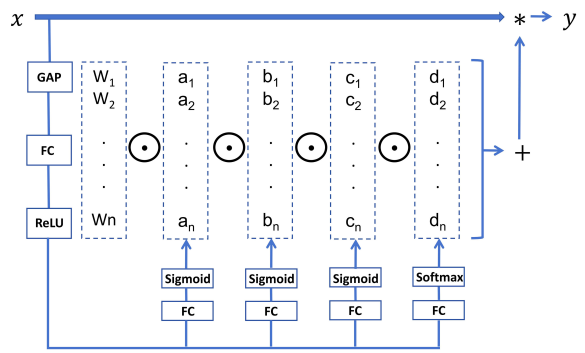


FIGURE 2
The architecture of the ODConv module.

information while disregarding irrelevant data, leading to resource conservation and augmented model performance. Several mainstream attention mechanisms, such as SE-Net (Hu et al., 2018), ECA-Net (Wang et al., 2020), BAM (Park et al., 2018), CBAM (Woo et al., 2018) and GAM (Liu et al., 2021), have been demonstrated to enhance the detection performance of models.

The GAM represents a form of global attention mechanism that curtails information loss and amplifies interactions across global dimensions. Consequently, the neural network's aptitude for extracting target features is bolstered. The schematic depiction of the GAM module structure is presented in Figure 3.

GAM employs a sequential channel-spatial attention mechanism with the aim of amplifying global inter-feature interactions while reducing information dispersion. In the channel attention submodule of GAM, a three-dimensional configuration is employed to preserve tridimensional information. The input feature map undergoes dimensional transformation and subsequently undergoes an MLP operation. The result is then reverted to the original dimension, and a sigmoid function is applied to produce the final output.

In the spatial attention submodule, aimed at intensifying focus on spatial information, two convolutional layers facilitate spatial data fusion. Initially, a convolution employing a kernel size of 7 is executed to diminish channel count and computational complexity. Subsequently, another convolution with a kernel size of 7 enhances the number of channels while maintaining uniform channel consistency. The resulting output is then processed through a sigmoid function.

In order to enhance the detection performance of the detection network, we introduced GAM modules at five distinct locations in the neck network. The architecture of the YOLOv7 network with added GAM modules, as well as the specific structures of individual sub-modules within the network, are illustrated in Figure 4.

3.4 Wise-IoU

The bounding box regression function holds a pivotal role in object detection by enhancing object localization accuracy, accommodating objects of varying scales, rectifying object orientations and shapes, and bolstering algorithmic robustness. This collective functionality contributes significantly to the advancement of object detection algorithms.

However, the majority of current research on Intersection over Union (IoU) (Yu et al., 2016) assumes that the training data consists of high-quality samples, with their primary focus being on enhancing the fitting capability of bounding box regression loss functions, such as Generalized-IoU (GIoU) (Rezatofighi et al., 2019), Distance-IoU (DIoU) (Zheng et al., 2020), Complete-IoU (CIoU) (Zheng et al., 2020), and Efficient-IoU (EIoU) (Zhang et al., 2022b), as shown in Table 1, where their advantages and disadvantages are compared. Yet, when dealing with datasets that

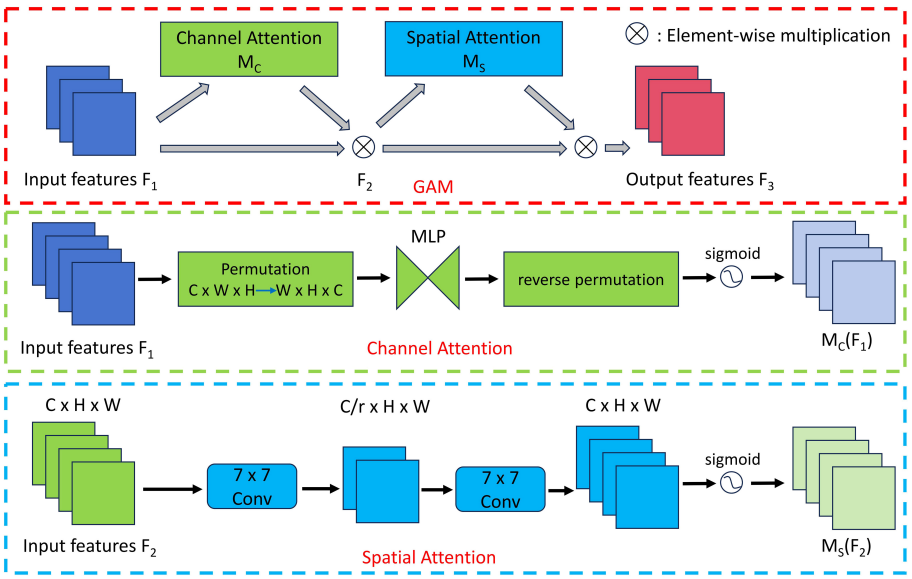


FIGURE 3
The structure diagram of the GAM module.

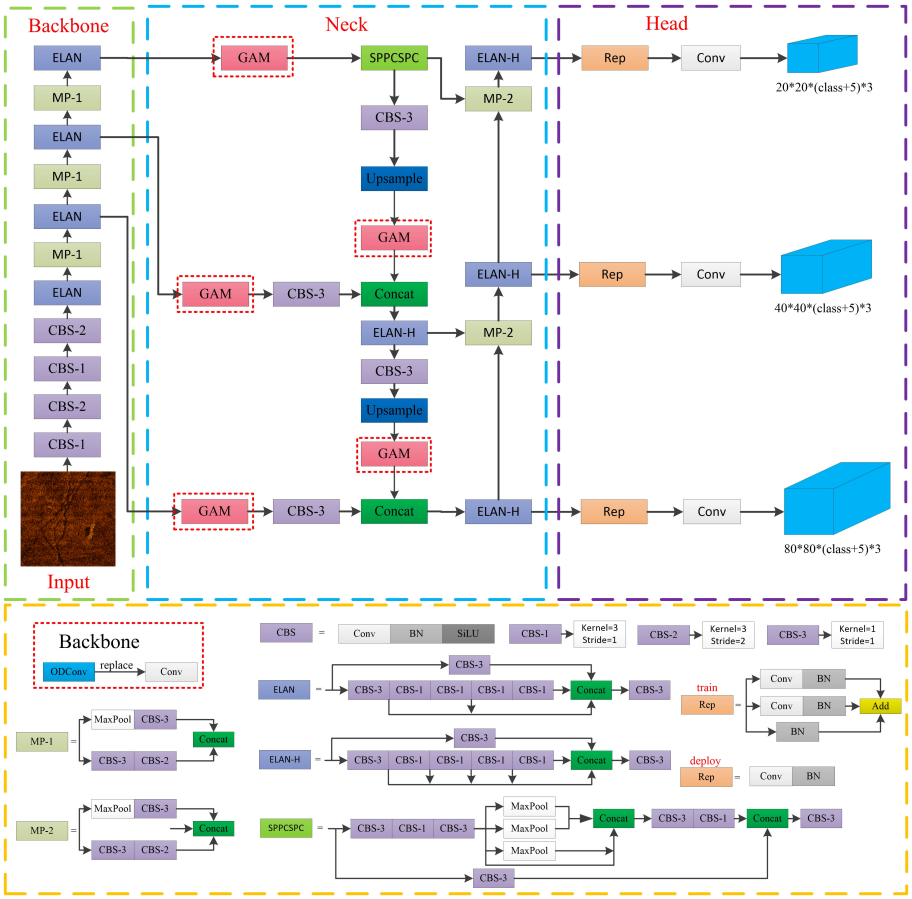


FIGURE 4
The network architecture diagram of the improved YOLOv7.

contain a significant amount of inaccurately annotated low-quality data, blindly intensifying the fitting ability of the bounding box regression loss function can have detrimental effects on the model's learning process.

In SSS imagery, targets are highly susceptible to noise interference in the generated images, which presents significant challenges for annotation. In the process of manual annotation, inaccuracies inevitably arise, as illustrated in Figure 5. If the annotation boxes are initially flawed, when an excellent detection model generates high-quality anchor boxes for low-quality sample data, the loss function L_{IoU} will have a relatively large value, leading to a substantial gradient gain. In such cases, the model will learn in an unfavorable direction. This phenomenon is particularly relevant in the context of scientific research and analysis for SSS imagery.

To address the issue of poor quality in underwater SSS data, we introduce the WIoU (Tong et al., 2023) as the bounding box loss

TABLE 1 Comparison of the advantages and shortcoming of different IoU methods.

	Overlapping	Center Point	Aspect Ratio	Advantage	Shortcoming
IoU	✓	×	×	Taking into account scale invariance and non-negativity.	If two boxes do not intersect, it cannot reflect the distance and cannot accurately reflect the degree of overlap between the two boxes.
GIoU	✓	×	×	Addressing the issue where the loss equals zero when there is no overlap between the detection box and the ground truth box.	When there is containment between the detection box and the ground truth box, GIOU degenerates into IOU, and when the two boxes intersect, convergence is slow in both the horizontal and vertical directions.
DIoU	✓	✓	×	Directly regressing the Euclidean distance between the centers of the two boxes to accelerate convergence.	Considering the aspect ratio of bounding boxes during the regression process, there is still room for further improvement in accuracy.

(Continued)

TABLE 1 Continued

	Overlapping	Center Point	Aspect Ratio	Advantage	Shortcoming
CIoU	✓	✓	✓	Introducing loss terms for the scale of the detection box, as well as for its length and width, which makes the predicted box better match the ground truth.	The aspect ratio describes relative values, introducing some degree of ambiguity and not considering the balance of difficulty levels among samples.
EIoU	✓	✓	✓	Calculating differences in width and height instead of aspect ratio, while also incorporating Focal Loss to tackle the problem of imbalanced difficulty levels among samples.	More attention is given to high-quality anchor boxes, with insufficient focus on low-quality anchor boxes.

function. This aims to alleviate the impact of low-quality anchor boxes generated during annotation. The WIoU function employs a dynamic non-monotonic focus mechanism that evaluates anchor box quality through outliers, instead of IoU. This approach furnishes a judicious gradient allocation strategy, curbing the competitiveness of high-quality anchor boxes while attenuating detrimental gradients arising from low-quality instances. Consequently, WIoU prioritizes anchor boxes of moderate quality, ameliorating detector performance overall.

The symbols defined in WIoU are illustrated as shown in Figure 6. In this figure, the blue box represents the smallest bounding box, and the red line represents the line connecting the centers of the true box and the predicted box, where the union area is denoted as $S_u = wh + w_{gt}h_{gt} - W_iHi$.

The WIoU methodology, founded on distance metrics, incorporates a two-tier attention mechanism known as WIoU v1. WIoU v1 can be represented by Equations 3, 4.

$$\mathcal{L}_{WIoUv1} = \mathcal{R}_{WIoU}\mathcal{L}_{IoU} \tag{3}$$

$$\mathcal{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{4}$$

where $\mathcal{L}_{IoU} \in [0, 1]$, W_g and H_g are the dimensions of the minimum bounding box, $(x, y)$and (x_{gt}, y_{gt}) represent the center coordinates of the predicted box and the ground truth box.

Subsequently, building upon WIoU v1, the incorporation of outliers is achieved through the Equation 5.

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\overline{\mathcal{L}_{IoU}}} \in [0, +\infty) \tag{5}$$

Finally, a non-monotonic focus coefficient β is formulated and integrated into WIoU v1. As a result, we obtain Equation 6.

$$\mathcal{L}_{WIoU} = r\mathcal{L}_{WIoUv1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{6}$$

A reduced outlier score implies a higher quality anchor box, yielding a diminished gradient gain assigned to it. Consequently, the bounding box regression concentrates on anchor boxes of intermediate quality. In contrast, anchor boxes exhibiting larger outlier scores are allocated lesser gradient gains, effectively curtailing the generation of significant harmful gradients from low-quality instances. Notably, as $\overline{\mathcal{L}_{IoU}}$ remains dynamic, the categorization threshold for anchor boxes' quality also remains adaptive. This adaptability empowers WIoU to judiciously allocate

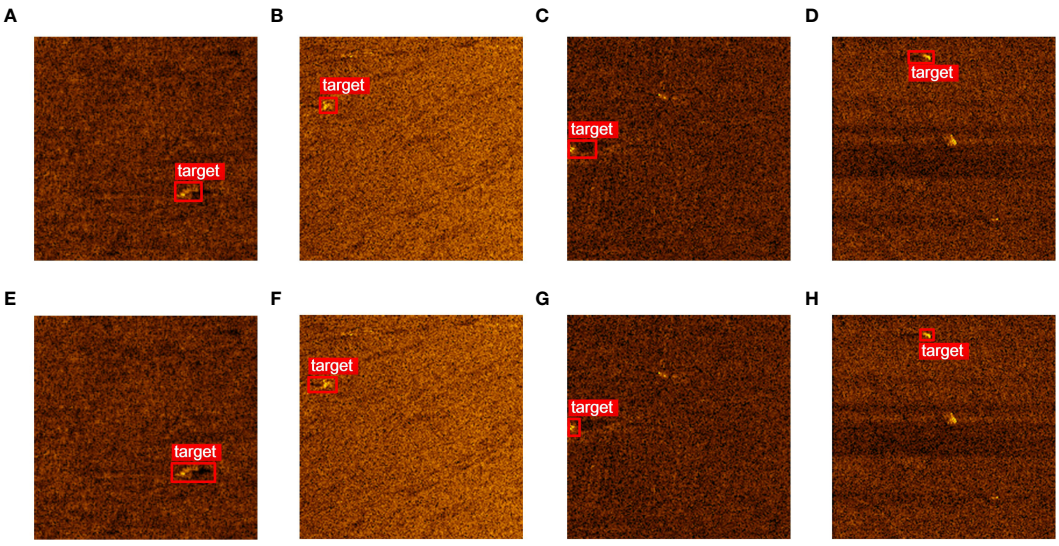


FIGURE 5
Low quality annotated samples. **(A–D)** are low-quality samples with inaccurate annotations, while **(E–H)** have accurate annotations.

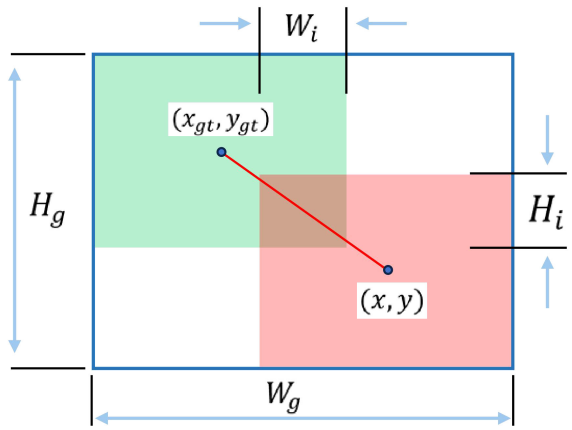


FIGURE 6
The symbol definitions in WIoU.

gradient gains that are suitable for real-time scenarios, enhancing its effectiveness in each instance.

4 Experiments

4.1 Experiment platform

The experiments presented in this study were carried out on an Ubuntu 20.04 system, serving to corroborate the efficacy of the proposed enhanced detection algorithm. Detailed configuration parameters of the system are provided in Table 2.

4.2 Model evaluation metrics

When evaluating the detection performance of the improved YOLOv7, we employed evaluation metrics including Recall (R), Precision (P), Average Precision (AP), and mean Average Precision (mAP). The calculation methods of these four indicators can be expressed by Equations 7–10 respectively.

$$R = TP/(TP + FN) \tag{7}$$

$$P = TP/(TP + FP) \tag{8}$$

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \sum_{i=1}^{N} AP_i/N \tag{10}$$

Within the array of evaluation metrics mentioned, True Positive (TP) signifies the tally of correctly identified positive samples, False Positive (FP) corresponds to the count of erroneously identified negative samples, and False Negative (FN) stands for the tally of positively labeled samples that remain undetected. The variable N represents the overall number of detected categories.

TABLE 2 Experimental environment settings.

Component	Specification
Operating system	Ubuntu 20.04(64-bit)
Deep learning framwork	Pytorch 1.11
Programming language	Python 3.9
GPU accelerated environment	CUDA 11.3
Graphics Card (GPU)	Nvidia GeForce RTX 3090
Processor (CPU)	Platinum 8255C CPU @ 2.50GHz

4.3 Dataset preparation

In the data acquisition phase, we deployed objects of two types, namely cylindrical and conical structures, as detection targets in the experimental marine area. We utilized the SS3060 dual-frequency SSS as the detector for data collection. The SSS and detection targets are illustrated in Figure 7. The size of the SSS is 100mm in diameter and 1250mm in length, with a weight of 25kg in air and 12kg in water. And the performance parameters of SSS are presented in Table 3. For the experiment's execution, the SSS was affixed beneath an unmanned boat. The utilization of GPS signals emanating from the unmanned boat enabled the verification of congruence between features visible in the SSS images and the physically predetermined targets. This methodology thereby facilitated the creation of a dataset characterized by high quality.

After deploying the targets, to ensure the diversity of the collected dataset, we employed two different survey paths in the target water area to perform a comprehensive scan of underwater targets. The placement of the targets and the scanning paths are illustrated in Figure 8. In the figure, the lateral distance between the targets is approximately 50 meters, and the longitudinal distance is approximately 100 meters. Due to the influence of underwater currents, some degree of deviation in this distance is inevitably present.

Due to the complex and variable underwater environment, as well as the susceptibility of images to noise interference, the images acquired using SSS also exhibit significant variations, as shown in Figure 9.

Discerning distinct target features within SSS images presents a formidable challenge. Manual annotation subsequent to data collection is arduous, making on-site, real-time labeling the optimal strategy. To attain the real-time processing of SSS images, we adopt a tactic wherein image segments are extracted from the sonar waterfall plot at intervals of 30 seconds, illustrated in Figure 10. This approach facilitates the annotation of targets on SSS images in real-time, while accounting for the field environment and GPS coordinates.

Furthermore, the targets occupy a minuscule proportion within the complete SSS image. Training the network directly with large-scale SSS images would generate an excessive number of negative samples, potentially impeding the training process and squandering computational resources. Moreover, considering practical applications, the network needs to be deployed on resource-

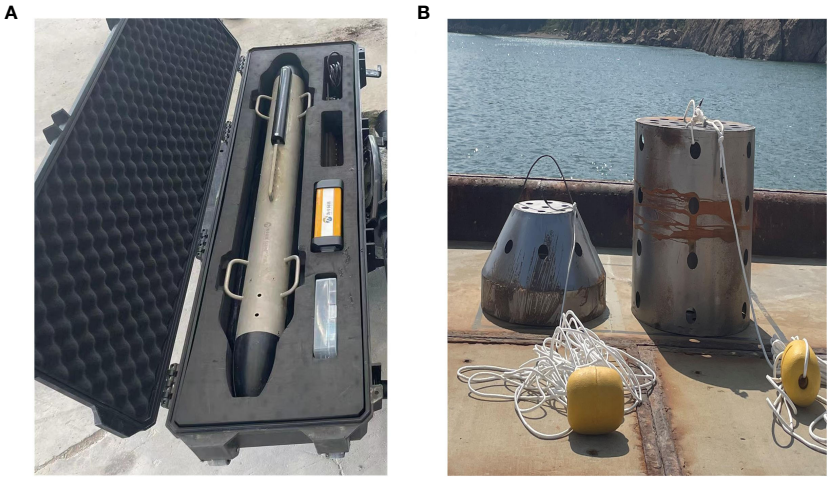


FIGURE 7
The SSS and preset targets. **(A)** SSS. **(B)** preset targets.

TABLE 3 Performance parameters of the SSS.

Frequency	300kHz	600kHz
Maximum range	150m	100m
Maximum slope distance	230m	200m
Horizontal beam width	0.5°	0.26°
Vertical beam width	50°	50°
Horizontal resolution	1.3m	0.45m
Vertical resolution	2.5m	1.25m

constrained underwater autonomous vehicles, making it imperative to restrict the image size fed into the detection network. To address these challenges, we partitioned the images into diminutive patches with dimensions of 200 × 200. Each patch features a 50-pixel overlap to prevent the loss of target characteristics. From these patches, we selectively identified those containing targets for training, significantly reducing the generation of irrelevant negative samples stemming from extraneous background information. Similarly, during the detection phase, we performed the same cropping operation before inputting the complete image into the detection network.

Finally, we filtered out unusable data and conducted data augmentation using high-quality data, yielding a total of 975 sample images. These images include 293 Cones, 318 Cylinders, and 364 Non-target instances. ("Non-target" refers to miscellaneous items on the seafloor, such as rocks or accidentally dropped artificial objects, which were not intentionally deployed by us. Despite not being the primary focus of the experiment, these Non-target items share certain similarities with the intentionally deployed targets. Including them in the dataset is essential, as their presence could potentially impact our ability to detect the deployed targets.) These samples were then randomly divided into training, validation, and test sets in a 7:1:2 ratio, with the specific number of samples for each set as shown in Table 4.

4.4 Experiment results

To validate the effectiveness of the algorithm proposed in this study for detecting small targets in SSS imagery, we tested the

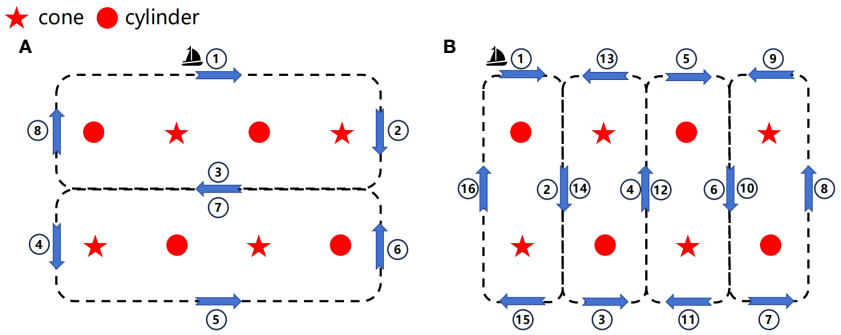


FIGURE 8
The target deployment locations and scanning paths. **(A)** Scanning path 1. **(B)** Scanning path 2.

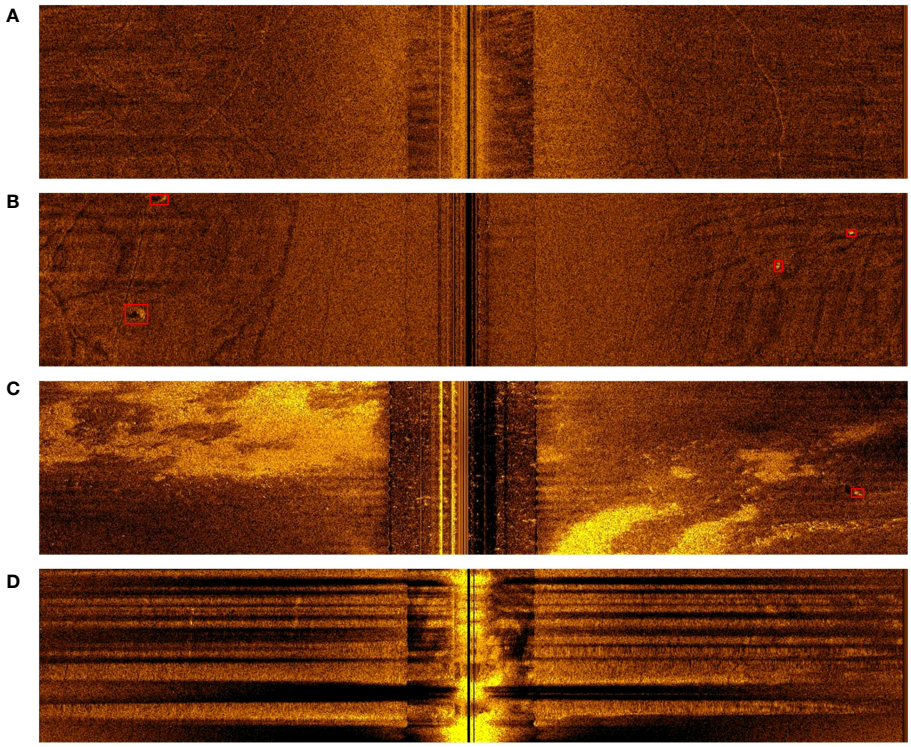


FIGURE 9
Acquired Sonar Images. **(A)** Background Images. **(B)** Images with Targets. **(C)** Target Images in Complex Environments. **(D)** Interfered Images.

algorithm on a real dataset collected during our sea trials. The variations in various loss functions and accuracy metrics during the training process are illustrated in Figure 11.

To ensure that all the introduced modifications exerted a positive influence on the network, a sequence of ablation experiments was carried out. The results of these experiments are presented in Table 5. In the table, *mAP*@0.5 represents the average precision at an IoU threshold of 0.5, while *mAP*@0.5: 0.95 represents the average of mAP values at IoU thresholds ranging from 0.5 to 0.95. It is apparent that the integration of k-means++, ODConv, GAM, and WIoU enhancements has resulted in an improved detection performance of the original YOLOv7 model on our assembled SSS dataset. The comparison of Precision-Recall (PR) curves on the test set between the improved YOLOv7 network and the original YOLOv7 network is shown in Figure 12, while the comparison of confusion matrices is shown in Figure 13. From Figure 12, it can be observed that the improved YOLOv7 network achieved an average precision improvement of 5.05% on the test set.

From Figures 12, 13, it can be observed that the improved YOLOv7 network demonstrates a noticeable enhancement in the detection performance of Non-target objects. In Figure 12, the PR curve of the enhanced YOLOv7 network for the Non-target

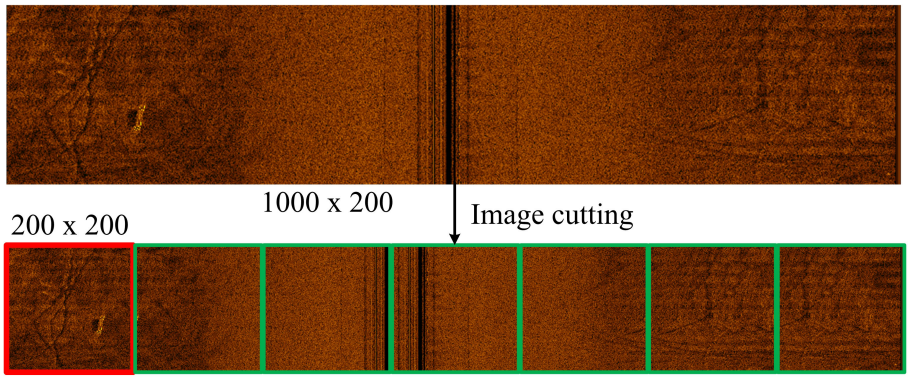


FIGURE 10
Preprocessing of SSS images. We partitioned the images into diminutive patches with dimensions of 200×200. Each patch features a 50-pixel overlap to prevent the loss of target characteristics.

TABLE 4 The actual dimensions of underwater targets and the final dataset sample size.

Category	Target			Dataset			
	Diameter	Height	Number	Train	Val	Test	Total
Cone	0.30m/0.50m	0.60m	4	205	29	59	293
Cylinder	0.50m	1.00m	4	223	31	64	318
Non-target	/	/	/	255	36	73	364

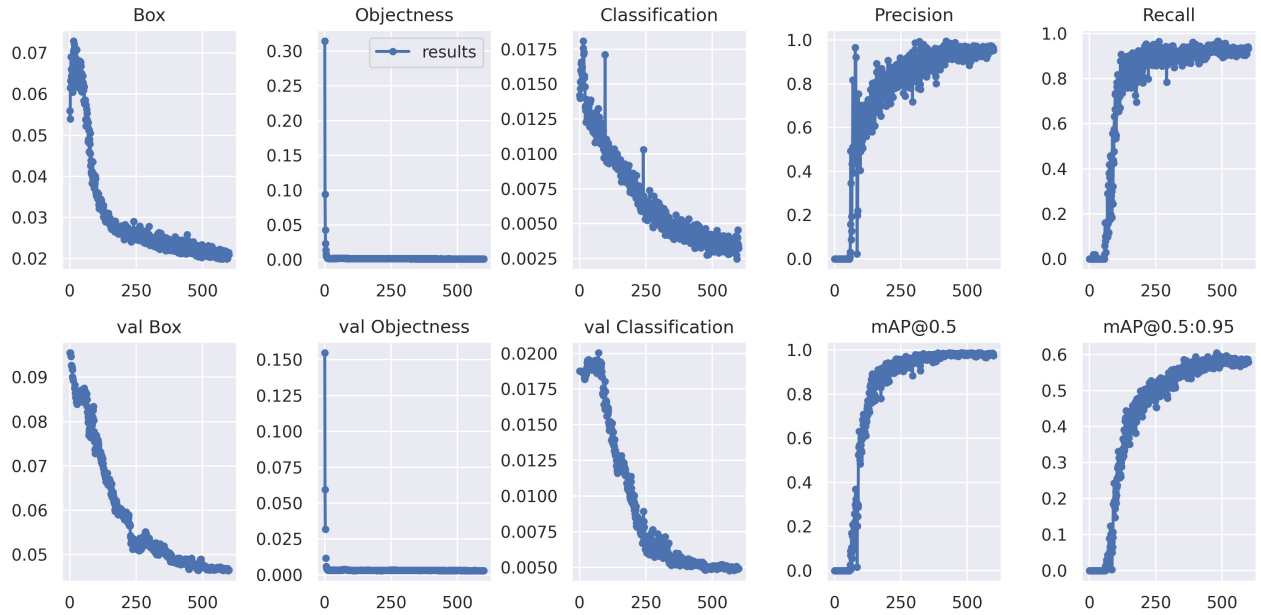


FIGURE 11
The loss function and relevant metrics during the training process of the improved YOLOv7. The horizontal axis in the figure represents the number of training epochs.

category shows a value of 0.909, which represents an improvement of 0.095 compared to the original network's 0.814. In Figure 13, within the improved YOLOv7's confusion matrix, the Non-target category registers a value of 0.94, as opposed to the original network's 0.92, marking a 0.02 improvement.

In addition, our experimental results provide evidence of the enhanced network's superior performance in detecting Non-target objects, as depicted in Figure 14. The original YOLOv7 network misclassified Non-target objects as Cylinder and Cone, whereas the improved YOLOv7 network can accurately identify Non-target categories. This advancement has reduced the false detection rate for Non-target, which holds significant practical significance in engineering applications. During the search process, it prevents wasting time on Non-target objects.

Furthermore, a comparative analysis was conducted between our enhanced detection algorithm and prominent detection networks to validate the efficacy of the proposed methodology. The comparative visualization of detection outcomes is illustrated in Figure 15. Detailed detection metrics are presented in Table 6. These findings collectively furnish compelling evidence for the superior performance of the approach proposed in this paper within the domain of small target detection using SSS.

TABLE 5 Ablation experiment.

Model	K-means++	ODConv	GAM	WIoU	*mAP*@0.5(%)	*mAP*@0.5: 0.95(%)
	×	×	×	×	90.73	49.78
	✓	×	×	×	91.77(1.04↑)	50.39(0.61↑)
YOLOv7	✓	✓	×	×	93.28(2.55↑)	51.17(1.39↑)
	✓	✓	✓	×	94.49(3.76↑)	51.79(2.01↑)
	✓	✓	✓	✓	95.78(5.05↑)	52.29(2.51↑)

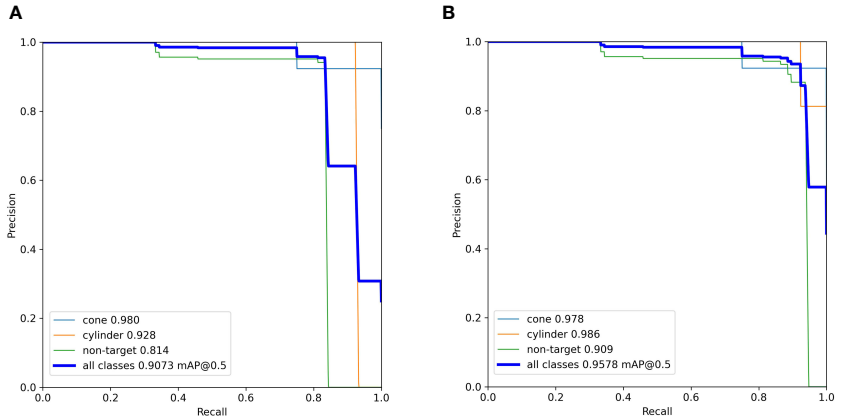


FIGURE 12
The PR curve on the test set. **(A)** initial YOLOv7 network. **(B)** improved YOLOv7 network.

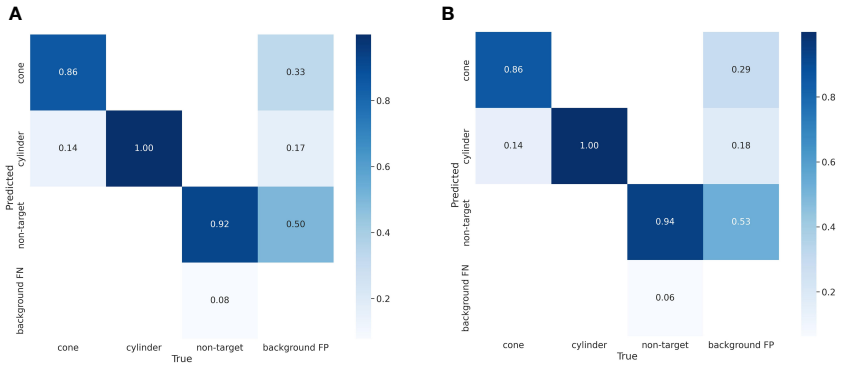


FIGURE 13
The Confusion Matrix on the test set. **(A)** initial YOLOv7 network. **(B)** improved YOLOv7 network.

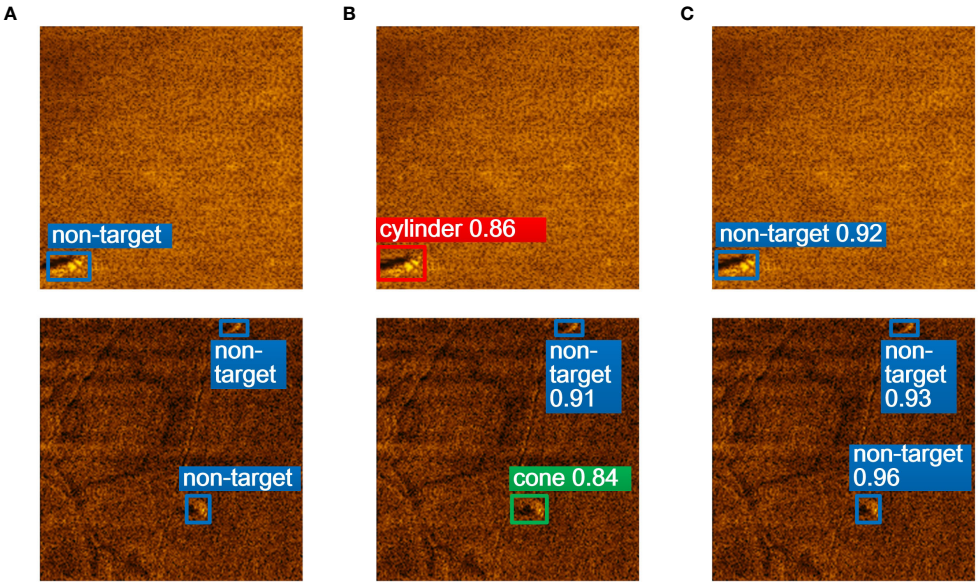


FIGURE 14
Comparison of Non-target category detection results between the improved YOLOv7 and the original YOLOv7 networks. **(A)** Labels. **(B)** Initial YOLOv7 network. **(C)** Improved YOLOv7 network.

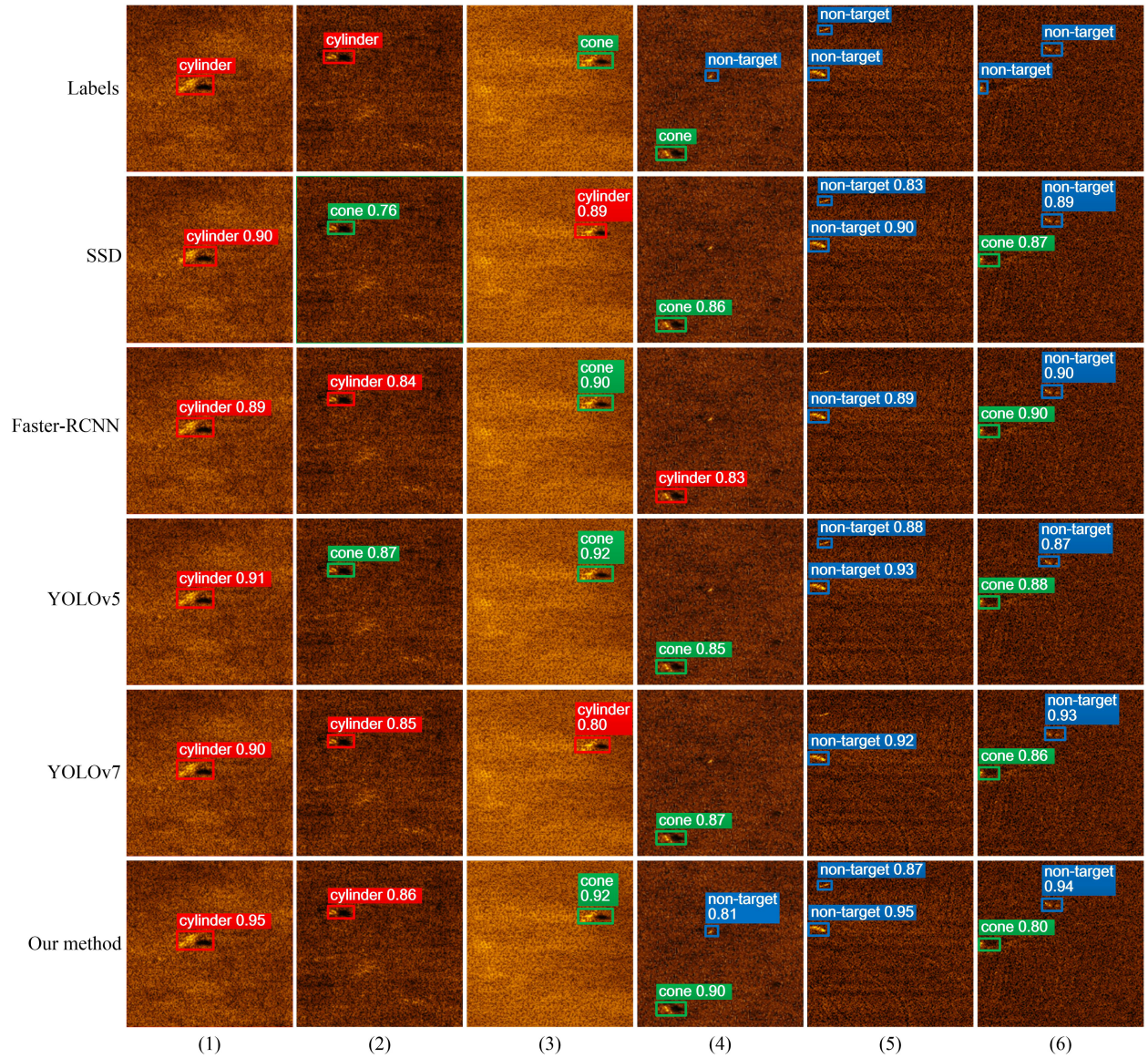


FIGURE 15
Comparison of detection results between our method and other detection networks. The first row in the figure represents the ground truth labels for different target categories, while the second to fifth rows depict the detection results of various algorithms.

The results illustrated in Figure 15 provide empirical validation of the efficacy of the approach introduced in this research. As demonstrated in columns (2), (3), and (4) of Figure 15, some mainstream detection networks often exhibit mis-detections when accurately distinguishing between the categories of cylindrical and conical objects. In contrast, the proposed method in this paper demonstrates accurate detection for objects that are challenging to differentiate, with higher

TABLE 6 Comparison of detection metrics between our method and other detection networks.

Method	Precision(%)	Recall(%)	*mAP*@0.5(%)	*mAP*@0.5: 0.95(%)
SSD	88.31	89.76	89.28	48.24
Faster-RCNN	85.33	83.91	87.19	46.73
YOLOv5	88.72	90.46	89.98	49.80
YOLOv7	93.56	89.12	90.73	49.78
Our method	92.99	89.10	95.78	52.29

probability values assigned. This highlights the superiority of the algorithm presented in this paper.

Nonetheless, it is important to note that the enhanced network in this study does exhibit certain limitations. For example, as depicted in column (6) of Figure 15, all networks misclassify a Non-target as a Cone. This misclassification arises due to the distinct shadow surrounding the Non-target and the similarity in the size of the bright spot to the Cone category, resulting in a false positive detection. At present, there is a lack of definitive solutions for scenarios in which acoustic image features exhibit extremely high similarity, yet the actual objects belong to different categories. Using a higher-precision device to acquire images with increased resolution may be beneficial for addressing this issue.

5 Conclusions

This study collected a dataset of small target SSS images during sea trials and proposed an enhancement method based on the YOLOv7 model for detecting small targets in SSS images. The method utilizes the k-means++ algorithm to obtain more accurate initial anchor box sizes. Subsequently, it employs ODConv to replace static convolution modules in the YOLOv7 backbone network and integrates a GAM attention mechanism into the YOLOv7 neck network, thereby enhancing the feature extraction capabilities of the detection network. In the loss function section, a WIoU loss function is introduced to balance the impact of high-quality and low-quality anchor boxes on gradients, enhancing the network's focus on average-quality anchor boxes. Experimental results demonstrate the effectiveness of the proposed YOLOv7-based enhancement algorithm, with *mAP*@0.5 and *mAP*@0.5: 0.95 metrics reaching 95.78% and 52.29%, respectively, representing improvements of 5.05% and 2.51% over the original YOLOv7 network. Furthermore, comparisons with mainstream underwater detection networks confirm the superiority of the proposed method in small target detection in SSS images.

The proposed method can be applied to autonomous target detection in Unmanned Underwater Vehicles (UUVs) and Unmanned Surface Vehicles (USVs), enhancing the autonomous operational capabilities of unmanned autonomous ocean observation platforms. In the future, we plan to collect more diverse small target data and continue researching SSS-based small target detection methods to further contribute to underwater exploration.

Data availability statement

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

Author contributions

CC: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. CW: Methodology, Writing – review & editing. DY: Software, Writing – review & editing. XW: Software, Writing – review & editing. WL: Project administration, Writing – review & editing. FZ: Conceptualization, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Key Research and Development Program (2023YFC2808400).

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Arthur, D., and Vassilvitskii, S. (2007). "K-means++ the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, SODA 2007*. (New Orleans, Louisiana, USA: ACM), 1027–1035. doi: 10.1145/1283383.1283494

Bhattacharya, S., Maddikunta, P. K. R., Pham, Q.-V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., et al. (2021). Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey. *Sustain. cities Soc.* 65, 102589. doi: 10.1016/j.scs.2020.102589

Chen, T., Wang, N., Chen, Y., Kong, X., Lin, Y., Zhao, H., et al. (2023). Semantic attention and relative scene depth-guided network for underwater image enhancement. *Eng. Appl. Artif. Intell.* 123, 106532. doi: 10.1016/j.engappai.2023.106532

Chen, Z., Wang, H., Shen, J., and Dong, X. (2014). "Underwater object detection by combining the spectral residual and three-frame algorithm," in *Lecture Notes in Electrical Engineering* (Berlin, Germany: Springer). 279, 1109–1114. doi: 10.1007/978-3-642-41674-3_154

Chen, T., Wang, N., Wang, R., Zhao, H., and Zhang, G. (2021). One-stage cnn detector-based benthonic organisms detection with limited training dataset. *Neural Networks* 144, 247–259. doi: 10.1016/j.neunet.2021.08.014

Fan, Z., Xia, W., Liu, X., and Li, H. (2021). Detection and segmentation of underwater objects from forward-looking sonar based on a modified mask rcnn. *Signal Image Video Process.* 15, 1135–1143. doi: 10.1007/s11760-020-01841-x

Hożyń, S. (2021). A review of underwater mine detection and classification in sonar imagery. *Electronics* 10, 2943. doi: 10.3390/electronics10232943

Hou, F., Lei, W., Li, S., and Xi, J. (2021). Deep learning-based subsurface target detection from gpr scans. *IEEE sensors J.* 21, 8161–8171. doi: 10.1109/JSEN.2021.3050262

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-Excitation Networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Salt Lake City, UT, USA: IEEE). pp. 7132–7141. doi: 10.1109/CVPR.2018.00745

Jia, Y., Wang, H., Chen, W., Wang, Y., and Yang, B. (2022). An attention-based cascade r-cnn model for sternum fracture detection in x-ray images. *CAAI Trans. Intell. Technol.* 7, 658–670. doi: 10.1049/cit2.12072

Jin, L., Liang, H., and Yang, C. (2019). Accurate underwater atr in forward-looking sonar imagery using deep convolutional neural networks. *IEEE Access* 7, 125522–125531. doi: 10.1109/ACCESS.2019.2939005

Kim, W.-K., Bae, H. S., Son, S.-U., and Park, J.-S. (2022). Neural network-based underwater object detection off the coast of the korean peninsula. *J. Mar. Sci. Eng.* 10, 1436. doi: 10.3390/jmse10101436

Le, H. T., Phung, S. L., Chapple, P. B., Bouzerdoum, A., Ritz, C. H., and Tran, L. C. (2020). Deep gabor neural network for automatic detection of mine-like objects in sonar imagery. *IEEE Access* 8, 94126–94139. doi: 10.1109/ACCESS.2020.2995390

Lee, S., Park, B., and Kim, A. (2018). Deep learning from shallow dives: Sonar image generation and training for underwater object detection. *arXiv preprint arXiv:1810.07990*. doi: 10.48550/arXiv.1810.07990

Li, J., Chen, L., Shen, J., Xiao, X., Liu, X., Sun, X., et al. (2023b). Improved neural network with spatial pyramid pooling and online datasets preprocessing for underwater target detection based on side scan sonar imagery. *Remote Sens.* 15, 440. doi: 10.3390/rs15020440

Li, L., Li, Y., Yue, C., Xu, G., Wang, H., and Feng, X. (2023c). Real-time underwater target detection for auv using side scan sonar images based on deep learning. *Appl. Ocean Res.* 138, 103630. doi: 10.1016/j.apor.2023.103630

Li, W., Wang, J., Zhao, X. F., Wang, Z., and Zhang, Q. J. (2021). "Target detection in color sonar image based on yolov5 network," in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. (Xi'an, China: IEEE), 1–5.

Li, C., Ye, X., Xi, J., and Jia, Y. (2023a). A texture feature removal network for sonar image classification and detection. *Remote Sens.* 15, 616. doi: 10.3390/rs15030616

Li, C., Zhou, A., and Yao, A. (2022). Omni-dimensional dynamic convolution. *arXiv preprint arXiv:2209.07947*. doi: 10.48550/arXiv.2209.07947

Liu, Y., Shao, Z., and Hoffmann, N. (2021). Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv preprint arXiv:2112.05561*. doi: 10.48550/arXiv.2112.05561

Mukherjee, K., Gupta, S., Ray, A., and Phoha, S. (2011). Symbolic analysis of sonar data for underwater target detection. *IEEE J. Oceanic Eng.* 36, 219–230. doi: 10.1109/JOE.2011.2122590

Neupane, D., and Seok, J. (2020). A review on deep learning-based approaches for automatic sonar target recognition. *Electronics* 9, 1972. doi: 10.3390/electronics9111972

Park, J., Woo, S., Lee, J.-Y., and Kweon, I. S. (2018). Bam: Bottleneck attention module. *arXiv preprint arXiv:1807.06514*. doi: 10.48550/arXiv.1807.06514

Raghuvanshi, D. S., Dutta, I., and Vaidya, R. (2014). "Design and analysis of a novel sonar-based obstacleavoidance system for the visually impaired and unmanned systems," in *2014 International Conference on Embedded Systems (ICES)*. (Coimbatore, India: IEEE), 238–243.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Long Beach, CA, USA: IEEE), 658–666.

Singh, D., and Valdenegro-Toro, M. (2021). "The marine debris dataset for forward-looking sonar semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (Montreal, BC, Canada: IEEE), 3741–3749.

Soeb, M. J. A., Jubayer, M. F., Tarin, T. A., Al Mamun, M. R., Ruhad, F. M., Parven, A., et al. (2023). Tea leaf disease detection and identification based on yolov7 (yolo-t). *Sci. Rep.* 13, 6078. doi: 10.1038/s41598-023-33270-4

Tang, Y., Wang, L., Jin, S., Zhao, J., Huang, C., and Yu, Y. (2023). Auv-based side-scan sonar real-time method for underwater-target detection. *J. Mar. Sci. Eng.* 11, 690. doi: 10.3390/jmse11040690

Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2023). Wise-iou: Bounding box regression loss with dynamic focusing mechanism. *arXiv preprint arXiv:2301.10051*. doi: 10.48550/arXiv.2301.10051

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023a). "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Vancouver, BC, Canada: IEEE), 7464–7475.

Wang, N., Chen, T., Kong, X., Chen, Y., Wang, R., Gong, Y., et al. (2023b). Underwater attentional generative adversarial networks for image enhancement. *IEEE Trans. Human-Machine Syst.* 53 (3), 490-500. doi: 10.1109/THMS.2023.3261341

Wang, N., Chen, T., Liu, S., Wang, R., Karimi, H. R., and Lin, Y. (2023c). Deep learning-based visual detection of marine organisms: A survey. *Neurocomputing* 532, 1–32. doi: 10.1016/j.neucom.2023.02.018

Wang, J., Feng, C., Wang, L., Li, G., and He, B. (2022). Detection of weak and small targets in forward-looking sonar image using multi-branch shuttle neural network. *IEEE Sensors J.* 22, 6772–6783. doi: 10.1109/JSEN.2022.3147234

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Seattle, WA, USA: IEEE), 11534–11542.

Wang, Z., Zhang, S., Huang, W., Guo, J., and Zeng, L. (2021). Sonar image target detection based on adaptive global feature enhancement network. *IEEE Sensors J.* 22, 1509–1530. doi: 10.1109/JSEN.2021.3131645

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision. (ECCV)* (Munich Germany: Springer), 3–19.

Xiao, T., Cai, Z., Lin, C., and Chen, Q. (2021). A shadow capture deep neural network for underwater forward-looking sonar image detection. *Mobile Inf. Syst.* 2021, 1–10. doi: 10.1155/2021/3168464

Yang, D., Cui, Y., Yu, Z., and Yuan, H. (2021). Deep learning based steel pipe weld defect detection. *Appl. Artif. Intell.* 35, 1237–1249. doi: 10.1080/08839514.2021.1975391

Yang, H., Liu, Y., Wang, S., Qu, H., Li, N., Wu, J., et al. (2023). Improved apple fruit target recognition method based on yolov7 model. *Agriculture* 13, 1278. doi: 10.3390/agriculture13071278

Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*. (Amsterdam, The Netherlands: ACM), 516–520.

Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022b). Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi: 10.1016/j.neucom.2022.07.042

Zhang, P., Tang, J., Zhong, H., Ning, M., Liu, D., and Wu, K. (2021a). Self-trained target detection of radar and sonar images using automatic deep learning. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2021.3096011

Zhang, P., Tang, J., Zhong, H., Ning, M., Liu, D., and Wu, K. (2021b). Self-trained target detection of radar and sonar images using automatic deep learning. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2021.3096011

Zhang, H., Tian, M., Shao, G., Cheng, J., and Liu, J. (2022a). Target detection of forward-looking sonar image based on improved yolov5. *IEEE Access* 10, 18023–18034. doi: 10.1109/ACCESS.2022.3150339

Zhao, W., Chen, F., Huang, H., Li, D., and Cheng, W. (2021). A new steel defect detection algorithm based on deep learning. *Comput. Intell. Neurosci.* 2021, 1–13. doi: 10.1155/2021/5592878

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020). "Distance-iou loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*. (New York, USA: AAAI), Vol. 3, 12993–13000.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 27 March 2024
DOI 10.3389/fmars.2024.1388617

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Maohan Liang,
National University of Singapore, Singapore
Xin Peng,
Tsinghua University, China

*CORRESPONDENCE
Anmin Zhang
zhanganmin@sina.com

RECEIVED 20 February 2024
ACCEPTED 18 March 2024
PUBLISHED 27 March 2024

CITATION
Kang Z, Gao M, Liao Z and Zhang A (2024) Collaborative communication-based ocean observation research with heterogeneous unmanned surface vessels.
Front. Mar. Sci. 11:1388617.
doi: 10.3389/fmars.2024.1388617

Collaborative communication-based ocean observation research with heterogeneous unmanned surface vessels

Zhen Kang[1,2], Miao Gao[1,2], Zihao Liao[1] and Anmin Zhang[1,2]*

[1]Tianjin University, School Marine Science and Technology, Tianjin, China, [2]Key Laboratory of Ocean Observation Technology (KLOOT), Ministry of Natural Resources (MNR), Tianjin, China

Unmanned surface vehicles (USVs) are crucial in ensuring maritime safety and observation, attracting widespread attention and research. However, a single USV exhibits limited performance and cannot effectively observe complex marine environments. In contrast, clusters of USVs can collaborate to execute complex maritime tasks, thereby enhancing the overall operational efficiency. USVs typically form heterogeneous clusters by combining vehicles with varying maneuverabilities and communication network capabilities. This has sparked an increased interest in cooperative communication research within heterogeneous USV clusters. The heterogeneous USVs discussed in this paper share the same dynamic model; however, they differ in dynamic parameters and communication capabilities. First, this study establishes a three-degree-of-freedom motion mathematical model for an underdriven USV considering environmental interference. It estimates the dynamic parameters of four USVs and evaluates their communication abilities, laying the foundation for researching the cooperative control of heterogeneous USV clusters and their application in Ocean Observation. Next, the communication capability of the USVs is assessed by studying the communication mode and signal transmission loss within the USV clusters. This study investigates the problem of cooperative communication in USV cluster formation, starting with the communication delay of USV clusters under a directed switching topology. Finally, a coherent formation controller is designed under a switching communication topology to address the dynamic transformation of communication topologies within heterogeneous USV clusters. This verifies that heterogeneous USV clusters can seamlessly form and maintain formation shapes during communication topology transformations through formation simulation experiments involving four heterogeneous USVs was 22% higher than that of dispersed control topology structures. This study provides a solid foundation for future investigations into the cooperative control of heterogeneous USV clusters and their applications in marine observations.

KEYWORDS

heterogeneous, USV, collaborative communication, ocean observation, topology optimization

1 Introduction

Unmanned surface vehicles (USVs) are pivotal tools for ocean exploration and are currently garnering global attention. In December 2018, the International Maritime Organization (IMO) Maritime Safety Committee approved a framework and methodology for a regulatory scoping exercise for autonomous surface vessels at sea during its 100th session (Schröder-Hinrichs et al., 2019). Owing to its capability to conduct autonomous navigation at sea, USVs offer advantages such as high autonomy, broad operating range, effective concealment, and substantial reductions in labour and maintenance costs. Equipped with various technologies, USVs can autonomously execute a diverse range of maritime tasks, significantly enhancing task execution efficiency (Zhang et al., 2019). Consequently, USVs are considered crucial platforms for future maritime operations. In addition, the technological advancements in this field are significant for the exploration, development, and protection of oceans.

When a USV performs autonomous navigation and operation at sea, the motion is characterised by typical underdrive, non-linearity, uncertainty, multiple constraints, unpredictable state, and limited communication (Gu et al., 2019). Furthermore, the superposition of multiple factors makes it difficult to control the USV, and the operating environment at sea is complex, with relatively severe perturbations of wind, waves, and currents, limited communication, and relatively complicated task allocation problems. The cooperative control of USV clusters under communication constraints also faces significant challenges (Wang et al., 2020).

USV clusters are connected via the communication network to form a whole, and the communication mode determines the control mode between clusters to a certain extent. The interaction of USV information can be divided into three types (Xie et al., 2021): centralised, decentralised, and distributed communications.

Centralised communication (Sun et al., 2022) is a multi-intelligence body cluster with a control centre that designs the global control protocol and communicates with all other intelligences to exchange information. Thus, the performance of centralised communication has a natural superiority, is easy to implement, and is currently the most commonly used communication method in cluster research. However, centralised communication has a high computational cost and poor robustness, is not easily scalable, and has poor responsiveness to environmental changes.

Decentralised communication (Chainho et al., 2017) means that each intelligence has a controller that can communicate directly with other intelligences, and each intelligence autonomously processes information and makes plans and decisions. This depends on local information, which reduces the computational burden and complexity; however, the lack of communication and cooperation between intelligences leads to a lower synergy efficiency and does not ensure the realisation of global goals.

Distributed communication (Lim et al., 2008) differs from centralised and decentralised communication and is a type of communication between the two. Intelligent bodies and adjacent intelligent bodies in distributed communication can realise mutual communication and coordinated action, and the information of intelligent bodies does not need to pass through the control centre. This apportions the communication pressure of the entire system to each intelligent body, improves the coordination of multi-intelligent body communication systems, and has the advantages of robustness, flexibility, and ease of expansion.

For a USV cluster adopting distributed communication, each USV interacts with the neighbouring USV with information to ensure that the state quantity or a certain variable of all USVs is eventually consistent. Moreover, it can effectively improve the efficiency of the system synergy and the robustness of the system under the premise of realising a global goal. This paper presents a method for assessing the communication capabilities of heterogeneous USVs. By comparing three communication methods among USVs clusters, stable and real-time radio communication has been chosen as the mode for information exchange between USVs clusters. A method for evaluating the communication capabilities of heterogeneous USVs based on a signal transmission loss model is proposed. The communication topology of USVs is abstracted into a graph data structure, and several important properties of Laplace matrices are provided. Based on the observation progress consistency controller, the speeds of heterogeneous USVs are adjusted through the optimization of communication topology to achieve consistent observation progress among all vessels. Finally, compared with dispersed control experiments, the optimization of communication topology based on the consistency controller has improved observation efficiency by 22%.

2 Mathematical modelling of heterogeneous USV motion

To study the motion simulation and cooperative communication problems of a USV, it is necessary to obtain its dynamic parameters. Currently, the kinetic parameters of a USV are identified through experiments, theoretical calculations, and approximate projections (Wu, 2011). Owing to the complexity and variability of the structure and working conditions of a USV (Xing, 2012), theoretical calculations and experimental measurements have certain errors and limitations. Therefore, this study adopts an integrated method to estimate the dynamic parameters of a USV using a combination of theoretical calculations and experiments. The method is based on the known parameters of the USV for theoretical calculation, and the principles of the theoretical calculation are as follows:

(1) Additional inertial mass term, $m_1 \approx 1.05m$, where m is the actual mass of the USV.

(2) Additional inertial mass term, $m_2 \approx m + 0.5(\rho\pi D^2 L)$, where ρ is the density of water, D is the average depth of immersion of the ship, and L is the effective length of the ship.

(3) Additional inertial mass term $m_3 \approx mL^2/8$. This formula is applicable to paddle rudder and dual-thruster USVs, $m_3 \approx (m(L^2 + W^2) + 0.5(0.1mB^2 + \rho\pi D^2 L^3))/12$ Fig, where W is the actual width of the ship and B is the distance between the two thrusters (Lim et al., 2008).

(4) The hydrodynamic damping term is $d_1 \approx m_1 g/(2u_0)$, where g is the gravitational acceleration and u_0 is the economical velocity of the USV.

(5) Hydrodynamic damping term $d_2 \approx -0.5\rho u_0 L^2 Y'v$, where $Y'v = -5(D/L)^2$.

(6) Hydrodynamic damping term $d_3 \approx 0.5\rho V L^4 N'r$, where $N'r = -0.65(D/L)^2$.

These kinetic parameters were partially calculated using Smit's formula. First, according to the above theoretical calculation, the calculation results are then substituted into the standard USV motion model to obtain a set of preliminary dynamic parameters Ship slewing simulation experiments are carried out on the computer, and the simulation results are compared and analysed with the measured data of the actual slewing test on the water surface to determine the differences between the two and the reasons for the differences. Moreover, according to the results of the analysis, the dynamic parameters are adjusted and optimised appropriately to ensure they are more aligned with the real situation. Finally, based on the analysis results, the dynamic parameters were adjusted and optimised to make them more consistent with the actual situation.

In the following, real ship data verify the abovementioned kinetic parameter identification method, and four USVs, named USV I-IV, developed by a 3I team personnel of the School of Marine Science and Technology of Tianjin University, are considered. The physical photos are shown in Figure 1.

A comprehensive approach combining theoretical calculations and experiments, as described above, was used to confirm the kinetic parameters of the USV; Table 1 lists the parameters of each USV and the kinetic parameters determined.

The accuracy of the established motion model and parameters of the USV were verified, and a constant-slewing motion simulation was performed using the USV dynamic parameters listed in Table 1. The USV sailed in the due north direction at an economical velocity with the paddle or vector thrusters turned to the right by 35° (USV III was turned off by controlling the right thrusters).

In static waters without environmental interference, the motion trajectories depicted by the centres of gravity of the four USVs are shown in Figure 2A. It is shown that the different USVs exhibit different longitudinal and transverse distances as well as different diameters of rotary gyration, reflecting the different manoeuvrability of each USV. In the presence of interference from sea winds and currents, the wind velocity $U_T = 2\ m/s$, wind direction $\psi_T = 0°$, current velocity $V_c = 0.15\ m/s$, and current direction $\psi_c = \pi/4$ were set, and the motion trajectories depicted by the centres of gravity of the four USVs are shown in Figure 2B. The experimental results show that the motion of each USV tends to drift in the direction of $\pi/4$, which indicates that the sea current has a significant influence on the USV motion state.

The simulation experiments show that the underdriven three-degree-of-freedom motion mathematical model established in this study can correctly describe the navigation of the USV and demonstrate the manoeuvrability of different USVs considering environmental perturbations.

3 USV cluster cooperative communications

3.1 Communication methods for USV clusters

The maritime Internet of Things (IoT) has recently emerged as a revolutionary communication paradigm where a large number of

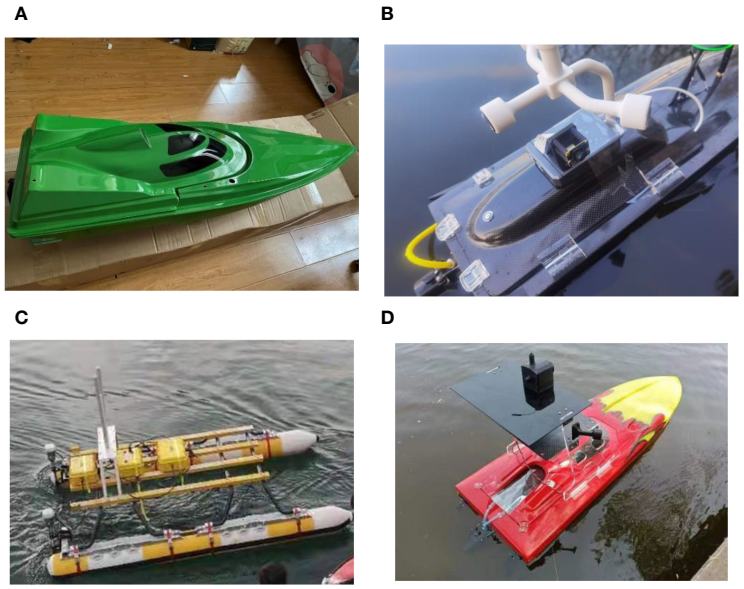


FIGURE 1
Physical drawing of four USV. **(A)** USV I, **(B)** USV II, **(C)** USV III, **(D)** USV VI.

TABLE 1 Physical and kinetic parameters of four USV.

Parameter	BISHENG(USV I)	YEYING (USV II)	DOLPHIN-1(USV III)	LIEYAN (USV IV)
length/m	1.23	0.75	3.20	1.37
width/m	0.38	0.25	2.20	0.38
mass/kg	12.4	16.6	90.0	10.1
Mode of advancement	oar and rudder	vector thruster	Dual Thrusters	vector thruster
m1/kg	13.02	17.43	94.5	10.61
m2/kg	27.28	20.95	256.93	13.63
m3/kg·m2	2.35	2.33	57.6	2.37
d1/kg·s-1	17.72	5.5	163.67	11.88
d2/kg·s-1	23.07	24.3	234.89	17.93
d3/kg·s-1	4.54	2.78	312.69	4.37
Economic velocity/m·s-1	3.6	4.1	2.8	4.4

moving vessels are closely interconnected in intelligent maritime networks. To promote smart traffic services in maritime IoT, it is necessary to accurately and robustly predict the spatiotemporal vessel trajectories. It is beneficial for collision avoidance, maritime surveillance, and abnormal behaviour detection, etc. Motivated by the strong learning capacity of deep neural networks, Ryan Wen Liu et al. (Liu et al., 2022; Liang et al., 2022; Liang et al., 2024) proposed an AIS data-driven trajectory prediction framework, whose main component is a long short-term memory (LSTM) network.

Maritime communication networks primarily include radio-, satellite-, and shore-based cellular network communication systems. All of these communication systems can be applied to USV-trunking communication. Figure 3 shows the widely used maritime communication networks.

Radio communication systems (Xia et al., 2017) are widely used in marine communications and can provide near-, medium-, and long-range communication. Typical frequencies are medium frequency (MF), high frequency (HF), and very high frequency (VHF). USVs can communicate with the surrounding USVs and are generally suitable for short-range communication. The radio communication system has a stable signal, low cost, and good real-time performance; however, with increased communication distance, the data transmission rate decreases, and it cannot realise global communication coverage.

Satellite communication systems (Xia et al., 2017) play an irreplaceable role in marine communication and provide global coverage. Typical examples are the maritime satellite system (INMARSAT), the Iridium system (Iridium), and China's Beidou satellite navigation system, which is one of the most common and reliable choices for USV communication. Moreover, USVs can communicate with ground stations or other USVs over long distances, thereby enabling remote control, data transmission, and other functions. Satellite communication systems cover a wide range of applications; however, they exhibit extremely high costs, signal delays, and information security problems.

Shore-based cellular network communication systems (Xia et al., 2017) suit offshore marine communication. USVs can use mobile communication technologies, such as 4G and 5G, to provide

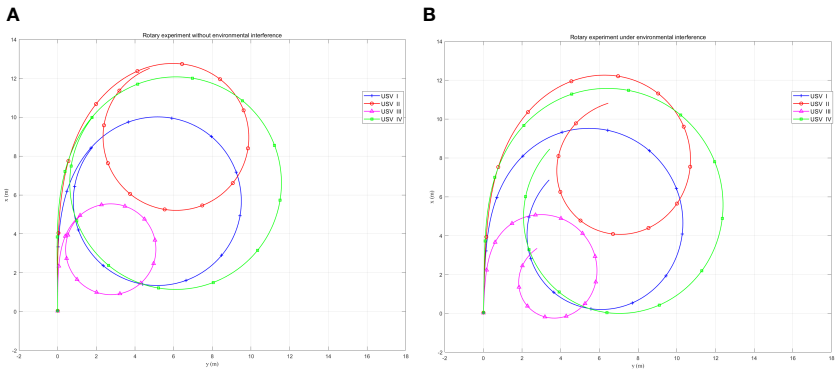


FIGURE 2
Steady rotation Experiment for four USVs. **(A)** Without environmental interference, **(B)** With environmental interference.

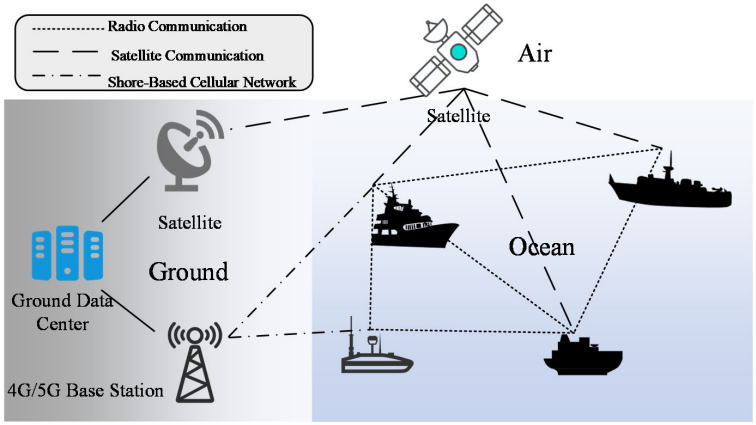


FIGURE 3
Schematic diagram of the maritime communications network.

high-velocity and stable offshore network services. However, the range of coastal sea areas covered by shore-based cellular networks is minimal.

As shown in Table 2, Radio communication systems (Alqurashi et al., 2023) have been widely used in the field of maritime communication, offering stable signals, low costs, and good real-time capabilities for near, medium, and short-distance communication. Shore-based cellular network communication systems are an excellent choice for near-sea maritime communication, while unmanned vessels can also utilize mobile communication technologies such as 4G and 5G to provide high-speed and stable network services in coastal areas. However, the coverage range of shore-based cellular networks in maritime areas is limited. Satellite communication systems hold an irreplaceable position in maritime communication, offering global communication coverage. However, they currently suffer from high costs, significant signal latency, and potential issues related to information security. In most cases, heterogeneous USV clusters require continuous information interaction. Owing to the shortcomings of satellite communication, such as signal delay and high cost, and the limitations of cellular networks, radio communication has become the optimal communication mode for real-time control and cooperative operation of USV clusters. Stable radio communication with good real-time performance can satisfy cluster information interaction requirements.

TABLE 2 Comparison of the characteristics of USV cluster communication methods.

Method	Advantages	Disadvantages
Radio communications	Good real-time, low-cost, and stable signal	Limited by distance, weather, and other factors
Satellite communications	Wide coverage and stable signal	Signal delays and high costs
Shore-based cellular network communications	Stable signal and high bandwidth	Limited to near-shore communications

3.2 Assessment of USV cluster communications capability

To ensure that the unmanned crafts within a cluster system communicate effectively with each other, an assessment of the communication capabilities of the unmanned craft is required to determine the communication connectivity of the cluster. The USV cluster is assumed to communicate via the HF/VHF frequency digital transmission radio, and the signals are transmitted along the line-of-sight channel. As shown in Figure 4, according to the radio transmission theory, the channel loss experienced when signals are transmitted at sea mainly consists of large-scale and small-scale fading, with large-scale fading including path transmission loss, shadow fading, and small-scale fading, including multipath fading. To confirm whether the two vessels can communicate, the performance of the communication systems of the two vessels, as well as the loss during signal transmission, can be evaluated.

First, a signal transmission loss model is established, assuming that two USVs, A and B, in the cluster interact with each other for information. In addition, the distance between the two USVs is d (km), the frequency of the signal is f (Mhz), and the loss of the signal

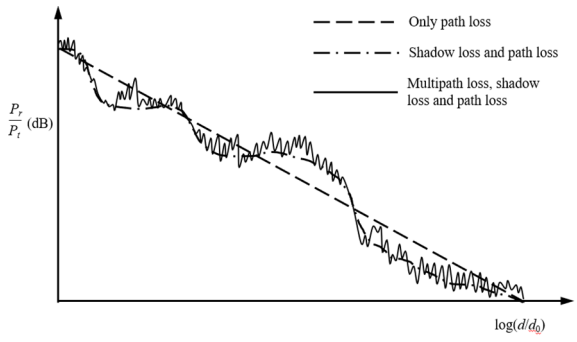


FIGURE 4
Maritime communications propagation loss.

in the propagation process is L_p (dB). According to the study (Jin, 2013), the formula is calculated as follows Equation 1:

$$L_p = 20\,log_{10}(f) + 20\,log_{10}(d) + 32.45 + 10\kappa\,log_{10}\left(\frac{d}{d_0}\right) \quad (1)$$

where κ is the path loss exponent and the specific value is determined by the communication environment factors at sea. Equation 1 considers marine environment correction compared to the signal-free space transmission model.

Second, the signal reception strength is calculated using the performance parameters of the communication systems of the two USVs and the signal transmission loss. The formula for calculating the signal reception strength of A and B USVs is as follows Equation 2:

$$RSS_A = P_{t,B} + G_{t,B} + G_{r,A} - L_p$$

$$RSS_B = P_{t,A} + G_{t,A} + G_{r,B} - L_p \quad (2)$$

RSS_A is the signal strength transmitted by USV B and received by USV A, RSS_B is the signal strength transmitted by USV A and received by USV B, $P_{t,A}$ and $P_{t,B}$ are the signal transmitting powers of USVs A and B, respectively. $G_{t,A}$ and $G_{t,B}$ denote the transmitting antenna gains of USVs A and B, respectively. $G_{r,A}$ and $G_{r,B}$ are the receiver antenna gains for USVs A and B, respectively.

Finally, the heterogeneous USVs cluster communication capability is evaluated using the wireless signal strength to measure communication effectiveness between two USVs. It is known that the reception sensitivity of USV A is $R_{s,A}$, whereas the receiving sensitivity of USV B is $R_{s,B}$. Considering Π_{BA}, Π_{AB} mean the success or failure status of communication from USV B to USV A and from USV A to USV B, respectively, then whether the two USVs can communicate or not can be determined using the following relationship:

(1) For communications from USVs B to A.

If $RSS_A - R_{s,A} \geq -83$ dBm, $\Pi_{BA} = 1$(Communication from USVs B to A is successful).

If $RSS_A - R_{s,A} < -83$ dBm or the two USVs do not communicate at the same frequency, $\Pi_{BA} = 0$ (Communication from USVs B to A is unsuccessful).

(2) For communication from USVs A to B.

If $RSS_B - R_{s,B} \geq -$ 83dBm, $\Pi_{AB} = 1$(Communication from USV A to USV B is successful).

If $RSS_B - R_{s,B} < -$ 83dBm or the two USVs do not communicate at the same frequency, $\Pi_{AB} = 0$ (Communication from USV A to USV B is unsuccessful).

The assessment of the communication capability of heterogeneous USVs shows that two USVs can interact with each other through the HF/VHF frequency digital transmission radios depending mainly on the frequency of the signals, the distance between the two USVs, the performance of the communication system, and the environmental interference of the communication scenario, among other factors. Furthermore, this can occur without considering the mutual communication interference between multiple USVs. Evaluating the communication capabilities of any two USVs in a cluster can determine the overall communication connectivity.

3.3 Unmanned vessel cluster communication topology

After completing the communication capability assessment of heterogeneous USV clusters, the communication topology between USVs is described using a graphical data structure (Zhou, 2011). Consider a cluster system consisting of a total of n USVs, which is represented by the graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots v_n\}$ is the set of vertices, and each vertex represents each USV. $E = V \cdot V = \{(v_i, v_j), i, j = 1 \ldots n\}$ is the set of edges, v_i is the starting point, v_j is the ending point, and $\boldsymbol{E}$ denotes whether a communication link is established between the USVs or not, and if there exists an edge from a vertex to reach v_i, the vertex is said to be a neighbour j of v_i, and the set of j is N_i. Describing the connectivity of a graph using a matrix of $A = [a_{ij}] \in R^{n\times n}$, $\boldsymbol{A}$ is the adjacency matrix, which represents the connectivity between vertices, and according to the evaluation of the communication capabilities discussed in the previous section and a_{ij} is as follows Equation 3:

$$a_{ij} = \begin{cases} 1 \; if \;\; \Pi_{ji} = 1 \\ 0 \; \text{otherwise} \end{cases} \quad (3)$$

Consider that the USV can process its own information directly such that $a_{ii} = 0, i = 1, \ldots, n$, that is, the main diagonal of the adjacency matrix, is zero. If communication does not consider direction, $\boldsymbol{G}$ is an undirected graph when $\boldsymbol{A}$ is a symmetric matrix.

The degree $d(v_i)$ of vertex v_i denotes the number of edges associated with the vertex. For directed graphs, the degree of vertex v_i is divided into the in-degree $d_{in}(v_i)$ and out-degree $d_{out}(v_i)$, that is, the number of directed edges from entering vertex v_i and the number of directed edges from exiting vertex v_i. Moreover, the in-degree matrix better captures the impact of neighbouring USVs on one's ship. For an undirected graph, the degree of vertex v_i is the number of edges and is equal to the number of neighbours. $D = diag\{d(v_1), d(v_2), \ldots, d(v_n)\}$ is a degree matrix that represents the communication connectivity of a USV.

The Laplace matrix $\boldsymbol{L}$ that defines the graph is given by Equation 4.

$$L = D - A \quad (4)$$

If the communication topology is a directed graph, $\boldsymbol{D}$ denotes the incidence matrix. For example, Figure 5 illustrates the communication topology of a cluster of four USVs.

The corresponding adjacency, degree, and Laplace matrices for the communication topology shown in Figure 5 are as follows:

$$\boldsymbol{D} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \boldsymbol{L} = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

The computed Laplace matrix reflects important information regarding the communication system, and several important

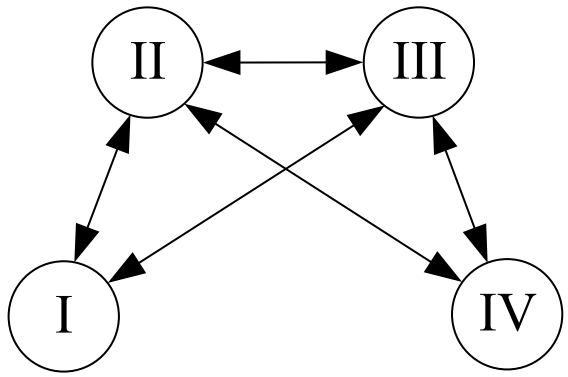


FIGURE 5
Communication topology with four nodes.

properties of the Laplace matrix $\boldsymbol{L}$ are described below (Liu and Huang, 2023).

(1) For undirected graphs, Laplacian matrix $\boldsymbol{L}$ is symmetric and semi-positive definite (given a real symmetric matrix A of size $n \times n$, matrix $\boldsymbol{A}$ is a semi-positive definite matrix if there is a constant $\mathbf{x}^T\mathbf{Ax} \geq 0$for any vector $\boldsymbol{x}$ of length n).

(2) For an undirected graph, the eigenvalues of the Laplacian matrix $\boldsymbol{L}$ are ordered to satisfy $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, where $\lambda_1 = 0$ and $\lambda_2 > 0$ if and only if the graph is connected.

(3) For a directed graph, a sufficient condition for containing a directed spanning tree is that $\boldsymbol{L}$ has one and only one zero eigenvalue, and all other eigenvalues have nonnegative real parts.

4 Coherent formation control of USV clusters under switching communication topology

4.1 Unmanned vessel cluster coherence control

Consistency control under a switching communication topology is a core problem in cooperative control. First, we introduce the basic concept of coherence control, analyse traditional coherence control strategies, and discuss their effects when applied to USV clusters with a single static communication topology.

The goal of consistency research is to design a consistency algorithm that relies only on the information of its neighbours to ensure that all intelligent bodies agree on a certain quantity of interest (Olfati-Saber et al., 2007). Consistency can be used to study clustered distributed architectures in which no individual of the cluster is overly dependent on any individual; therefore, the system as a whole is more robust. Designing consistency algorithms is an important aspect of consistency studies; for a first-order continuous system, it is modelled as follows Equation 5:

$$\dot{x}_i(t) = u_i(t) \tag{5}$$

where x_i denotes the state of the ith intelligence and u_i denotes the control input of the ith intelligence. The most common continuous time consistency algorithm is shown in Equation 6:

$$u_i(t) = -\sum_{j=1}^{n} a_{ij}(t)[x_i(t) - x_j(t)] \tag{6}$$

where $a_{ij}(t)$ denotes the value of the connection state between the intelligence i and neighbour j in the communication topology in a first-order system, and if $t \to \infty$ and $\left|x_i(t) - x_j(t)\right| \to 0, i, j = 1, 2, \ldots, n, \forall\, i \neq j$ exist, then the cluster of intelligences is said to have reached agreement. Suppose the communication topology graph is undirected and is connected under the control of Equation 6. In that case, the states of the individual intelligences gradually converge to those of the neighbouring intelligences until the states of all the intelligences converge to become consistent. The convergence time depends specifically on the structure of the communication topology graph, and good control must converge within the shortest possible time.

Furthermore, Equation 6 is written in matrix form as follows Equation 7:

$$\dot{x}(t) = -\boldsymbol{L}_n(t)\boldsymbol{x}(t) \tag{7}$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]$ is the cluster state vector and $\boldsymbol{L}_n(t)$ is the Laplace matrix of the cluster communication topology.

If x_i denotes the displacement of the ith intelligence, Equation 7 is referred to as a first-order consistency algorithm. For second-order intelligence considering displacements and velocities, the second-order multi-intelligence model is defined by Equation 8:

$$\begin{cases} \dot{x}_i(t) = v_i(t) \\ \dot{v}_i(t) = u_i(t) \end{cases}, i = 1, 2, \ldots, n \tag{8}$$

where v_i denotes the velocity of the ith intelligence, and the second-order consistency algorithm is given as follows Equation 9:

$$u_i(t) = -\sum_{j=1}^{n} a_{ij}(t)[(x_i(t) - x_j(t)) + \gamma(v_i(t) - v_j(t))] \tag{9}$$

γ denotes the coupling strength coefficient $\gamma > 0$ in a second-order system if $t \to \infty$ or $\left|x_i(t) - x_j(t)\right| \to 0$, $\left|v_i(t) - v_j(t)\right| \to 0, i, j = 1, 2, \ldots, n, \forall\, i \neq j$. The cluster of intelligence is then said to have reached an agreement.

Similarly, Equation 9 can be written in matrix form as follows Equation 10:

$$\begin{bmatrix} \dot{\boldsymbol{x}}(t) \\ \dot{\boldsymbol{v}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{n\times n} & \boldsymbol{I}_n \\ -\boldsymbol{L}_n(t) & -\gamma\boldsymbol{L}_n(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(t) \\ \boldsymbol{v}(t) \end{bmatrix} \tag{10}$$

where $\mathbf{v}(t)$ denotes the cluster velocity vector. A sufficient condition for the second-order consistency Equation 9 to achieve consistency is that the communication topology graph has a spanning tree, and the coupling strength coefficient γ satisfies (Gao et al., 2017) Equation 11:

$$\gamma > \max_{i=2,\cdots,n} \sqrt{\frac{2}{|\lambda_i(-\boldsymbol{L})|\cos\left(\frac{\pi}{2} - \tan^{-1}\frac{-\mathrm{Re}(\lambda_i(-\boldsymbol{L}))}{Im(\lambda_i(-\boldsymbol{L}))}\right)}} \tag{11}$$

10.3389/fmars.2024.1388617

where λ_i denotes the characteristic root of the Laplace matrix $\boldsymbol{L}$. It is evident that, unlike the first-order model, parameter γ also influences whether the second-order system can achieve consistency.

To illustrate the usefulness of the consistency algorithm, a cluster of $n = 4$ USVs is selected. The simplified USV model is given by Equation 8, with the cluster having the following undirected communication topology, as shown in Figure 6.

The degree, adjacency, and Laplace matrices of the communication topology graph were obtained from Figure 6:

$$\boldsymbol{D} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{L} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

Obviously this communication topology graph contains a spanning tree that sets the $\boldsymbol{x}(0) = [4.5, 3.0, 4.0, 5.7]^{\mathrm{T}}$ and $\boldsymbol{v}(0) = [1.0, 1.5, 1.2, 1.7]^{\mathrm{T}}$.

Using the second-order consistency controller (10), according to the coupling strength coefficient condition (11), it is necessary to satisfy the $\gamma > 0.71$. In this study we used $\gamma = 1$. The simulation was performed for 30 s, and Figure 7 depicts graphs of the positions and velocities of the four USVs under the control of a second-order consistency algorithm.

As shown in Figure 7, the initial positions and velocities of the four USVs are different; however, under the control of the second-order consistency algorithm, the positions of the USVs eventually converge to the same position at approximately 5 s, and later. Subsequently, the velocities of the USVs converge to the same velocity at approximately 7 s, which ultimately converge to the average value of the initial velocities. Therefore, the simulation results verify the correctness of Equation 10, and the system exhibited good convergence.

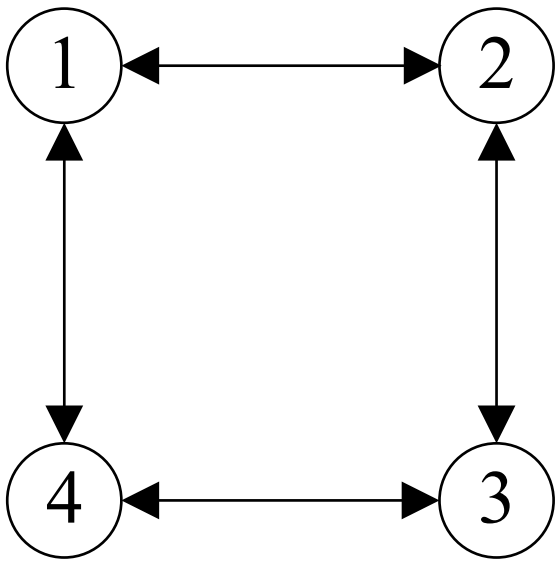


FIGURE 6
Quad-Intelligent body cluster communication topology.

4.2 Study of coherent formation based on cooperative communication of USVs

Unmanned USV clusters can cooperate with each other and adapt to the formation environment; that is, the offset is superimposed based on consistency, as shown in Figure 8. In the realisation of USV formation control, considering the communication capability of heterogeneous USV clusters, the communication distance between cluster individuals and the limitations and perturbations of the communication environment may interrupt the communication link. This results in a change in the communication topology map of the USV clusters. To ensure the consistency of the switching topology, the corresponding mechanisms and algorithms were studied to ensure that the system can adaptively deal with topology changes, maintain a high degree of reliability and stability of the cluster, and enhance the robustness and fault tolerance of the system.

Starting with the communication delay, the problem of USV cluster formation was investigated under a directed switching topology. The feedback linearisation of the USV dynamics model into a second-order integrator model considers a USV cluster system consisting of n USVs moving in a two-dimensional plane, and the motion model of the ith USV can be expressed as follows Equation 12:

$$\begin{cases} \dot{x}_{xi}(t) = v_{xi}(t) \\ \dot{v}_{xi}(t) = u_{xi}(t) \\ \dot{x}_{yi}(t) = v_{yi}(t) \\ \dot{v}_{yi}(t) = u_{yi}(t) \end{cases} \tag{12}$$

where x_{xi}, v_{xi}, and u_{xi} denote the position, velocity, and control input of the ith USV in the x-direction, respectively, and x_{yi}, v_{yi}, and u_{yi} denote the position, velocity, and control input of the ith USV in the y-direction, respectively.

To realise consistent control of the USV clusters, the controller of the ith USV is proposed as follows Equation 13:

$$u_i(t) = -\partial_1 (v_i(t) - v_d(t)) - \sum_{j \in N_i} a_{ij}(t)\{\partial_2[(x_i(t) - (x_j(t - \tau_t) - d_{ij}))] + \partial_3[(v_i(t) - v_j(t - \tau_t)]\} \tag{13}$$

where τ_t denotes the communication delay between individual USVs; ∂_1, ∂_2, and ∂_3 denote the control gains; $v_d(t)$ is the reference velocity for USV cluster navigation; and d_{ij} is the desired relative position of USV j concerning USV i.

The condition for the system to agree in a finite amount of time is the existence of $t_0 \in [0, +\infty)$, then,

$$\begin{cases} \lim\limits_{t \to t_0} || x_i(t) - (x_j(t - \tau_t) - d_{ij}) || = 0 \\ \lim\limits_{t \to t_0} || x_i(t) - v_j(t) || = 0 \end{cases} \tag{14}$$

This controller can prove the stability of the system by constructing Lyapunov-Krasovskii generalised functions Equation 14 (Zhang et al., 2023); eventually, the system is globally asymptotically stabilised, and controller (13) maintains the system consisting of model (12) in formation and brings the velocity to

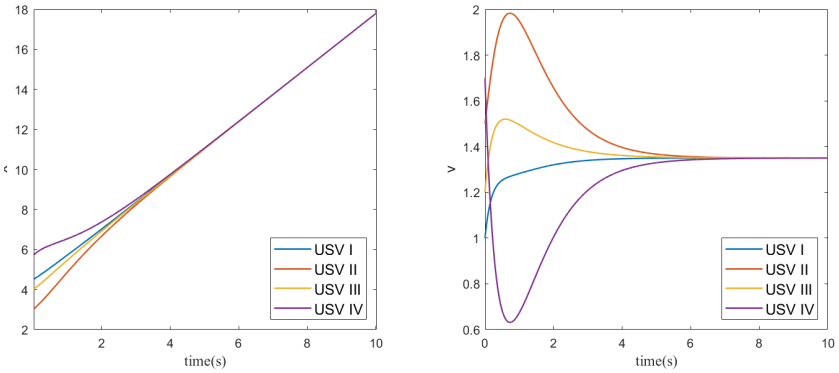


FIGURE 7
Position and velocity versus time curves of four USVs.

unity. Considering the USV as moving on a two-dimensional plane, the equation of motion of USV i is expressed as follows Equation 15:

$$\begin{cases} \dot{x}_{xi}(t) = v_{xi}(t) \\ u_{xi}(t) = -\partial_1 (v_{xi}(t) - v_{xd}(t)) \\ -\sum_{j\in N_i} a_{ij}(t)\{\partial_2[(x_{xi}(t) - (x_{xj}(t-\tau_t) - d_x(i,j)))] + \partial_3[(v_{xi}(t) - v_{xj}(t-\tau_t)]\} \\ \dot{x}_{yi}(t) = v_{yi}(t) \\ u_{yi}(t) = -\partial_1 (v_{yi}(t) - v_{yd}(t)) \\ -\sum_{j\in N_i} a_{ij}(t)\{\partial_2[(x_{yi}(t) - (x_{yj}(t-\tau_t) - d_y(i,j)))] + \partial_3[(v_{yi}(t) - v_{yj}(t-\tau_t)]\} \end{cases} \tag{15}$$

$d_x(i,j)$ and $d_y(i,j)$ are the elements in the relative position matrices $\boldsymbol{d}_x$, and $\boldsymbol{d}_y$ in the x- and y-direction, respectively.

Consider a USV cluster system consisting of four USVs with the initial state of each USV as follows:

$$\begin{aligned} (x_{x1}, x_{y1}, v_{x1}, v_{y1}) &= (0m, \quad 0m, \quad 0\tfrac{m}{s}, 0\tfrac{m}{s}) \\ (x_{x2}, x_{y2}, v_{x2}, v_{y2}) &= (0m, \quad 100m, \quad 0\tfrac{m}{s}, 0\tfrac{m}{s}) \\ (x_{x3}, x_{y3}, v_{x3}, v_{y3}) &= (0m, \quad 200m, \quad 0\tfrac{m}{s}, 0\tfrac{m}{s}) \\ (x_{x4}, x_{y4}, v_{x4}, v_{y4}) &= (0m, \quad 300m, \quad 0\tfrac{m}{s}, 0\tfrac{m}{s}) \end{aligned}$$

The communication capability is simplified to a maximum communication distance of 110 m, the parameters in the controller are set to ∂_1= 1.4, ∂_2= 1.1, and ∂_3= 1.2. Furthermore, the communication delay is set to a constant τ = 0.5 s, the desired velocity of the cluster is v_d = 2 m/s, and the setup is to form a positively oriented formation, with a relative position matrix of:

$$\boldsymbol{d}_x = \begin{bmatrix} 0 & 100 & 100 & 0 \\ -100 & 0 & 0 & -100 \\ -100 & 0 & 0 & -100 \\ 0 & 100 & 100 & 0 \end{bmatrix} \quad \boldsymbol{d}_y = \begin{bmatrix} 0 & 0 & 100 & 100 \\ 0 & 0 & 100 & 100 \\ -100 & -100 & 0 & 0 \\ -100 & -100 & 0 & 0 \end{bmatrix}$$

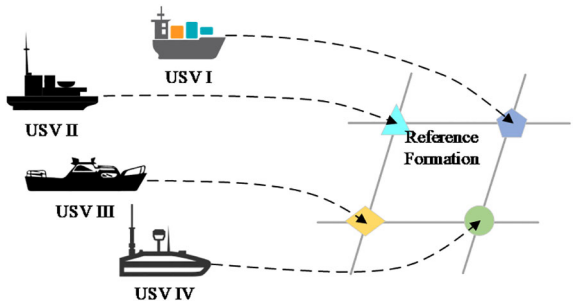


FIGURE 8
Schematic diagram of USV clusters forming a formation.

At the beginning of the formation, the USVs are lined up in a single line. The communication topology of the cluster is shown in Figure 9A, and the formation is formed after 32.8 s. The communication topologies are shown in Figure 9B, where both communication topologies contain spanning trees.

Figure 10 shows the position change diagram of the USV cluster that gradually formed under the control of the coherent-formation controller. In the formation process, owing to the change in the relative positions of each USV, the communication topology of the USV cluster changes accordingly. The communication topology changes from Figure 9A to Figure 9B and maintains the shape of the formation under the communication topology Figure 9B. As shown in Figure 10, at the initial position (0m), the four USVs are arranged in a line formation to begin the observation mission. As the USVs carry out different observation tasks, the relative positions of each USV change, causing the dynamic topology of the USV formation to change. Under the optimization of communication topology, the USVs cluster will restore its square formation shape at 80 meters. It can be observed that the USV cluster also forms and maintains its shape smoothly under changes in the communication topology.

Figures 11, 12 show the velocity changes of the four USVs in the x- directions and y-directions during the communication topology switching process. It can be observed that from the initial position (0 meters) to the position at 30 meters, the velocities of the USVs change due to the variations in their relative positions. Under the transformation of communication topology, the communication topology optimization algorithm based on a consensus controller enables the velocities of the four USVs in the x and y directions to quickly converge at the position of 30 meters, maintaining the formation of the USVs cluster. This verifies the effectiveness and stability of the control algorithm.

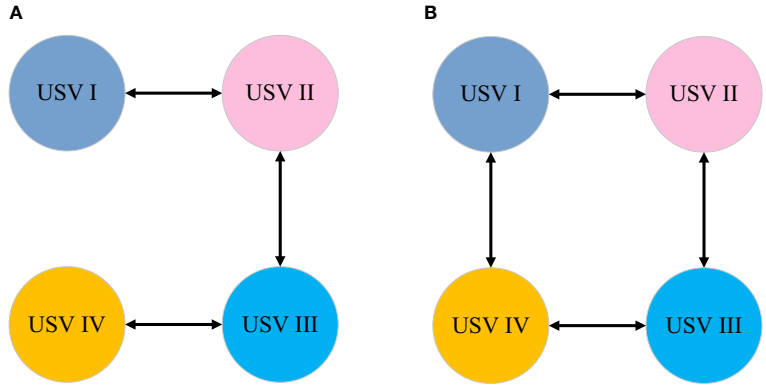


FIGURE 9
Communication topology before and after switchover. **(A)** Initial communication topology, **(B)** Optimized communication topology.

After a period under the control of the ocean observation progress consistency controller, the ocean observation progress of the four USVs gradually became consistent, synchronously completing observation tasks until the end. The variation in the velocity of each unmanned vessel, as shown in Figure 13 under the drive of the ocean observation progress consistency controller, the progress of oceanic observation based on communication topology optimization dynamically optimizes the topology structure according to the observation velocity of the USV and the ability to complete observation tasks. This ensures the completion of oceanic observation tasks while improving the efficiency of oceanic observation. From the figure, it can be seen that the slower vessels (such as: USV III) accelerate to "catch up" with the observation progress while ensuring task completion. Conversely, the faster vessel (such as: USV IV) decelerates to "slow down" the observation progress. Eventually, the progress of the final task was consistent with the constraints of the USV model.

The observation coordination of the collaborative consensus controller based on communication topology optimization reduces the observation time by 22% compared to dispersed control. Because each USV initially has the same position, and dispersed control entails each USV independently completing its respective tasks based on its observation capabilities, resulting in task intersections and reduced efficiency. In contrast, the collaborative consensus controller based on communication topology optimization coordinates control according to changes in communication topology structure, as well as the velocity and observation capabilities of each USV. They complement each other's advantages and synchronize completion until the end of the observation tasks. A comparison of test results, as shown in Table 3, demonstrates a 22% increase in search efficiency for observation progress coordination compared to dispersed control.

5 Conclusion

To maintain the formation of heterogeneous USV clusters under dynamic changes in communication topology, the feedback of the heterogeneous USV model was linearised into a second-order

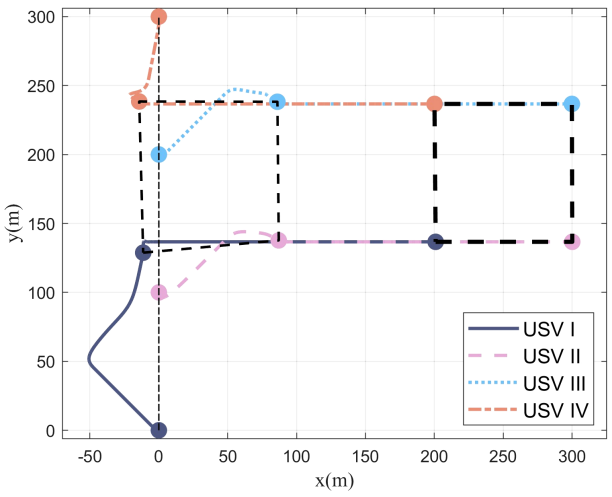


FIGURE 10
Map of changes in the position of USV clusters.

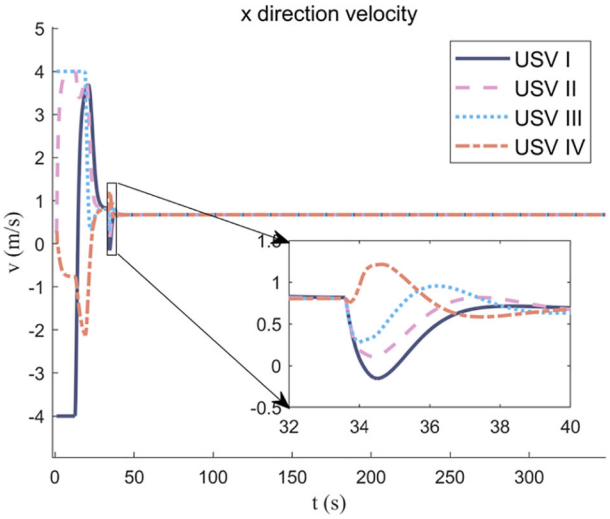


FIGURE 11
Velocity in the x-direction of each USV under switching topology.

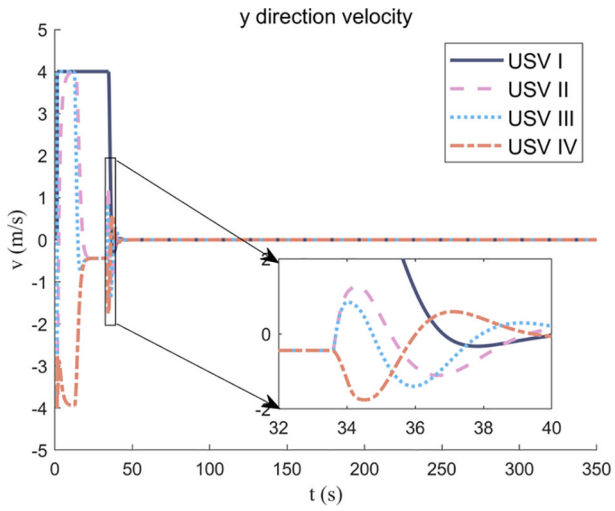


FIGURE 12
Velocity in the y-direction of each USV under switching topology.

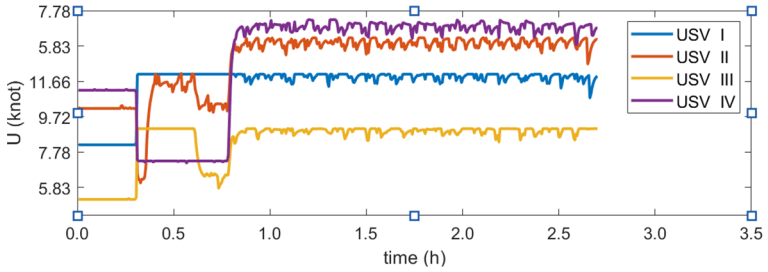


FIGURE 13
Navigation velocity of each USV.

integrator model. Starting from the communication delay, the formation generation simulation experiments with four heterogeneous USV formations verified that the communication topology transformation can smoothly form and maintain the formation shape of heterogeneous USV clusters. The efficiency of marine observation by USV clusters based on topology optimisation of consistent communication was 22% higher than that of dispersed control topology structures. This study provides a solid foundation for future investigations into the cooperative control of heterogeneous USV clusters and their applications in marine observations. Currently, the research has not considered the influence of marine environmental factors on the coordinated formation of USV clusters. The next step will involve studying factors such as wind and currents.

TABLE 3 Simulation experiment results for search progress consistency.

No.	Allocated area (nmi^2)	Collaborative Consensus controller Time (h)	Dispersed Control Time (h)
USV I	4.49	3.34	2.70
USV II	3.38	3.13	2.70
USV III	7.61	3.48	2.70
USV IV	6.33	2.98	2.70

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

ZK: Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. MG: Formal analysis, Resources, Writing – review & editing. AZ: Writing – review & editing, Conceptualization, Funding acquisition, Methodology, Supervision. ZL: Writing – review & editing, Validation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Guangdong Special Fund for Promoting Economic Development (Guangdong Natural Resources Cooperation) (Grant No. [2022119), Tianjin Research Innovation Project for Postgraduate Students (No. 2021YJSB177), and National Natural Science Foundation of China (52201414).

Acknowledgments

Special thanks go to the funding support provided for this research.

Conflict of interest

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

Publisher's note

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

References

Alqurashi, F. S., Trichili, A., Saeed, N., Ooi, B. S., and Alouini, M. S.. (2023). Maritime communications: A survey on enabling technologies, opportunities, and challenges. *IEEE Internet things J.* doi: 10.1109/JIOT.2022.3219674

Chainho, P., Drüsedow, S., Pereira, R. L., Chaves, R., and Portabales, A. R. (2017). Decentralized communications: trustworthy interoperability in peer-to-peer networks. European Conference on Networks and Communications. IEEE. (IEEE), 1–5.

Gao, Y., Dai, L., Xia, Y., and Liu, Y. (2017). Decentralized communications: trustworthy interoperability in peer-to-peer networks. European Conference on Networks and Communications. IEEE, 1–5

Gu, N., Wang, D., Peng, Z., and Liu, L. (2019). Distributed containment maneuvering of uncertain under-actuated USVs guided by multiple virtual leaders with a formation. *Ocean Eng.* 187 (Sep.1), 105996.1-105996.10. doi: 10.1016/j.oceaneng.2019.04.077

Jin, M. (2013). *Analysis and research on propagation models and related algorithms in mobile communication* (Beijing: Beijing University of Posts and Telecommunications).

Liang, M., Liu, R. W., Zhan, Y., Li, H., Zhu, F., and Wang, F.-Y. (2022). Fine-grained vessel traffic flow prediction with a spatio-temporal multigraph convolutional network. *IEEE Trans. Intell. Transport. Syst.* 23, 23694–23707. doi: 10.1109/TITS.2022.3199160

Liang, M., Weng, L., Gao, R., Li, Y., and Du, L. (2024). Unsupervised maritime anomaly detection for intelligent situational awareness using AIS data. *Knowledge-Based Syst.* 284. doi: 10.1016/j.knosys.2023.111313

Lim, I. H., Kim, Y. I., Lim, H. T., Choi, M., Hong, S., Lee, S., et al. (2008). Distributed Restoration system applying Multi-Agent in distribution automation system. doi: 10.1109/PES.2008.4596084

Liu, K., and Huang, Z. (2023). Research on the Hermitian Laplacian matrix of directed graphs. *Comp. Sci.* 50, 7.

Liu, R. W., Liang, M., Nie, J., Lim, W. Y. B., Zhang, Y., and Guizani, M. (2022). "Deep learning-powered vessel trajectory prediction for improving smart traffic services in maritime internet of things," in *IEEE Transactions on Network Science and Engineering*, Vol. 9. 3080–3094.

Olfati-Saber, R., Fax, J. A., and Murray, R. M. (2007). "Consensus and cooperation in networked multi-agent systems," in *Proceedings of the IEEE*, Vol. 95. 215–233.

Schröder-Hinrichs, J. U., Song, D. W., Fonseca, T., Lagdami, K., and Loer, K. (2019). *Transport 2040: Automation, Technology, Employment - The Future of Work* (Sverige: World Maritime University), 83–115.

Sun, Q., Yao, Y., Yi, P., Hu, Y. J., Yang, Z., and Yang, G. (2022). Learning controlled and targeted communication with the centralized critic for the multi-agent system. *Appl. Intell.* 53, 14819–14837. doi: 10.1007/s10489-022-04225-5

Wang, D., Fu, M., Ge, S. S., and Li, D. (2020). Velocity free platoon formation control for USVs with output constraints and model uncertainties. *Appl. Sci.* 10 (3), 1118. doi: 10.3390/app10031118

Wu, G. (2011). *Research on maneuverability and intelligent control technology of unmanned vessels* (Harbin: Harbin Engineering University).

Xia, M., Zhu, Y., Chen, E., Xing, C., Yang, T., and Wen, W. (2017). Development status and challenges of maritime communication. *Sci. China: Ser. F* 47, 677–695.

Xie, W., Tao, H., Gong, J., Luo, W., Yin, F., and Liang, X. (2021). Development status and research progress of key technologies for maritime unmanned systems clusters. *Chin. J. Ship Res.* 16, 12.

Xing, J. (2012). Research on parameter identification method of dynamic positioning ship motion mathematical model. *Doctoral dissertation*. Harbin: Harbin Eng. Univ.

Zhang, W., Liao, Y., Jiang, F., and Zhao, T. (2019). Review and trend analysis of unmanned surface vessel technology development. *Unmanned Syst. Tech* 2, 1–9.

Zhang, X., Zhou, L., Xing, W., and Yao, S. (2023). Consensus control of multi-autonomous underwater vehicles formation under switching topologies. *J. Harbin Eng. Univ.* 44, 587–593.

Zhou, T. (2011). Research on target detection probability in maritime observation at sea. *Doctoral dissertation*. Dalian: Dalian Maritime Univ.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 05 April 2024
DOI 10.3389/fmars.2024.1357144

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Carlos Pérez-Collazo,
University of Vigo, Spain
Changqing Jiang,
University of Duisburg-Essen, Germany

*CORRESPONDENCE
Bowen Zhao
zhbwmail@163.com

RECEIVED 17 December 2023
ACCEPTED 11 March 2024
PUBLISHED 05 April 2024

CITATION
Zhang D, Zhao B, Zhang Y and Zhou N (2024) Numerical simulation of hydrodynamics of ocean-observation -used remotely operated vehicle. *Front. Mar. Sci.* 11:1357144. doi: 10.3389/fmars.2024.1357144

Numerical simulation of hydrodynamics of ocean-observation-used remotely operated vehicle

Dapeng Zhang[1], Bowen Zhao[2*], Yi Zhang[1] and Nan Zhou[3]

[1]Ship and Maritime College, Guangdong Ocean University, Zhanjiang, Guangdong, China, [2]Department of Applied Mathematics and Mathematical Modeling, Saint-Petersburg State Marine Technical University, Saint-Petersburg, Russia, [3]Harbin Engineering University, Qingdao Innovation and Development, Qingdao, Shandong, China

Remotely operated vehicle is the most widely used underwater robot and can work safely and steadily in complex environments compared to autonomous underwater vehicle and other types. It has obvious advantages in operation time and plays a significant function in marine engineering equipment. Hydrodynamic coefficients are the coefficients of ROV motion equation. In order to simulate the motion and predict the performance of a ROV, the hydrodynamic coefficients must be determined first. The motion mathematical model of remotely operated vehicles is also established, and the hydrodynamic dynamics of the vehicles are simulated using the finite volume method by combining overset mesh technology and governing equations. Finally, a simulation and verification of the standard model SUBOFF model and the calculation process's dependability are also conducted. The primary hydrodynamic coefficients of the ROV were derived through the simulation data fitting process. The results showed that the ROV's asymmetry results in an obvious disparity in pressure resistance between the forward and backward sailing, ascending and descending motions, and this disparity becomes significantly greater as the velocity increases. This method confirmed the accuracy of the hydrodynamic simulation computation of the remotely operated vehicle and served as a guide for the maneuverability and design of the vehicle.

KEYWORDS

ROV, hydrodynamic calculation, overset mesh, numerical simulation, hydrodynamic coefficient

1 Introduction

The underwater robots offer a significant role in maritime engineering equipment. In addition to having a more flexible operating mode and the ability to be outfitted with various operating instruments for a variety of activities and working settings, underwater robots also provide a number of benefits over other types of equipment in terms of

operational efficiency. Additionally, using underwater robots, which can function at depths that divers are unable to reach, may speed up human progress into the deep sea. In addition to being able to function in unique settings that divers are unable to access, underwater robots are also often used in nuclear waste cleaning operations in nuclear power plant reservoirs. This efficiently preserves practitioner safety and enhances nuclear power plant safety. Underwater vehicles have been classified into many categories, including autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), and underwater gliders (UGs). Among them, ROVs are divided into three types: self-propelled in water, towed, and crawling (Dalibor and Marcin, 2024). AUVs are not bound by cables, have a large range of activities, and have good concealment performance, but their underwater operation time is affected by the amount of energy they carry. The energy of the ROV is provided by the mother ship through cables, and it is capable of carrying out complex underwater operations for a long time, and is currently the most widely used underwater robot (Zhang et al., 2023).

The United States created the first ROV in 1960 and called it "CURV1". First shown to the public at large, this ROV type was instrumental in retrieving a hydrogen bomb that had been abandoned by the US in Spanish seas (Whitcomb and Yoerger, 1993; Fan et al., 2012; Cepeda et al., 2023). During this time, ROVs were just starting out; the primary function of these early ROVs was to assist the military with recovery operations. Owing to the effects of the Middle East oil crisis in the 1970s, nations all over the globe dedicated significant resources to the study and creation of machinery for the extraction of subterranean oil. The offshore oil sector is expanding, and with it, so is the need for ROVs to monitor offshore oil platforms. The major working waters of this kind of ROV are in the North Sea oilfield, and its birth also signals the market acknowledgment of the ROV business and builds a firm market basis for the long-term growth of the ROV industry (Christophe, 2023; Selig et al., 2023).

The performance of ROVs has significantly improved between the early 1980s of the 20th century and the early 21st century. The most notable improvements have been in the operating depth, operating range, and operating duration of ROVs, which have been developed by various nations. The development of large-scale "operational-grade" ROVs at this time was mainly driven by the exploration of seabed resources by different nations, which improved human knowledge of the kinds and composition of resources in the deep sea and on the seabed. As a result of this era's progress, ROVs are now in the large-scale manufacturing phase and have amassed the necessary technology to enter the large-scale commercial usage phase (Ren and Hu, 2023). Since the beginning of the 21st century, the evolution of ROVs has been characterized by functional diversity and miniaturization. The future development trend for ROVs is miniaturization and intelligence, which will drive industry upgrading and further development of ROVs in aquaculture, military reconnaissance, underwater equipment maintenance, and marine resource development.

In order to save design expenses and increase efficiency, it is crucial to accurately determine the ROV's hydrodynamic coefficients. These coefficients are then used to make selections about the ROV's layout, propeller model, and handling performance. At present, the following methods are typically used to determine the hydrodynamic coefficient of underwater vehicles: system identification, experimentation with constraint models, empirical formula technique, and CFD software simulation calculation method. The empirical formula is the empirical or semi-empirical formula produced by synthesizing the data acquired by a vast number of previous model experiments or ROV actual excursion. By considering the analysis of control systems that are designed to maximize the operability limits for launch and recovery of a ROV from a small unmanned surface vessel (USV), Ahsan Tanveer and Sarvat Mushtaq Ahmad (Tanveer and Ahmad, 2023) use numerical simulation for the analysis, where the method combines recent approaches for wave compensating dynamic positioning, active heave compensation, and positioning control of the ROV with multi-body dynamic simulation of the surface vessel and ROV, including hydrodynamic forces and dynamic interactions from wires that depend on the ROV depth and moonpool. A fuzzy adaptive controller considering thruster dynamics is proposed by Mingjie et al. to improve the trajectory tracking performance of work-class ROVs (Mingjie et al., 2023). The system identification approach involves using the ROV's motion data from real navigation or the experimental data from the constraint model to build a mathematical model. This model is then used to estimate the hydrodynamic coefficients that describe the ROV's hydrodynamic performance.

The rapid advancement and widespread adoption of high-speed computers in recent years have allowed several scientific researchers to notice that computational fluid dynamics (CFD) software can rapidly calculate the hydrodynamic force of ROV. The previous approaches' drawbacks may be successfully addressed by the CFD software simulation technique, which can also assist researchers in increasing the effectiveness of their R&D and confirming if the hydrodynamic performance of ROV satisfies design specifications. SKORPA (Skorpa, 2012) performed hydrodynamic numerical simulation calculations on the WR-200 ROV model by simplifying it and using CFD software to analyze the results. The findings indicate that the pitch torque of the ROV can be effectively decreased by adjusting whether the water flow passes through the middle of the ROV. Chin and Lau (2012) carried out hydrodynamic numerical computations on the ROV model using ANSYS-CFX software. The findings demonstrated that the hydrodynamic coefficients acquired by CFD software may successfully help designers enhance the structural design of ROVs.

With the wide application of underwater robots in the development of marine resources, people have more stringent requirements for the performance of underwater robots. The hydrodynamic performance of the ROV is the basis of underwater positioning, path planning and maneuvering control, and the quality of the hydrodynamic performance directly determines the success of an ROV design (Manimaran, 2022). Due to the complex shape and structure of the ROV, the ROV is affected by various complex forces such as thrust of the propulsion mechanism, water flow resistance, buoyancy, gravity and tensile force in the water, so the calculation of the hydrodynamic force of the ROV is a complex kinematic and dynamic problem.

Understanding the hydrodynamic performance of ROVs is essential for several reasons. Firstly, it ensures efficient and precise operation of the vehicle, allowing operators to navigate through challenging underwater environments with accuracy. This is especially important when performing delicate tasks such as manipulating equipment, collecting samples, or conducting repairs. Furthermore, studying ROV hydrodynamic performance helps in enhancing safety. By understanding how the vehicle responds to different conditions, operators can mitigate potential risks and avoid collisions with underwater obstacles or hazards. This not only protects the ROV but also minimizes the potential damage to the surrounding environment. Moreover, the hydrodynamic performance of ROVs directly affects the speed and efficiency of tasks. By optimizing the vehicle's maneuverability, operators can reduce the time required to complete missions, saving resources and improving overall productivity.

The motion of the ROV is a spatial motion with six degrees of freedom. According to the movement force and moment, the maneuverability mathematical model can be constructed to determine the optimal control rule and control system (Zhao et al., 2022; Zhao et al., 2023). Hydrodynamic coefficients are the coefficients of ROV motion equation. In order to simulate the motion and predict the performance of a ROV, the hydrodynamic coefficients must be determined first. The existing hydrodynamic computation techniques for ROV mainly include the following methods: system identification, experimental constraint model method, empirical formula method, and CFD simulation calculation method. However, the hydrodynamic calculation theory of AUV is more mature than that of ROV, and the hydrodynamic calculation of ROV requires further improvement and verification of the dependability of the simulation calculation technique. Thus, the hydrodynamic forces of the ROV model during straight-line and planar motion mechanism (PMM) need to be primarily determined. The purpose of this study is to show the approach and method to obtain these hydrodynamic coefficients using CFD method, and to investigate the hydrodynamic characteristics of the ROV during turning maneuver. The rest of the paper is organized as follows. Section 2 presented the mathematical models for hydrodynamics and maneuverability of ROV. Section 3 carried out a verification study of hydrodynamic numerical methods. Section 4 and 5 simulated the hydrodynamic performance of ROV in steady and unsteady motions and fitted a large number of hydrodynamic coefficients. Finally, the conclusion draw from this paper are presented in Section 6.

2 Computational theory

2.1 Mathematical models

A motion coordinate system consisting of two right-hand coordinate systems $E-\xi\eta\zeta$ and $O-xyz$ was established, as shown in Figure 1.

The linear velocity and angular velocity of the ROV in the moving coordinate system can be expressed as $V_1 = [V_1 \ V_2]^T$, linear velocity V_1 is expressed as $V_1 = [u \ v \ w]^T$, angular velocity V_2, expressed as $V_2 = [p \ q \ r]^T$. The forces and moments experienced by the ROV in the dynamic coordinate system can be expressed as $A = [F \ M]^T$, ROV is expressed as the force $F = [X \ Y \ Z]^T$, the moment exerted is expressed as: $M = [K \ M \ N]^T$. The force and velocity of the ROV are positive in the direction of the coordinate axis of the dynamic coordinate system, and the moment and angular velocity are determined by the right-hand rule. Table 1 shows the six ROV degrees of freedom, while Table 2 shows the motion parameters and coordinate components.

The dynamic equation of the ROV must be established before analyzing its motion. The following assumptions are applied to the ROV model to answer the equation of motion: the ROV is a rigid body with a constant form, mass, and centroid. The hydrodynamic force of ROV is considered independent of the impact of the seabed environment and umbilical cable. The theory proposes that the center of gravity coincides with the origin of the secondary coordinate system, and the three axes of the follower coordinate system represent the inertial main axes of the ROV. The dynamic model of the ROV could be developed using the moment of inertia and the rigid body motion hypothesis, as shown in Equation (1).

$$M_{RB}\dot{V} + C_{RB}(V)V = F \tag{1}$$

In Equation (1), M_{RB} is ROV quality matrix; C_{RB} is centripetal force and coriolis matrix of ROV, $C_{RB}(V) = -C_{RB}{}^{T}(V)$ coefficient is related to the speed of movement, whereas; F displays overall torque in the ROV. The mass-matrix of the ROV is given in Equation (2).

$$M_{RB} = \begin{bmatrix} mI_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_g \end{bmatrix} \tag{2}$$

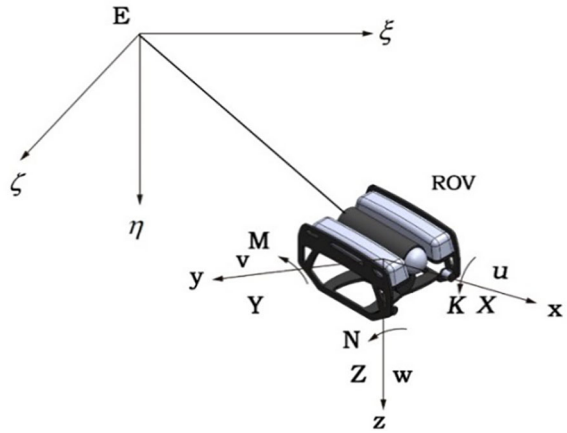


FIGURE 1
ROV system coordinate system.

TABLE 1 The six degrees of freedom of ROV.

Type of Motion	X axis	Y axis	Z axis
Translation	Surge	Sway	Heave
Rotation	Roll	Pitch	Yaw

10.3389/fmars.2024.1357144

TABLE 2 Motion parameters and coordinate components.

Vector	X axis	Y axis	Z axis
Velocity V_1	u (Longitudinal velocity)	v (Lateral velocity)	w (Vertical velocity)
Angular velocity V_2	p (Longitudinal angular velocity)	q (Lateral angular velocity)	r (Vertical angular velocity)
External force F	X (Longitudinal force)	Y (Lateral force)	Z (Vertical force)
Moment M	K (Roll moment)	M (Pitch moment)	N (Yaw moment)

Herein, m is mass; I_g is inertial matrix; $I_g = diag[I_x\ I_y\ I_z]$ in I_x, I_y, I_z, as shown in Equations (3–5):

$$I_x = \int_V \rho(y_i^{\ 2} + z_i^{\ 2})dx_i dy_i dz_i \tag{3}$$

$$I_y = \int_V \rho(z_i^{\ 2} + x_i^{\ 2})dx_i dy_i dz_i \tag{4}$$

$$I_z = \int_V \rho(x_i^{\ 2} + y_i^{\ 2})dx_i dy_i dz_i \tag{5}$$

Coriolis and centripetal matrices of $C_{RB}(V)$ are depicted in Equation (6).

$$C_{RB}(V) = \begin{bmatrix} mS(V_2) & 0_{3\times3} \\ 0_{3\times3} & -S(I_g V_2) \end{bmatrix} \tag{6}$$

where S is the vector multiplication operator, which can be represented by the following matrix.

$$S(\lambda) = -S^T(\lambda) \begin{bmatrix} 0 & -\lambda_3 & \lambda_2 \\ \lambda_3 & 0 & -\lambda_1 \\ -\lambda_2 & \lambda_1 & 0 \end{bmatrix} \tag{7}$$

By combining Equations (2) and (6), the Equation (7) is simplified as shown in Equation (8).

$$\begin{cases} X = m(\dot{u} - vr + wq) \\ Y = m(\dot{v} - wp + ur) \\ Z = m(\dot{w} - uq + vp) \\ K = I_x\dot{p} + (I_z - I_y)qr \\ M = I_y\dot{q} + (I_x - I_z)rp \\ N = I_z\dot{r} + (I_y - I_x)pq \end{cases} \tag{8}$$

To improve an analysis of the dynamical features of ROVs, it is critical to develop a hydrodynamic model that serves as the foundation for forecasting their maneuverability. In order to reduce the impact of variables like size, speed, and mass, a non-factor approach was used. The following hydrodynamic coefficients are unitless, and their unitless guidelines are shown in Table 3. By using the dimensionless rule outlined in Table 3, the dimensionless model of the ROV may be derived by applying the dimensionless hydrodynamic coefficients (Zhao et al., 2022; Zhao et al., 2023), as shown in Equations (9–14).

$$X_D = \frac{\rho}{2}L^4(X'_{qq}q^2 + X'_{rr}r^2 + X'_{rp}rp) + \frac{\rho}{2}L^3(X'_{vr}vr + X'_{wq}wq) + \frac{\rho}{2}L^2(X'_{uu}u^2 + X'_{vv}v^2 + X'_{ww}w^2) \tag{9}$$

$$Y_D = \frac{\rho}{2}L^4(Y'_{vq}vq + Y'_{wq}wq + Y'_{wr}wr) + \frac{\rho}{2}L^3(Y'_{ur}ur + Y'_{up}up + Y'_{v|r|}\frac{v}{|v|}\left|(v^2+w^2)^{\frac{1}{2}}\right||r|) + \frac{\rho}{2}L^2(Y'_{uu}u^2 + Y'_{uv}uv + Y'_{v|v|}v\frac{v}{|v|}\left|(v^2+w^2)^{\frac{1}{2}}\right| + Y'_{vw}vw) \tag{10}$$

$$Z_D = \frac{\rho}{2}L^4(Z'_{pp}pp + Z'_{pr}pr + Z'_{rr}rr) + \frac{\rho}{2}L^2(Z'_{wq}\frac{w}{|w|}\left|(v^2+w^2)^{\frac{1}{2}}\right||q| + Z'_{q}uq) + \frac{\rho}{2}L^2(Z'_{uu}u^2 + Z'_{uw}uw + Z'_{w|w|}w\left|(v^2+w^2)^{\frac{1}{2}}\right| + Z'_{|w|}uw + Z'_{ww}w\left|(v^2+w^2)^{\frac{1}{2}}\right| + Z'_{vv}v^2) + \frac{\rho}{2}L^3(Z'_{vp}vp + Z'_{vr}vr) \tag{11}$$

$$K_D = \frac{\rho}{2}L^5(K'_{p|p|}p|p| + K'_{pq}pq + K'_{pr}pr) + \frac{\rho}{2}L^4(K'_{up}up + K'_{ur}ur + K'_{vp}vp + K'_{wq}wq + K'_{wr}wr) + \frac{\rho}{2}L^4(K'_{uu}u^2 + K'_{uv}uv + K'_{v|v|}v\left|(v^2+w^2)^{\frac{1}{2}}\right| + K'_{vw}vw) \tag{12}$$

$$M_D = \frac{\rho}{2}L^3(M'_{uw}uw + M'_{uu}u^2 + M'_{w|w|}w\left|(v^2+w^2)^{\frac{1}{2}}\right| + M'_{|w|}uw + M'_{w|w|}w\left|(v^2+w^2)^{\frac{1}{2}}\right|) + \frac{\rho}{2}L^5(M'_{pp}p^2 + M'_{q|q|}q|q| + M'_{pr}pr + M'_{rr}r^2) + \frac{\rho}{2}L^4(M'_{uq}uq + M'_{vp}vp + M'_{vr}vr + M'_{|w|q}\left|(v^2+w^2)^{\frac{1}{2}}\right|q) \tag{13}$$

TABLE 3 Non-dimensioning rules.

Item	Non- dimensionless
Time	$t' = tU/L$
Velocity	$(u', v', w') = (u, v, w)/U$
Mass	$m' = m/(\frac{1}{2}\rho L^3)$
Length	$l' = l/L$
Angular velocity	$(p', q', r') = (p, q, r)/U$
Moment of inertia	$I' = I/(\frac{1}{2}\rho L^5)$
Acting force	$(X', Y', Z') = (X, Y, Z)/(\frac{1}{2}\rho U^2 L^2)$
Moment	$(K', M', N') = (K, M, N)/(\frac{1}{2}\rho U^2 L^3)$

$$N_D = \frac{\rho}{2}L^4(N'_{up}up + N'_{ur}ur + N'_{vp}vp + N'_{wq}wq + N'_{v|r|}\frac{v}{v}\left|(v^2+w^2)^{\frac{1}{2}}\right||r|) + \frac{\rho}{2}L^3(N'_{uu}u^2 + N'_{uv}uv + N'_{v|v|}v\left|(v^2+w^2)^{\frac{1}{2}}\right| + N'_{vw}vw) + \frac{\rho}{2}L^5(N'_{r|r}r\left|r + N'_{pq}pq + N'_{qr}qr\right) \quad (14)$$

2.2 CFD basic theory

The basic theory of the governing equation is the three conservation laws: the law of conservation of momentum, the law of conservation of mass, and the law of conservation of energy. Since the solution process in this paper does not involve energy, the law of conservation of energy is not considered in the governing equation in this paper. The basic form of the equation for the conservation of momentum is the N-S equation, which was proposed by Navier-Stokes and needs to be satisfied in general fluid systems, the N-S equation form, as shown in Equation (15).

$$\begin{cases} \rho\frac{du}{dt} = \rho g_x - \frac{\partial p}{\partial x} + \mu(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x^2}) \\ \rho\frac{du}{dt} = \rho g_y - \frac{\partial p}{\partial y} + \mu(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial y^2}) \\ \rho\frac{du}{dt} = \rho g_z - \frac{\partial p}{\partial z} + \mu(\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 u}{\partial z^2}) \end{cases} \quad (15)$$

where, ρ is water density, p is pressure and μ is kinematic viscosity coefficient.

The standard representation of mass conservation is the continuity equation, which states that the mass of matter in a given space remains constant. The change in mass within a control volume is determined by the difference between the inflow and outflow of mass from that volume. This work examines an incompressible fluid with a constant density. The continuity equation is further simplified, as seen in Equation (16).

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (16)$$

Herein, u, v, w are fluid velocity components.

The standard adopted in this article is k-ε turbulence model. At present, the RANS approach is the most frequently employed approach in engineering applications to analyze turbulence models, and it is also the method utilized in this research.

$$\rho\frac{\partial \bar{u}_i}{\partial t} + \rho\bar{u}_l\frac{\partial \bar{u}_i}{\partial x_j} = \rho\bar{f}_l + \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}(\mu\frac{\partial \bar{u}_i}{\partial x_j} - \rho\overline{u'_i u'_j}) \quad (17)$$

In Equation (17), ρ -Fluid density; $\bar{p}$ -Pressure average; μ —Dynamic viscosity; $\rho\overline{u'_i u'_j}$ —Reynolds stress.

In computational fluid dynamics, the basic principle of numerical solving is to solve for each discrete node to obtain an approximation of the overall flow field. The finite volume method (FVM) is used to discretize the governing equations. The wall function approach is employed to address the flow field in proximity to the wall. The development of the overset mesh approach accelerates the resolution of intricate flow fields. Its fundamental idea involves breaking down complex flow fields into smaller, independent sub-regions, with each sub-region generating a distinct mesh. When simulating complicated motions, there is no need to renew the mesh since each sub-region's mesh shape is fixed. Complex motions could be accomplished by specifying motions inside each sub-region's mesh.

As shown in Figure 2, the calculation steps of overset meshes mainly include: (1) sub-region division and mesh generation, (2) determining overset boundary conditions, (3) determining the interpolation type between sub-regions, and (4) calculating the flow field. In the calculation, the first three steps need to be continuously adjusted to ensure the convergence of the flow field. The difficulty of the overset mesh technique is that as the mesh position of each sub-region changes, the position of the boundary and the position of the hole area need to be determined repeatedly.

3 Verification of hydrodynamic numerical methods

3.1 SUBOFF Model

3.1.1 Proposal of verification methods

In order to verify the reliability of the theory proposed in this paper in the numerical calculation of underwater vehicles and the rationality of the numerical calculation model and meshing form, the standard model of underwater submersibles, the SUBOFF model, was selected for the verification of hydrodynamic numerical calculations.

3.1.2 SUBOFF Parameters of the model

The SUBOFF model has been accepted as a standard model by the ITTC. This standard model can be built according to the shape formula in ITTC. The SUBOFF model used in this paper is shown in Figure 3. The main model parameters are shown in Table 4.

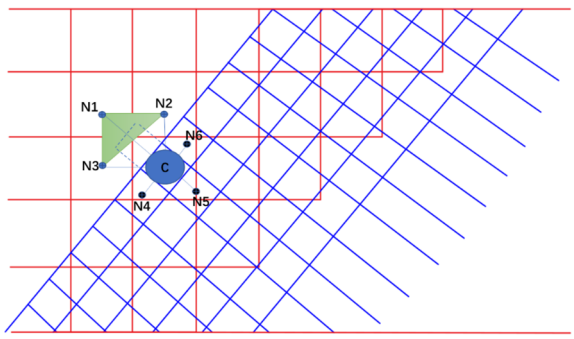


FIGURE 2
Overset mesh.

FIGURE 3
SUBOFF model.

TABLE 4 SUBOFF model parameters.

Parameter	Value
Overall length	4.356m
Forebody length	1.016m
Parallel body length	2.229m
Afterbody length	1.111m
Maximum Diameter	0.508m
Volume of displacement	$0.718m^3$
Longitudinal center of buoyancy	2.012m

3.2 Compute domain determination and meshing

3.2.1 Computational domain

A computational domain needs to consider whether the area is capable of ensuring fully convergent computational results and limit the percentage of invalid regions to increase computational efficiency. The size of the computational domain for this paper is calculated based on the partition of the computational domain in prior SUBOFF simulation studies. The domain length is determined by multiplying the length by 6. The height is determined by multiplying the length by 3. Similarly, the width is determined by multiplying the length by 3. Furthermore, for an adequately accelerated flow field, the submersible is situated at a distance equal to two-hull dimensions. The length of the hull from the velocity inlet is twice the length of the boat, while the length of the boat is three times the length from the pressure outlet.

3.2.2 Meshing

According to the results of the division of the calculation area, this paper divides the meshing of different regions. When meshing, the basic principle of meshing is strictly followed, and the mesh refinement is carried out in the area close to the hull, and the mesh density is reduced layer by layer in the area far away from the hull. In this paper, we select tetrahedral mesh, cut mesh, and prismatic layer mesh model, and set the basic mesh size to 0.3m, set the boundary layer to 5 layers, and set the minimum mesh size to 3.125% of the basic mesh size. Figure 4 is the longitudinal meshing result in the calculation domain, Figures 5, 6 are the meshing result of the bow and stern of the hull, respectively.

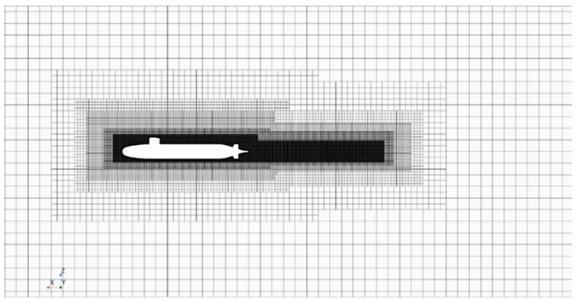


FIGURE 4
Vertical overall meshing results.

3.3 Simulation calculation and verification

In order to enhance computational efficiency, it is critical to reduce the proportion of invalid regions in the computational domain and ascertain whether the computational area will deliver fully convergent results when determining the computational domain. The size of the computational domain to be employed in this research is established by referring to the partition of the computational domain in the previous SUBOFF simulation studies. The width is three times the length, the height is three times the length, and the computed domain length is six times the length. Additionally, the SUBOFF is positioned at the second-hull length, the hull length is tripled from the pressure outlet, and the hull length from the velocity inlet is doubled from the hull length in order to be sure that the computed flow field has a viable acceleration area.

3.3.1 Boundary and computation conditions

The boundary conditions should be set separately in every surface of calculation domain. The boundary conditions are determined by designating the incoming flow as the velocity inlet, the outlet as the pressure outlet, the hull surface as the wall, and the surrounding surface as the wall. The standard $k-\epsilon$ model is selected as the turbulence model. Five working conditions with

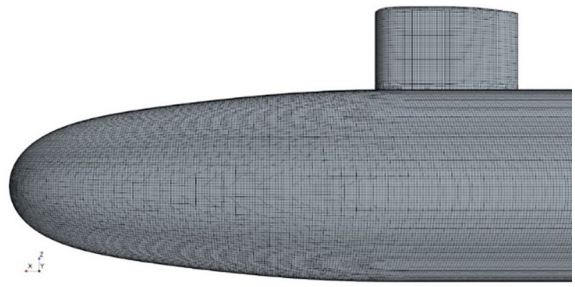


FIGURE 5
Bow mesh.

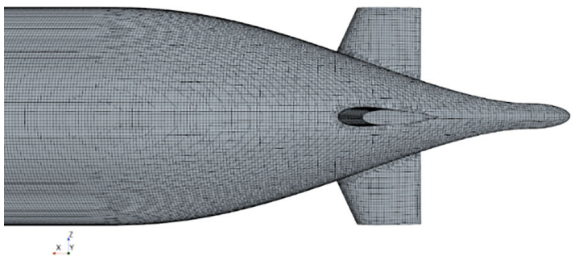


FIGURE 6
Stern mesh.

velocities of 3.045, 5.144, 6.091, 7.161 and 8.231 m/s were selected for straight-line simulation experiments.

3.3.2 Calculation results and validity analysis

According to the boundary conditions under different working conditions, five straight-line simulations are carried out in this section, and the simulation results and the relative error results of the simulated calculation are shown in Table 5.

In addition, the simulation results are compared with the experimental results, as shown in Figure 7. Through the comparison curve between the experimental values and the simulated values, it can be clearly seen that the straight-line resistance value gradually increases with the increase of velocity, but the relationship between resistance and velocity is not linear. At a velocity of 3.045m/s, the error between the simulated value and the experimental value is the largest, and the reliability of the simulated value at low velocity is slightly lower than that at medium and high velocities. Overall, the simulation data are basically consistent with the experimental data, and the relative error between the two is 1%-3%, which can meet the actual standards of the project.

By analyzing the SUBOFF straight-line simulation experiment, the engineering standard can be satisfied, and then it is confirmed that the calculation theory, the meshing method and the selected calculation model proposed in this paper are reasonable and reliable, so the calculation method can be popularized and used in the subsequent simulation experiments.

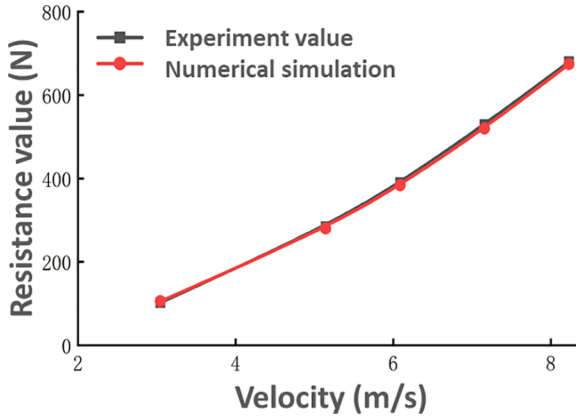


FIGURE 7
Comparison of simulation and experimental data.

FIGURE 8
Three-dimensional model.

4 Simulation calculation of the steady motion of the ROV

4.1 ROV model simplification and computational domain division

4.1.1 Simplified ROV model

Figure 8 shows that the computational model used in this paper is based on a small ROV. In order to do simulation calculations using CFD software, it is necessary to simplify the model of the ROV by removing the complex surfaces that do not contribute to the computation. The principle of simplification encompasses several key characteristics. Firstly, the simplified model must align with the primary scale of the original model, while also preserving the essential components of the ROV. Furthermore, the internal components of the original model do not impact the simulation experiment and can be excluded. This study integrates the pertinent components of the ROV to assure the consistency of the reduced model. A basic ROV model is seen in Figure 9. The simplified model is derived by undergoing the process of simplification, and the fundamental parameters of the simplified model are shown in Table 6.

4.1.2 Computational domain division and boundary layer setting

In the simulation calculation, the length of the calculated domain is 9L (L is the length of ROV), the length of the ROV from the velocity inlet is 3L, the length from the pressure outlet is 5L, the height and width of the calculated domain are taken as 5L, and the ROV is placed on the third L. In order to easily control the

TABLE 5 Resistance calculation value and error.

Velocity(m/s)	3.045	5.144	6.091	7.161	8.231
Simulation values(N)	105.10	279.96	383.96	519.68	674.35
Experimental values(N)	102.30	283.80	389.93	528.89	680.14
Relative error(%)	2.737%	-1.353%	-1.531%	-1.741%	-0.851%

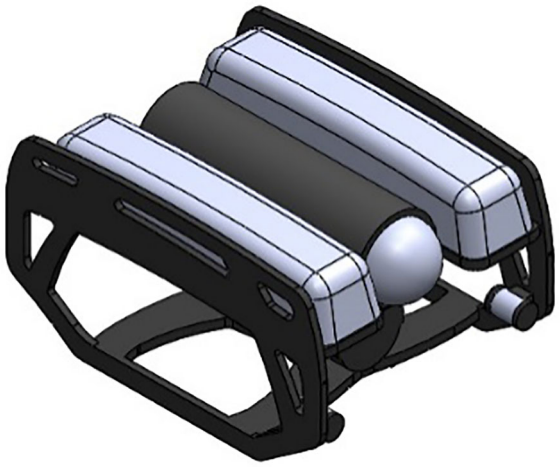


FIGURE 9
Simplified model.

mesh size of different regions, it is necessary to set up an internal calculation domain in the overall calculation domain. Referring to the division of the internal computing domain of the standard model in Section 3, this paper divides the internal computing domain into two modules: hull refinement area and wake refinement area. The overall computational domain in the ROV stationary motion simulation experiment is shown in Figure 10, and the internal computational domain is shown in Figure 11.

Based on the empirical formula, the thickness of the boundary layer is 5 mm. The number of boundary layers is generally determined according to the Reynolds number of the calculated working case, and for the movement under the high Reynolds number, the number of boundary layers is set to 5-10 layers. The mesh quality under different boundary layers is compared, and the number of boundary layers is finally determined to be 5.

4.2 Mesh type and mesh independence validation

4.2.1 Mesh type

The ROV model is relatively simple in the simulation calculations to be carried out in this section; the hexahedral mesh type is used. From the tank wall to the ROV surface, the produced

TABLE 6 ROV basic parameters.

Parameters	Numeric value
Length (mm)	475.2
Width (mm)	338.05
Altitude (mm)	253.85
surface area (mm^2)	918600
Lateral profile area (mm^2)	105600
Front view section area (mm^2)	48100
Top view section area (mm^2)	134600

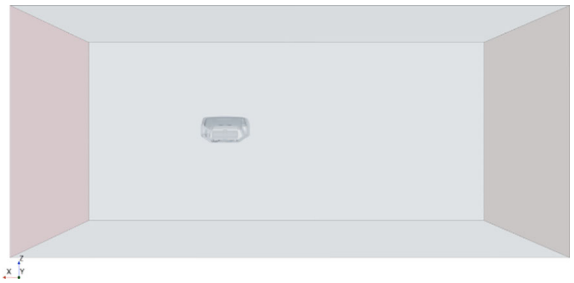


FIGURE 10
Overall calculation domain.

mesh is encrypted layer by layer, which reduces computation time and guarantees simulation calculation accuracy. Hexahedral mesh is also used in this paper's internal calculation domain, as the ROV model does not need complicated maneuvers in the stationary motion simulation calculation.

4.2.2 Mesh independence validation

In order to ensure the accuracy and precision of the simulation results, in addition to considering the mesh type, the number of meshes is also an important influencing factor after the mesh type and calculation domain size are determined. In this paper, a mesh independence verification is designed to determine the appropriate number of meshes. Seven mesh quantities of 780,000, 1.06 million, 1.54 million, 2 million, 2.21 million, 2.53 million and 3 million were selected for independent verification, and the ROV velocity was set to 1.5m/s. The ROV forward motion simulation experiments were carried out under different mesh numbers, and the resistance values under different mesh numbers were counted, as shown in Table 7. The data in the table is generated into a resistance graph, as shown in Figure 12.

Figure 12 illustrates that the resistance value rises with a decreasing number of meshes and tends to stabilize at a certain number of meshes. Due to the large mesh size, the reproduction degree of the ROV is limited when the number of meshes is fewer than one million, which causes the ROV simulation experiment to have a high resistance value. When there are more than 2.5 million meshes, the mesh size is smaller, the ROV form is better restored, and the ROV resistance value is more in line with the experimental value. Furthermore, the resistance value tends to remain stable when the mesh count exceeds 2.5 million. This is

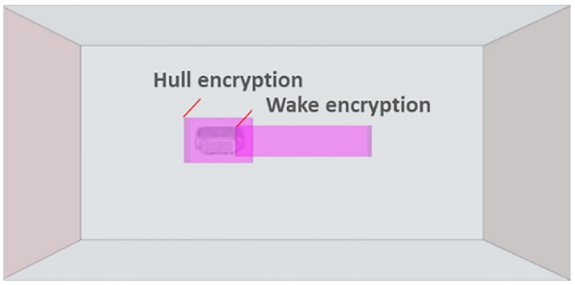


FIGURE 11
Internal computing domain.

TABLE 7 Different mesh resistance values.

Number of meshes/(10,000)	78	106	154	200	221	253	300
Resistance value/(N)	48.2	46.48	46.25	46.02	45.85	45.59	45.52
Relative error	—	3.567%	0.480%	0.493%	0.388%	0.560%	0.156%

because, beyond this threshold, further reductions in mesh size no longer enhance the reduction degree of ROV shape. Consequently, the resistance values obtained through simulation calculation under identical working conditions are congruent with the experimental values, and the resistance values remain stable. The number of meshes used in this article is 2.53 million in order to prevent having too many meshes affect the simulation computation and to take computer performance into consideration. The mesh for simulation calculations is completed based on the size of the calculation domain and the meshing form that was previously established. Figure 13 displays the longitudinal overall mesh; Figure 14 displays the head mesh of the RVO model; and Figure 15 displays the surface pressure distribution in the ROV.

4.3 Hydrodynamic calculations

4.3.1 Hydrodynamic calculations

The ROV resistance under different velocities was simulated and calculated. The simulation results at different velocities are shown in Table 8. The resistance values at different velocities in Table 7 are represented by graph lines, as shown in Figures 4–9.

Figure 16 shows that the difference between total resistance and compressive resistance grows with increasing velocity, but the values of both resistances increase linearly with increasing velocity. The fraction of pressure resistance in the overall resistance of the progressive voyage is much higher than the shear resistance. The pressure differential between the front and rear surfaces of the ROV rises as velocity increases, which also increases the pressure resistance since the compressive resistance is produced by this pressure difference between the surfaces during movement. In addition, as the velocity increases, the proportion of pressure resistance in total resistance is decreasing. Because the velocity gradient of the ROV surface flow field increases with increasing velocity, the value of the shear resistance is closely related to the velocity gradient of the flow field. In general, the increase in positive resistance is mainly due to the pressure resistance, and it is necessary to pay close attention to the shape of the front and rear surfaces of the ROV in the design of the ROV, and minimize the lateral water frontal area.

4.3.2 Hydrodynamic calculations for reverse navigation

Given the asymmetrical form of the ROV in relation to the mid-cross profile, it is necessary to treat the numerical modeling of the reverse journey condition as a distinct consideration in the straight-line experiment. Similar to the actual journey, the simulation trials are conducted at five different velocities. The boundary requirements for the rear simulation are identical to those for the positive course. However, the calculation region must be partitioned again to ensure that the tail of the ROV model is oriented towards the direction of the entering flow. Figure 17 displays the longitudinal overall mesh after resetting the computation domain and mesh in the reverse navigation scenario. Figure 18 displays the distribution of surface pressure. The computed outcomes for the five operational scenarios are shown in Table 9. The information presented in the table is visually represented in the form of a graph, as seen in Figure 19.

The shear resistance and pressure resistance in Tables 7, 8 are plotted and compared, and the results are shown in Figures 20, 21.

In Figure 20, the shear resistances during forward and backward sailing are seen to grow as the velocity increases. Additionally, the shear resistance values for both positive and backward sailing states are almost identical at the same velocity. This is due to the similarity in velocity gradient and shear resistance along the wall of the ROV, which remains consistent at the same velocity. Figures 21, 22 clearly

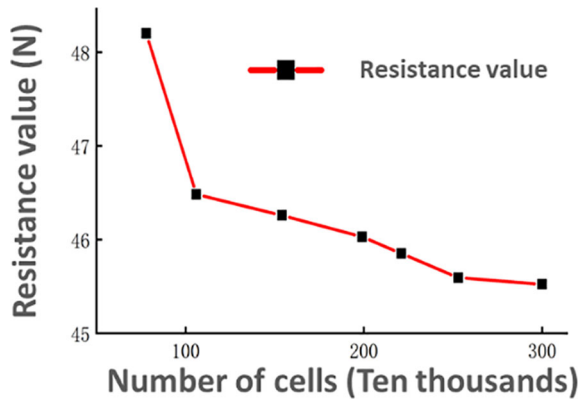


FIGURE 12
Resistance value changes with the number of meshes.

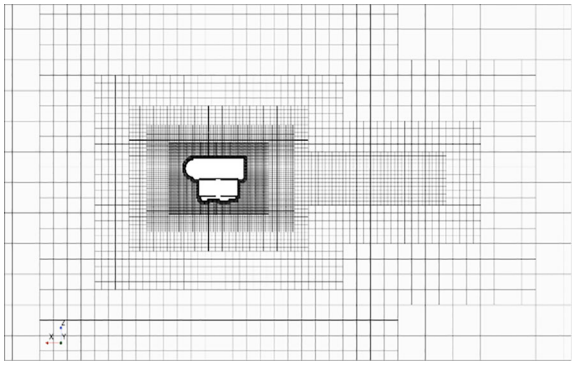


FIGURE 13
Longitudinal overall mesh of forward motion.

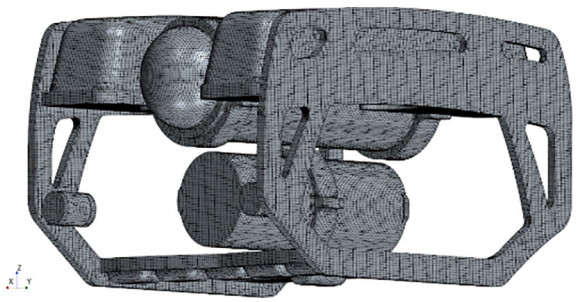


FIGURE 14
Header mesh.

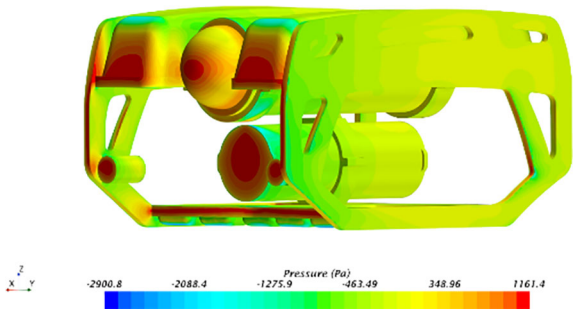


FIGURE 15
Surface pressure distribution of forward motion.

demonstrate a noticeable disparity in pressure resistance between the positive and backward navigation modes. Furthermore, the pressure resistance in both states escalates as the velocity rises. The compressive resistance of the ROV is influenced by its shape, particularly the asymmetry between the front and rear parts. As a result, the disturbance caused by the ROV to the surrounding flow field is different when it is moving forward compared to when it is moving in reverse. The reverse pressure resistance is greater than the positive pressure resistance at the same velocity, primarily due to the poor streamlining of the ROV tail. It can be concluded that when designing the ROV, the symmetry of the front and rear parts of the ROV should be ensured as much as possible while ensuring that the ROV shape is fluid. This shape distribution is conducive to reducing the straight-line resistance of the ROV and improving the propulsion efficiency.

4.3.3 Hydrodynamic calculations of ascent motion

The underwater submersible is capable of executing six degrees of freedom of motion underwater. In addition to the direct motion along the x-axis, the simulation computation for stable motion should also take into account the direct motion along the y-axis and z-axis directions. This chapter focuses on simulating and calculating the direct motion of the negative direction of the z-axis. The first step involves dividing the calculation domain and creating a mesh. The meshing results for the negative direction of the z-axis are shown in Figure 23. Furthermore, the insignificance of the mesh count is confirmed, and it is established that there are precisely 2.58 million meshes without upward motion. Figure 24 displays the surface pressure distribution of the ascending motion.

After setting the boundary conditions of the calculation model, four working conditions were selected for simulation calculation, and the simulation results are shown in Table 10.

The data in Table 9 are plotted as graphs, as shown in Figure 24.

Figure 25 demonstrates that the pressure resistance and shear resistance both escalate as the velocity of upward motion increases. During the upward motion, the ratio of pressure resistance to total resistance is higher compared to the motion along the x-axis, and the shear resistance is much reduced. Since the shape of the ROV is cuboidal, the waterfront area in the z-axis direction is larger than the waterfront area in the x-axis direction, the proportion of the pressure resistance caused by the pressure difference between the upper and lower surfaces in the z-axis direction in the total resistance is much greater than the ratio of the pressure resistance in the x-axis direction to the total resistance. When designing the ROV, it is necessary to comprehensively consider the ratio of the z-axis waterfront area to the x-axis waterfront area according to the working conditions of the ROV, which can effectively improve the hydrodynamic performance of the ROV.

4.3.4 Hydrodynamic calculations of sinking movements

The asymmetry of the top and lower sections of the ROV is contemplated, comparable to the simulation along the x-axis. This study presents a numerical calculation of motion along the axis in both positive and negative directions. Initially, the computational domain undergoes re-meshing, and the outcome of the meshing process is shown in Figure 26. Figure 27 displays the surface pressure distribution of the descending motion.

After the boundary conditions of the calculation model are set, the simulation of the four working conditions is carried out, and the simulation results are shown in Table 11.

The data in Table 10 are plotted as shown in Figure 28. Comparing the data in Tables 9, 10, Figures 29–31 is obtained.

By analyzing the data in Figures 29–31, it is assessed that in the ascending and sinking motions, the resistance of the ascending motion is less than that of the sinking motion. Different from the experimental data of forward and reverse sailing, the shear resistance

TABLE 8 Numerical calculation results of forward motion.

Velocity(m/s)	1.0	1.25	1.5	1.75	2.0
Pressure resistance(N)	-19.33	-30.20	-43.64	-59.74	-78.24
Shear resistance(N)	-0.93	-1.40	-1.96	-2.67	-3.42
Total resistance(N)	-20.26	-31.60	-45.60	-62.41	-81.66

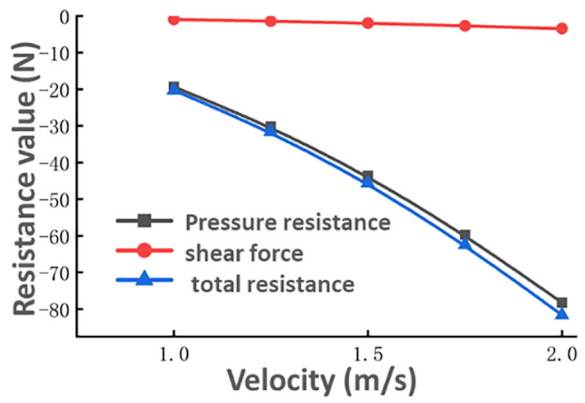


FIGURE 16
Forward resistance.

and pressure resistance in the ascending and descending motion have obvious changes at the same velocity. The change of shear resistance relative to the pressure resistance is small, and the pressure resistance accounts for a large proportion of the difference between the two states. In this regard, the causes of the upward and downward movements are analyzed in depth, and due to the asymmetry of the upper and lower parts of the ROV shape, the pressure difference between the upper and lower surfaces in the upward and downward movements is different, which makes the difference in the downward pressure resistance of the two working conditions. In addition, because the water facing area in the z-axis heave motion is greater than the forward navigation motion of the x-axis, the motion of the z-axis has a great influence on the flow field around the ROV, and the velocity gradient near the wall of the ROV changes significantly in the ascending and sinking motions, which leads to the difference between the shear resistances of the two motion states.

4.3.5 Hydrodynamic calculations for lateral movements

Considering that the left and right parts of the ROV are symmetrical, only one direction is selected for numerical calculation of the movement along the y-axis. The meshing and computational domains were re-meshed, and the meshing results are shown in Figure 32. The surface pressure distribution is shown in Figure 33.

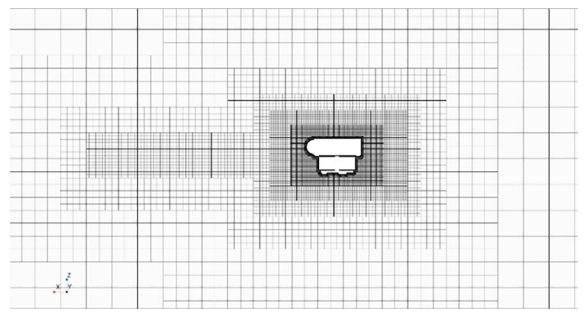


FIGURE 17
Inverted longitudinal overall mesh.

The simulation results of the four calculation cases are shown in Table 12.

The graph is plotted based on the data in Table 11, as shown in Figure 34.

In Figure 34, the resistance of the ROV transverse motion increases with the increase of velocity, and the pressure resistance occupies the main part of the total resistance. In lateral motion, there is a significant pressure difference between the frontal and backwater surfaces of the ROV, which is the main cause of the pressure resistance. Since the ROV is moving at a low velocity, the velocity gradient of the surrounding flow field is smaller, and the shear resistance caused by the velocity gradient of the flow field near the wall of the ROV is much smaller than the pressure resistance.

4.4 Data processing

4.4.1 Principles of data processing for stationary motion

Because the stationary motion is a uniform motion, the hydrodynamic term moving in all directions of the extension only has a velocity term and has nothing to do with the acceleration term. Simplifying the kinematic equations on the three axes yields Equation (18).

$$\begin{cases} F_x = -X_{uu}u^2 \\ F_y = -Y_{v|v|}\left|(v^2)^{\frac{1}{2}}\right| \\ F_z = -\left[Z'_{w|w|}w\left|(w^2)^{\frac{1}{2}}\right| + Z'_{ww}w\left|(w^2)^{\frac{1}{2}}\right|\right] \end{cases} \tag{18}$$

In Equation (18): F_x, F_y, F_z are numerical calculation of the resistance value; X'_{uu}, Z'_{ww}, $Y'_{v|v|}$, $Z'_{w|w|}$ are nonlinear hydrodynamic coefficients.

In order to facilitate data expression, the variables in the formula need to be dimensionless. Several variables in Equation (18) are dimensionless, as shown in Equation (19).

$$X_{uu} = \frac{\rho}{2}L^2X'_{uu},\quad Y_{v|v|} = \frac{\rho}{2}L^2Y'_{v|v|}$$

$$Z_{w|w|} = \frac{\rho}{2}L^2Z'_{w|w|},\quad Z_{ww} = \frac{\rho}{2}L^2Z'_{ww} \tag{19}$$

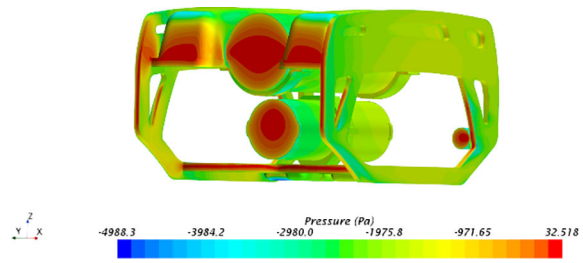


FIGURE 18
Surface pressure distribution of reverse motion.

TABLE 9 Calculation results of reverse flight value.

Velocity(m/s)	1.0	1.25	1.5	1.75	2.0
Pressure resistance(N)	-21.79	-34.01	-48.92	-66.59	-87.02
Shear resistance(N)	-0.93	-1.40	-1.96	-2.62	-3.36
Total resistance(N)	-22.72	-35.41	-50.88	-69.21	-90.38

In Equation: L for ROV length; ρ for the density of water. Bringing Equation (19) into Equation (18) gives a dimensionless expression, as shown in Equation (20).

$$\begin{cases} F_x = -\frac{\rho}{2}L^2 X'_{uu}u^2 \\ F_y = -\frac{\rho}{2}L^2 Y'_{v|v|}\left|(v^2)^{\frac{1}{2}}\right| \\ F_z = -\frac{\rho}{2}L^2\left[Z'_{w|w|}w\left|(w^2)^{\frac{1}{2}}\right| + Z'_{ww}w\left|(w^2)^{\frac{1}{2}}\right|\right] \end{cases} \tag{20}$$

Considering the external characteristics of the ROV, when moving laterally, its resistance values in the positive and negative directions are the same. However, when the z-axis is moving, the resistance values of the rising (w<0) and sinking (w>0) movements are different, and the resistance coefficients are also different. The resistance to the z-axis motion is expressed as shown in Equation (21).

$$= Z_{w|w|}w|w| + Z_{ww}ww = \begin{cases} (Z_{w|w|} + Z_{ww})w|w|, (w > 0) \\ (Z_{w|w|} - Z_{ww})w|w|, (w < 0) \end{cases} \tag{21}$$

Further analysis can be obtained the expression of the coefficient $Z_{w|w|}$ and Z_{ww} in Equation (22).

$$Z_{ww} = \frac{1}{2}\left(Z_{w|w|}^{(+)} + Z_{w|w|}^{(-)}\right)$$

$$Z_{w|w|} = \frac{1}{2}(Z_{w|w|}^{(+)} - Z_{w|w|}^{(-)}) \tag{22}$$

4.4.2 Fitting of the data

According to the above equation of motion, the data fitting results of hydrodynamic calculation are carried out by the least squares method, and MATLAB software is used in the data fitting, compared with the resistance curve in the simulation calculation, the fitting curve obtained by the least squares method is more accurate, as shown in Figures 35–37.

The values of each hydrodynamic coefficient could be derived based on the functional expression of the curve fitted to the data. The hydrodynamic coefficients of a statistical type are shown in Table 13.

5 Simulation calculation of the unsteady motion of the ROV

In the direct motion state, only the hydrodynamic coefficient related to the velocity can be obtained, and the hydrodynamic coefficient related to the acceleration cannot be obtained. In order to comprehensively analyze the motion of the ROV, it is not comprehensive to obtain only the coefficient related to the acceleration, and it is necessary to obtain the hydrodynamic coefficient of the acceleration term in order to accurately analyze the motion of the ROV.

5.1 The ROV model simplifies the computational domain division and boundary layer setting

5.1.1 Meshing and mesh-independent verification

5.1.1.1 Meshing

To analyze the unstable motion, it is essential to use overset meshes, which consist of two distinct sets of meshes that are separated independently and then nested and merged. When the

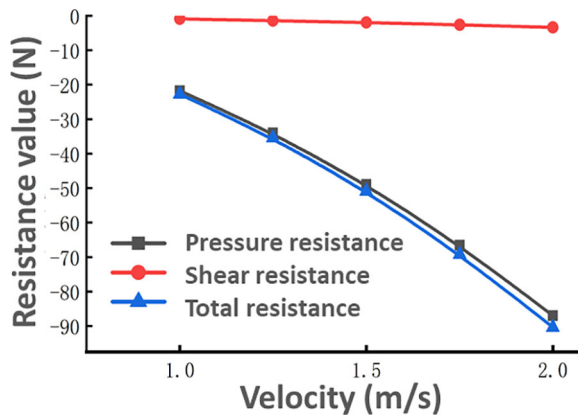


FIGURE 19
Inversion resistance.

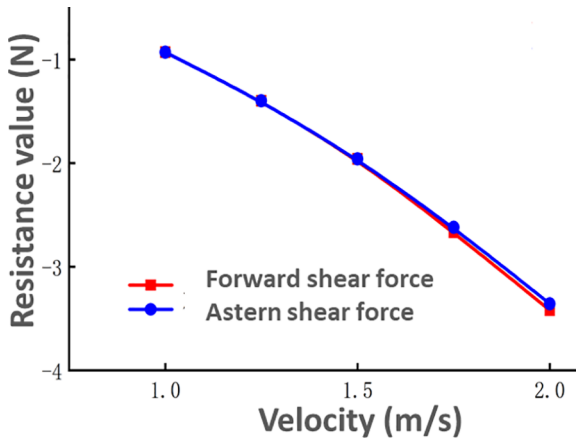


FIGURE 20
Shear resistance comparison.

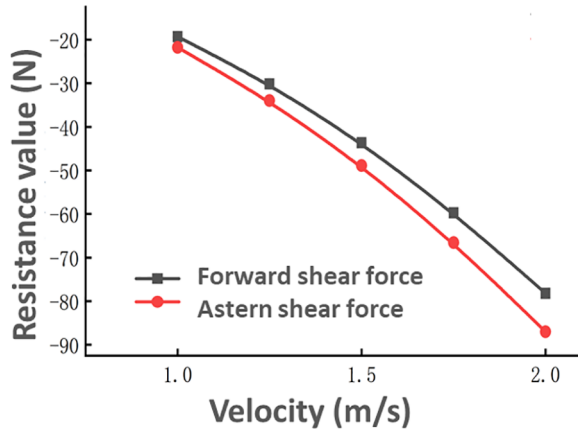


FIGURE 21
Comparison of pressure resistance.

overset mesh replicates intricate movement, the inner mesh propels the ROV to move in unison, while the outside mesh emulates the stationary motion inside the complicated movement. These two movements are interconnected to achieve the complex motion of the ROV. However, the use of an overset mesh necessitates the inclusion of two sets of meshes, resulting in a larger number of meshes. This increase is constrained by the limitations of computer performance. When employing the overset mesh calculation method, it is necessary to divide the mesh area into smaller sections in order to achieve greater precision. Ultimately, the determination of the number of meshes and the size of each mesh section should be based on a comprehensive assessment of calculation accuracy and computer performance. The mesh is divided into two areas, the inner mesh and the outer mesh, and in order to avoid the interpolation of the mesh affecting the calculation accuracy, the inner mesh must be set to ensure that it has a certain distance from the ROV surface. In this article, this distance is set to 0.25L. The outer mesh needs to be divided into multiple regions, which has the advantage of ensuring the calculation accuracy and effectively controlling the number of

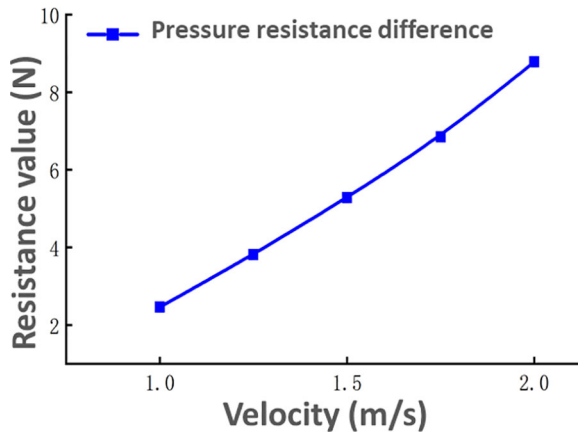


FIGURE 22
Differential pressure resistance.

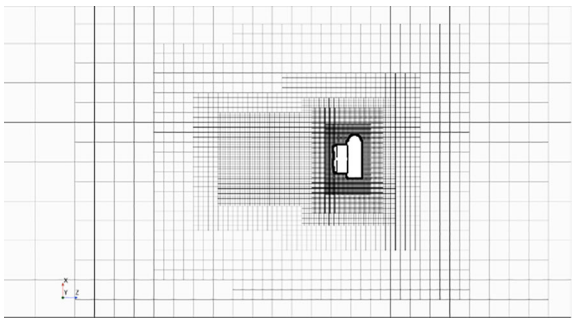


FIGURE 23
Vertical overall mesh of ascending motion.

meshes. Here, the outer mesh is divided into four parts: the motion area, the encryption area, the transition area, and the outer mesh area. When meshing, you need to set the mesh base size for different areas, as shown in Table 14.

5.1.1.2 Mesh independence validation

After determining the mesh type and calculation domain, in order to ensure the accuracy of the simulation results and the calculation accuracy, in addition to considering the mesh type, the number of meshes is also an important influencing factor. Before performing numerical simulation calculations, mesh independence verification is required to determine the appropriate number of meshes. In this paper, six mesh quantities of 950,000, 1.17 million, 1.54 million, 1.95 million, 2.46 million, and 3 million are selected for verification. According to the set mesh type and boundary conditions, a working condition in the heave motion is selected for verification, the velocity is set to V=1.5m/s, the frequency is f=0.3125, and the simulation values of the force under different mesh numbers are counted, and the calculation results are shown in Table 15.

According to the analysis of the change of force and mesh number in Figure 38, the Lateral force decreases with the increase of the number of meshes, and the resistance value tends to be flat when the number of meshes is greater than 2.46 million. When the number of meshes is less than 2 million, the mesh size is large,

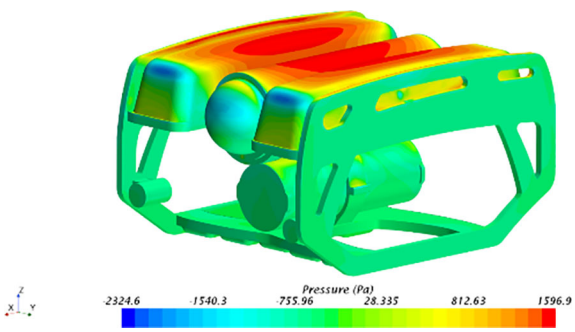


FIGURE 24
Pressure distribution on the surface of ascending motion.

TABLE 10 Numerical calculation results of ascending movement.

Velocity(m/s)	1.0	1.25	1.5	1.75
Pressure resistance(N)	64.161	100.264	145.230	198.090
Shear resistance(N)	0.627	0.928	1.276	1.668
Total resistance(N)	64.788	101.192	146.506	199.758

the surface restoration degree of the mesh to the ROV is poor, and the gap between the simulated value and the experimental value is large. When the number of meshes is greater than 2.67 million, the mesh size is small, and the mesh can better restore the shape and flow field of the ROV, and the resistance value obtained is also more accurate. Considering the requirements of computer performance and computational accuracy, the final number of meshes in this paper is 2.67 million.

5.2 Hydrodynamic calculations of planar motion mechanism

5.2.1 Definition and description of pure lateral motion

In the plane motion, there are two kinds of motion states, pure transverse motion and pure bow roll, in which the pure transverse motion is formed by the coupling of the uniform motion in the direction of the extended x-axis and the translational movement in the direction of the extended y-axis, and the motion is sinusoidally oscillating in the horizontal plane, and the angle between the bow and the x-axis is always zero. In the setting of this paper, according to the calculation results of the direct course resistance, the incoming flow velocity in the x-axis direction of the pure lateral motion is selected as V=1.5m/s. A schematic diagram of the motion of the pure lateral motion state, as shown in Figure 39.

The equation of motion for a pure lateral motion is shown in Equation (23).

$$\begin{cases} \eta = a\sin\omega t \\ \psi = \dot{\psi} = 0 \\ v = \dot{\eta} = a\omega\cos\omega t \\ \dot{v} = -a\omega^2\sin\omega t \end{cases} \tag{23}$$

In Equation (23), η—ROV lateral shift;
a —ROV Pure transverse amplitude;
ω —ROV Pure horizontal swing circle frequency;
ψ、 $\dot{\psi}$ —ROV Tilt angle and angular velocity about the z-axis;
v、 $\dot{v}$ —ROV Transverse velocity and acceleration.

The Lateral force Y and the yaw moment N are expressed using the velocity and acceleration terms, and the force and moment expressions are shown in Equation (24).

$$\begin{cases} Y = Y_{\dot{v}}\dot{v} + Y_v v + Y_0 \\ N = N_{\dot{v}}\dot{v} + N_v v + N_0 \end{cases} \tag{24}$$

Bringing Equation (23) into Equation (24), the expression of force and moment can be expressed by Equation (25).

$$\begin{cases} Y = Y_0 - a\omega^2 Y_{\dot{v}}\sin\omega t + a\omega Y_v\cos\omega t \\ N = N_0 - a\omega^2 N_{\dot{v}}\sin\omega t + a\omega N_v\cos\omega t \end{cases} \tag{25}$$

In order to simplify the writing and facilitate the subsequent data fitting, Equation (25) is further simplified to obtain Equation (26).

$$\begin{cases} Y = Y_0 + Y_a\sin\omega t + Y_b\cos\omega t \\ N = N_0 + N_a\sin\omega t + N_b\cos\omega t \end{cases} \tag{26}$$

The expression of the relationship between (25) and the coefficient in Equation (26) is represented by Equation (27).

$$Y_a = -a\omega^2 Y_{\dot{v}},\ Y_b = a\omega Y_v$$

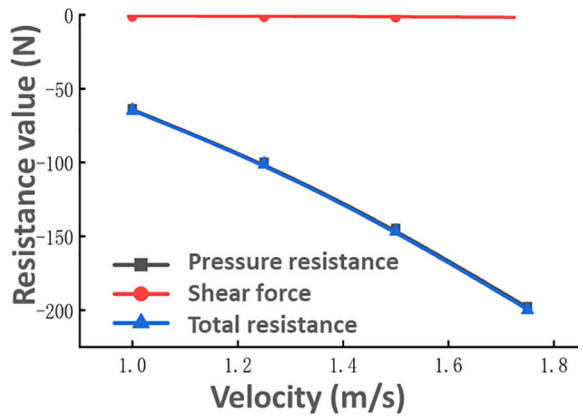


FIGURE 25
Ascending motion resistance.

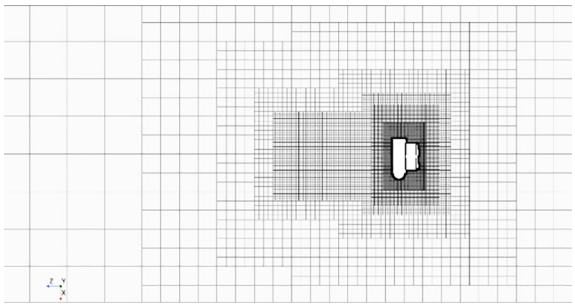


FIGURE 26
Vertical overall mesh of sinking motion.

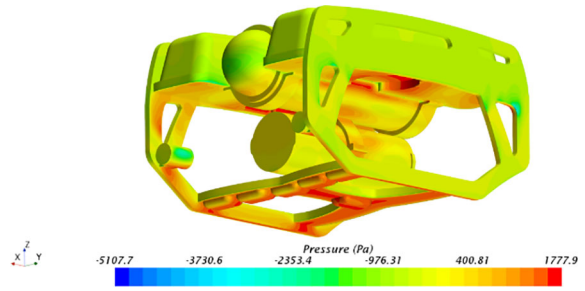


FIGURE 27
Surface pressure distribution of sinking motion.

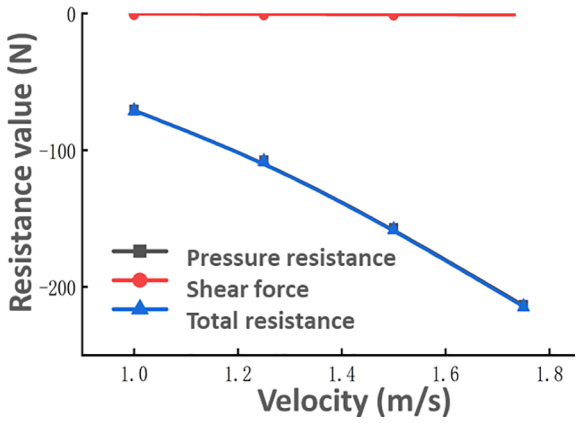


FIGURE 28
Sinking resistance.

$$N_a = -a\omega^2 N_{\dot{v}}, N_b = a\omega N_v \tag{27}$$

The hydrodynamic coefficients of the pure lateral motion are dimensionless, and the results are shown in Equation (28).

$$Y_{\dot{v}}' = \frac{Y_{\dot{v}}}{\frac{1}{2}\rho L^3}, Y_v' = \frac{Y_v}{\frac{1}{2}\rho L^2 U}, N_{\dot{v}}' = \frac{N_{\dot{v}}}{\frac{1}{2}\rho L^4}, N_v' = \frac{N_v}{\frac{1}{2}\rho L^3 U} \tag{28}$$

The hydrodynamic coefficients that need to be obtained in pure lateral motion are shown in Table 16.

5.2.2 Post-processing and data analysis of pure lateral movements

In the pure traverse motion simulation, according to the motion analysis of the ROV, the motion of the ROV needs to be expressed through the functional equation. Pure transverse motion is formed by superimposing a constant velocity motion in the x-axis direction and a variable velocity motion in the y-axis. In this paper, the flow velocity V=1.5m/s is obtained, and the motion of the y-axis can be defined by the field function. The motion of the ROV in the y-axis direction is programmed, and the field function is written according to the equation of motion of the ROV to prepare for the subsequent ROV motion setting. The frequencies were f=0.2, 0.25, 0.3125 and 0.4 respectively. In this paper, the flow velocity is taken as V=1.5m/s, and for the amplitude a, a=0.15m is selected based on referring to the predecessors. In this paper, the simulation time is 4T (T is the motion period). For the time step, T/500-T/300 is generally taken in the simulation experiment, and the time step in this paper is set to T/400 considering the performance limitation of the computer. The step sizes and calculation times for the four working conditions are listed in Table 17.

For the simulation data in unsteady motion, the corresponding hydrodynamic coefficient can be obtained through data processing, and it is not appropriate to directly use the least squares method to process the simulation data because the lateral force and the yaw moment are periodically varying in the pure lateral motion. Previous studies have shown that accurate hydrodynamic coefficients can be obtained by processing the periodic data by Fourier expansion. The Fourier expansion is shown in Equations (29), (30).

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\left(\frac{n\pi t}{l}\right) + b_n \sin\left(\frac{n\pi t}{l}\right)\right) \tag{29}$$

$$\begin{cases} a_n = \frac{1}{l}\int_{-l}^{l} f(t)\cos\left(\frac{n\pi t}{l}\right)dt, (n = 0,\ 1, 2\ldots) \\ b_n = \frac{1}{l}\int_{-l}^{l} f(t)\sin\left(\frac{n\pi t}{l}\right)dt, (n = 0,\ 1, 2\ldots) \end{cases} \tag{30}$$

Suppose the period is2l, because $\omega = \frac{2\pi}{T} = \frac{\pi}{l}$, then Equation (29) can be reduced to Equation (31).

$$f(t) = \frac{a_0}{2} + a_1 \cos(\omega t) + b_1 \sin(\omega t) + \sum_{n=2}^{\infty}(a_n \cos(n\omega t) + b_n \sin(n\omega t)) \tag{31}$$

Because n=2 and later terms have much less than the coefficient values a_1 and a_2, in omitted n=2 and after the item, Equation (32) can be obtained.

$$f(t) = \frac{a_0}{2} + a_1 \cos(\omega t) + b_1 \sin(\omega t) \tag{32}$$

Using the Fourier theorem to expand Equation (30), we get Equations (33), (34).

TABLE 11 Numerical calculation results of sinking motion.

Velocity(m/s)	1.0	1.25	1.5	1.75
Pressure resistance(N)	70.862	107.930	157.455	213.711
Shear resistance(N)	0.332	0.504	0.699	0.941
Total resistance(N)	71.194	108.434	158.154	214.652

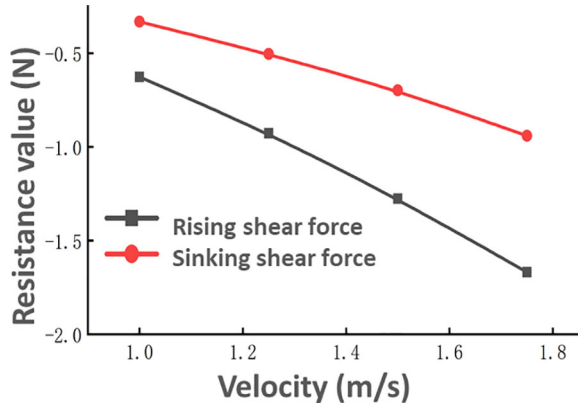


FIGURE 29
Shear resistance comparison.

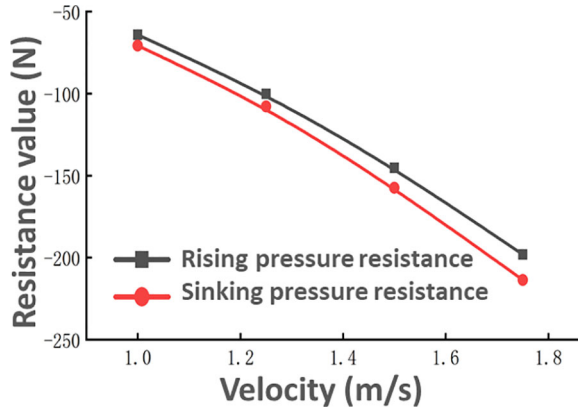


FIGURE 30
Comparison of pressure resistance.

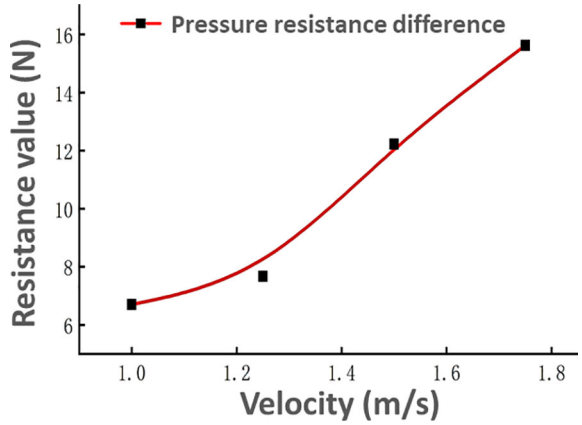


FIGURE 31
Pressure resistance difference.

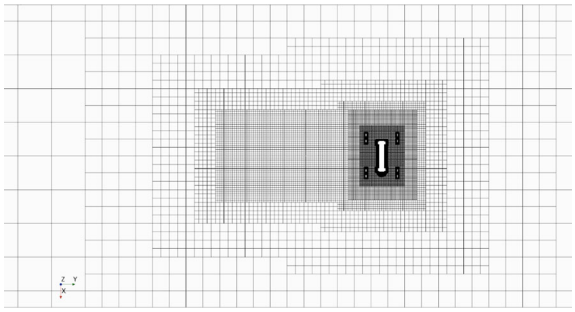


FIGURE 32
Vertical overall mesh of lateral motion.

For Lateral force Y:

$$\begin{cases} b_1 = Y_a = -a\omega^2 Y_v = \frac{1}{l}\int_{-l}^{l} f(t)\sin(\omega t)dt \\ a_1 = Y_b = a\omega Y_v = \frac{1}{l}\int_{-l}^{l} f(t)\cos(\omega t)dt \end{cases} \tag{33}$$

For yaw moment N:

$$\begin{cases} b_1' = N_a = -a\omega^2 N_v = \frac{1}{l}\int_{-l}^{l} f(t)\sin(\omega t)dt \\ a_1' = N_b = a\omega N_v = \frac{1}{l}\int_{-l}^{l} f(t)\cos(\omega t)dt \end{cases} \tag{34}$$

The simulation calculation of working conditions at different frequencies is carried out, and the stable data that can be used for data processing are selected from the simulation data, and the simulation data can be found to be stable after the second cycle through experiments. In this paper, the data of the third cycle are selected for analysis and processing, and the simulation data of the third cycle is fitted by Fourier series using MATLAB software. The coefficients under the Fourier series were obtained by fitting the curves, and the fitting coefficients under different working conditions were counted, and the statistical results are shown in Table 18.

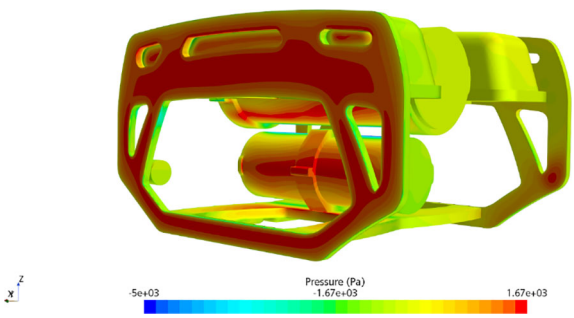


FIGURE 33
Pressure distribution on surface of lateral motion.

TABLE 12 Numerical calculation results of lateral movement.

Velocity(m/s)	1.0	1.25	1.5	1.75
Pressure resistance(N)	-55.385	-87.262	-125.136	-171.153
Shear resistance(N)	-0.485	-0.747	-1.078	-1.467
Total resistance(N)	-55.87	-88.009	-126.213	-172.62

According to Equations (33), (34), the data in the table are fitted by MATLAB software, and the hydrodynamic coefficients of the ROV in the unsteady state can be obtained through the quadratic fitting, and the fitting results are shown in Figures 40–43.

The hydrodynamic coefficients obtained are shown in Table 19.

5.2.3 Definition and description of pure yaw motion

Pure yaw motion refers to superimposing a bow motion on the basis of pure lateral motion, and ensuring that the direction of the velocity of the ROV movement is tangent to the trajectory. And in the follower coordinate system, the lateral motion velocity and lateral acceleration of the follower coordinate system are zero. Through the pure yaw motion experiment, the hydrodynamic coefficients of the underwater robot in relation to angular velocity and angular acceleration can be obtained. As with the pure lateral motion, the incoming velocity is set to V=1.5m/s in this paper. A schematic diagram of a pure yaw motion is shown in Figure 44.

The pure yaw motion can be expressed by the equation of motion, which can be expressed as Equation (35).

$$\begin{cases} \psi = \psi_0 \sin(\omega t) \\ r = \dot{\psi} = \psi_0 \omega \cos(\omega t) \\ \dot{r} = -\psi_0 \omega^2 \sin(\omega t) \\ v = \dot{v} = 0 \end{cases} \tag{35}$$

In Equation (35):

ψ—amplitude of pure yaw motion;

ψ_0—amplitude of pure bow shake movement;

ω—the circular frequency of the pure yaw motion;

$r, \dot{r}$—the angle of rotation about the z-axis and the angular velocity;

$v, \dot{v}$—Lateral velocity and acceleration.

After analyzing the pure yaw motion of the ROV, the lateral force Y and the yaw moment N are expressed by angular velocity and angular acceleration, and the expressions of force and moment are obtained, as shown in Equation (36).

$$\begin{cases} Y = Y_0 + Y_{\dot{r}}\dot{r} + Y_r r \\ N = N_0 + N_{\dot{r}}\dot{r} + N_r r \end{cases} \tag{36}$$

If the parametric expression in Equation (54) is brought into Equation (55), then the expression of force and moment can be

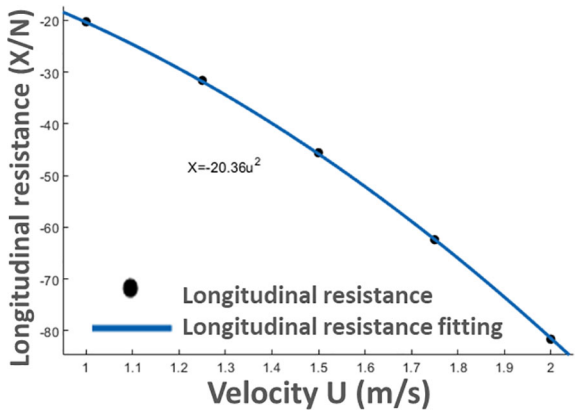


FIGURE 35
Resistance fitting curve.

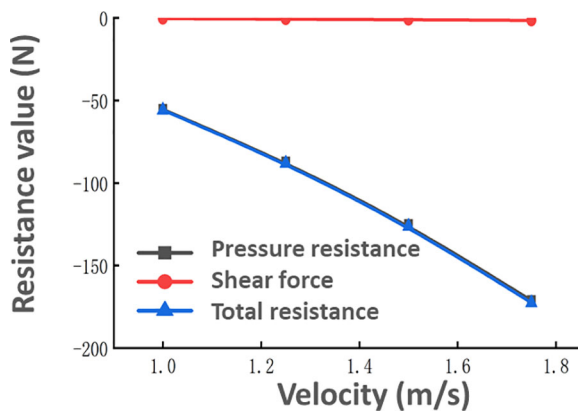


FIGURE 34
Lateral motion resistance.

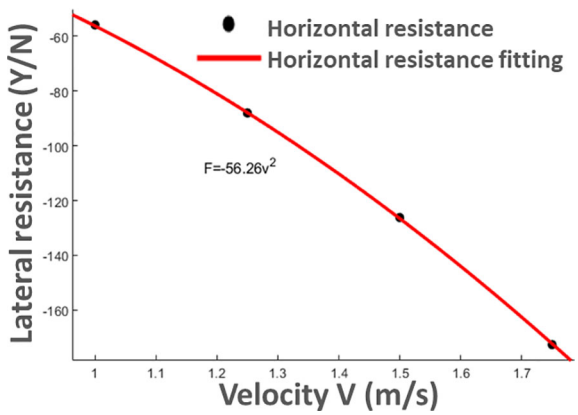


FIGURE 36
y-axis resistance fitting curve in the y-axis direction.

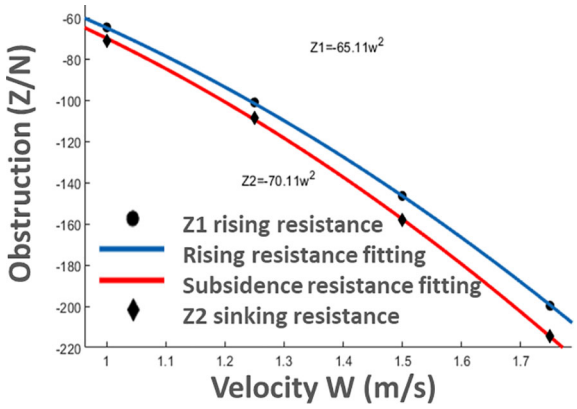


FIGURE 37
Fitting curve of ascending and sinking resistance.

expressed as Equation (37).

$$\begin{cases} Y = Y_0 - \psi_0 \omega^2 Y_{\dot{r}} \sin \omega t + \psi_0 \omega Y_r \cos \omega t \\ N = N_0 - \psi_0 \omega^2 N_{\dot{r}} \sin \omega t + \psi_0 \omega N_r \cos \omega t \end{cases} \quad (37)$$

In order to simplify the writing and facilitate the subsequent data processing, Equation (37) is simplified to obtain Equation (38).

$$\begin{cases} Y = Y_0 + Y_c \sin \omega t + Y_d \cos \omega t \\ N = N_0 + N_c \sin \omega t + N_d \cos \omega t \end{cases} \quad (38)$$

The relationship between Equation (36) and the coefficient of the corresponding term in Equation (38) can be expressed by Equation (39).

$$Y_c = -\psi_0 \omega^2 Y_{\dot{v}}, Y_d = \psi_0 \omega Y_v \quad (39)$$

The hydrodynamic coefficients were dimensionless respectively, and the results are shown in Equation (40).

$$Y_{\dot{r}}' = \frac{Y_{\dot{r}}}{\frac{1}{2}\rho L^4}, Y_r' = \frac{Y_r}{\frac{1}{2}\rho L^3 U}, N_{\dot{r}}' = \frac{N_{\dot{r}}}{\frac{1}{2}\rho L^5}, N_r' = \frac{N_r}{\frac{1}{2}\rho L^4 U} \quad (40)$$

The hydrodynamic coefficients that need to be obtained in a pure bow motion are shown in Table 20.

5.2.4 Post-processing and data analysis of pure bow motion

Initially, it is necessary to establish a fresh coordinate system whereby the point of origin coincides with the location of the center of gravity. The equation of motion of the pure bow jolt should be programmed first, as with the pure yaw motion, and the code acquired by programming is saved in the field function to prepare for the succeeding motion settings. Following the completion of the overlap area's motion configuration, simulation and calculation are required for the four working conditions that were selected: f=0.2, 0.25, 0.3125, and 0.4. The entering velocity in pure bow motion is fixed at V=1.5m/s. The time step size and calculation time remain the same as those used for pure sideways motion. For a pure yaw motion, the Lateral force and the yaw moment are expanded in Fourier series, as shown in Equations (41), (42).

For Lateral force Y:

$$\begin{cases} c_1 = Y_c = -\psi_0 \omega^2 Y_{\dot{r}} = \frac{1}{l} \int_{-l}^{l} f(t) \sin(\omega t) dt \\ d_1 = Y_d = \psi_0 \omega Y_r = \frac{1}{l} \int_{-l}^{l} f(t) \cos(\omega t) dt \end{cases} \quad (41)$$

For yaw moment N:

$$\begin{cases} c_1' = N_c = -\psi_0 \omega^2 N_{\dot{r}} = \frac{1}{l} \int_{-l}^{l} f(t) \sin(\omega t) dt \\ d_1' = N_d = \psi_0 \omega N_r = \frac{1}{l} \int_{-l}^{l} f(t) \cos(\omega t) dt \end{cases} \quad (42)$$

As with the pure lateral motion, the data of the third period are selected for data fitting, and the fitting coefficients under different working conditions in the Fourier fitting are counted, as shown in Table 21.

TABLE 13 Statistics of hydrodynamic coefficients of steady motion.

Hydrodynamic coefficient	Numeric value	Non -sub-traction coefficient	Numeric value
X_{uu}	-20.36	X_{uu}'	-0.1955
$Y_{v\|v\|}$	-56.26	$Y_{v\|v\|}'$	-0.5401
$Z_{w\|w\|}$	-67.61	$Z_{w\|w\|}'$	-0.6490
Z_{ww}	-2.5	Z_{ww}'	-0.0240

TABLE 14 Meshing settings.

Mesh Area	Mesh Name	Mesh Size
Inner Mesh	Overset mesh	3.125%X
Outer mesh	Movement area	3.125%X
	Encrypted area	6.25%X
	Transition area	12.5%X
	Exterior area	400%X

The mesh size is expressed as a percentage of the base size, and X represents the base size value in the mesh settings.

TABLE 15 Vertical force of different meshes.

Number of meshes/(10,000)	95	117	154	195	267	300
Lateral force/(N)	-44.57	-44.81	-45.23	-45.71	-45.98	-46.05
Relative error	—	-0.538%	-0.937%	-1.061%	-0.591%	-0.152%

According to Equations (41), (42), the data in the table are quadratically fitted by MATLAB software, and the hydrodynamic coefficients of pure yaw motion can be obtained through quadratic fitting, and the results of quadratic fitting are shown in Figures 44–48.

The hydrodynamic coefficients and dimensionless coefficients obtained by quadratic fitting are counted, and the statistical results are shown in Table 22.

5.3 Hydrodynamic calculation of the vertical plane motion mechanism

5.3.1 Definition and description of pure heave motion

The motion of the vertical plane mechanism includes two kinds of motions: pure heave and pure pitching, and the pure heave motion refers to the combined motion formed by the superposition of the uniform motion of the x-axis and the variable velocity motion in the z-axis direction, and the trajectory of the pure heave motion in the xoz plane is a sine wave, and the angle between the bow and the x-axis of the ROV is always kept at zero degrees. In this paper, the constant motion of the extended x-axis is set to V=1.5m/s. A schematic diagram of the pure heave motion is shown in Figure 49.

The pure heave motion can be expressed by a parametric equation as shown in Equation (43).

$$\begin{cases} \zeta = a\sin(\omega t) \\ \theta = \dot{\theta} = 0 \\ w = \dot{\zeta} = a\omega\cos(\omega t) \\ \dot{w} = -a\omega^2\sin(\omega t) \end{cases} \tag{43}$$

In Equation (43):

ζ—ROV vertical displacement;

a—penchant amplitude;

ω—circular frequency;

θ, $\dot{\theta}$—angular velocity about the y-axis;

w, $\dot{w}$—Vertical velocity and acceleration.

The vertical force Z and the pitching moment M are expressed by the terms velocity and acceleration, and the expression for the force and moment is shown in Equation (44).

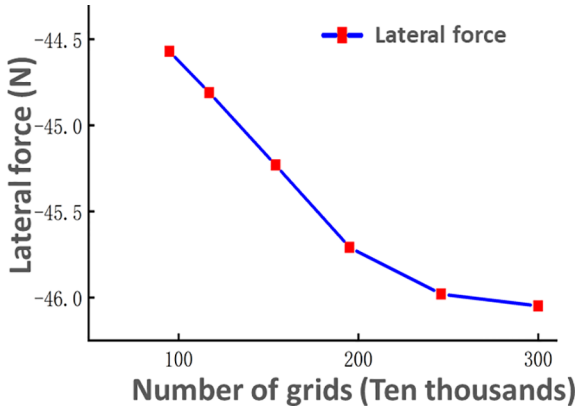


FIGURE 38
Force changes with the number of meshes.

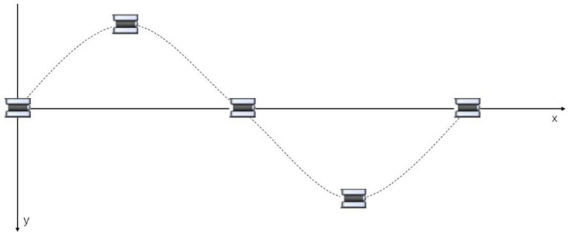


FIGURE 39
Schematic diagram of pure swaying motion.

TABLE 16 Hydrodynamic coefficients of pure transverse motion.

Lateral force coefficients		Yaw moment coefficients	
Y_v	$Y_{\dot{v}}$	N_v	$N_{\dot{v}}$
$Y_v^{'}$	$Y_{\dot{v}}^{'}$	$N_v^{'}$	$N_{\dot{v}}^{'}$

$$\begin{cases} Z = Z_0 + Z_{\dot{w}}\dot{w} + Z_w w \\ M = M_0 + M_{\dot{w}}\dot{w} + M_w w \end{cases} \tag{44}$$

Transporting Equation (43) into Equation (44) allows for a pure equation of motion for heave and heave, as shown in Equation (45).

$$\begin{cases} Z = Z_0 - a\omega^2 Z_{\dot{w}}\, sin\,(\omega\, t) + a\omega Z_w\, cos\,(\omega\, t) \\ M = M_0 - a\omega^2 M_{\dot{w}}\, sin\,(\omega\, t) + a\omega M_w\, cos\,(\omega\, t) \end{cases} \tag{45}$$

In order to simplify the writing and facilitate subsequent data processing, Equation (45) is simplified, and the simplified equation is shown in Equation (46).

$$\begin{cases} Z = Z_0 + Z_1\, sin\,\omega\, t + Z_2\, cos\,\omega\, t \\ M = M_0 + M_1\, sin\,\omega\, t + M_2\, cos\,\omega\, t \end{cases} \tag{46}$$

The expression of the relationship between Equation (45) and the coefficient of the corresponding term in Equation (46) is represented by Equation (47).

$$Z_1 = -a\omega^2 Z_{\dot{w}}, Z_2 = a\omega Z_w$$

$$M_1 = -a\omega^2 M_{\dot{w}}, W_2 = a\omega M_w \tag{47}$$

The hydrodynamic coefficients in Equation (45) are dimensionless respectively, and the dimensionless mode is shown in Equation (48).

$$Z_{\dot{w}}^{'} = \frac{Z_{\dot{w}}}{\frac{1}{2}\rho L^3}, Z_w^{'} = \frac{Z_w}{\frac{1}{2}\rho L^2 U}, M_{\dot{w}}^{'} = \frac{M_{\dot{w}}}{\frac{1}{2}\rho L^4}, M_w^{'} = \frac{M_w}{\frac{1}{2}\rho L^3 U} \tag{48}$$

The hydrodynamic coefficients that need to be obtained in the pure heave and heave are shown in Table 23.

5.3.2 Post-processing and data analysis of pure heave motion

In pure heave motion, the trajectory of the ROV in the xoz plane is a sine wave. As with pure transverse motion, the sinusoidal motion is decomposed into a uniform motion in the direction of the x-axis and a variable velocity motion in the direction of the z-axis. For the constant velocity motion in the x-axis direction, the velocity V=1.5m/s, and the variable velocity motion in the z-axis direction is represented by a mathematical function.

In order to facilitate the setting of the velocity of the variable velocity movement, it is necessary to encode in the field function and write the expression of the function that controls the variable velocity motion. Once the code has been programmed in the field function, you can continue to set the variable velocity motion in the z-axis direction of the overlap area. After the motion setting of the overlap area is completed, four working conditions with frequencies of f=0.2, 0.25, 0.3125 and 0.4 are selected for calculation. The incoming velocity is set to V=1.5m/s, and the amplitude is consistent with the pure lateral motion, and a=0.15m is selected. In order to obtain stable data, the calculation period is selected as four periods, and the data of the third period is selected for data fitting. Time step, set to T/400.

Considering the periodicity of the data, it is necessary to use the Fourier series expansion for the data obtained in the simulation calculation, and the Fourier series expansion for the vertical force Z and the pitching moment M is shown in Equations (48), (59).

For the vertical force Z:

$$\begin{cases} e_1 = Z_1 = -a\omega^2 Z_{\dot{w}} = \frac{1}{l}\int_{-l}^{l} f(t)\, sin\,(\omega\, t)dt \\ f_1 = Z_2 = a\omega Z_w = \frac{1}{l}\int_{-l}^{l} f(t)\, cos\,(\omega\, t)dt \end{cases} \tag{49}$$

TABLE 17 Initial value setting.

Frequency(f/HZ)	Cycle (T/s)	$\omega = 2\pi f$ (1/s)	Step(s)	Simulation time
0.2	5	1.2566	0.01250	20
0.25	4	1.5708	0.01000	16
0.3125	3.2	1.9635	0.00800	12.8
0.4	2.5	2.5133	0.00625	10

TABLE 18 Pure sway calculation data.

Frequency f(HZ)	$\omega = 2\pi f$ (1/s)	$a\omega$	$-a\omega^2$	Y_a	Y_b	N_a	N_b
0.2000	1.2566	0.1885	-0.2369	3.9270	-16.3500	0.2140	0.0502
0.2500	1.5708	0.2356	-0.3701	5.5970	-21.5800	0.2925	0.0692
0.3125	1.9635	0.2945	-0.5783	8.3060	-28.0900	0.3872	-0.0921
0.4000	2.5133	0.3770	-0.9475	13.0800	-38.2000	0.4806	-0.5153

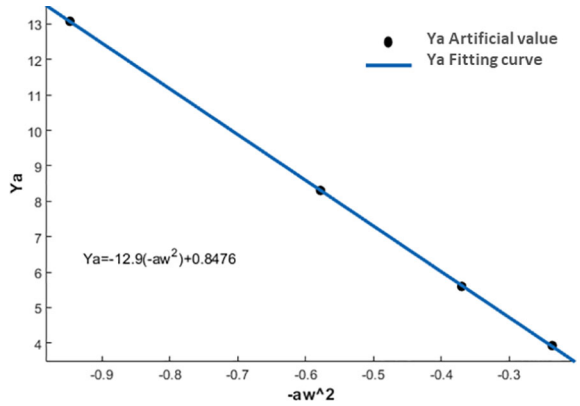


FIGURE 40
Ya fitting curve.

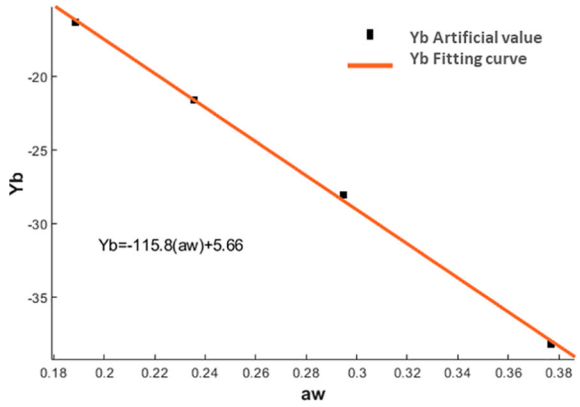


FIGURE 41
Yb fitting curve.

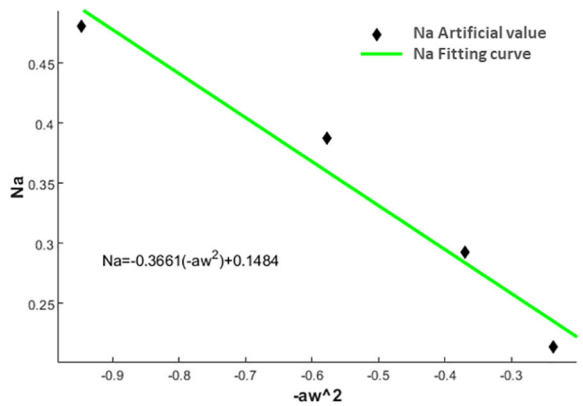


FIGURE 42
Na fitting curve.

For the yaw moment M:

$$\begin{cases} e_1^{'} = M_1 = -a\omega^2 M_{\dot{w}} = \frac{1}{l}\int_{-l}^{l} f(t)\,sin\,(\omega\,t)dt \\ f_1^{'} = M_2 = a\omega M_w = \frac{1}{l}\int_{-l}^{l} f(t)\,cos\,(\omega\,t)dt \end{cases} \tag{50}$$

After calculating the working conditions at different frequencies, the data of the third period were selected from the obtained data for processing, and the data were fitted by Fourier series using MATLAB software. The coefficients of the fitting curves under different working conditions are statistically shown in Table 24.

According to Equations (49), (50), the data in the table are fitted by MATLAB software, and the hydrodynamic coefficients in the pure heave motion can be obtained through the secondary fitting, and the results of the second fitting are shown in Figures 50–53.

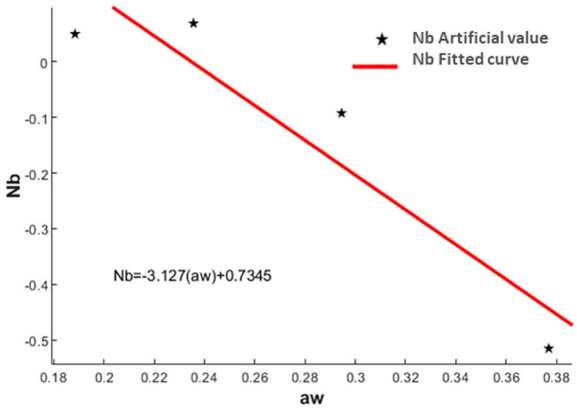


FIGURE 43
Nb fitting curve.

TABLE 19 Statistics of pure transverse hydrodynamic coefficients.

Lateral force coefficient				Yaw moment coefficient			
Y_v	-115.8	$Y_{\dot{v}}$	-12.9	N_v	-3.127	$N_{\dot{v}}$	-0.3661
$Y_v^{'}$	-0.7411	$Y_{\dot{v}}^{'}$	-0.2710	$N_v^{'}$	-0.0438	$N_{\dot{v}}^{'}$	-0.0168

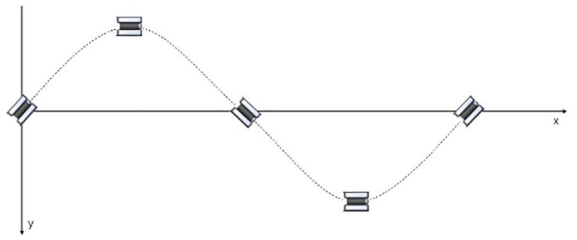


FIGURE 44
Pure yow motion.

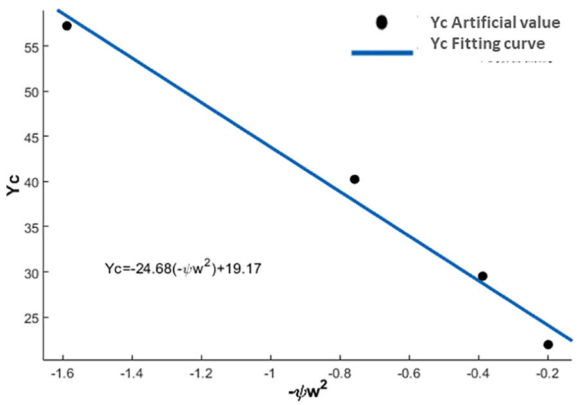


FIGURE 45
Yc fitting curve.

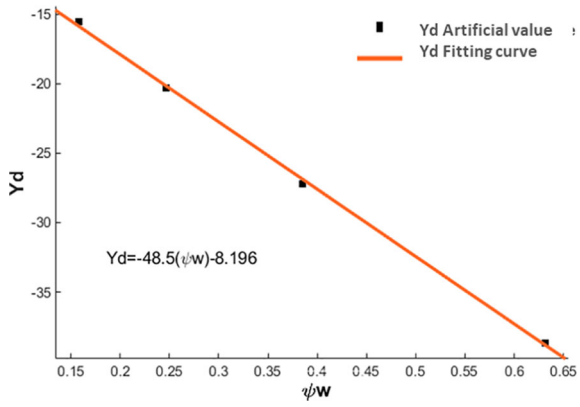


FIGURE 46
Yd fitting curve.

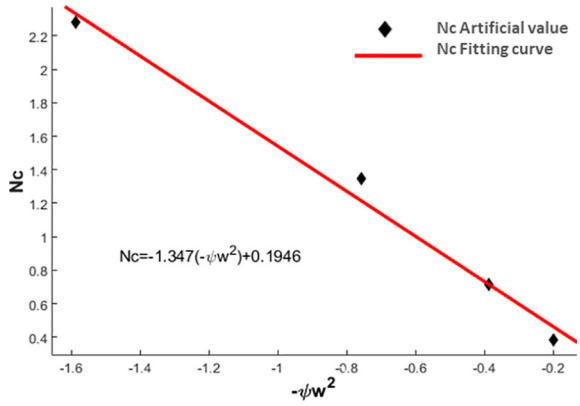


FIGURE 47
Nc fitting curve.

The hydrodynamic coefficients obtained by quadratic fitting and their dimensionless values are statistically analyzed, and the statistical results are shown in Table 25.

5.3.3 Definition and description of pure pitching motion

Pure pitch motion is a combination of a rotational motion about the y-axis on the basis of pure lateral motion. In motion, the direction of velocity of the ROV is tangent to the sinusoidal trajectory. In the follow-up coordinate system, the vertical velocity and acceleration of the ROV are zero. With pure pitching motion, the hydrodynamic coefficients related to angular velocity and angular acceleration of the underwater robot can be obtained, assuming that the velocity of the moving flow is V=1.5m/s. A schematic diagram of pure pitch motion is shown in Figure 54.

TABLE 20 Pure yaw motion coefficients.

Lateral force coefficient		Yaw moment coefficient	
Y_r	$Y_{\dot{r}}$	N_r	$N_{\dot{r}}$
Y_r'	$Y_{\dot{r}}'$	N_r'	$N_{\dot{r}}'$

The pure pitch motion of the ROV can be expressed by the parametric equation, as shown in Equation (51).

$$\begin{cases} \theta = \theta_0 \sin(\omega t) \\ w = \dot{w} = 0 \\ q = \dot{\theta} = \theta_0 \omega \cos(\omega t) \\ \dot{q} = -\theta_0 \omega^2 \sin(\omega t) \end{cases} \tag{51}$$

In Equation (51):
θ—Pitch angle;
θ_0—Pitching motion amplitude;
ω—Circular frequency;
q, $\dot{q}$—Angle and angular velocity about the y-axis;
w, $\dot{w}$—Vertical velocity and acceleration.

The equations for the ROV pitch moment M and the vertical force Z can be expressed by the pitch angle and the pitch angular velocity, as shown in Equation (52).

TABLE 21 Calculation data of pure yaw motion.

Frequency f (HZ)	$\omega = 2\pi f$ (1/s)	$\psi_0\omega$	$-\psi_0\omega$	Y_c	Y_d	N_c	N_d
0.2000	1.2566	0.1579	-0.1984	21.9400	-15.5500	0.3853	-0.2365
0.2500	1.5708	0.2467	-0.3876	29.5200	-20.3400	0.7122	-0.7061
0.3125	1.9635	0.3855	-0.7570	40.2600	-27.2000	1.3480	-1.4570
0.4000	2.5133	0.6317	-1.5876	57.2700	-38.6500	2.2790	-2.8750

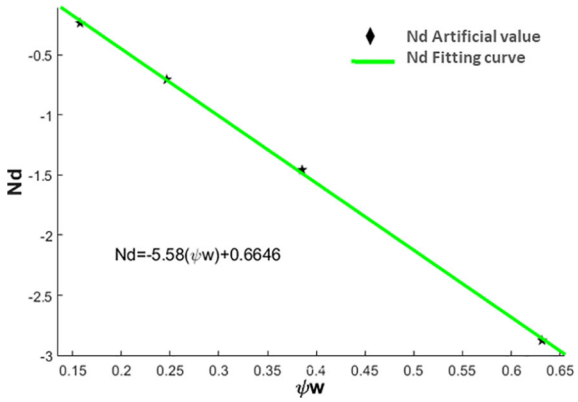


FIGURE 48
Nd fitting curve.

TABLE 22 Statistics of pure motion hydrodynamic coefficient.

Lateral force coefficient				Yaw moment coefficient			
Y_r	-48.500	$Y_{\dot{r}}$	-24.680	N_r	-5.5800	$N_{\dot{r}}$	-1.3470
$Y_r^{'}$	-0.6792	$Y_{\dot{r}}^{'}$	-1.1344	$N_r^{'}$	-0.1710	$N_{\dot{r}}^{'}$	-0.1355

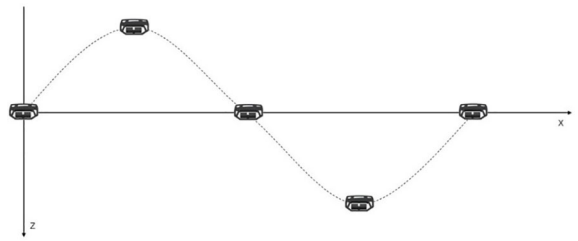


FIGURE 49
Pure heave motion.

TABLE 23 Pure heave hydrodynamic coefficient.

Vertical force coefficient		Pitch moment coefficient	
Z_w	$Z_{\dot{w}}$	M_w	$M_{\dot{w}}$
$Z_w^{'}$	$Z_{\dot{w}}^{'}$	$M_w^{'}$	$M_{\dot{w}}^{'}$

$$\begin{cases} Z = Z_0 + Z_{\dot{q}}\dot{q} + Z_q \\ M = M_0 + M_{\dot{q}}\dot{q} + M_q \end{cases} \tag{52}$$

Bringing the equation of pure pitch motion into Equation (52) gives Equation (53).

$$\begin{cases} Z = Z_0 - \theta_0\omega^2 Z_{\dot{q}}\, sin\,(\omega\, t) + \theta_0\omega Z_q\, cos\,(\omega\, t) \\ M = M_0 - \theta_0\omega^2 M_{\dot{q}}\, sin\,(\omega\, t) + \theta_0\omega M_q\, cos\,(\omega\, t) \end{cases} \tag{53}$$

TABLE 24 Pure heave calculation data.

Frequency f(HZ)	ω = 2πf (1/s)	aω	$-a\omega^2$	Z_1	Z_2	M_1	M_2
0.2000	1.2566	0.1885	-0.2369	7.8060	-21.4300	-0.2524	1.6190
0.2500	1.5708	0.2356	-0.3701	11.6100	-28.4800	-0.3763	2.0220
0.3125	1.9635	0.2945	-0.5783	15.9700	-38.4700	-0.5651	2.5040
0.4000	2.5133	0.3770	-0.9475	23.4300	-54.2500	-0.8938	3.2090

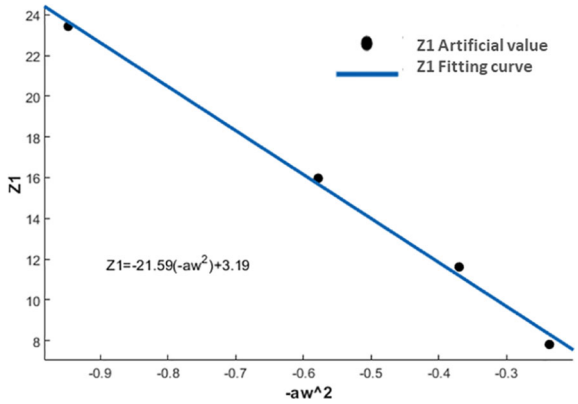


FIGURE 50
Z1 fitting curve.

In order to simplify the writing and facilitate the subsequent data processing, Equation (53) is simplified to obtain Equation (54).

$$\begin{cases} Z = Z_0 + Z_3 \sin \omega t + Z_4 \cos \omega t \\ M = M_0 + M_3 \sin \omega t + M_4 \cos \omega t \end{cases} \tag{54}$$

The relationship between Equation (53) and the coefficient of the corresponding term in Equation (54) can be expressed by Equation (55).

$$Z_3 = -\theta_0 \omega^2 Z_{\dot{q}}, Z_4 = \theta_0 \omega Z_q$$

$$M_3 = -\theta_0 \omega^2 M_{\dot{q}}, W_4 = \theta_0 \omega M_q \tag{55}$$

The hydrodynamic coefficient of pure pitching motion and its dimensionless value are shown in Equation (56).

$$Z'_{\dot{q}} = \frac{Z_{\dot{q}}}{\frac{1}{2}\rho L^4}, Z'_q = \frac{Z_q}{\frac{1}{2}\rho L^3 U}, M'_{\dot{q}} = \frac{M_{\dot{q}}}{\frac{1}{2}\rho L^5}, M'_q = \frac{M_q}{\frac{1}{2}\rho L^4 U} \tag{56}$$

The hydrodynamic coefficients that need to be obtained in pure pitching and sinking are shown in Table 26.

5.3.4 Post-processing and data analysis of pure pitching movements

Pure pitch motion is made by superimposing a rotational angular velocity on the basis of pure heave motion, and in pure pitch motion, the

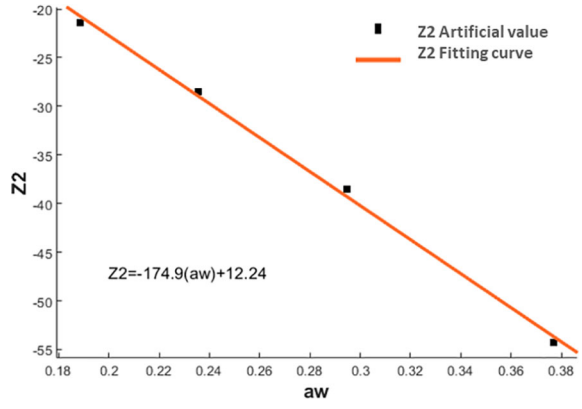


FIGURE 51
Z2 fitting curve.

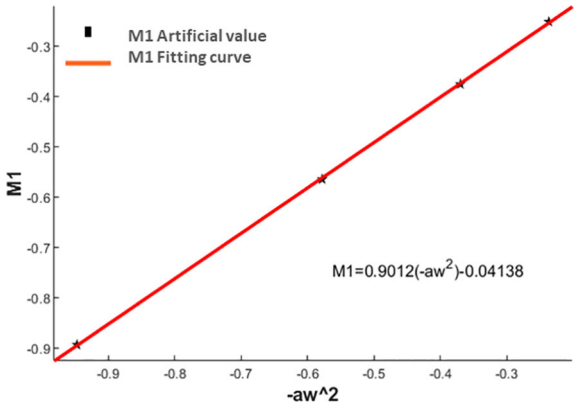


FIGURE 52
M1 fitting curve.

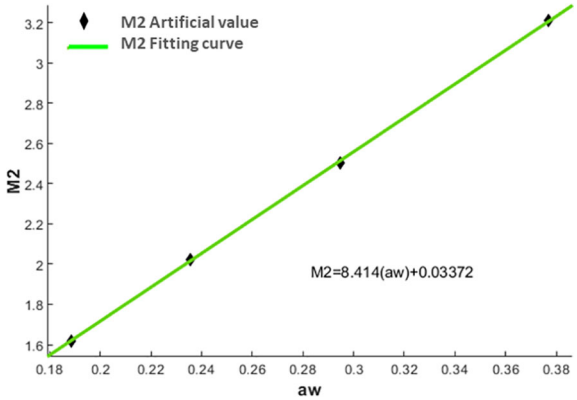


FIGURE 53
M2 fitting curve.

trajectory of the center of gravity is a sine wave. In this paper, the flow velocity is set to V=1.5m/s, and the variable velocity motion in the z-axis direction is set to be the same as the pure heave motion. In addition, an angular acceleration needs to be superimposed on the ROV, and the angular acceleration expression is shown in Equation (57).

$$\begin{cases} \theta_0 = \frac{a\omega}{V} \\ q = \dot{\theta} = \theta_0 \omega \cos(\omega t) = \frac{a\omega^2}{V} \cos(\omega t) \end{cases} \tag{57}$$

In the original pitch motion simulation calculation, as with the pure heave motion, the pure pitch equation of motion must be programmed first. The code that results from this programming is then saved in the entry function to set up the future motion settings. Following the completion of the overlap area's motion setting procedure, four operating conditions—f=0.2, 0.25, 0.3125, and 0.4—are chosen for simulation computations. Since the amplitude is consistent with the pure heave motion and the incoming flow velocity of V=1.5 m/s, a= 0.15 m is chosen. The calculating period is chosen to consist of four periods in order to provide steady data; the third period's data is chosen for data fitting. Set the time step to T/400. The step size and calculation period parameters are shown in Table 25. The extension of the Fourier series of the data received via simulation is required due to the periodic character of the data: the vertical force Z and pitching moment M Fourier series expansions are presented in Equations (58), (59).

For the vertical force Z:

$$\begin{cases} e_2 = Z_3 = -\theta_0 \omega^2 Z_{\dot{q}} = \frac{1}{l} \int_{-l}^{l} f(t) \sin(\omega t) dt \\ f_2 = Z_4 = \theta_0 \omega Z_q = \frac{1}{l} \int_{-l}^{l} f(t) \cos(\omega t) dt \end{cases} \tag{58}$$

TABLE 25 Statistics of pure heave hydrodynamic coefficient.

Vertical force coefficient				Pitch moment coefficient			
Z_w	-174.9	$Z_{\dot{w}}$	-21.59	M_w	8.414	$M_{\dot{w}}$	0.9012
Z_w'	-1.1193	$Z_{\dot{w}}'$	-0.4535	M_w'	0.1178	$M_{\dot{w}}'$	0.0414

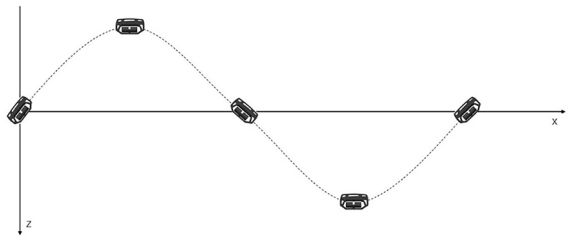


FIGURE 54
Pure pitch motion.

TABLE 26 Pure pitch hydrodynamic coefficients.

Vertical force coefficients		Pitch moment coefficients	
Z_q	$Z_{\dot{q}}$	M_q	$M_{\dot{q}}$
$Z_q^{'}$	$Z_{\dot{q}}^{'}$	$M_q^{'}$	$M_{\dot{q}}^{'}$

For yaw moment M:

$$\begin{cases} e_2^{'} = M_3 = -\theta_0\omega^2 M_{\dot{q}} = \frac{1}{l}\int_{-l}^{l} f(t)\sin(\omega t)dt \\ f_2^{'} = M_4 = \theta_0\omega M_q = \frac{1}{l}\int_{-l}^{l} f(t)\cos(\omega t)dt \end{cases} \quad (59)$$

Following the calculation of operating conditions at various frequencies, the data from the third cycle had been selected for analysis and processing. The MATLAB program was then used to fit the data using Fourier series. The statistical analysis of the Fourier fitting yields the coefficients of the fitting curves, which are shown in Table 27.

According to Equations (58), (59), the data in the table were re-fitted by MATLAB software, and the fitting results are shown in Figures 55–58.

The statistically solved hydrodynamic coefficient of pure pitching motion and its dimensionless value, and the numerical statistical results are shown in Table 28.

6 Conclusion

In summary, the ROV's asymmetry results in an obvious disparity in pressure resistance between the forward and backward sailing, ascending and descending motions, and this disparity becomes significantly greater as the velocity increased.

TABLE 27 Pure heave calculation data.

Frequency f (HZ)	ω = 2πf (1/s)	$\theta_0\omega$	$-\theta_0\omega^2$	Z_3	Z_4	M_3	M_4
0.2000	1.2566	0.1579	-0.1984	-17.8200	-28.6100	1.4680	1.4010
0.2500	1.5708	0.2467	-0.3876	-24.8400	-37.7400	1.7900	1.6290
0.3125	1.9635	0.3855	-0.7570	-34.4200	-49.3000	2.1910	1.8710
0.4000	2.5133	0.6317	-1.5876	-50.0500	-65.9000	2.9300	2.0750

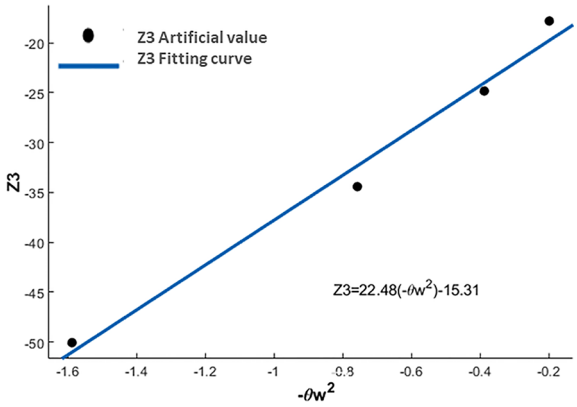


FIGURE 55
Z3 fitting curve.

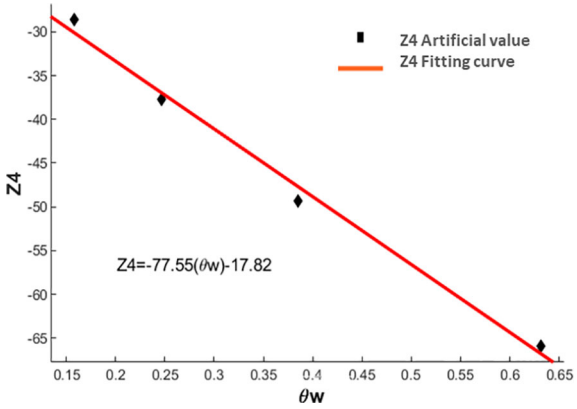


FIGURE 56
Z4 fitting curve.

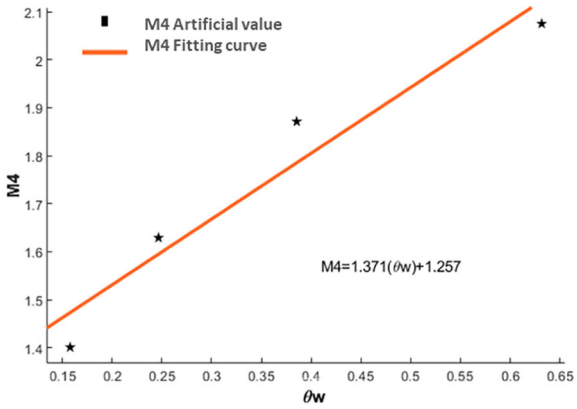


FIGURE 57
M3 fitting curve.

The relationship between pressure resistance and pressure resistance is closely related to the shape of the ROV model. Therefore, when designing the ROV, attention should be taken to ensure that the shape is as symmetrical as possible in order to achieve optimal hydrodynamic performance. Through a comparative and analytical analysis of the calculated and experimental values of the SUBOFF model, the reliability of the simulation process used in this paper is confirmed. Furthermore, the process is extended to the simulation calculation of the ROV model, enabling it to complete the ROV model simulation

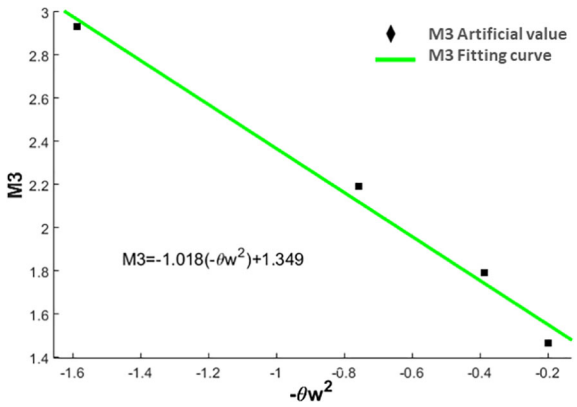


FIGURE 58
M4 fitting curve.

TABLE 28 Statistics of pure pitch hydrodynamic coefficients.

Vertical force coefficients				Pitch moment coefficients			
Z_q	-77.55	$Z_{\dot{q}}$	22.48	M_q	1.371	$M_{\dot{q}}$	-1.018
$Z_q^{'}$	-1.086	$Z_{\dot{q}}^{'}$	1.0333	$M_q^{'}$	0.042	$M_{\dot{q}}^{'}$	-0.1024

experiment and yield high-quality experimental results. This work also confirms that the unstable motion of the ROV can be simulated using the technique of superimposing the field function of the overset mesh, with satisfactory simulation results.

By analyzing data, we can derive some of the ROV's hydrodynamic coefficients; they will serve as a foundation for future maneuverability tests and will cut down on the time it takes to develop the ROV. In ROV hydrodynamic modeling simulations, simulations can help predict the performance of ROVs under various constraint conditions and optimize their design and operation before actual deployment. They can also assist in evaluating the impact of different constraints on ROVs, such as maneuverability and stability under varying depths, water flow conditions, and workloads. By conducting simulation experiments, the costs and risks associated with physical testing can be reduced, while providing reliable data to guide the design and operation of ROVs.

Although the hydrodynamic calculation of ROV is performed in this study, no approximate formula for the hydrodynamic calculation of ROV is developed in this paper owing to experimental equipment limitations and computation time limitations. Further comparisons between the experimental data and the hydrodynamic coefficients found in this study are necessary. Future research may investigate the impact of various mesh types and mesh numbers on hydrodynamic computing efficiency; nevertheless, this paper's meshing verification still has some holes due to computational restrictions.

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

DZ: Writing – original draft, Writing – review & editing. BZ: Writing – original draft, Writing – review & editing. YZ: Investigation, Validation, Visualization, Writing – review & editing. NZ: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by Program for Scientific Research Start-up Funds of Guangdong Ocean University (060302072101), Comparative Study and Optimization of Horizontal Lifting of Subsea Pipeline (2021E05011), China Scholarship Council (CSC202306320084).

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

Cepeda, S. F. M., MaChado, F. S. D. M., Barbosa, S. H. F., Moreira, D. S. S., Almansa, M. J. L., Souza, M. L., et al. (2023). Exploring autonomous and remotely operated vehicles in offshore structure inspections. *J. Mar. Sci. Eng.* 11, 2172. doi: 10.3390/jmse11112172

Chin, C., and Lau, M. (2012). Modeling and testing of hydrodynamic damping Model for a complex-shaped remotely-operated vehicle for control. *J. Mar. Sci. Appl.* 11, 150–163. doi: 10.1007/s11804-012-1117-2

Christophe, V. (2023). Unmotorize ROV gripper to catch profiling floats. *Ocean Eng.* 288, 116013. doi: 10.1016/j.oceaneng.2023.116013

Dalibor, I., and Marcin, S. (2024). Assessing damage and predicting future risks: A study of the Schilling Titan 4 manipulator on work class ROVs in offshore oil and gas industry. *Ocean Eng.* 291, 116282. doi: 10.1016/j.oceaneng.2023.116282

Fan, S.-b., Lian, L., and Ren, P. (2012). Research on hydrodynamics model test for deep sea open- framed remotely operated vehicle. *China Ocean Eng.* 26, 329–339. doi: 10.1007/s13344-012-0025-1

Manimaran, R. (2022). Hydrodynamic investigations on the performance of an underwater remote operated vehicle under the wave using OpenFOAM. *Ships Offshore Structures* 17, 2186–2202. doi: 10.1080/17445302.2021.1979921

Mingjie, L., Caoyang, Y., Xiaochao, Z., Chunhu, L., and Lian, L. (2023). Fuzzy adaptive trajectory tracking control of work-class ROVs considering thruster dynamics. *Ocean Eng.* 267, 113232. doi: 10.1016/j.oceaneng.2022.113232

Ren, F., and Hu, Q. (2023). ROV sliding mode controller design and simulation. *Processes* 11, 2359. doi: 10.3390/pr11082359

Selig, G. M., Drazen, J. C., Auster, P. J., Mundy, B. C., and Kelley, C. D. (2023). Distribution and structure of deep-sea demersal fish assemblages across the central and western Pacific Ocean using data from undersea imagery#13. *Front. Mar. Sci.* 10, 3389. doi: 10.3389/fmars.2023.1219368

Skorpa, S. (2012). *Numerical Simulation of Flow Around Remotely Operated Vehicle (ROV)* (Trondheim: Norwegian University of Science and Technology).

Tanveer, A., and Ahmad, M. S. (2023). Genetic-algorithm-based proportional integral controller (GAPI) for ROV steering control†. *Eng. Proc.* 32, 3390. doi: 10.3390/engproc2023032004

Whitcomb, L. L., and Yoerger, D. R. (1993). "A new distributed real-time control system for the Jason underwater robot," in *Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems*, vol. 1. (IEEE, Yokohama, Japan), 368–374.

Zhang, D., Wang, X., Zhao, M., Hong, L., and Li, X. (2023). Numerical investigation on hydrodynamic characteristics and drag influence of an open-frame remotely operated underwater vehicle. *J. Mar. Sci. Eng.* 11, 2143. doi: 10.3390/jmse11112143

Zhao, B., Yun, Y., Hu, F., Sun, J., Wu, D., and Huang, B. (2022). Hydrodynamic coefficients of the DARPA SUBOFF AFF-8 in rotating arm maneuver: Part I: Test technology and validation. *Ocean Eng.* 266, 113148. doi: 10.1016/j.oceaneng.2022.113148

Zhao, B., Yun, Y., Hu, F., Sun, J., Wu, D., and Huang, B. (2023). Hydrodynamic coefficients of the DARPA SUBOFF AFF-8 in rotating arm maneuver-Part II: Test results and discussion. *Ocean Eng.* 268, 113466. doi: 10.1016/j.oceaneng.2022.113466

10.3389/fmars.2024.1357144

Glossary

u	Longitudinal velocity
p	Roll angular velocity
X	Longitudinal force
v	Lateral velocity
q	Pitch angular velocity
Y	Lateral force
w	Vertical velocity
r	Yaw angular velocity
Z	Vertical force
K	Roll moment
m	Mass
α	Angle of attack
M	Pitch moment
I	Moment of inertia
β	Drift angle
N	Yaw moment
ρ	Water density
ϕ	Roll angle
θ	Pitch angle
ψ	Yaw angle

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 09 April 2024
DOI 10.3389/fmars.2024.1382147

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Man Zhou,
Nanyang Technological University, Singapore
Pengpeng Li,
Nanjing University of Science and Technology, China

*CORRESPONDENCE
Xin He
hexin6770@163.com

RECEIVED 05 February 2024
ACCEPTED 20 March 2024
PUBLISHED 09 April 2024

CITATION
He X, Jia T and Li J (2024)
Learning degradation-aware visual prompt for maritime image restoration under adverse weather conditions.
Front. Mar. Sci. 11:1382147.
doi: 10.3389/fmars.2024.1382147

Learning degradation-aware visual prompt for maritime image restoration under adverse weather conditions

Xin He[1*], Tong Jia[2] and Junjie Li[3]

[1]School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China, [2]School of Art and Design, Yantai Institute of Science and Technology, Yantai, China, [3]School of Electromechanical and Automotive Engineering, Yantai University, Yantai, China

Adverse weather conditions such as rain and haze often lead to a degradation in the quality of maritime images, which is crucial for activities like navigation, fishing, and search and rescue. Therefore, it is of great interest to develop an effective algorithm to recover high-quality maritime images under adverse weather conditions. This paper proposes a prompt-based learning method with degradation perception for maritime image restoration, which contains two key components: a restoration module and a prompting module. The former is employed for image restoration, whereas the latter encodes weather-related degradation-specific information to modulate the restoration module, enhancing the recovery process for improved results. Inspired by the recent trend of prompt learning in artificial intelligence, this paper adopts soft-prompt technology to generate learnable visual prompt parameters for better perceiving the degradation-conditioned cues. Extensive experimental results on several benchmarks show that our approach achieves superior restoration performance in maritime image dehazing and deraining tasks.

KEYWORDS

maritime image, image restoration, image deraining, image dehazing, prompt learning, deep learning, visual transformer

1 Introduction

Adverse weather conditions, including rain and haze, frequently occur in our everyday environment. These conditions result in diminished visual quality in captured images and significantly affect the effectiveness of numerous maritime vision systems, such as autonomous ships for ocean observation (Zheng et al., 2024). In maritime navigation and transportation, correctly identifying and interpreting environmental information from images is vital for safety. Figure 1 shows the physical imaging process of different adverse weather conditions. Thus, image processing under adverse weather conditions contributes

to enhancing the safety of maritime traffic and navigation by reducing accidents and collisions (Lu et al., 2021).

To solve image restoration under adverse weather conditions, early algorithms are predominantly based on traditional prior models. In the context of image dehazing, one common approach was the atmospheric scattering model (He et al., 2010), which assumed that haze in an image could be represented as a result of light scattering due to atmospheric particles (Li et al., 2018a). These algorithms typically aim to estimate and remove the haze from images, enhancing visibility. On the other hand, for image deraining, a prevalent technique was the linear superposition model. This model assumed that the observed image under rainy conditions could be expressed as a linear combination of the clean background scene and the rain streaks (Chen et al., 2023b). These early deraining algorithms focus on separating the rain streaks from the desired scene, thus improving the clarity of the image. However, these early prior-based algorithms struggled to adapt to complex and rapidly changing scenes, as they relied heavily on predefined models that could not effectively account for the wide range of scenarios encountered in real-world environments.

With the rise of big data and artificial intelligence, a plethora of image restoration methods based on deep learning have emerged. These techniques aim to learn the mapping relationship between degraded images and their corresponding clear counterparts. Convolutional Neural Networks (CNNs) have emerged as a powerful tool for image restoration due to their inherent ability to capture and learn complex hierarchical features from data. We have witnessed the rapid advancement of CNNs in image dehazing and deraining (Li et al., 2020; Zhou et al., 2021; Chen et al., 2022). However, due to the inherent characteristics of convolution operations, specifically the use of local receptive fields and the independence of input content, CNNs struggle to effectively model spatially-long feature dependencies of images (Chen et al., 2023c).

Later, Transformer-based models (Vaswani et al., 2017) originally bring significant breakthroughs to the natural language processing (NLP) field. The vision Transformer (ViT), as a new network backbone, has been widely applied to various tasks. It has also been utilized in image restoration tasks Zamir et al. (2022) and has achieved better performance compared to CNNs due to its ability to model non-local features effectively. Albeit these approaches have achieved commendable restoration performance in the given weather situation, they often exhibit suboptimal results when applied to maritime images.

The reason behind this can be summarized as follows: (1) Maritime images are often captured over large bodies of water, which introduces additional complexities due to the presence of reflective surfaces, varying water conditions, and dynamic backgrounds. These factors can exacerbate the impact of weather-related degradation. (2) Weather conditions at sea can change rapidly, with haze and rain appearing and dissipating quickly. This dynamic nature poses challenges for image restoration, as algorithms must adapt to evolving weather conditions. Thus, effective image restoration techniques tailored to the unique characteristics of maritime environments are essential for ensuring safe and efficient maritime activities (Zheng et al., 2020).

This raises a question: how to better help image restoration in adapting to the complex and ever-changing maritime scenes under adverse weather conditions? Recent trends of prompt learning (Wang et al., 2022) in artificial intelligence, may offer a potential solution. Prompt learning empowers deep models to adapt swiftly to complex and dynamically changing environments. It allows for the creation of tailored prompts that can capture the intricacies of specific situations, ensuring the model's responsiveness to various challenges. Therefore, this motivates us to introduce prompt learning to better encode degradation features of different weather conditions. This paper proposes a prompt-based learning method with degradation perception for maritime image restoration. The proposed method comprises two essential components: a restoration module and a prompting module. The restoration module is employed for image restoration, while the prompting module encodes weather-related degradation-specific information to modulate the restoration module. Specifically, the main contributions of this paper are as follows:

- This paper presents a new solution for image restoration in adverse weather conditions for maritime images. By incorporating prompt learning into the Transformer-based restoration network, it enhance the adaptability of deep networks to various weather degradation characteristics, enabling our model to adaptively learn more useful features to facilitate better restoration.
- This paper employs a prompt creation block to generate a set of learnable parameters by implicitly predicting degradation-conditioned soft prompts. In addition, this paper further introduces a prompt fusion block to guide the restoration process by interacting with the network backbone.
- Quantitative and qualitative experiments demonstrate that our proposed method achieves favorable performance on multiple benchmark datasets, and can better reconstruct clear images and restore image details compared to previous methods.

2 Related work

In this section, this paper presents a review of recent work related to maritime image restoration and prompt learning.

2.1 Maritime image restoration

To deal with the uncertain task of maritime image restoration, considerable efforts have been made. Existing approaches can be categorized into strategies based on priors and learning-based strategies. Hu et al (Hu et al., 2019). proposed a haze removal method based on illumination decomposition. This method decomposes the hazy image into a haze layer by separating the glow layer. It estimates the transmission rate using haze-line prior, thereby restoring the haze-free image. Luo et al (Lu et al., 2021).

proposed a novel CNN-based visibility dehazing framework aimed at enhancing the visual quality of images captured by maritime cameras under hazy conditions. This framework comprises two subnetworks: the coarse feature extraction module and the fine feature fusion module. Hu et al. (Hu et al., 2021) proposed a deep learning-based variational optimization method for reconstructing haze-free images from observed hazy images. This method fully leverages a unified denoising framework and strong deep learning representation capabilities. Guo et al. (Guo et al., 2021) designed a heterogeneous twin birth haze removal network, HTDNet, to enhance maritime surveillance capabilities in haze environments. The network consists of a twin feature extraction module for learning coarse haze features and a feature fusion module for integration and enhancement.

Van et al. (Van Nguyen et al., 2021) proposed a haze removal algorithm for maritime environment images based on texture and structure priors in illumination decomposition. This method utilizes a haze removal algorithm to eliminate the haze component from the glow-free layer and employs illumination compensation to restore natural illumination in the glow layer. Yang et al. (Yang et al., 2022) proposed a multi-head pyramid large kernel encoder-decoder network (LKEDN-MHP) for denoising tasks in maritime images. This method utilizes the transmission map extracted from the guidance image as an additional input to improve the network performance. Liu et al. (Liu et al., 2022) proposed a CNN-based dual-channel two-stage image dehazing network, which utilizes an attention mechanism to achieve adaptive fusion of multi-channel features. Hu et al. (Hu et al., 2022) proposed a maritime video dehazing algorithm based on spatiotemporal information fusion and improved dark channel prior. This method utilizes an enhanced dark channel prior model to restore each frame image, thereby achieving video dehazing.

Recently, Huang et al. (Huang et al., 2023) proposed an improved convex optimization model based on an atmospheric scattering model to achieve image dehazing. This method integrates simplified atmospheric light value estimation and the V channel in the HSV color space to obtain more local information. He et al. (He and Ji, 2023) improved MID-GAN is capable of training with non-paired adversarial learning. It consists of a CycleGAN cycle framework with two constraint branches. And it introduced an effective attentionrecursive feature extraction module to gradually extract haze components in an unsupervised manner. Chen et al. (Chen et al., 2022) introduced a contrastive learning mechanism based on the CycleGAN framework to improve dehazing performance. However, due to the limited performance of the aforementioned methods in the task of maritime image restoration and the relative saturation of model capabilities, there is a need to explore an effective approach to address these issues.

2.2 Prompt learning

Prompt learning was initially introduced in the field of natural language processing (NLP) and has proven to be highly effective, it has been applied to various vision-related tasks. Prompt learning is divided into two different methods: hard prompts and soft prompts. Hard prompts refer to explicit and predefined instructions given to the model during training. These prompts provide specific information and guide the model to produce the desired output. Soft prompts are more flexible and adaptive. They are not explicitly defined but rather generate prompt information based on input data or learn from the training process. Soft prompts allow the model to dynamically adjust its behavior based on input and context, enabling it to capture more specific and nuanced contextual information.

Recently, Zhou et al. (Zhou et al., 2022) demonstrated that a simple design based on conditional prompt learning performs exceptionally well in various problem scenarios, including generalization from base classes to novel classes, cross-dataset prompt transfer, and domain generalization. Potlapalli et al. (Potlapalli et al., 2023) demonstrated the effectiveness of their designed prompt block in integrated image restoration by integrating it into state-of-the-art restoration models. The prompt block can interact with input features, dynamically adjust representations, and adapt the restoration process to the relevant degradation. Li et al. (Li et al., 2023b) proposed a novel prompt-in-prompt learning for universal image restoration. The method involves simultaneous learning of high-level degradation-aware prompts and low-level basic restoration prompts to generate effective universal restoration prompts. By utilizing a selective promptfeature interaction module to modulate features most relevant to the degradation. Ai et al. (Ai et al., 2023) proposed a multi-modal prompt learning method called MPerceiver, which includes cross-modal adapters and image restoration adapters to learn holistic and multiscale detail representations. The adaptability of text and visual prompts is dynamically adjusted based on degradation prediction, enabling effective adaptation to various unknown degradations. Kong et al. (Kong et al., 2024) proposed sequential learning strategy and prompt learning strategy, respectively. These two strategies are effective for both CNN and Transformer backbones, and they can complement each other to learn effective image representations.

Inspired by these methods, this paper proposes a prompt-based learning approach to guide the maritime image restoration process, facilitating the integration and communication of information.

3 Proposed method

In this section, this paper first describes the overall pipeline of the model. Then, this paper provides details of the restoration module and prompting module, which serve as the fundamental building blocks of the approach. The restoration module mainly comprises two key elements: multi-head self-attention (MHSA) and dual gated feed-forward network (DGFN). The prompting module mainly consists of two key elements: prompt creation block (PCB) and prompt fusion block (PFB).

3.1 Overall pipeline

The overall pipeline of the proposed model, as illustrated in the Figure 2, is based on a hierarchical encoder-decoder framework (Chen et al., 2023a). Given a maritime degraded image $I_{\text{rain}} \in \mathbb{R}^{H\times W\times 3}$, where $H \times W$ denotes the spatial resolution of the feature maps, and C represents the channels, this paper performs feature projection embedding using a 3×3 convolution. On the network backbone, this paper stacks 4 levels of hierarchical encoder-decoders, where the encoder-decoder serves as the restoration module of the model, extracting rich spatially variant degradation distribution features. To extract multiscale representations from degradation information, each level of the restoration module covers its specific spatial resolution and channel dimensions. Beginning with high-resolution input, the restoration module aims to progressively decrease spatial resolution while enhancing channel capacity, resulting in a low-resolution latent representation $F \in \mathbb{R}^{H/8\times W/8\times 8C}$. During the stage of high-resolution image restoration, this paper incorporates a prompting module into the framework to generate prompts and enrich input features for dynamically guiding the restoration process of the restoration module. This paper also introduces skip connections (Li et al., 2023a) to bridge consecutive intermediate features, ensuring stable training. Next, this paper provides a detailed description of the proposed restoration module, prompting module, and their core building blocks.

3.2 Restoration module

This paper develops a restoration module as a feature extraction unit, which can be used to encode degradation information to recover output clean restored images. Formally, given the input features of the $(l-1)$-th block X_{l-1}, the encoding of the restoration module process can be represented as Equations (1, 2):

$$X_l' = X_{l-1} + MHSA(LN(X_{l-1})), \tag{1}$$

$$X_l = X_l' + DGFN(LN(X_l')), \tag{2}$$

where LN denotes the layer normalization, X_l' and X_l represent the outputs of MHSA and DGFN, which are described below.

3.2.1 Multi-head self-attention

In reviewing the standard self-attention mechanism in Transformers (Zamir et al., 2022), given queries Q, keys K, and values V, the output of dot-product attention is typically represented as Equation (3):

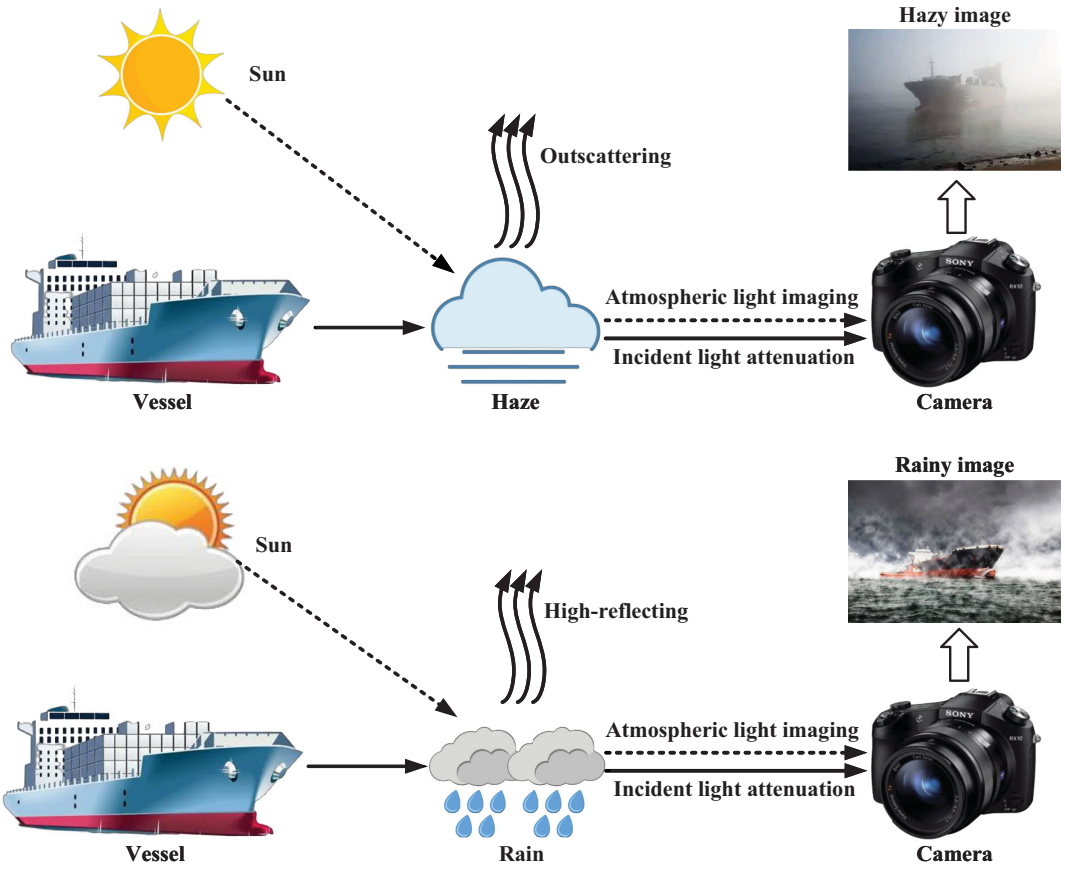


FIGURE 1
The physical imaging process of different adverse weather conditions.

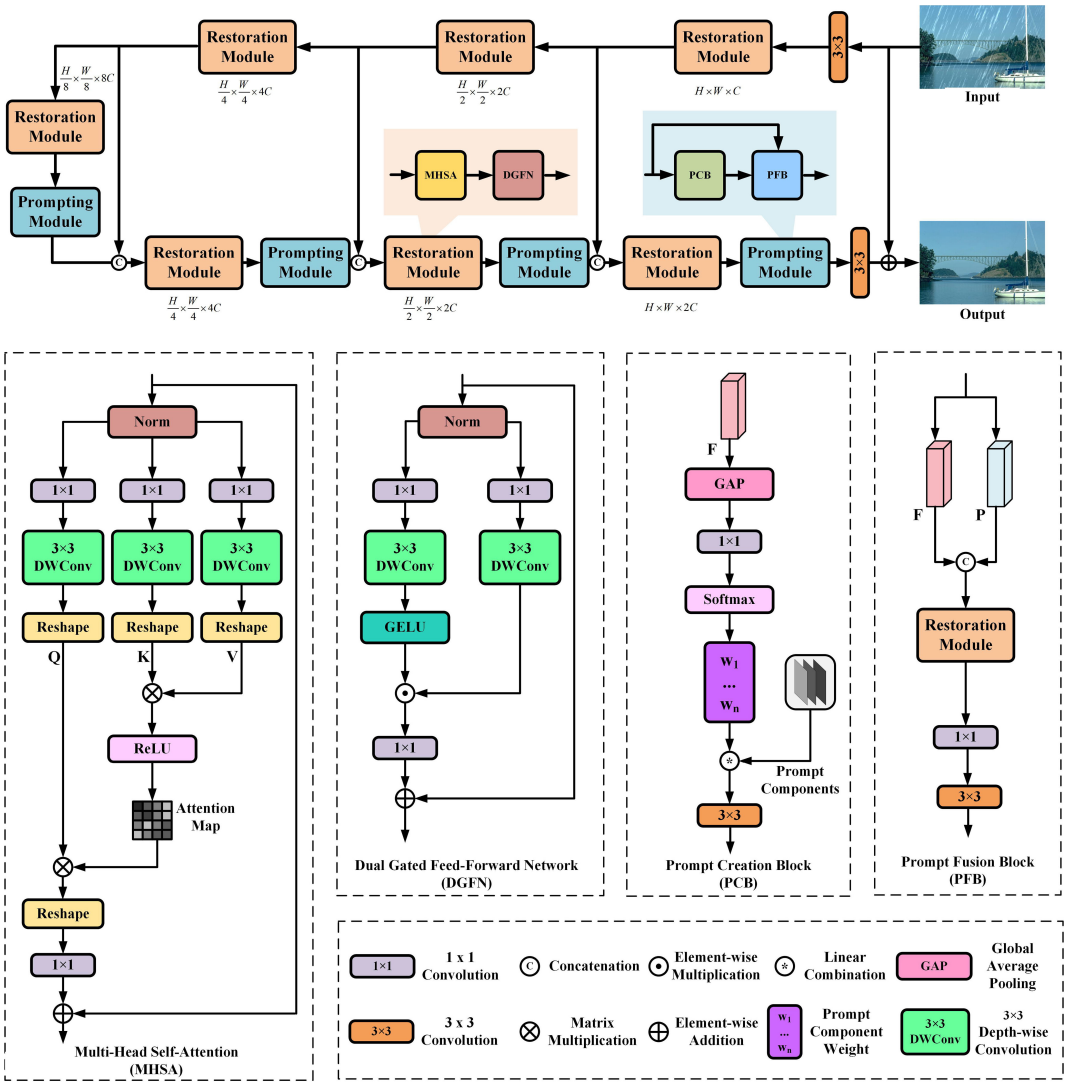


FIGURE 2
The overall architecture of the proposed network.

$$Attention' = Softmax\left(\frac{Q^{\top}K}{\alpha}\right) \cdot V, \tag{3}$$

where α is a learnable parameter; Q, K, and V represent the matrix forms of Q, K, and V, respectively. It is noted that computing self-attention using the Softmax function may lead to unstable gradients due to the presence of exponential functions, which could also limit the network's ability for nonlinear fitting. This work replaces it with the ReLU activation function, which can alleviate the issues of gradient vanishing or exploding, and aid in learning better feature representations. Specifically, this work starts by aggregating pixel-level cross-channel context through the application of a 1×1 convolution. Subsequently, a 3×3 depthwise convolution is applied to encode channel-wise context. This work employs bias-free convolution layers in the network. Next, it reshapes the projections of queries and keys to allow their dot product interaction to generate a transposed attention map of size $\mathbb{R}^{\hat{C} \times \hat{C}}$, rather than a massive regular attention map of size $\mathbb{R}^{\hat{H}\hat{W} \times \hat{H}\hat{W}}$. Then, the attention map is further interacted with the reshaped projections of values to complete the self-attention computation. Overall, the MHSA process is defined as Equations (4, 5):

$$Attention = ReLU\left(\frac{Q^{\top}K}{\alpha}\right) \cdot V, \tag{4}$$

$$\hat{X} = \underset{1\times1}{\text{Conv}}(Attention(Q, K, V)) + X, \tag{5}$$

where X and $\hat{X}$ are the input and output feature maps, $Conv_{1\times1}(\cdot)$ denotes 1×1 convolution.

3.2.2 Dual gated feed-forward network

To enhance the enrichment of contextual information, this paper introduces a dual gated feed-forward network that operates on each pixel. It incorporates two branches based on a gating mechanism. They initially undergo feature Y transformation by using 1×1 convolutions, followed by 3×3 depth-wise convolutions

to encode information from spatially adjacent pixel positions, facilitating the learning of local image details for effective restoration. One branch extends the feature channels, while the other branch, activated with GELU non-linearity, reduces the channels back to the original input dimensionality, enabling the discovery of non-linear contextual information in hidden layers. The DGFN is formulated as Equations (6, 7):

$$Gated = GELU(\mathrm{Conv}_{3\times3}(\mathrm{Conv}_{1\times1}(X))) \odot \mathrm{Conv}_{3\times3}(\mathrm{Conv}_{1\times1}(Y)), \tag{6}$$

$$\hat{Y} = \mathrm{Conv}_{1\times1}(Gated(Y)) + Y, \tag{7}$$

where $\hat{Y}$, $Conv_{3\times3}(\cdot)$, $\odot$ denote outputs, 3×3 depthwise convolution and element-wise multiplication, respectively. Overall, compared to MHSA, DGFN plays a distinctly different role by governing the flow of information across various levels in our pipeline, thereby enabling each level to focus on fine details complementary to other levels.

3.3 Prompting module

Different from (Zhou et al., 2023) that models global features, this paper further proposes a prompting module designed to perceive features of interfering information in degraded images and dynamically generate valuable prompts to guide high-quality maritime image restoration. Given the input features F, the prompting module first employs PCB to generate prompts for the distribution of degraded features. Subsequently, PFB collaboratively fuses the input features F with the generated prompt information to obtain the output features F_b. The overall procedure of prompting module is defined as Equation (8):

$$\hat{F} = PFB(PCB(F, F_{pc}), F), \tag{8}$$

where F_{pc} represents the learnable prompt components, the prompt creation block and prompt fusion block are described below.

3.3.1 Prompt creation block

The PCB dynamically captures the distribution characteristics in degraded maritime images. This capability enhances the provision of useful prompt information for the restoration process. Here, this paper employs a soft prompt (Potlapalli et al., 2023) to generate a set of learnable parameters, which encode distinctive deteriorative features related to various weather conditions. For input features F, the network first applies global average pooling, followed by a 1×1 convolution to obtain a compact feature vector. Subsequently, a Softmax function is applied to derive the prompt weights $W \in \mathbb{R}_N$, where the value of N is determined by the prompt length. In soft prompt learning, the length of the prompt determines the amount of information and guidance provided to the model. In fact, it's essential to strike a balance between the impact of prompt length, choosing an appropriate length that balances information, guidance, and diversity of generation to achieve satisfactory results. This paper will analyzes its impact in Section 4.4.2. Next, the weights are adjusted within the prompting components using a linear combination. Finally, a 3×3 convolution is applied to obtain the conditional prompt P. The overall procedure of PCB is defined as Equations (9, 10):

$$W_i = Softmax(\mathrm{Conv}_{1\times1}(GAP(F))), \tag{9}$$

$$P = \mathrm{Conv}_{3\times3}(Linear(W_i \cdot F_{pc})), \tag{10}$$

where $GAP(\cdot)$, $Linear(\cdot)$ denote the global average pooling and linear combination, respectively.

3.3.2 Prompt fusion block

To facilitate more effective interaction between the prompt information and input features for better guidance in the restoration process, this paper designs a PFB. In this module, the input features F are concatenated with the degradation distribution prompt P along the channel dimension. Subsequently, the concatenated information is further processed by the restoration module to generate the transformation of the input features. Finally, through operations of 1×1 convolution and 3×3 convolution, the features are smoothed and mapped to the output features $\hat{F}$. The overall procedure of PFB is defined as Equation (11):

$$\hat{F} = \mathrm{Conv}_{3\times3}(\mathrm{Conv}_{1\times1}(\mathcal{RM}(\mathcal{C}(F \cdot P)))), \tag{11}$$

where $\mathcal{RM}(\cdot)$ and $\mathcal{C}(\cdot)$ denote the process of involving restoration module and concatenation, respectively.

3.4 Loss function

To supervise the training progress of our network, this paper employs the L_1 pixel loss function. The final output is the restored image I_{rec}. It is obtained by adding the residual image I_{res} to the input degraded image I_{deg}, where $I_{res} \in \mathbb{R}^{H\times W\times3}$. During training, the network minimizes the loss function ,which is defined as Equation (12):

$$L_{pixel} = \left\| I_{rec} - I_{gt} \right\|_1, \tag{12}$$

where I_{gt} represents the ground truth image, and $\| \cdot \|_1$ denotes the L_1-norm.

4 Experiments

In this section, this paper evaluates our method on benchmarks for image dehazing and image deraining tasks. The main experiments are conducted using PyTorch and trained on 4 TESLA V100 GPUs.

4.1 Experimental settings

Following (Zamir et al., 2022), our method employs a 4-level encoder-decoder framework. From level-1 to level-4, the number of

restoration module is [4,6,6,8], attention heads are [1,2,4,8], and number of channels is [48,96,192,384]. The prompt length in the prompting module is 5. The batch size and patch size are configured as 16 and 128, respectively. Our model is trained using the AdamW optimizer for a total of 300,000 iterations with cosine annealing scheme (Loshchilov and Hutter, 2016) to gradually decrease the initial learning rate from 3×10^{-4} to 1×10^{-6}. Specifically, the learning rate is initially set at 3×10^{-4} for the 92,000 iterations and subsequently reduced to 1×10^{-6} over the next 208,000 iterations.

4.2 Experimental results on image dehazing

To evaluate the dehazing performance of the method, this paper trains on the RESIDE SOTS-Outdoor (Li et al., 2018a), but since this dataset lacks maritime scenes, this paper conducts testing on real maritime hazy images. It should be noted that there are no ground truth images for real maritime hazy images. Here, this paper compares our method with 6 popular image dehazing algorithms, including DCP (He et al., 2010), DehazeNet (Cai et al., 2016), AODNet (Li et al., 2017), GridDehazeNet Liu et al. (2019), MSBDN (Dong et al., 2020), and DeHamer (Guo et al., 2022). For fair comparison, all comparison algorithms use consistent pre-trained weights trained on the training set. In the case of hazy maritime images without ground truth data, this paper employs non-reference metrics such as NIQE (Naturalness Image Quality Evaluator) (Mittal et al., 2012b) and BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal et al., 2012a). When the NIQE or BRISQUE scores are lower, it means that the image is considered to have higher quality in terms of naturalness (for NIQE) or overall spatial quality (for BRISQUE).

As presented in Table 1, our proposed method achieves notably lower NIQE and BRISQUE scores, indicating that it produces high-quality outputs characterized by clearer content and superior perceptual quality when compared to other models in maritime scenarios. This underscores the effectiveness of our model in enhancing image quality within maritime contexts. To provide compelling evidence, this paper illustrates a visual quality comparison between two samples generated by recent methods in Figure 3. It is observed that the performance of DCP (He et al., 2010) is suboptimal, particularly in the sky region, where undesirable halo effects occur. DehazeNet (Cai et al., 2016) and AODNet (Li et al., 2017), are found to exhibit insufficient learning capabilities, resulting in their inability to effectively remove haze from images. The limitations in their learning capabilities become particularly evident when dealing with challenging and complex hazy scenes, such as those with dense fog or severe haze. The dehazing results obtained from GridDehazeNet Liu et al. (2019), MSBDN (Dong et al., 2020), and DeHamer (Guo et al., 2022) still exhibit residual haze, indicating limitations in the generalization of their algorithms to maritime images. These residual haziness issues suggest that their models may struggle to effectively adapt to the unique challenges posed by maritime scenarios. In contrast, our method demonstrates the capability to recover significantly clearer images, particularly in the sailboat regions. This suggests that the introduction of prompt learning cues can substantially improve the adaptation of dehazing algorithms to more challenging maritime images. The enhanced performance underscores the effectiveness of incorporating prompt learning, which enables better haze removal and results in visually superior outcomes in maritime scenes.

4.3 Experimental results on image deraining

To evaluate the deraining performance of the method, this paper carries out comprehensive experiments using the Rain13K dataset (Jiang et al., 2020), comprising 13,700 pairs of clean and rainy images. To evaluate our approach, this paper employs 4 benchmarks [Test100 (Zhang et al., 2019), Rain100H (Yang et al., 2017), Rain100L (Yang et al., 2017), and Test2800 (Fu et al., 2017b)] for testing purposes. Here, this paper compares our method with 9 popular image deraining algorithms, including DerainNet (Fu et al., 2017a), SEMI (Wei et al., 2019), DIDMDN (Zhang and Patel, 2018), UMRL (Yasarla and Patel, 2019), RESCAN (Li et al., 2018b), PReNet (Ren et al., 2019), MSPFN (Jiang et al., 2020), MPRNet (Zamir et al., 2021), and IDT (Xiao et al., 2022). Here, this paper uses full-reference image evaluation metrics, as there are ground truth images available. This paper quantitatively assesses our results by computing both PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) scores (Wang et al., 2004), specifically using the Y channel within the YCbCr color space. This allows us to make objective comparisons, focusing on luminance information, which is crucial for evaluating image quality accurately. These metrics provide valuable insights into the fidelity and structural similarity between the deraining images and their corresponding ground truth references.

Table 2 presents the quantitative results of various algorithms for image deraining. In comparison to the recent IDT method (Xiao et al., 2022), our approach demonstrates an average improvement of 1dB across the four datasets. This significant enhancement highlights our method's ability to adapt more effectively to diverse rainy conditions. Our results suggest that our approach outperforms existing methods by providing superior deraining performance across a range of challenging rain scenarios.

TABLE 1 Quantitative comparisons of different methods on hazy maritime images.

Method	DCP	DehazeNet	AODNet	GridDehazeNet	FFA Net	DeHamer	Ours
NIQE	3.464	3.466	3.433	3.504	3.530	3.483	3.015
BRISQUE	22.149	25.884	25.006	24.113	24.847	25.054	22.028

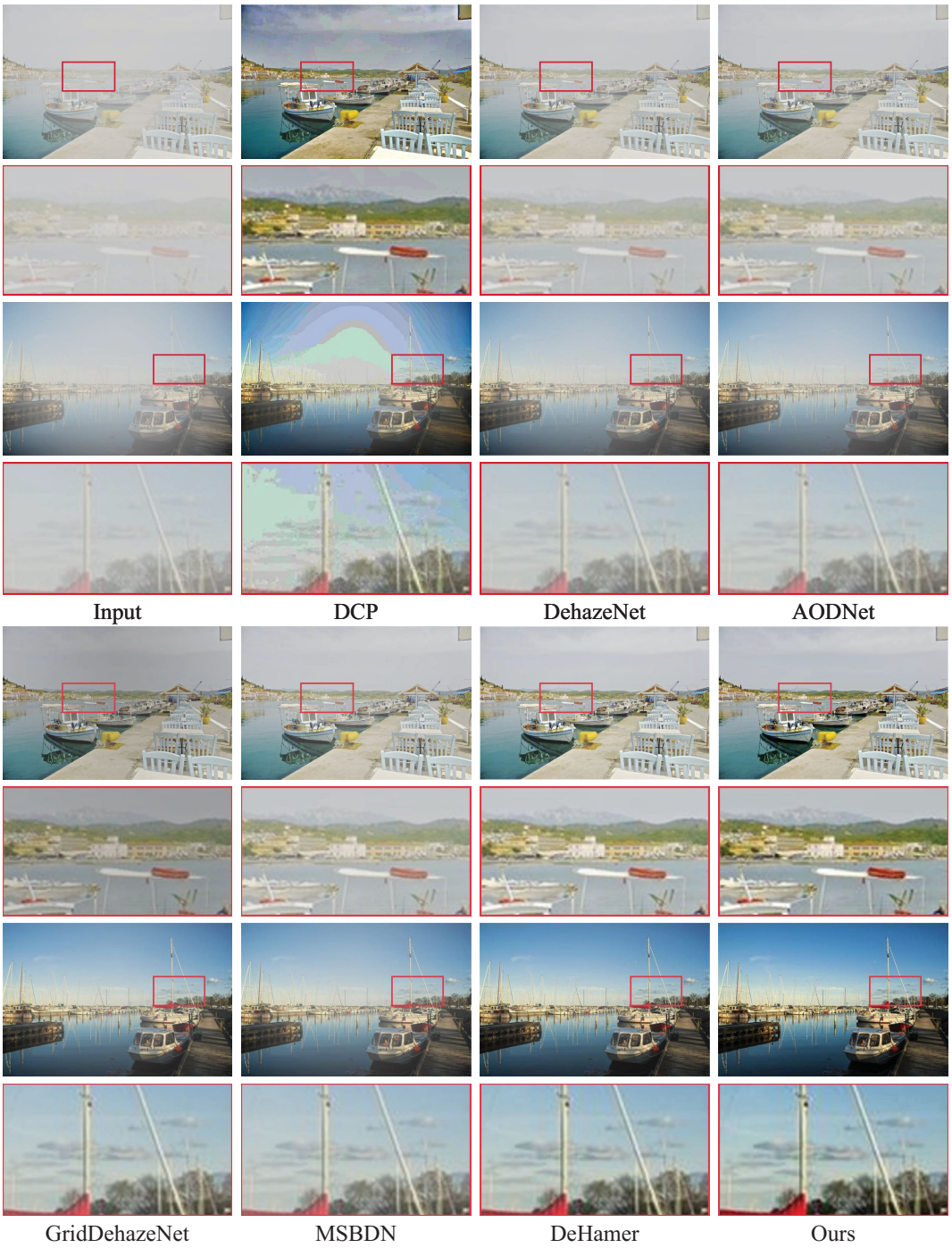


FIGURE 3
Image dehazing comparisons for different methods on hazy maritime images.

Figures 4 and 5 show the visual results on the Rain100H and Test100 datasets. Observing these visual results, it becomes evident that DerainNet (Fu et al., 2017a), SEMI (Wei et al., 2019), and DIDMDN (Zhang and Patel, 2018) struggle to effectively remove heavy rain artifacts. UMRL (Yasarla and Patel, 2019), RESCAN (Li et al., 2018b), and PReNet (Ren et al., 2019) still leave residual rain streaks in their recovery results. MSPFN (Jiang et al., 2020), MPRNet (Zamir et al., 2021), and IDT (Xiao et al., 2022) exhibit shortcomings in preserving local image details, especially in regions such as the ship's hull. In contrast, our approach stands out due to its incorporation of prompt learning, enabling adaptive feature extraction. As a result, it excels in eliminating rain streaks while effectively retaining intricate image structures. This demonstrates the robustness and effectiveness of our method in addressing the challenges posed by rainy conditions and preserving fine-grained image details.

TABLE 2 Quantitative comparisons of different methods on the Rain13K dataset.

Datasets Method	Test100		Rain100H		Rain100L		Test2800		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	22.26	0.787
SEMI	22.35	0.788	16.56	0.486	25.03	0.842	24.43	0.782	22.09	0.725
DIDMDN	22.56	0.818	17.35	0.524	25.23	0.741	28.13	0.867	23.32	0.738
UMRL	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	27.39	0.872
RESCAN	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	28.11	0.852
PReNet	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	28.94	0.894
MSPFN	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	30.35	0.900
MPRNet	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.68	0.923
IDT	29.69	0.905	29.95	0.898	37.01	0.971	33.38	0.937	32.51	0.928
Ours	31.08	0.907	30.85	0.900	38.30	0.974	34.02	0.941	**33.56**	**0.930**

Bold indicates the best results.

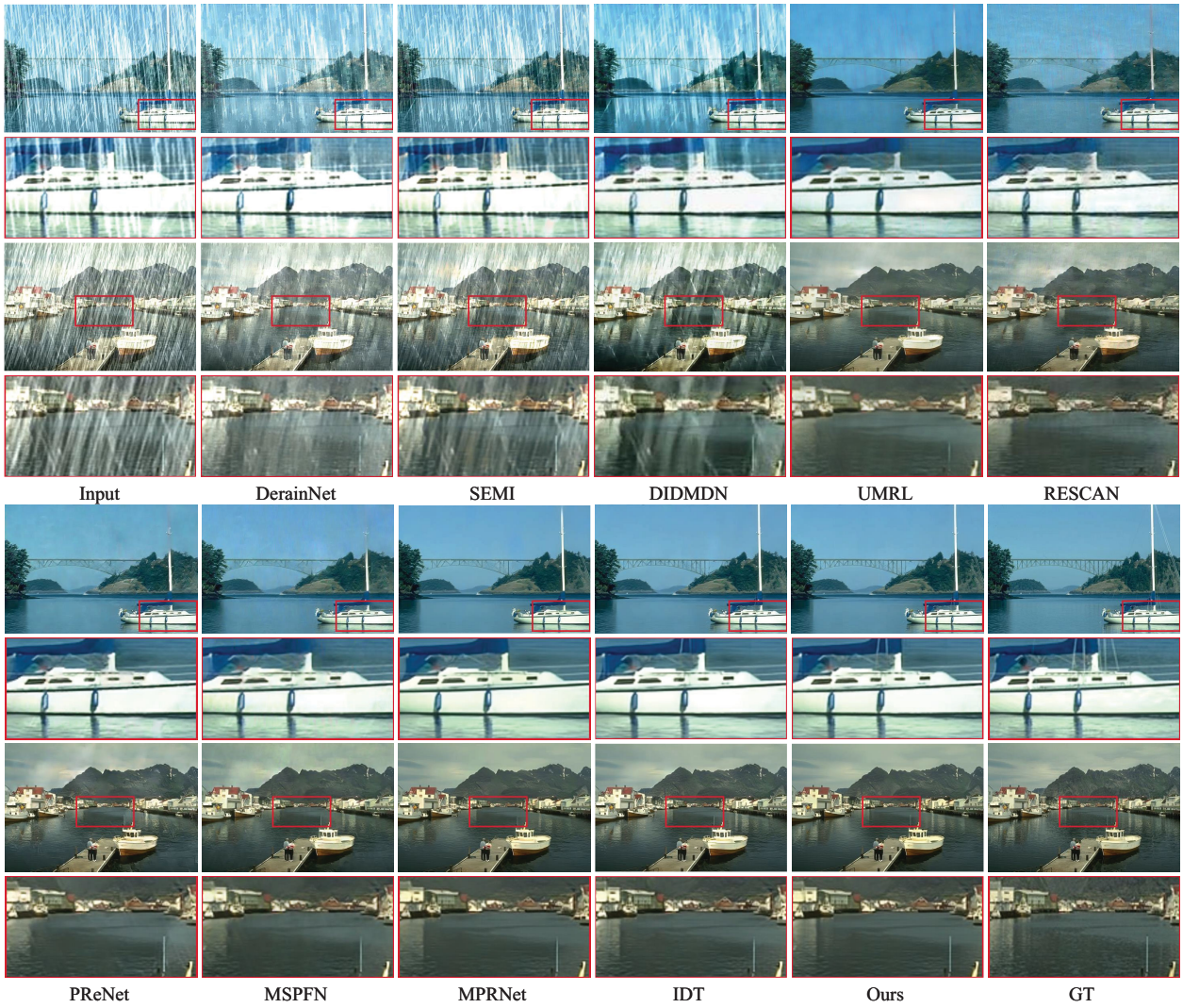


FIGURE 4
Image deraining comparisons for different methods on the Rain100H dataset.

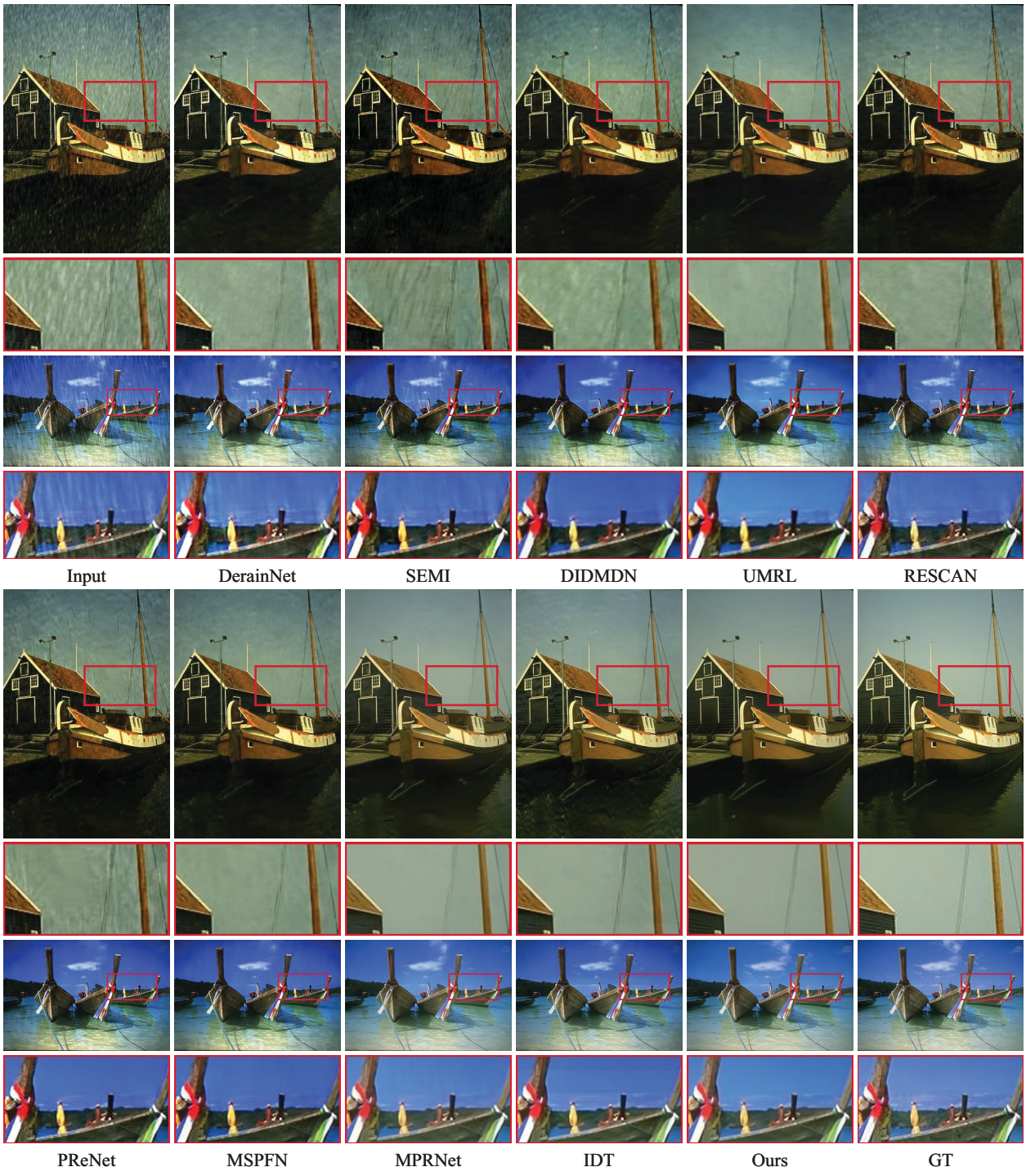


FIGURE 5
Image deraining comparisons for different methods on the Test100 dataset.

4.4 Ablation experiment and analysis

To further analysis the effectiveness of our proposed method, this paper conducts additional ablation experiments. This paper focuses on evaluating the performance of our approach in the image deraining task, specifically examining the effectiveness of prompting modules and the effect of prompt length in prompting modules.

4.4.1 Effectiveness of prompting modules

In this experiment, this paper aims to gauge the contributions of the prompting modules within our methodology. This paper designs three distinct model variants to comprehensively analyze the impact of prompting modules on our approach. These variants encompass models with or without prompting modules, models with or without PFB, and models with prompting modules placed at

TABLE 3 Ablation experimental result on the effectiveness of prompting modules.

Model	Prompting module			Resto-ration module	PSNR/ SSIM
	PCB	PFB	Positi-on		
(a)	×	×	–	✓	33.15/ 0.924
(b)	✓	×	Enc.	✓	33.39/ 0.927
(c)	✓	✓	Enc.	✓	33.48/ 0.929
(d)	✓	✓	Dec.	✓	**33.56/ 0.930**

Bold indicates the best results.

TABLE 4 Ablation experimental result on the effect of prompt length in prompting modules.

Length	3	4	5	6	7
PSNR/ SSIM	33.41/ 0.928	33.49/ 0.929	**33.56/ 0.930**	33.52/ 0.930	33.44/ 0.927

Bold indicates the best results.

different positions within the network architecture. The quantitative results for these various model variants are documented in Table 3, shedding light on their respective contributions to the image restoration task.

Starting with Model (a), the performance starkly deteriorates in the absence of prompting modules. This striking decline underscores the pivotal role that prompting modules play in the image restoration process. They serve as critical components in guiding the model's understanding of the task and assisting it in achieving high-quality restoration results. Comparing Model (b) and Model (c), it becomes evident that the absence of PFB in the prompting modules results in a disconnect between the guidance provided by the prompts and the features processed in the main restoration modules. This lack of synchronization hinders the model's ability to effectively utilize the provided prompts, consequently restricting its overall performance. Further comparison between Model (c) and Model (d) reveals the advantage of introducing prompting modules during the network decoding phase. This strategic placement allows the model to better harness the clear image features, facilitating improved performance and yielding the best network performance among the variants.

4.4.2 Effect of prompt length in prompting modules

To delve deeper into the impact of prompts in our prompting modules, this paper conducts experiments to investigate the effect of prompt length. This paper varies the length of prompts while keeping other parameters constant and examined how it influenced the model's performance. Table 4 presents the quantitative results. Our findings reveal that prompt length plays a crucial role in image restoration process. Specifically, shorter prompts tend to yield faster convergence and better overall results, while longer prompts sometimes lead to overfitting or increased computational complexity. This analysis allows us to optimize the prompt length within our prompting modules to achieve the best balance between performance and efficiency. Finally, this paper determines that a prompt length of 5 is the most suitable configuration for our model.

4.5 User study

This paper conducts a user study to assess the outcomes of various image restoration techniques. These studies are based on restoring real foggy maritime images. A portion of the participants invited to our user study are professionals engaged in maritime

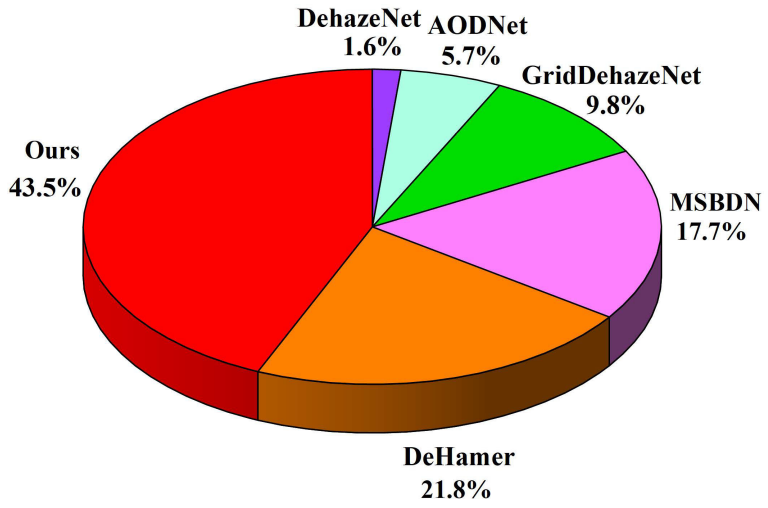


FIGURE 6
Averaged selection percentage of user study.

TABLE 5 Comparison of model efficiency.

	Uformer	Restormer	IDT	DRSformer	Ours
Paramaters (M)	50.8	26.1	16.4	33.7	35.6
Flops (G)	45.9	174.7	61.9	242.9	43.2

activities to ensure the professionalism of our work. Additionally, in order to enhance the universality of our work, this paper has also invited non-maritime professionals to participate in our user study. Participants are presented with a set of images and asked to select the one with the best visual clarity. To maintain fairness, the methods remain anonymous, and the images within each set are randomly ordered. This paper distributes the questionnaire widely to online users and collect responses. Finally, this paper receives responses from 46 human evaluators. Figure 6 depicts the average selection percentage for each method. Based on the majority of human evaluators' feedback, our method consistently outperforms the others.

4.6 Limitations

While our proposed method introduces prompt learning to enhance restoration performance under adverse weather conditions, it comes with relatively high parameter count and computational complexity. The comparison results are presented in the Table 5, compared to other methods, our approach has a higher parameter count. As a result, significant computational resources are also required, which to some extent limits the applicability of the model. In order to enable our method to be more rapidly and conveniently applied in maritime operations, this paper plans to address this issue through strategies such as model compression and pruning, aiming to make the model more suitable for maritime vision applications.

5 Conclusions

This paper has proposed a prompting image restoration approach by learning degradation-aware visual prompt for maritime surveillance. Our proposed approach possesses the capability to interact with input features, allowing for dynamic adjustments in weather-related representations. This adaptability ensures that the restoration process is tailored to the specific degradation being addressed. This paper has validated the effectiveness of our method on extensive experimental datasets, enhancing its restoration performance in various weather conditions, including rain removal and haze removal in maritime images. In future work, this paper plans to explore leveraging text models such as CLIP as alternative prompts to further guide the image restoration process.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://sites.google.com/view/reside-dehaze-datasets, https://www.deraining.tech/benchmark.html.

Author contributions

XH: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. TJ: Data curation, Software, Visualization, Writing – original draft. JL: Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by the Research Project of the Naval Staff Navigation Assurance Bureau (2023(1)).

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, Y., Huang, H., Zhou, X., Wang, J., and He, R. (2023). Multimodal prompt perceiver: Empower adaptiveness, generalizability and fidelity for all-in-one image restoration. arXiv [preprint] arXiv:2312.02918.

Cai, B., Xu, X., Jia, K., Qing, C., and Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. image Process.* 25, 5187–5198. doi: 10.1109/TIP.2016.2598681

Chen, X., Li, H., Li, M., and Pan, J. (2023a). "Learning a sparse transformer network for effective image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Canada: IEEE), 5896–5905. doi: 10.1109/CVPR52729.2023.00571

Chen, X., Pan, J., Dong, J., and Tang, J. (2023b). Towards unified deep image deraining: A survey and a new benchmark. arXiv [preprint] arXiv:2310.03535.

Chen, X., Pan, J., Jiang, K., Li, Y., Huang, Y., Kong, C., et al. (2022). "Unpaired deep image deraining using dual contrastive learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 2017–2026

Chen, X., Pan, J., Lu, J., Fan, Z., and Li, H. (2023c). "Hybrid cnn-transformer feature fusion for single image deraining," in *Proceedings of the AAAI Conference on Artificial Intelligence*. (USA: AAAI Press), Vol. 37. 378–386. doi: 10.1609/aaai.v37i1.25111

Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., et al. (2020). "Multi-scale boosted dehazing network with dense feature fusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 2157–2167.

Fu, X., Huang, J., Ding, X., Liao, Y., and Paisley, J. (2017a). Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* 26, 2944–2956. doi: 10.1109/TIP.2017.2691802

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., and Paisley, J. (2017b). "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 3855–3863. doi: 10.1109/CVPR.2017.186

Guo, C.-L., Yan, Q., Anwar, S., Cong, R., Ren, W., and Li, C. (2022). "Image dehazing transformer with transmission-aware 3d position embedding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 5812–5820.

Guo, Y., Lu, Y., Liu, R. W., Wang, L., and Zhu, F. (2021). "Heterogeneous twin dehazing network for visibility enhancement in maritime video surveillance," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. (USA: IEEE), 2875–2880.

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353.

He, X., and Ji, W. (2023). Single maritime image dehazing using unpaired adversarial learning. *Signal Image Video Process.* 17, 593–600. doi: 10.1007/s11760-022-02265-5

Hu, H.-M., Guo, Q., Zheng, J., Wang, H., and Li, B. (2019). Single image defogging based on illumination decomposition for visual maritime surveillance. *IEEE Trans. Image Process.* 28, 2882–2897. doi: 10.1109/TIP.83

Hu, X., Wang, J., Zhang, C., and Tong, Y. (2021). Deep learning-enabled variational optimization method for image dehazing in maritime intelligent transportation systems. *J. Advanced Transportation* 2021, 1–18. doi: 10.1155/2021/6658763

Hu, Q., Zhang, Y., Liu, T., Liu, J., and Luo, H. (2022). Maritime video defogging based on spatialtemporal information fusion and an improved dark channel prior. *Multimedia Tools Appl.* 81, 24777–24798. doi: 10.1007/s11042-022-11921-4

Huang, H., Li, Z., Niu, M., Miah, M. S., Gao, T., and Wang, H. (2023). A sea fog image defogging method based on the improved convex optimization model. *J. Mar. Sci. Eng.* 11, 1775. doi: 10.3390/jmse11091775

Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., et al. (2020). "Multi-scale progressive fusion network for single image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 8346–8355.

Kong, X., Dong, C., and Zhang, L. (2024). Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. arXiv [preprint] arXiv:2401.03379.

Li, Z., Lei, Y., Ma, C., Zhang, J., and Shan, H. (2023b). Prompt-in-prompt learning for universal image restoration. arXiv [preprint] arXiv:2312.05038.

Li, Y., Lu, J., Chen, H., Wu, X., and Chen, X. (2023a). "Dilated convolutional transformer for highquality image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Canada: IEEE), 4198–4206. doi: 10.1109/CVPRW59228.2023.00442

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. (2017). "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*. (Italy: IEEE), 4770–4778.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2018a). Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* 28, 492–505. doi: 10.1109/TIP.83

Li, R., Tan, R. T., and Cheong, L.-F. (2020). "All in one bad weather removal using architectural search," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3175–3185.

Li, X., Wu, J., Lin, Z., Liu, H., and Zha, H. (2018b). "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proceedings of the European conference on computer vision (ECCV)*. (Germany: Springer), 254–269.

Liu, X., Ma, Y., Shi, Z., and Chen, J. (2019). "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE/CVF international conference on computer vision*. Korea (South): (IEEE), 7314–7323.

Liu, T., and Zhou, B. (2022). Dual-channel and two-stage dehazing network for promoting ship detection in visual perception system. *Math. Problems Eng.* 2022. doi: 10.1155/2022/8998743

Loshchilov, I., and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. arXiv [preprint] arXiv:1608.03983.

Lu, Y., Guo, Y., and Liang, M. (2021). Cnn-enabled visibility enhancement framework for vessel detection under haze environment. *J. advanced transportation* 2021, 1–14. doi: 10.1155/2021/7649214

Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012a). No-reference image quality assessment in the spatial domain. *IEEE Trans. image Process.* 21, 4695–4708. doi: 10.1109/TIP.2012.2214050

Mittal, A., Soundararajan, R., and Bovik, A. C. (2012b). Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* 20, 209–212.

Potlapalli, V., Zamir, S. W., Khan, S., and Khan, F. S. (2023). Promptir: Prompting for all-in-one blind image restoration. arXiv [preprint] arXiv:2306.13090.

Ren, D., Zuo, W., Hu, Q., Zhu, P., and Meng, D. (2019). "Progressive image deraining networks: A better and simpler baseline," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3937–3946.

Van Nguyen, T., Mai, T. T. N., and Lee, C. (2021). Single maritime image defogging based on illumination decomposition using texture and structure priors. *IEEE Access* 9, 34590–34603. doi: 10.1109/ACCESS.2021.3060439

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wang, Z., Zhang, Z., Lee, C.-Y., Zhang, H., Sun, R., Ren, X., et al. (2022). "Learning to prompt for continual learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 139–149.

Wei, W., Meng, D., Zhao, Q., Xu, Z., and Wu, Y. (2019). "Semi-supervised transfer learning for image rain removal," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3877–3886.

Xiao, J., Fu, X., Liu, A., Wu, F., and Zha, Z.-J. (2022). Image de-raining transformer. *IEEE Trans. Pattern Anal. Mach. Intell.*

Yang, W., Gao, H., Jiang, Y., and Zhang, X. (2022). A novel approach to maritime image dehazing based on a large kernel encoder–decoder network with multihead pyramids. *Electronics* 11, 3351. doi: 10.3390/electronics11203351

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., and Yan, S. (2017). "Deep joint rain detection and removal from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 1357–1366.

Yasarla, R., and Patel, V. M. (2019). "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 8405–8414.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M.-H. (2022). "Restormer: Efficient transformer for high-resolution image restoration," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 5728–5739.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., et al. (2021). "Multi-stage progressive image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, virtual*. (virtual: IEEE) 14821–14831.

Zhang, H., and Patel, V. M. (2018). "Density-aware single image de-raining using a multi-stream dense network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 695–704.

Zhang, H., Sindagi, V., and Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. *IEEE Trans. circuits Syst. video Technol.* 30, 3943–3956. doi: 10.1109/TCSVT.76

Zheng, S., Sun, J., Liu, Q., Qi, Y., and Zhang, S. (2020). "Overwater image dehazing via cycle-consistent generative adversarial network," in *Proceedings of the Asian Conference on Computer Vision*. (Japan: Springer), doi: 10.3390/electronics9111877

Zheng, K., Zhang, X., Wang, C., Li, Y., Cui, J., and Jiang, L. (2024). Adaptive collision avoidance decisions in autonomous ship encounter scenarios through rule-guided vision supervised learning. *Ocean Eng.* 297, 117096. doi: 10.1016/j.oceaneng.2024.117096

Zhou, M., Huang, J., Guo, C.-L., and Li, C. (2023). "Fourmer: An efficient global modeling paradigm for image restoration," in *International conference on machine learning*. (USA: PMLR), 42589–42601.

Zhou, M., Xiao, J., Chang, Y., Fu, X., Liu, A., Pan, J., et al. (2021). "Image de-raining via continual learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (virtual: IEEE), 4907–4916.

Zhou, K., Yang, J., Loy, C. C., and Liu, Z. (2022). "Conditional prompt learning for vision-language models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 16816–16825.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 25 April 2024
DOI 10.3389/fmars.2024.1389553

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Feihu Zhang,
Northwestern Polytechnical University, China
Xuebo Zhang,
Northwest Normal University, China

*CORRESPONDENCE
Xin He
hexin6770@163.com

RECEIVED 21 February 2024
ACCEPTED 08 April 2024
PUBLISHED 25 April 2024

CITATION
He X, Li J and Jia T (2024) Learning hybrid dynamic transformers for underwater image super-resolution.
Front. Mar. Sci. 11:1389553.
doi: 10.3389/fmars.2024.1389553

Learning hybrid dynamic transformers for underwater image super-resolution

Xin He[1]*, Junjie Li[2] and Tong Jia[3]

[1]School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China, [2]School of Electromechanical and Automotive Engineering, Yantai University, Yantai, China, [3]School of Art and Design, Yantai Institute of Science and Technology, Yantai, China

Underwater image super-resolution is vital for enhancing the clarity and detail of underwater imagery, enabling improved analysis, navigation, and exploration in underwater environments where visual quality is typically degraded due to factors like water turbidity and light attenuation. In this paper, we propose an effective hybrid dynamic Transformer (called HDT-Net) for underwater image super-resolution, leveraging a collaborative exploration of both local and global information aggregation to help image restoration. Firstly, we introduce a dynamic local self-attention to adaptively capture important spatial details in degraded underwater images by employing dynamic weighting. Secondly, considering that visual transformers tend to introduce irrelevant information when modeling the global context, thereby interfering with the reconstruction of high-resolution images, we design a sparse non-local self-attention to more accurately compute self-similarity by setting a top-k threshold. Finally, we integrate these two self-attention mechanisms into the hybrid dynamic transformer module, constituting the primary feature extraction unit of our proposed method. Quantitative and qualitative analyses on benchmark datasets demonstrate that our approach achieves superior performance compared to previous CNN and Transformer models.

KEYWORDS

underwater image, image super-resolution, local self-attention, sparse self-attention, deep learning, visual transformer

1 Introduction

Underwater imaging poses distinct challenges owing to the natural attenuation, scattering, and color distortion of light within aquatic environments. These factors contribute to degraded image quality, thereby constraining the effectiveness of underwater observation, exploration, and surveillance systems (refer to Figure 1). Consequently, underwater image enhancement techniques, notably super-resolution, have attracted considerable attention in recent years. Super-resolution aims to reconstruct high-resolution images from low-resolution counterparts, thereby improving

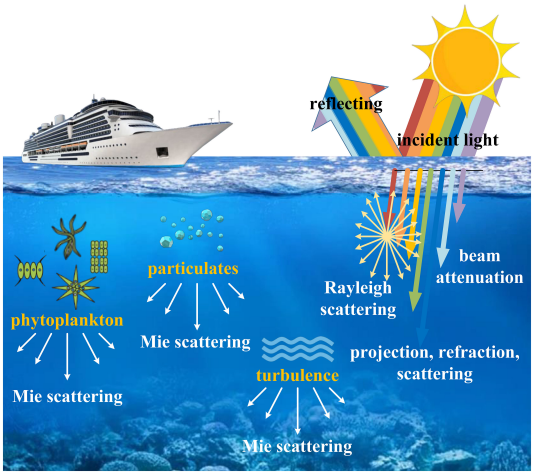


FIGURE 1
The physical imaging process of underwater conditions.

image clarity and detail. It holds immense potential for enhancing the visual quality of underwater scenes, facilitating better analysis, interpretation, and decision-making in various marine applications, such as ocean-observation and offshore engineering (Liu et al., 2024).

Despite recent advancements, underwater image super-resolution remains an active area of research, with ongoing efforts to develop robust and efficient algorithms capable of addressing the specific challenges.

posed by underwater environments (Islam et al., 2020a). Early efforts in underwater image super-resolution predominantly relied on traditional interpolation algorithms such as bicubic and bilinear. These approaches, while widely used in conventional image processing tasks, often yielded suboptimal results when applied to underwater imagery due to the unique characteristics of underwater environments.

In recent years, significant strides have been made in leveraging deep learning techniques, particularly convolutional neural networks (CNNs), for underwater image super-resolution. Unlike conventional interpolation methods, CNN-based approaches harness the capabilities of deep learning to discern intricate mappings between low-resolution and high-resolution image pairs directly from data. Various architectures, such as SRCNN (Super-Resolution Convolutional Neural Network) (Dong et al., 2015), VDSR (Very Deep Super-Resolution) (Kim et al., 2016), and EDSR (Enhanced Deep Super-Resolution) (Lim et al., 2017), have demonstrated remarkable performance improvements over those of conventional approaches. Subsequent research tends to focus on developing larger and deeper CNN models to enhance learning capabilities. However, due to the extensive number of network parameters, the computational cost of these methods is considerably high, limiting their applicability in real-world underwater scenarios (Jiang et al., 2021).

Later, Transformer-based architectures (Vaswani et al., 2017) have emerged as promising alternatives for underwater image super-resolution, offering unique advantages over CNN-based approaches. Unlike CNNs, Transformers leverage self-attention mechanisms to capture global dependencies and long-range dependencies within the input data (Han et al., 2022). For example, SwinIR (Liang et al., 2021) employs the window-based attention mechanism to better solve image super-resolution. Although the self-attention mechanism in the sliding window approach enables the extraction of local features, the discontinuity of the windows limits the ability to model local features within each window. In other words, these window-based image super-resolution methods are unable to aggregate information from outside the window, thus limiting the capability to model global information (Li et al., 2023a).

Indeed, the complexity and variability inherent in underwater environments elevate the challenges associated with underwater image super-resolution beyond those encountered in natural image superresolution tasks. The Transformer model, renowned for its adeptness in capturing global features, tends to introduce noticeable redundancy during the modeling process. Regrettably, this aspect has often been neglected in prior Transformer-based super-resolution approaches (Xiao et al., 2024). Therefore, developing a method to explore the characteristics of Transformers, aiming to better integrate both local and global features for joint modeling to achieve high-quality image reconstruction while reducing computational costs, holds significant promise.

To this end, we develop an effective hybrid dynamic Transformer (called HDT-Net) to solve underwater image super-resolution. The proposed method combines dynamic local self-attention with sparse non-local self-attention to synergistically enhance the representation capability of the Transformer model. The former dynamically explores local feature relationships based on a fully CNN model, mitigating errors induced by discontinuous windows. The latter aggregates features by selecting the most useful similarity values, alleviating redundancy caused by small self-attention weights. These strategies are carefully designed to address the challenges of complex underwater environments, thereby leveraging more effective feature information to improve the quality of image super-resolution. Finally, experimental validation on benchmark datasets confirms the effectiveness of the proposed approach.

In summary, the main contributions of this paper are as follows:

- We propose a lightweight deep model based on a hybrid Transformer for underwater image super-resolution tasks, aiming to enhance the quality of image reconstruction by jointly exploiting local and global features representation.
- We integrate a dynamic local self-attention and a sparse non-local self-attention to enable better capture of local and global feature information respectively, making the Transformer more effective and compact in long-range modeling.
- Experimental evaluation on commonly used benchmark datasets for underwater image super-resolution demonstrates that our method outperforms previous CNN and Transformer-based approaches both quantitatively and qualitatively.

2 Related work

In this section, we present a review of recent work related to underwater image super-resolution and vision transformer.

2.1 Underwater image super-resolution

Underwater image super-resolution is an uncertain task, and numerous studies have been conducted to explore suitable methods to address this challenge. Among the deep learning-based underwater image super-resolution models, CNN is one of the most common techniques. Shin et al. (Shin et al., 2016). proposed a CNN-based framework for estimating environmental light and transmission, featuring a versatile convolutional structure designed to mitigate haze in underwater images. Wang et al. (Wang et al., 2017). proposed a CNN-based underwater image enhancement framework called UIE-Net, comprising two sub-networks: CC-Net and HR-Net. CC-Net outputs color absorption coefficients for different channels to correct color distortion in underwater images. HR-Net outputs light attenuation transmission maps to enhance the contrast of underwater images. Li et al. (Li et al., 2017). proposed a novel generator network structure that combines the underwater image formation process to generate high-resolution output images. Subsequently, a dense pixel-level model learning pipeline is employed to perform color correction on monocular underwater images trained based on RGB-D and their corresponding generated images. The methods describe above address some aspects of underwater image super-resolution, yet they still exhibit a lack of robustness when handling highly complex underwater scenes.

Li et al. (Li et al., 2019). constructed an underwater image enhancement benchmark dataset, which provides a large-scale collection of real underwater images along with their corresponding reference images. This benchmark dataset facilitates comprehensive research on existing underwater image enhancement methods and enables easy training of CNNs for underwater image enhancement. But it lacks novelty in terms of algorithmic advancements compared to other methods. Guo et al. (Guo et al., 2019). proposed an underwater image enhancement method based on GAN. Additionally, the introduced MSDB combined with residual learning can improve network performance, while multiple loss functions can generate visually satisfactory enhancement results. Islam et al. (Islam et al., 2020b). proposed a simple yet effective underwater image enhancement model based on conditional genetic algorithms. This model evaluates image quality by incorporating global color, content, local texture, and style information to establish a perceptual loss function. Additionally, they provided a large-scale dataset consisting of paired and unpaired underwater image collections for supervised training. Chen et al. (Chen et al., 2020). proposed an improved deep reinforcement convolutional neural network based on deep learning principles. The main innovation involves incorporating wavelet bases into turbulence-based deep learning convolutional kernels, introducing an improved dense block structure. Further investigation is needed to assess the generalization of the methods utilized in the aforementioned studies to different underwater conditions.

Recently, Li et al. (Li et al., 2021). proposed a deep underwater image enhancement model. This model learns feature representations from different color spaces and highlights the most discriminative features through channel attention modules. Additionally, domain knowledge is integrated into the network by utilizing inverse media transmission maps as attention weights. Li et al. (Li et al., 2023b). proposed a novel method for realistic underwater image enhancement and super-resolution called RUIESR. Its purpose is to obtain paired data consistent with realistic degradation for training and to accurately estimate dual degradation to assist in reconstruction. In deep-sea or heavily polluted waters, the degradation characteristics may differ from those observed in the training data, potentially affecting the performance of the above methods. Dharejo et al. (Dharejo et al., 2024). investigated the integration of a typical Swin transformer with wave attention modules and reversible downsampling to achieve efficient multiscale self-attention learning with lossless downsampling. As a potential improvement over SwinIR, this model allows for faster training and convergence, as well as greater capacity and resolution. The computational complexity and resource requirements of this Transformer-based method may pose challenges.

2.2 Vision transformer

Vision Transformer (ViT) (Vaswani et al., 2017) is a model based on the Transformer architecture, initially proposed by Dosovitskiy et al. (Dosovitskiy et al., 2020). in 2020 to address image classification tasks in the field of computer vision. The introduction of ViT signifies the expansion of Transformer models from the domain of natural language processing to computer vision, ushering in a new paradigm for image processing tasks. Liang et al. (Liang et al., 2021). proposed an image restoration model called SwinIR. This model consists of three modules: shallow feature extraction, deep feature extraction, and HR reconstruction. It emphasizes the content-based interaction between image content and attention weights, achieved through a shifting window mechanism for long-range dependency modeling. The IPT (Chen et al., 2021) employs a multi-head, multi-tail, shared transformer body design, aiming to maximize the potential of the transformer architecture in serving various image processing tasks such as image super-resolution and denoising. The high computational complexity arising from this Transformer design may limit scalability to high-resolution images.

DRSAN (Park et al., 2021) proposes a dynamic residual network solution for lightweight super-resolution systems, leveraging different combinations of residual features considering input statistics. Additionally, it introduces residual self-attention, which, in collaboration with residual structures, enhances network performance without adding modules. Zamir et al. (Zamir et al.,

2022). introduced Restormer, an image restoration transformer model known for its high computational efficiency in handling high-resolution images. They made critical design adjustments to the core components of the transformer block to enhance feature aggregation and transformation. To integrate the robustness of CNNs into the Transformer model, Restormer incorporates deep convolutions for encoding spatial local context. ELAN (Zhang et al., 2022) utilizes shift convolution (shift-conv) to effectively extract local structural information from the image. Subsequently, it introduces an intra-group multi-scale self-attention (GMSA) module to leverage the long-range dependency of the image. Further acceleration of the model's computation is achieved by employing a shared attention mechanism. In the task of image super-resolution, the effectiveness of integrating local and global feature representations in the aforementioned methods still requires further improvement.

Diverging from current approaches, we introduce a lightweight deep model rooted in a hybrid dynamic Transformer (HDT-Net). The goal is to bolster the quality of image reconstruction by synergizing local and global feature representations.

3 Proposed method

In this section, we first describe the overall pipeline of the model. Then, we provide details of the hybrid dynamic transformer module (HDTM), which serve as the fundamental building modules of the approach. HDTM is composed of four identical hybrid dynamic transformer blocks (HDTBs) connected end to end, as illustrated in the Figure 2. The HDTB mainly comprises three key elements: dynamic local self-attention (DLSA), sparse non-local self-attention (SNSA), and feed-forward network (FFN).

3.1 Overall pipeline

Figure 2 illustrates an overview of the proposed HDT-Net for underwater image super-resolution. Specifically, the low-resolution underwater image is first processed through a convolutional layer with a filter size of 3×3 pixels for shallow feature extraction. Subsequently, the feature information is sequentially processed through six identical HDTMs for deep feature extraction and fusion, both locally and globally. Within each HDTM, four internal HDTBs are connected end to end for processing, and the extracted features are finally passed to the next module through a 3×3 convolution. After the completion of HDTM processing, the features are further projected using a convolutional layer with a filter size of 3×3 pixels. Following that, high-resolution image reconstruction is performed through a 3×3 convolution and upsampling operation using PixelShuffle (Shi et al., 2016).

The process of the overall pipeline can be represented as Equations 1-4:

$$X^{'} = \mathrm{Conv}_{3\times3}(X), \tag{1}$$

$$HDTM_s = HDTM_6(\ldots(HDTM_1(\mathbf{X}'))), \tag{2}$$

$$X_{\mathbf{low}} = \mathbf{X}' + HDTM_s(\mathbf{X}'), \tag{3}$$

$$X_{\mathrm{high}} = P(Conv_{3\times3}(X_{low})), \tag{4}$$

where $X, Conv3 \times 3, P(\cdot), X_{low}, X_{high}$ represent the input features and 3×3 convolution, upsampling operation using PixelShuffle, low resolution image features and high resolution image features, respectively. The process of HDTM in the overall pipeline can be expressed as Equations 5, 6:

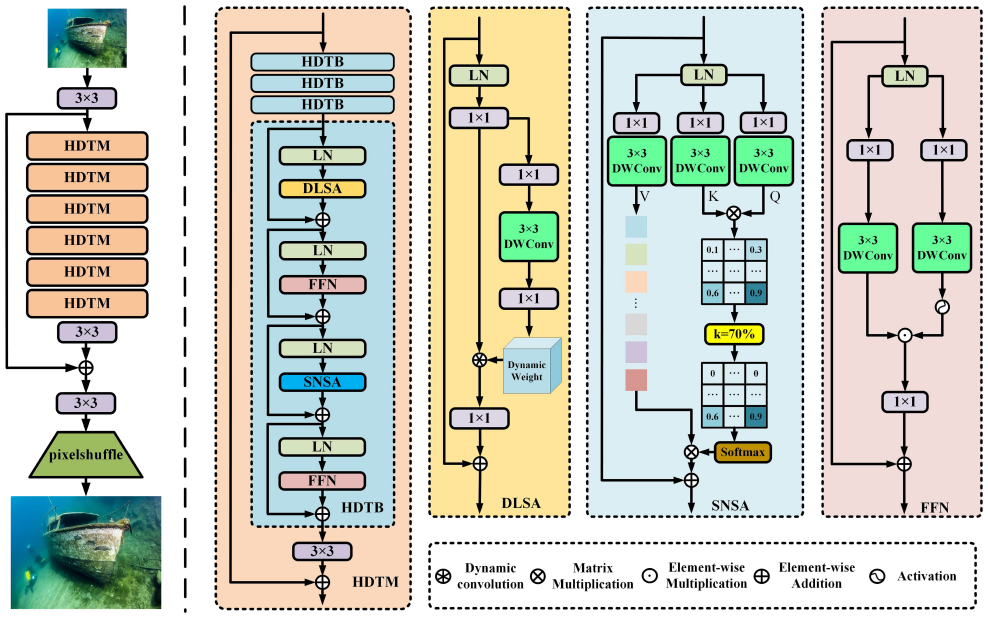


FIGURE 2
The overall architecture of the proposed network.

$$HDTB_s = HDTB_4(\ldots(HDTB_1(X))), \tag{5}$$

$$HDTM = X + Conv_{3\times3}(HDTB_s(X)) \tag{6}$$

3.2 Hybrid dynamic transformer block

We propose a hybrid dynamic transformer block consisting of DLSA, SNSA, and FFN. By combining DLSA and SNSA, the hybrid self-attention mechanism effectively weights each position against others in the input data, facilitating the integration of global information into each position's representation. Moreover, it enables the capturing of both global and local feature relationships at different positions in the image, allowing the model to capture long-range dependencies in the data. After each self-attention computation, the representation at each position undergoes non-linear transformation through FFN, mapping it to a new representation space to enhance the model's expressiveness. Formally, given the input features of the $(l-1)$-th block $\mathbf{X}_{l-1}$, the encoding of the HDTB process can be represented as Equations 7–10:

$$\mathbf{X}_l^d = \mathbf{X}_{l-1} + DLSA(LN(\mathbf{X}_{l-1})), \tag{7}$$

$$\mathbf{X}_l^f = \mathbf{X}_l^d + FFN(LN(\mathbf{X}_l^d)), \tag{8}$$

$$\mathbf{X}_l^s = \mathbf{X}_l^f + SNSA(LN(\mathbf{X}_l^f)), \tag{9}$$

$$\mathbf{X}_l = \mathbf{X}_l^s + FFN(LN(\mathbf{X}_l^s)), \tag{10}$$

where LN denotes the layer normalization, $\mathbf{X}_l^d$. d $\mathbf{X}_l^s$ represent the outputs of DLSA and SNSA, $\mathbf{X}_l^f$. d $\mathbf{X}_l$ represent the outputs of FFN, which are described below.

3.2.1 Dynamic local self-attention

To enhance the extraction and fusion of local features, we introduce a DLSA method aimed at capturing spatial relationships within an image, while also accommodating variable receptive fields. In contrast to conventional self-attention mechanisms, DLSA functions uniformly across the entire image. This dynamic approach empowers each spatial location to selectively attend to its nearby regions based on contextual cues. Specifically, given input features $X_{in} \in \mathbb{R}^{H\times W\times C}$ generated by layer normalization, 1 × 1 convolution is performed for feature aggregation. Similar to (Li et al., 2023a), we introduce a squeeze and excitation network (SENet) (Hu et al., 2018) as our dynamic weight generation network, which has no normalization layers and non-linear activations. Additionally, we employ a 3×3 depth-wise convolutional layer in SENet to encode features, ensuring better calculation of dynamic attention for local attention.

The proposed dynamic weight generation formula is as Equations 11-13:

$$\mathrm{X}_1 = \mathrm{DConv}_{3\times3}(\mathrm{Conv}_{1\times1}(\mathrm{X}_{\mathrm{in}})), \tag{11}$$

$$\mathrm{X}_2 = \mathrm{Conv}_{1\times1}(\mathrm{X}_1), \tag{12}$$

$$\mathrm{W}(\mathrm{x}) = \mathcal{R}(\mathrm{X}_2), \tag{13}$$

where $\mathcal{R}(\cdot)$ represents the reshaping function. In DLSA, we utilize learnable dynamic convolutions. Unlike traditional fixed kernels, learnable dynamic convolutional kernels offer greater flexibility and adaptability. Each pixel has a corresponding $K \times K$ dynamic kernel for dynamic convolution. We divide the number of feature channels into G heads, and learn separate dynamic weights in parallel. For the generated pixel-wise weights $\mathbf{W}$, we obtain the aggregated features using the following formula as Equation 14:

$$DLSA(\mathbf{X}) = \mathbf{W} \circledast \mathrm{X}_{\mathbf{in}}, \tag{14}$$

where $\circledast$ denotes the dynamic convolution operation using weight sharing across each channel.

3.2.2 Sparse non-local self-attention

Due to the fact that the dynamic estimated features generated by DLSA are based on fully convolutional operations, the efficiency of modeling global features is relatively low. To better perceive global features, we revisit the standard dot-product self-attention in Transformer (Zamir et al., 2022). However, this algorithm calculates attention maps based on fully connected operations for all query-key pairs. In our work, we develop SNSA to replace it, which leverages sparsity by selecting the top-k tokens (Chen et al., 2023) most relevant to the query, thus obtaining the most crucial information for computation. This approach avoids involving irrelevant information in the feature interaction process.

Specifically, we first perform feature aggregation by applying a 1 × 1 convolution, followed by a depthwise convolution with filter size of 3×3 pixels to encode per-channel contexts. This allows for self-attention computation across the three dimensions of query Q, key K, and value V, rather than spatial dimensions. Utilizing channel-wise similarity helps reduce memory consumption for efficient inference. Next, we compute the similarity between all pairs of queries and keys, and employ a selection strategy to mask out values with lower similarity, retaining those with higher similarity.

As shown in the Figure 2, k represents an adjustable parameter for dynamically setting the sparsity level. When k=70%, only the top 70% of elements with the highest scores are retained for activation, while the remaining 30% of elements are masked as 0. Finally, softmax is applied to normalize elements larger than the top-k, ensuring the output is a probability distribution. For elements with scores less than top-k, we use a scatter function to replace their probability at the given index with 0. This dynamic selection results in attention following a sparse distribution. Finally, matrix multiplication is used to multiply softmax with Value, which is then connected to the input residual through feature projection to obtain the final result.

The derivation formula for SNSA is as Equation 15:

$$SNSA(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathcal{S}\left(\mathcal{M}_k \odot \frac{\mathbf{Q}\mathbf{K}^{\top}}{\lambda}\right)\mathbf{V}, \tag{15}$$

where $S(\cdot)$ represents the softmax operation, λ is an optional temperature factor defined by $\lambda = \sqrt{d}$. Typically, multi-head attention is applied to each of the k new Q, K, and V, resulting in $d = C/k$ channel dimension outputs, which are then concatenated and projected linearly to obtain the final result of all heads.

$$[\mathcal{M}_k]_{ij} = \begin{cases} 1, \mathcal{M}_{ij} \in & \text{top} - \text{k} \quad (\text{row} \quad \text{j}) \\ 0, & \text{otherwise} \end{cases}, \tag{16}$$

where $\mathcal{M}_k$ denotes the top-k selection operator in Equation 16.

3.2.3 Feed-forward network

To extract sophisticated features from both the local and global self-attention data of the model and facilitate the learning of abstract representations, we introduce the FFN following the DLSA and SNSA modules. Specifically, we design two branches based on gating mechanisms. It first uses 1×1 convolutions for feature transformation and then employs 3 × 3 depth-wise convolutions to encode information from spatially adjacent pixel positions. One branch is used to expand feature channels, while in the other branch, it is activated along with the Gelu nonlinearity to reduce the channels back to the original input dimension and search for nonlinear contextual information in the hidden layers.

The FFN is formulated as Equations 17-19:

$$X_1 = GELU(\text{Conv}_{3\times3}(\text{Conv}_{1\times1}(X))), \tag{17}$$

$$X_2 = \text{Conv}_{3\times3}(\text{Conv}_{1\times1}(X)), \tag{18}$$

$$\hat{X} = \text{Conv}_{1\times1}(X_1 \odot X_2) + X\,. \tag{19}$$

In general, FFN plays a distinctly different role compared to self-attention. It controls the flow of information passing through various levels of our pipeline, allowing each level to focus on complementary contextual information to other levels.

3.3 Loss function

Building upon existing methods, we adopt the L1 loss function as the loss function for our model. The expression for the L1 loss function is defined as Equation 20 :

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \| y_i - \hat{y}_i \|_1, \tag{20}$$

where N is the number of samples in the dataset. y_i represents the ground truth value for the i-th sample. $\hat{y}_i$ represents the predicted value for the i-th sample.

The L1 loss function calculates the mean absolute error between the predicted values and the ground truth values, providing a measure of the average magnitude of the errors.

4 Experiments

In this section, we first introduce the implementation details, datasets and evaluation metrics. Then, we compare the proposed HDT-Net with 10 baseline methods, including bicubic, SRCNN (Dong et al., 2015), DSRCNN (Mao et al., 2016), SRGAN (Ledig et al., 2017), SRDM-GAN (Islam et al., 2020a), RFDN (Liu et al., 2020), LatticeNet+ (Luo et al., 2020), SMSR (Wang et al., 2021), IPT (Chen et al., 2021), and SwinIR (Liang et al., 2021). Finally, ablation experiments are conducted to validate the effectiveness of the proposed method. The experiments are trained on a server with two NVIDIA GeForce RTX 3090 GPUs.

4.1 Experimental settings

4.1.1 Implementation details

In the proposed SNSA, the threshold for top-k is set to 70%. We will analyze its impact in the ablation study. During the training, the batch size and patch size are configured as 16 and 64, respectively. The number of multi-head self-attention is set to be 6, and the number of feature is set to be 90. We utilize the Adam optimizer (Kingma and Ba, 2014) with default parameter configurations to train our model. The initial learning rate is established at 5×10^{-4}, employing a multi-step scheduler over 500K iterations.

4.1.2 Datasets and evaluation metrics

We validate the performance of various methods using the classic underwater image super-resolution benchmark datasets, USR-248 and UFO-120 (Liu et al., 2024). Each dataset showcases distinct underwater degradation characteristics, enabling comprehensive evaluation across diverse underwater imaging scenarios. Consistent with previous studies, we utilize PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) scores (Wang et al., 2004) to quantitatively compare the restoration results of different algorithms, enabling performance evaluation. In addition, we conduct evaluation calculations on the model parameter quantities of different deep networks.

4.2 Quantitative evaluation

Following (Dharejo et al., 2024), Table 1 presents the quantitative results of various methods on the USR-248 and UFO-120 datasets, including experimental setups with three different super-resolution scaling factors: ×2, ×4, and ×8. As shown, the experimental results demonstrate that our proposed HDT-Net consistently achieves the best quantitative performance. Compared to the state-of-the-art method SwinIR (Liang et al., 2021), our approach shows an average improvement of 0.5dB in PSNR, with a reduction in parameters by 58%. This indicates that our proposed hybrid transformer,

TABLE 1 Quantitative comparisons of different methods on the USR-248 and UFO-120 datasets.

Methods	Scale	USR-248		UFO-120		Average		Params(M)
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
Bicubic	x2	26.78	0.8263	27.01	0.8465	26.89	0.8364	–
SRCNN	x2	27.89	0.8467	27.12	0.8654	27.50	0.8560	0.067
DSRCNN	x2	28.12	0.8584	27.88	0.8731	28.00	0.8657	0.361
SRGAN	x2	28.41	0.8612	28.54	0.8815	28.47	0.8713	1.54
SRDM-GAN	x2	28.51	0.8592	28.58	0.8823	28.54	0.8707	0.586
RFDN	x2	28.72	0.8633	28.81	0.8841	28.76	0.8737	0.528
LatticeNet+	x2	28.74	0.8714	28.85	0.8854	28.79	0.8784	0.75
SMSR	x2	28.88	0.8712	28.91	0.8862	28.89	0.8787	0.985
IPT	x2	29.33	0.8831	29.05	0.8921	29.19	0.8876	11.3
SwinIR	x2	29.88	0.9018	30.01	0.9021	29.94	0.9019	11.45
Ours	x2	**31.23**	**0.9217**	**31.54**	**0.9168**	**31.38**	**0.9192**	4.71
Bicubic	x4	25.07	0.7823	25.12	0.8165	25.09	0.7994	–
SRCNN	x4	25.17	0.7978	25.21	0.8157	25.19	0.8067	0.067
DSRCNN	x4	25.78	0.8064	26.81	0.8177	26.29	0.8120	0.361
SRGAN	x4	26.09	0.8178	26.14	0.8188	26.11	0.8183	1.54
SRDM-GAN	x4	26.19	0.8211	26.51	0.8247	26.35	0.8229	0.586
RFDN	x4	26.66	0.8216	26.81	0.8350	26.73	0.8283	0.528
LatticeNet+	x4	26.78	0.8239	26.85	0.8245	26.81	0.8242	0.75
SMSR	x4	27.07	0.8296	27.15	0.8310	27.11	0.8303	0.985
IPT	x4	27.11	0.8626	27.16	0.8632	27.13	0.8629	11.3
SwinIR	x4	27.18	0.8634	27.27	0.8644	27.22	0.8639	11.45
Ours	x4	**27.69**	**0.8712**	**27.82**	**0.8745**	**27.75**	**0.8728**	4.71
Bicubic	x8	23.46	0.7684	23.84	0.7781	23.65	0.7732	–
SRCNN	x8	24.07	0.7877	24.12	0.7981	24.09	0.7929	0.067
DSRCNN	x8	24.12	0.7987	24.18	0.8031	24.15	0.8009	0.361
SRGAN	x8	24.22	0.8021	24.29	0.8024	24.25	0.8022	1.54
SRDM-GAN	x8	24.41	0.8162	24.47	0.8178	24.44	0.8170	0.586
RFDN	x8	24.55	0.8178	24.67	0.8218	24.61	0.8198	0.528
LatticeNet+	x8	25.08	0.8321	25.11	0.8324	25.09	0.8322	0.75
SMSR	x8	25.16	0.8344	25.23	0.8354	25.19	0.8349	0.985
IPT	x8	25.22	0.8353	25.34	0.8411	25.28	0.8382	11.3
SwinIR	x8	25.82	0.8555	26.04	0.8559	25.93	0.8557	11.45
Ours	x8	**26.37**	**0.8662**	**26.48**	**0.8655**	**26.42**	**0.8658**	4.71

Bold indicates the best results.

as opposed to window-based transformers, can better capture feature correlations. Particularly challenging is the task of image super-resolution at a scaling factor of ×8. In contract, our proposed solution, leveraging the efficient fusion of local and global information, exhibits robust performance advantages in complex underwater scenes.

4.3 Qualitative evaluation

Figures 3, 4 illustrate the visual comparison results of different methods on the USR-248 and UFO-120 datasets, respectively. Note that we do not compare RFDN (Liu et al., 2020) and LatticeNet+ (Luo et al., 2020) as their visual results are not available. It is evident

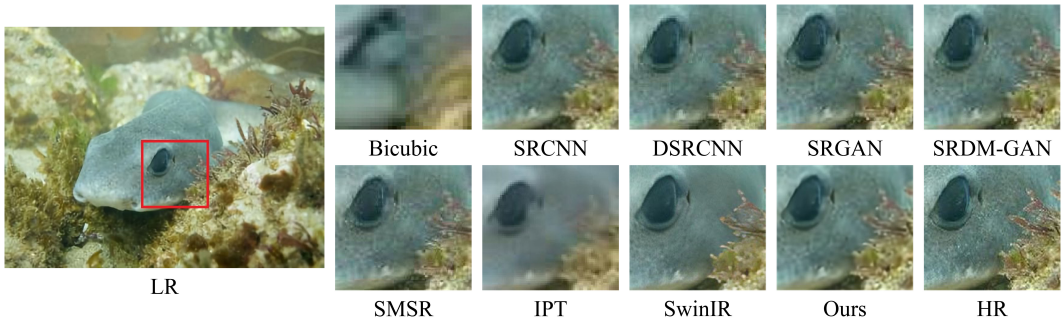


FIGURE 3
Image super-resolution comparisons for different methods on the USR-248 dataset.

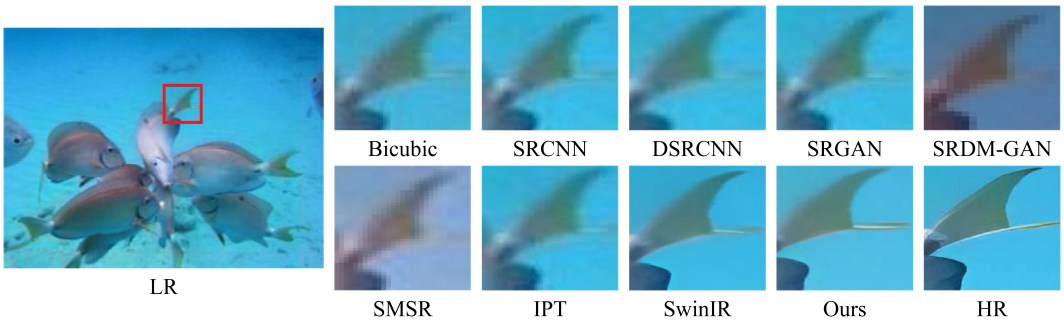


FIGURE 4
Image super-resolution comparisons for different methods on the UFO-120 dataset.

that effectively enhancing image resolution quality in complex underwater environments presents a formidable challenge compared to natural images. We find that the restoration results of most Transformer-based approaches tend to smooth out the details and textures of the images, which is attributed to the dense pattern of self-attention mechanisms. Furthermore, window-based self-attention global modeling methods fail to effectively aggregate information outside the window, thus affecting the quality of the restored images, as observed in SwinIR (Liang et al., 2021). In contrast, our proposed method achieves better image restoration by exploring the aggregation of local and global information. These quantitative and qualitative results indicate the effectiveness of the proposed hybrid dynamic Transformers, providing new insights into the challenging task of underwater image super-resolution.

4.4 Ablation study

In this section, we conduct a further analysis of the impact of the components proposed in our method and compare it against baseline models. To ensure a fair comparison, we employ the same settings used to train all baseline models as those of the proposed method. Here, we conduct ablation experiments with ×2 super-resolution on the USR-248 dataset. Specifically, the ablation study includes (1) effectiveness of the DLSA and SNSA, (2) effect of top-k values in the SNSA, and (3) effect of the number of HDTMs.

4.4.1 Effectiveness of the DLSA and SNSA

First, we analyze the effectiveness of the two key components proposed in the method, including DLSA and SNSA. To do this, we separately remove one of the components for comparative analysis. Table 2 presents the quantitative results of different variant models. It can be seen that our approach combining DLSA and SNSA achieves the best performance. Figure 5 illustrates the visual comparison results of different ablation models. It can be observed that, compared to using only a single self-attention mechanism for feature modeling, our proposed method can better restore the structure and detail regions of underwater images. The combination of local and non-local self-attention mechanisms enables the model to strike a balance between enhancing local details and preserving the overall scene context, resulting in more accurate and coherent super-resolved images.

4.4.2 Effect of top-k values in the SNSA

Next, we analyze the impact of the top-k value in SNSA. Regarding the choice of sparsity value, it also plays a crucial role in determining the performance of the model. A smaller sparsity value may result in a dense attention map, which could lead to

TABLE 2 Quantitative comparison of ablation results about the effectiveness of DLSA and SNSA.

Models	w/o DLSA	w/o SNSA	Ours
PSNR/SSIM	30.48/0.9060	29.63/0.8958	**31.23/0.9217**

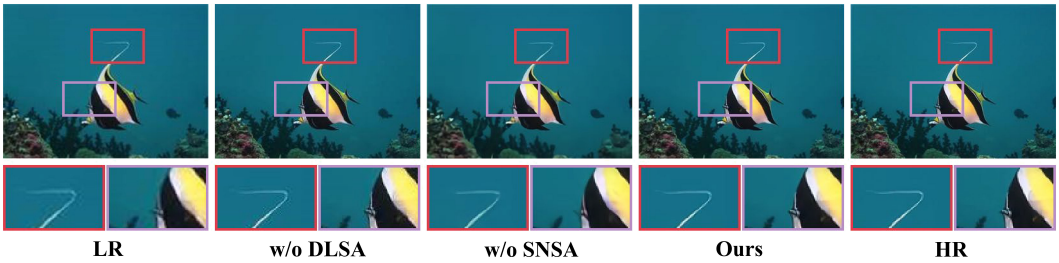


FIGURE 5
Visual comparison of ablation results about the effectiveness of DLSA and SNSA.

increased computational overhead and potential overfitting to noisy or irrelevant features. On the other hand, a larger sparsity value may cause the model to miss important global context or relevant features. Therefore, selecting an optimal sparsity value, such as k=70% in Figure 6, strikes a balance between capturing sufficient global information and maintaining computational efficiency, ultimately contributing to improved performance in underwater image super-resolution tasks.

4.4.3 Effect of the number of HDTMs

Finally, we analyze the impact of the number of HDTMs in the network backbone. Figure 7 presents the quantitative results using different numbers of HDTMs. It can be observed that when the number ranges from 6 to 8, the growth of PSNR value gradually converges. Therefore, to balance model efficiency and performance, we ultimately choose $N = 6$ as the configuration for the final network.

4.5 Limitations

While our proposed method demonstrates superior performance on classical underwater image super-resolution datasets (visible data) (Liu et al., 2024), its applicability is currently somewhat limited. The model's performance is significantly affected in scenarios with low light conditions, such as deep-sea environments or areas with poor visibility, where methods utilizing sonar (Yang, 2023; Zhang et al., 2024) for detection are more prevalent. To adapt our method to a wider range of underwater scenarios, we will explore the potential applications of the proposed method in sonar images.

5 Conclusions

In this paper, we have proposed an effective hybrid dynamic Transformer for underwater image super-resolution. We demonstrate the crucial importance of jointly exploring local features and global information in underwater image reconstruction for achieving high-quality results. At the technical level, we integrate dynamic local self-attention and sparse non-local self-attention to stack into the hybrid dynamic transformer module, forming the backbone of our proposed method. The former effectively captures details in underwater image regions, while the latter aids in the recovery of global image structure and color. Our proposed method achieves satisfactory reconstruction results on benchmark datasets. In future work, we will explore the extension of this hybrid transformer approach to other navigation-related visual tasks.

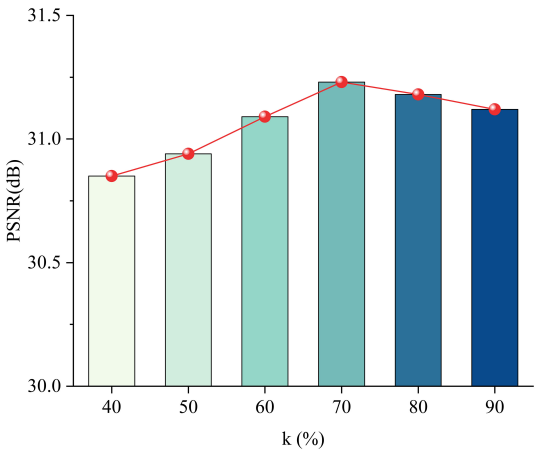


FIGURE 6
Comparison of ablation results about the effect of top-k values in the SNSA.

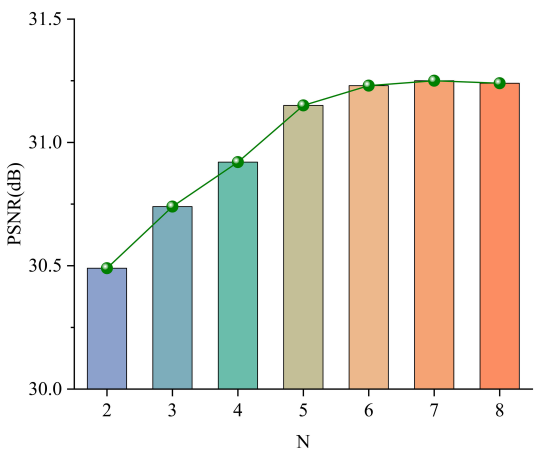


FIGURE 7
Comparison of ablation results about the effect of the number of HDTMs.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://drive.google.com/drive/folders/1dCe5rlw3UpzBs25UMXek1JL0wBBa697Q; https://www.v7labs.com/open-datasets/ufo-120.

Author contributions

XH: Data curation, Formal analysis, Methodology, Writing – original draft. JL: Investigation, Writing – review & editing. TJ: Data curation, Software, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by the Research Project of the Naval Staff Navigation Assurance Bureau (2023(1)).

Conflict of interest

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

Publisher's note

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

References

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., et al. (2021). "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Virtual: IEEE), 12299–12310.

Chen, X., Li, H., Li, M., and Pan, J. (2023). "Learning a sparse transformer network for effective image deraining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Canada: IEEE), 5896–5905. doi: 10.1109/CVPR52729.2023.00571

Chen, Y., Niu, K., Zeng, Z., and Pan, Y. (2020). "A wavelet based deep learning method for underwater image super resolution reconstruction," in *IEEE Access* (IEEE), 8, 117759–117769.

Dharejo, F. A., Ganapathi, I. I., Zawish, M., Alawode, B., Alathbah, M., Werghi, N., et al. (2024). Swinwave-sr: Multi-scale lightweight underwater image super-resolution. *Inf. Fusion* 103, 102127. doi: 10.1016/j.inffus.2023.102127

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). "Image super-resolution using deep convolutional networks," in *IEEE transactions on pattern analysis and machine intelligence* (IEEE), 38, 295–307.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Guo, Y., Li, H., and Zhuang, P. (2019). "Underwater image enhancement using a multiscale dense generative adversarial network," in *IEEE Journal of Oceanic Engineering* (IEEE), 45, 862–870.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., et al. (2022). "A survey on vision transformer," in *IEEE transactions on pattern analysis and machine intelligence* (IEEE), 45, 87–110.

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 7132–7141.

Islam, M. J., Enan, S. S., Luo, P., and Sattar, J. (2020a). "Underwater image super-resolution using deep residual multipliers," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. (Virtual: IEEE), 900–906. doi: 10.1109/ICRA40945.2020

Islam, M. J., Xia, Y., and Sattar, J. (2020b). "Fast underwater image enhancement for improved visual perception," in *IEEE Robotics and Automation Letters*. (Virtual: IEEE), 5, 3227–3234. doi: 10.1109/LSP.2016.

Jiang, N., Chen, W., Lin, Y., Zhao, T., and Lin, C.-W. (2021). "Underwater image enhancement with lightweight cascaded network," in *IEEE transactions on multimedia*. (IEEE) 24, 4301–4313.

Kim, J., Lee, J. K., and Lee, K. M. (2016). "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE) 1646–1654.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 4681–4690.

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). "Underwater image enhancement via medium transmission-guided multi-color space embedding," in *IEEE Transactions on Image Processing*. (IEEE) 30, 4985–5000.

Li, X., Dong, J., Tang, J., and Pan, J. (2023a). "Dlgsanet: lightweight dynamic local and global self-attention networks for image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (France: IEEE), 12792–12801. doi: 10.1109/ICCV51070.2023.01175

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). "An underwater image enhancement benchmark dataset and beyond," in *IEEE Transactions on Image Processing*. (IEEE) 29, 4376–4389.

Li, Y., Shen, L., Li, M., Wang, Z., and Zhuang, L. (2023b). "Ruiesr: Realistic underwater image enhancement and super resolution," in *IEEE Transactions on Circuits and Systems for Video Technology*. (IEEE). doi: 10.1109/TCSVT.2023.3328785

Li, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2017). "Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images," in *IEEE Robotics and Automation letters*. (IEEE) 3, 387–394.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. (2021). "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*. (Virtual: IEEE), 1833–1844.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. (2017). "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. (USA: IEEE), 136–144.

Liu, B., Ning, X., Ma, S., and Yang, Y. (2024). Multi-scale dense spatially-adaptive residual distillation network for lightweight underwater image super-resolution. *Front. Mar. Sci.* 10, 1328436. doi: 10.3389/fmars.2023.1328436

Liu, J., Tang, J., and Wu, G. (2020). "Residual feature distillation network for lightweight image superresolution," in *Computer Vision–ECCV 2020 Workshops*, UK, 2020. 41–55, Proceedings, Part III 16 (UK: Springer).

Luo, X., Xie, Y., Zhang, Y., Qu, Y., Li, C., and Fu, Y. (2020). "Latticenet: Towards lightweight image superresolution with lattice block," in *Computer Vision–ECCV 2020: 16th European Conference*, 2020. 272–289, Proceedings, Part XXII 16 (UK: Springer).

Mao, X.-J., Shen, C., and Yang, Y.-B. (2016). Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*..

Park, K., Soh, J. W., and Cho, N. I. (2021). "Dynamic residual self-attention network for lightweight single image super-resolution," in *IEEE Transactions on Multimedia*. 25, 907–918.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., et al. (2016). "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 1874–1883.

Shin, Y.-S., Cho, Y., Pandey, G., and Kim, A. (2016). "Estimation of ambient light and transmission map with common convolutional architecture," in *OCEANS 2016 MTS/IEEE Monterey (IEEE)*. (IEEE) 2016, 1–7.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). "Image quality assessment: from error visibility to structural similarity," in *IEEE transactions on image processing*. (IEEE) 13, 600–612.

Wang, L., Dong, X., Wang, Y., Ying, X., Lin, Z., An, W., et al. (2021). "Exploring sparsity in image super-resolution for efficient inference," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 4917–4926.

Wang, Y., Zhang, J., Cao, Y., and Wang, Z. (2017). "A deep cnn method for underwater image enhancement," in *2017 IEEE international conference on image processing (ICIP)*. (IEEE) 2017, 1382–1386.

Xiao, Y., Yuan, Q., Jiang, K., He, J., Lin, C.-W., and Zhang, L. (2024). "Ttst: A top-k token selective transformer for remote sensing image super-resolution," in *IEEE Transactions on Image Processing*. (IEEE). doi: 10.1109/TIP.2023.3349004

Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimedia Tools Appl.* 1–17. doi: 10.1007/s11042-023-16757-0

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M.-H. (2022). "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 5728–5739.

Zhang, X., Yang, P., Wang, Y., Shen, W., Yang, J., Wang, J., et al. (2024). "A novel multireceiver sas rd processor," in *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/TGRS.2024.3362886

Zhang, X., Zeng, H., Guo, S., and Zhang, L. (2022). "Efficient long-range attention network for image super-resolution," in *European Conference on Computer Vision*, (Israel: Springer), 649–667.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 05 June 2024
DOI 10.3389/fmars.2024.1390931

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Maohan Liang,
National University of Singapore, Singapore
Hongchu Yu,
Wuhan University of Technology, China

*CORRESPONDENCE
Qiang Mei
meiqiang@jmu.edu.cn
Peng Wang
wangp@ict.ac.cn

RECEIVED 24 February 2024
ACCEPTED 13 May 2024
PUBLISHED 05 June 2024

CITATION
Li Y, Liu X, Wang Z, Mei Q, Xie W, Yang Y and Wang P (2024) Construction of a large-scale maritime element semantic schema based on knowledge graph models for unmanned automated decision-making.
Front. Mar. Sci. 11:1390931.
doi: 10.3389/fmars.2024.1390931

Construction of a large-scale maritime element semantic schema based on knowledge graph models for unmanned automated decision-making

Yong Li[1], Xiaotong Liu[1], Zhishan Wang[1], Qiang Mei[2,3]*, Wenxin Xie[1], Yang Yang[4] and Peng Wang[2,5]*

[1]School of Software Engineering, Beijing University of Technology, Beijing, China, [2]Merchant Marine College, Shanghai Maritime University, Shanghai, China, [3]Navigation College, Jimei University, Xiamen, China, [4]School of Geographic Science, East China Normal University, Shanghai, China, [5]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

In maritime logistics optimization, considerable research efforts are focused on the extraction of deep behavioral characteristics from comprehensive shipping data to discern patterns in maritime vessel behavior. The effective linkage of these characteristics with maritime infrastructure, such as berths, is critical for the enhancement of ship navigation systems. This endeavor is paramount not only as a research focus within maritime information science but also for the progression of intelligent maritime systems. Traditional methodologies have primarily emphasized the analysis of navigational paths of vessels without an extensive consideration of the geographical dynamics between ships and port infrastructure. However, the introduction of knowledge graphs has enabled the integration of disparate data sources, facilitating new insights that propel the development of intelligent maritime systems. This manuscript presents a novel framework using knowledge graph technology for profound analysis of maritime data. Utilizing automatic identification system (AIS) data alongside spatial information from port facilities, the framework forms semantic triplet connections among ships, anchorages, berths, and waterways. This enables the semantic modeling of maritime behaviors, offering precise identification of ships through their diverse semantic information. Moreover, by exploiting the semantic relations between ships and berths, a reverse semantic knowledge graph for berths is constructed, which is specifically tailored to ship type, size, and category. The manuscript critically evaluates a range of graph embedding techniques, dimensionality reduction methods, and classification strategies through experimental frameworks to determine the most efficacious methodologies. The findings reveal that the maritime knowledge graph significantly enhances the semantic understanding of unmanned maritime equipment, thereby improving decision-making capabilities. Additionally, it establishes a semantic foundation for the development of expansive maritime models, illustrating the potential of knowledge graph technology in advancing intelligent maritime systems.

10.3389/fmars.2024.1390931

KEYWORDS

knowledge graph, graph embedding, intelligent maritime, ship classification, similar berth recommendation

1 Introduction

The advent of smart maritime systems was marked by the integration of distributed devices and the application of artificial intelligence (AI) and machine learning (ML) technologies. These systems extensively utilize sensors, including GPS, radar, and meteorological sensors, to capture real-time positioning and vital status information of vessels. However, the maritime domain is characterized by intricate spatiotemporal relationships, influenced by factors such as ship trajectories, weather variations, and port activities. Traditional AI and ML techniques face challenges in accurately modeling these relationships due to their limited knowledge representation capabilities, which fail to generalize the dynamic and multidimensional nature of the maritime environment. Additionally, the heterogeneity of maritime data, arising from disparate sources and formats, poses significant challenges in data standardization and interoperability in initial smart maritime systems.

To address these limitations, this study introduces knowledge graphs as a sophisticated method for exploring complex spatiotemporal relationships within the maritime domain. By defining entities and their interconnections, knowledge graphs offer a robust mechanism for semantic consistency, facilitating the resolution of semantic discrepancies during the information fusion process and ensuring data standardization and consistency. This knowledge graph-based approach, which transcends traditional distributed device and ML methods, enables advanced intelligent queries and reasoning with its graph data representation. Moreover, it exhibits remarkable scalability, allowing for the seamless integration of new entities and relationships (Fensel et al., 2020). Within the maritime sector, this methodology enables comprehensive analyses through the rules and relationships defined in the knowledge graph, thereby offering flexible adaptation to the evolving informational landscape and supporting a wider array of application scenarios.

A knowledge graph constitutes a network of interconnected entities and their relationships, where each entity and its attributes represent distinct knowledge points, and the relationships describe the associations among these points. Formally, a knowledge graph G is represented as $G \in (E, R, S)$, where $E = \{e_1, e_2, \ldots, e_n\}$ denotes the entity set in the knowledge base, encompassing $|E|$ distinct entities; $R = \{r_1, r_2, \ldots, r_n\}$ signifies the relationship set, comprising $|R|$ different relationships; and $S \subseteq E \times R \times E$ encapsulates the triple set within the knowledge base. A triplet, the fundamental unit of this representation, is denoted as $s = (h, r, t)$, where $h, t \in E$ represent the head and tail entities in the triplet, and $r \in R$ denotes the relationship connecting these two entities. For instance, in the triplet (*Paris, the Capital of, France*), the head "*Paris*" and the tail "*France*" are real-world entities connected through the relationship "*the Capital of*". Graph data in knowledge graphs, characterized by numerous nodes and edges, represent a high-dimensional data structure. However, processing such data can introduce computational complexity, the curse of dimensionality, and challenges in interpretability and visualization. Graph embedding techniques, therefore, are crucial for transforming high-dimensional graph data into manageable, lower-dimensional vector representations, capturing the intrinsic relationships, similarities, and structural characteristics among graph nodes (Wang et al., 2017). Prominent embedding methodologies like TransE (Bordes et al., 2013), TransD (Wang et al., 2014), and RESCAL (Nickel et al., 2011), among others, facilitate dimensionality reduction and vectorization of graphs through diverse principles. Knowledge graphs and graph embedding techniques have found application across intelligent transportation and maritime sectors (Ahmed et al., 2022; Liu et al., 2023a; Wang et al., 2023a; Liu et al., 2023b; Li et al., 2024) for functions such as traffic anomaly detection, marine environmental monitoring, and maritime geographic analysis, presenting their versatility and utility in addressing industry-specific challenges.

In shipping, ships form similar geographical relationships with geographic elements such as waterways, anchorages, and berths. Meanwhile, berths also consider the size and cargo type of ships in order to accommodate them, forming a relationship with the attributes of the ships. However, traditional recommendation methods are manual and cannot empower decision-making machines to meet the requirements of shipping. Therefore, the following two problems need to be addressed in research: (1) How to characterize the behavior of ships and the profile of berths using an understandable semantic pattern; (2) How to assist unmanned devices in making accurate decisions based on this pattern, thereby ensuring the safety of jurisdictional waters and efficiency of the port.

This study introduces the "Maritime Heterogeneous Knowledge Graph Brain", a concept based on the heterogeneous knowledge graph's construction. This system capitalizes on the heterogeneous knowledge graph data to autonomously perform tasks such as node classification, clustering, link prediction, and multimodal fusion, which integrates node and edge features, graph structure, and contextual characteristics. The "Maritime Heterogeneous Knowledge Graph Brain" employs diverse maritime graph data to enhance ship dispatch efficiency, optimize maritime decision-making, and improve the utilization of shore-based facilities, thereby boosting emergency response capabilities in ports and maritime environments. As depicted in Figure 1, the maritime heterogeneous knowledge graph is stored across multiple distributed databases. Information is then transmitted to various

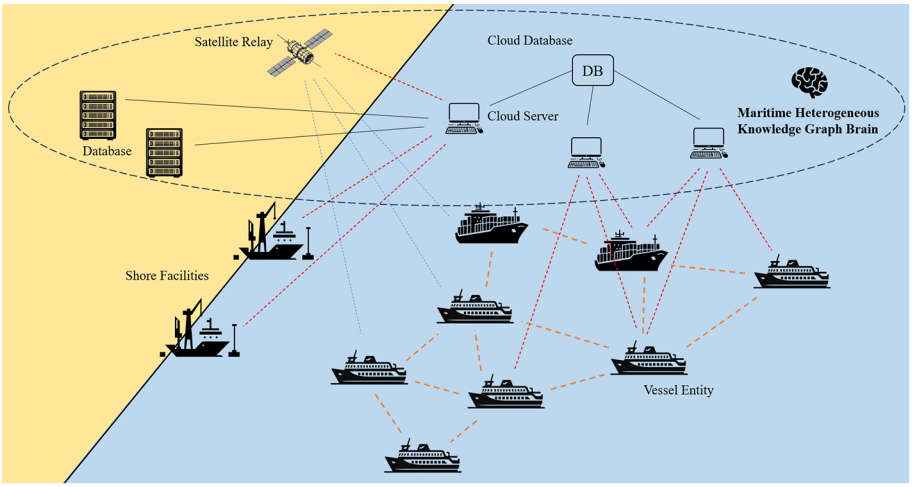


FIGURE 1
Illustration of the "Maritime Heterogeneous Knowledge Graph Brain" concept.

task-specific terminals, such as ships and shore-based facilities, via servers dedicated to different tasks. This arrangement facilitates real-time processing and response, advancing the development of maritime intelligence. The study conducts experiments and analyzes the effects based on the previously constructed maritime heterogeneous knowledge graph, focusing on typical applications such as ship classification and similar berth recommendation.

This manuscript combines knowledge graphs with maritime domain expertise to introduce a pioneering intelligent maritime application framework predicated on knowledge graph technology. The principal contributions of this manuscript are shown below:

1. The manuscript proposes pragmatic solutions to real-world challenges within the maritime industry. By using knowledge graph embedding technology, it addresses the uncertainties inherent in ship type prediction and the complexities surrounding similar berth recommendations for these two pivotal business scenarios. These solutions significantly contribute to enhancing the operational efficiency of shipping companies, optimizing berth allocations, and minimizing waiting times.
2. An innovative intelligent maritime application framework, grounded in knowledge graphs and graph embedding technology, is proposed. This framework not only integrates complex data pertaining to ships and infrastructure but also elucidates deep semantic relationships between entities via graph embedding algorithms. Designed with the maritime industry's specific requisites in mind, its modular architecture ensures both scalability and adaptability for prospective applications.
3. The paper investigates the utilization of this framework in facilitating specific downstream tasks, specifically ship type prediction and similar berth recommendation. Through the optimization of embedding models, classification models, and other components, significant enhancements in task accuracy were achieved. Moreover, experimental validation of the influence of varied weight distributions within the combined model on the performance of the berth recommendation system provides novel insights for the intelligent recommendation system design.

The structure of this manuscript is methodically organized as follows: Section 1 introduces the prevailing challenges in the intelligent maritime domain and advocates for the integration of knowledge graph methodologies. Section 2 delves into related works concerning knowledge graphs and the analysis of ship behaviors. Section 3 introduces the proposed intelligent maritime application framework that leverages knowledge graphs and graph embedding techniques. Section 4 validates the framework's efficacy and optimal performance for various downstream tasks through comparative experiments. Section 5 discusses the framework's applicability scope, alongside its algorithmic constraints. Section 6 concludes the manuscript, outlining future research trajectories.

2 Related work

2.1 Overview of related work on knowledge graphs

Knowledge graphs have become instrumental in the realm of knowledge storage and representation, exerting a significant influence across diverse domains. The emergence of big data coupled with advancements in AI technologies has propelled the exploration and application of knowledge graphs into a prominent position within various disciplines. These fields include computer science, natural language processing, information retrieval, ML, and human-computer interaction (Fensel op. cit.; Wang op. cit.).

In exploring algorithms related to knowledge graphs, current research trends focus primarily on enhancing the performance of downstream tasks and refining the accuracy of knowledge graph

embeddings. While traditional embedding models have shown commendable results in tasks like knowledge graph completion, they often overlook temporal dynamics and the integration of background knowledge. Li et al. (2023a) put forward an innovative rule-based embedding technique that extracts attributes from entities, employing logical rules to augment datasets, which in turn enhances the precision of knowledge graph completion endeavors. Jiang et al. (2024) unveiled a cutting-edge link prediction framework that leverages a multi-source hierarchical neural network based on knowledge graph embeddings, aimed at overcoming challenges in extracting intricate graph information and fostering the fusion of multiple feature knowledge semantics.

Owing to their robust knowledge representation capabilities, knowledge graphs have found applications in diverse fields, including recommendation systems (Bertram et al., 2023), intelligent healthcare (Yang et al., 2024), and industrial production (Zhengyu et al., 2022). For instance, Chen et al. (2022) utilized migration data from official provincial and city websites to craft a knowledge graph that maps the activities of COVID-19 patients, facilitating tracking, visualization, and reporting efforts in managing the pandemic. In the maritime sphere, Liu et al. (2023c) developed a Transformer-based model employing multi-convolution bidirectional encoders for extracting essential information from maritime pollution prevention regulations and laws, aiding in port state control inspections to minimize ship pollution. Furthermore, Gan et al. (2023) analyzed ship collision investigation reports issued by the China Maritime Safety Administration, proposing a novel approach for constructing knowledge graphs to elucidate the underlying factors of maritime accidents, thereby aiming to improve maritime traffic safety.

Despite the extensive utility of knowledge graphs across various sectors, challenges remain in the domain of geographical information, especially concerning maritime transportation. One primary issue is the need for real-time updates; maritime data, such as ship positions and weather conditions, are subject to rapid changes, necessitating knowledge graphs that can accommodate immediate updates to mirror the current realities. Additionally, the integration of disparate data sources to achieve data consistency and connectivity for a comprehensive knowledge graph poses significant hurdles. The diversity in data standards and formats across sources calls for the development of robust data integration and cleansing methodologies to ensure seamless knowledge graph construction.

2.2 Overview of related work on ship behavior mining

The advent of deep neural networks, computer vision, and natural language processing has spotlighted ship behavior mining as a key area within the intelligent maritime domain. Focused research endeavors in this area include ship trajectory detection (Deng et al., 2023; Zhang et al., 2023), intention prediction (Gao and Shi, 2020; Murray and Perera, 2021), and classification (Zhou et al., 2019), employing advanced computational models like convolutional neural networks (CNNs) (Chen et al., 2020, 2024) and generative adversarial networks (GANs) (Jia et al., 2023). Wang et al. (2023b) proposed a ship trajectory prediction model based on a sparse multi-graph convolutional hybrid network. This model simulates the dynamics and movement patterns of ships across spatial and temporal dimensions, incorporating features of ship trajectories that are sensitive to time into the prediction framework. Ma et al. (2021), by conducting statistical analysis on recorded ship movement trajectories, discovered that ship movements frequently show a strong correlation with their long-term historical trajectories. Consequently, they proposed an augmented long short-term memory network (ALSTM), which incorporates skip connections and adaptive memory modules into the traditional LSTM structure. This enhancement enables current memory units to engage with past data, facilitating a more sophisticated representation of the uncertainty and varied movements of individual ships. Zheng et al. (2022) introduced a proximal policy optimization and route guidance (PPORG) algorithm for autonomous ship navigation, focusing on collision avoidance and route optimization.

Ship classification research has also seen significant advancements. Escorcia-Gutierrez et al. (2022) employed masked convolutional neural networks for the detection of small ships and used a collision body optimization algorithm alongside a weighted regularized extreme learning machine method for the efficient classification of the detected ships. Similarly, Liang et al. (2021) introduced a multi-view feature fusion network (MVFFNet) for ship classification in the context of imbalanced data. This approach begins by extracting various multi-view features from automatic identification system (AIS)-based ship trajectories, followed by the application of a bidirectional gated recurrent unit network to amalgamate these multi-view features, thereby producing the ship classification results.

Multimodal approaches represent a prominent method for mining ship behavior by extracting features from various sensing modes and integrating these features to gain a more thorough understanding and enhanced data mining capabilities. Guo et al. (2023) developed a multimodal ship trajectory prediction approach through pattern distribution modeling. This approach utilizes a vector, randomly sampled from a multivariate Gaussian distribution, as the representation of trajectory patterns to generate multiple predicted trajectories. It employs adversarial learning to allow this Gaussian distribution to effectively capture ship trajectory patterns. Xiao et al. (2023) introduced an adaptive data fusion model that leverages multi-source AIS data for ship trajectory prediction. This model merges maritime mobile service identifiers and timestamps with multi-source AIS data, utilizing deep learning techniques for feature extraction and to enhance adaptability. Wang et al. (2023c) introduced a ship trajectory prediction model utilizing a sparse multi-graph convolutional hybrid network (SMCHN), which simulates interactions and movement trends among ships across temporal and spatial dimensions. This model enhances its predictive capability by integrating multi-source information and adjusting weights, and incorporates a temporal convolutional network with a gating mechanism for future trajectory predictions. Concurrently, Wang et al. (2023d) proposed a propagation trajectory interval prediction

framework that employs upper and lower bound estimation and attention-modified long short-term memory (LSTM) networks optimized through Bayesian techniques. This framework adopts dual optimization strategies tailored to interval prediction of ship trajectories, focusing on differentiated interval widths for longitude and latitude and a hyperparameter optimization to minimize coverage width criterion. Wang et al. (2024) presented an approach that merges spatial and temporal models to extract high-level features of ships from historical trajectory data. Employing a spatio-temporal graph convolutional network (ST-GCN) followed by RT-CNN, this method captures temporal dependencies of spatial interaction features, yielding refined spatio-temporal trajectory predictions. Shin et al. (2024) developed the AIS-ACNet, a deep learning framework for predicting AIS data that utilizes auxiliary tasks and a convolutional encoder. This model leverages diverse AIS data features and effectively integrates ship dynamics to enhance trajectory prediction accuracy. Zhang et al. (2024) introduced TrajBERT-DSSM, a novel method for ship destination prediction that analyzes AIS records to compare navigational trajectories of vessels with historical data, focusing on spatio-temporal correlations, geometric properties, and motion patterns. This method integrates a hierarchical geospatial coding system, geohash, with TrajBERT and the deep structured semantic model (DSSM) to assess trajectory similarity and predict destinations. Liu H et al. (2023) proposed a mechanism for monitoring and detecting abnormal ship behavior using a graph attention predictive and reconstructive network. This mechanism employs a sliding window technique for consistent data input and utilizes the peak over threshold (POT) method to dynamically adjust anomaly detection thresholds based on environmental changes in marine settings. Similarly, Li et al. (2022) presented a method for the semantic recognition of ship entry and exit movements based on a probabilistic topic model. This method is capable of uncovering ship movement patterns from vast amounts of trajectory data in an unsupervised way, thereby rendering the results more interpretable.

Furthermore, some studies have extended intelligent systems to other aspects of the maritime domain. Li X et al. (2023) introduced the temporal fusion transformer (TFT), a forecasting model based on attention mechanisms, taking the Tianjin Port maritime area as a case study to achieve multi-period, multi-feature forecasts of pollutant emissions, providing data references for management decisions of relevant departments. Xie et al. (2023) utilized the TFT, a deep learning model for time series forecasting based on attention mechanisms, to predict the spatiotemporal characteristics of ship emissions over multiple periods, achieving fine-grained traceability of ship emissions.

Our maritime prediction framework distinguishes itself from existing research by its foundation on a heterogeneous knowledge graph, which incorporates various data processing methods tailored to different downstream tasks. This design not only enhances the framework's scalability but also improves its reusability across diverse tasks.

3 Technical framework and dataset

3.1 Technical framework

This paper introduces an application framework leveraging the versatility of knowledge graphs within the maritime domain. The framework is designed around two key application scenarios: ship classification and similar berth recommendations. It integrates knowledge graph embeddings using comprehensive AIS data from ports and ships, utilizing the resulting vectors for specific downstream tasks. Additionally, the framework includes a visualization component, facilitating the graphical representation of maritime knowledge graphs derived from the dataset.

3.1.1 Process flowchart

The operational flow of this framework is illustrated in Figure 2. The initial phase involves constructing ship and berth knowledge graphs from triplet data. Following this, a knowledge graph embedding model trains on this data to learn the representative vectors for ships and berths. To improve computational efficiency, these learned vectors undergo dimensionality reduction. The dimensionally reduced vectors are then employed in data analysis for downstream tasks, specifically focusing on ship classification models and similar berth recommendation strategies.

3.1.2 Graph embedding model

The data within knowledge graphs consist of an extensive array of nodes and edges, embodying a form of high-dimensional data. While processing such data is plausible, it often incurs challenges including computational complexity, the curse of dimensionality, and impediments in interpretability and visualization. To address these challenges, graph embedding methods are employed to transform high-dimensional graph data into more manageable, lower-dimensional vector representations. This process entails converting nodes and edges within knowledge graphs into vectors within a vector space, enabling these vectors to encapsulate the relationships, similarities, and structural characteristics among nodes. Such transformation facilitates the application of knowledge graphs across diverse data analysis and ML endeavors. In our framework, we adopt various graph embedding models including TransE (Bordes op. cit.), TransD (Wang op. cit.), TransH (Lin et al., 2015), TransR (Ji et al., 2017), Analogy (Liu et al., 2017), ComplEx (Trouillon et al., 2016), and DistMult (Yang et al., 2014) to capture and elucidate the latent features inherent in the relationships and entities associated with berths. This methodology allows for encoding each berth into vectors of uniform dimensionality, thereby furnishing a quantifiable and analyzable feature representation for berths. Here is a brief overview of the graph embedding model implemented in this study.

TransE is a seminal approach to knowledge graph embedding. Central to TransE is the notion of transforming entities and their interrelations into vectors within a unified, low-dimensional vector

10.3389/fmars.2024.1390931

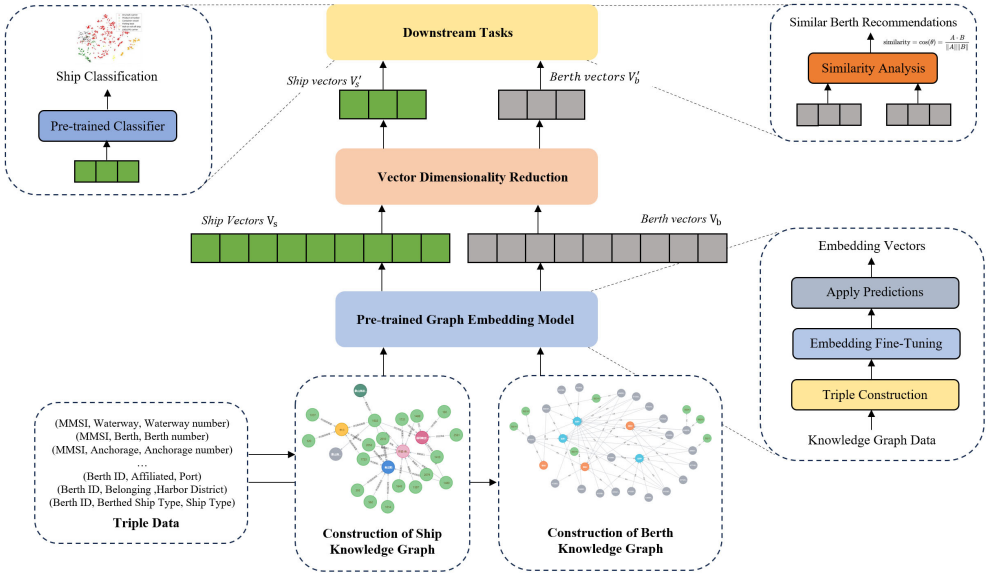


FIGURE 2
Framework flowchart.

space, effectively capturing the relational semantics between them. For each fact triple (*h, r, t*) in the knowledge graph, TransE treats this as a translational operation where the vector of the head entity *h*, through the action of the relation vector *r*, aligns with the vector of the tail entity *t*. Figure 3A illustrates the TransE model schematic. The model is trained to minimize the translational distance for positive sample embeddings while maximizing it for negative ones. This approach not only enhances the model's efficacy in tasks like link prediction but also enhances computational efficiency. However, TransE has limitations in addressing complex relationships, such as one-to-many, many-to-one, and many-to-many scenarios.

To overcome these limitations, the TransD model extends the capabilities of TransE. Illustrated in Figure 3B, TransD introduces a dynamic mapping matrix for each entity-relation pair, which facilitates the generation of new embeddings for head and tail entities. This matrix, derived from the original embedding vectors of the entities and relations, allows TransD to flexibly model a variety of relationship types. Given a fact triple (*h, r, t*), In TransD, the mapped head and tail entity vectors are denoted as $h' = h + M_r h$ and $t' = t + M_r t$, respectively, with M_r being the mapping matrix calculated based on relation *r*. Like TransE, TransD employs translational operations to link the head and tail entities, but its introduction of a dynamic mapping matrix significantly augments

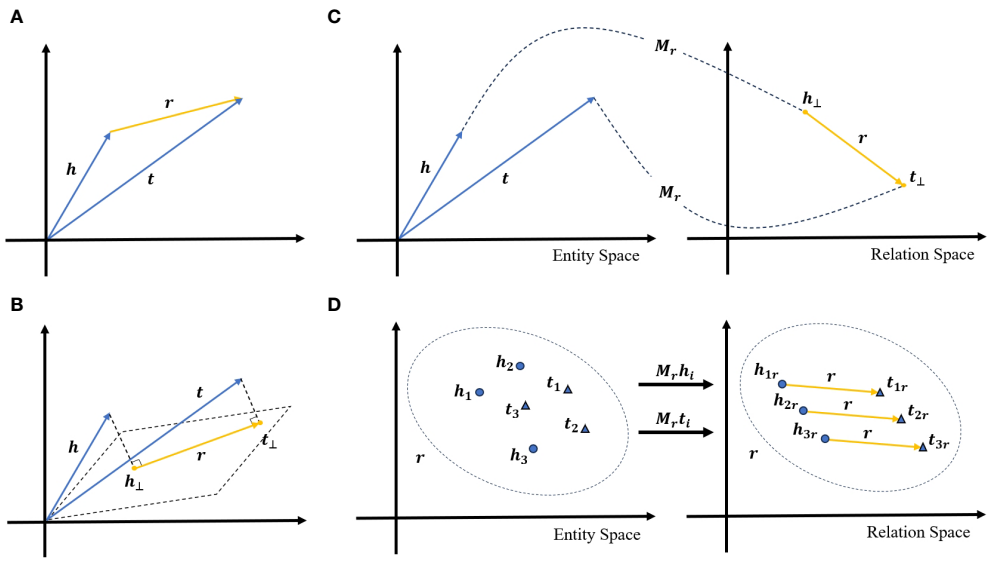


FIGURE 3
(A–D) Schematic diagram of the graph embedding model based on vector decomposition.

the model's ability to handle complex relational structures, building upon the foundational principles established by TransE.

The TransH model builds upon the foundations of the TransE model by introducing a novel approach that models each relation as a hyperplane. In this setup, entities are projected onto this hyperplane using relation-specific projection vectors, followed by a translational operation similar to TransE. Figure 3C illustrates the schematic diagram of the TransH model. For a given fact triplet (*h, r, t*), TransH first computes the projection vectors $h_\perp$ and $t_\perp$ of the head and tail entities onto the hyperplane corresponding to relation *r*. Then, on this hyperplane, it ensures that $h_\perp$ is brought close to $t_\perp$ through the translation of relation *r*, i.e., $h_\perp + r \approx t_\perp$. TransH utilizes two principal mapping functions: the projection function that maps entities from the original space onto the hyperplane, and a translation function that carries out the translational operation within the hyperplane. This dual-function approach allows TransH to model various types of relations with enhanced flexibility.

Similarly, the TransR model evolves from TransE to address more complex relational structures within knowledge graphs. TransR segregates entity and relation embeddings into separate vector spaces, transforming entity vectors from their native entity space to the relation-specific space using a dedicated transformation matrix. The model's schematic diagram is shown in Figure 3D. For a fact triplet (*h, r, t*), TransR first transforms the entities *h* and *r* into the relation space using the corresponding transformation matrix M_r, yielding transformed head entity h_r and tail entity t_r. In this relation space, TransR's goal is to align the transformed head entity h_r closely with the transformed tail entity t_r through the translation of relation *r*, i.e., $h_r + r \approx t_r$. While TransR substantially enhances the handling of complex relationships by utilizing separate spaces, the requisite transformation matrices add computational and relational complexity, resulting in increased parameterization and certain operational constraints.

Analogy is a graph embedding model that emphasizes linear structures to enhance the representation of semantic information in knowledge graphs. This model strives to maintain the linear structure of entity and relation embeddings. For a fact triplet (*h, r, t*), Analogy approximates the tail entity *t* by computing the Hadamard product of the head entity *h* with the relation *r*, denoted as $h \circ r \approx t$, where $\circ$ represents the Hadamard product between vectors. One of the notable advantages of Analogy is its simplicity and computational efficiency. With its streamlined architecture, Analogy can be trained at a lower computational cost, yet it still manages to learn high-quality embedding vectors. Additionally, by preserving the linear structure of embeddings, Analogy effectively captures the complex semantic relationships inherent in knowledge graphs, making it a robust model for understanding and processing graph-based data.

The ComplEx model leverages the properties of complex numbers to enhance the representation of entities and relations in knowledge graphs, particularly focusing on symmetric and anti-symmetric relationships. In this model, each entity and relation is depicted as a vector in the complex space. For a triplet (*h, r, t*) in the knowledge graph, ComplEx represents the entities *h* and *t*, and the relation *r*, as complex vectors $h, r, t \in C^d$. A distinctive scoring function is employed by ComplEx to assess the validity of a triplet, defined as $score(h, r, t) = Re(\sum_{i=1}^{d} h_i \cdot r_i \cdot \bar{t}_i)$, where $Re(\cdot)$ denotes taking the real part of the complex number and $\bar{t}_i$ represents the conjugate of t_i. One of the primary strengths of ComplEx is its adeptness at articulating symmetrical and anti-symmetrical relationships via the utilization of complex spaces. For example, within an anti-symmetrical relation *r*, if (*h*, *r*, *t*) is valid, then (*t*, *r*, *h*) is typically not valid, highlighting the model's capacity to discern and represent such relational nuances effectively. Additionally, the ComplEx model is noted for its simplicity and streamlined parameter set, which contributes to greater efficiency during training, especially advantageous when handling large-scale knowledge graphs.

The DistMult graph embedding model utilizes a simplified matrix factorization approach to learn representations of entities and relations. In this model, relations are modeled as diagonal matrices, which significantly reduces the number of parameters and boosts computational efficiency compared to other matrix factorization-based graph embedding models. DistMult employs a direct scoring function to evaluate the validity of a triplet. For a given triplet (*h, r, t*), the score is computed by $score(h, r, t) = h^\intercal diag(r) t$, where $diag(r)$ transforms the relation vector *r* into a diagonal matrix, and $\intercal$ denotes the vector transpose. The primary advantage of the DistMult model is its simplicity and operational efficiency. By modeling relations as diagonal matrices, DistMult reduces its parameter load, thereby enhancing its efficiency for training on large-scale knowledge graphs. However, this simplicity comes with certain drawbacks. The reliance on diagonal matrices to represent relations limits DistMult's ability to capture complex relational patterns, particularly asymmetric relations. Consequently, DistMult may underperform in tasks that require modeling of intricate relationship dynamics compared to more complex models.

3.1.3 Dimensionality reduction model

Despite the initial transformation of graph data into vector representations via the aforementioned graph embedding models, further refinement is crucial to enhance computational efficiency and focus on the pivotal features of each berth representation vector. This refinement process aims to filter out noise and redundant information, thereby sharpening the subsequent analyses. To achieve this, our framework incorporates established dimensionality reduction techniques, namely principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). These methods are adept at condensing the representation vectors while preserving essential information, thus facilitating more precise analyses. Here is a brief acknowledgment of the dimensionality reduction techniques applied in this study.

PCA reduces the dimensionality of data by identifying the directions that maximize variance within the data. The process begins with standardizing the data features, followed by calculating the covariance matrix to analyze correlations between these features. This analysis helps identify the principal directions of variance, known as principal components. The original data are then projected onto the new space defined by these principal components, effectively reducing its dimensionality.

t-SNE is an ML-based dimensionality reduction algorithm that maps high-dimensional data points into a two-dimensional or three-dimensional space. It simulates the distribution of similarities among high-dimensional data points in the low-dimensional space. t-SNE maintains proximity between similar points and ensures distances between dissimilar ones through an optimization process, thus preserving both the local and global structures of the data in the low-dimensional space.

3.1.4 Classification model

Leveraging the dimensionally reduced ship representation vectors, the framework employs multiple classification models to categorize ships effectively. This array includes K-nearest neighbor (KNN), decision tree, random forest, support vector machine (SVM), and Gaussian naive Bayes, among others. The diversity of ML classification models enriches the analysis by providing multiple analytical perspectives and employing varied classification strategies. This approach ensures the identification of the most effective model for the intended downstream tasks, optimizing the classification process.

3.1.5 Similarity calculation

For the purpose of similarity analysis, particularly in discerning the similarity between berths, the framework adopts the cosine similarity metric. The calculation method of cosine similarity is shown in Equation (1):

$$\textit{similarity} = \cos(\theta) = \frac{A \cdot B}{||A||\ ||B||} \tag{1}$$

where A and B denote the vectors under comparison, with the denominator representing the product of their magnitudes and the numerator being the dot product of the vectors. This metric's independence from vector dimensionality and insensitivity to vector scale render the similarity results straightforward and interpretable.

3.2 Dataset

The dataset used in this study consists of two primary components: the vessel entry and departure dataset and the berth statistics dataset. The vessel entry and departure data were gathered from the AIS records of ships entering and exiting Tianjin Port, China, from January to May 2022, as depicted in Figure 4. Tianjin Port was selected as the focus of this research due to its extensive navigational channels, anchorages, and berth resources, which offer a wealth of data and a variety of scenarios for analysis. This selection enables a thorough examination of vessel behavior and berth utilization across diverse conditions.

To determine whether the AIS data collected encompassed ship berthing activities and to eliminate noise data, this study performed a behavioral analysis on the gathered AIS data. Within this dataset, each docked ship was identified and cataloged by its Maritime Mobile Service Identity (MMSI), which facilitated the construction of a behavioral sequence for the ships based on their MMSI. Notably, berthing events displayed unique behavioral patterns, such as a reduction in the ship's speed to nearly zero during docking and adherence to the International Maritime Organization (IMO) standards regarding the ship's minimum queue length. Utilizing a sliding window algorithm, as illustrated in Figure 5, this study grouped points that exhibited latitude and longitude differences below a specified threshold within a certain timeframe into single trajectory points. By linking these trajectory points, the ship's AIS trajectory was accurately reconstructed to reflect its behavioral patterns. From the extensive AIS dataset, a

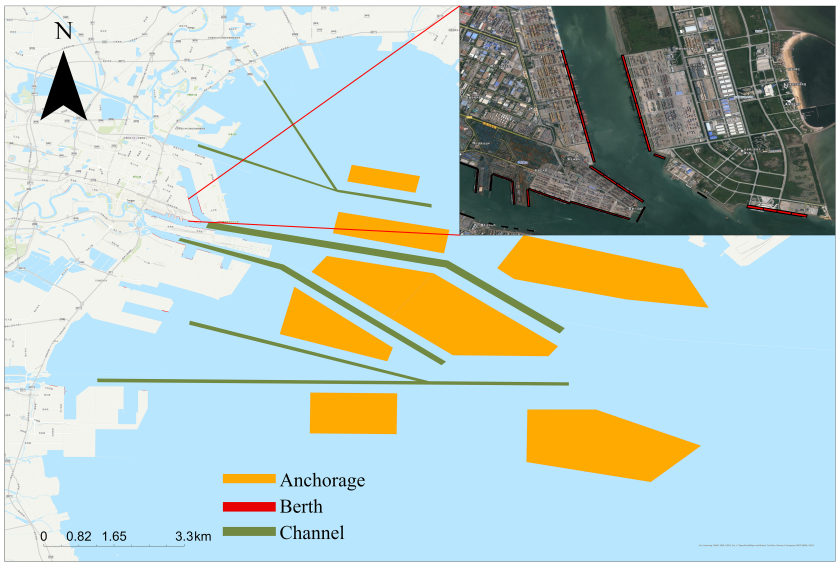


FIGURE 4
The scope of the Tianjin Port vessel entry and departure dataset is shown in the figure, where the orange represents the anchorage area. The green indicates the entry and departure channels, and the red represents the berthing areas.

10.3389/fmars.2024.1390931

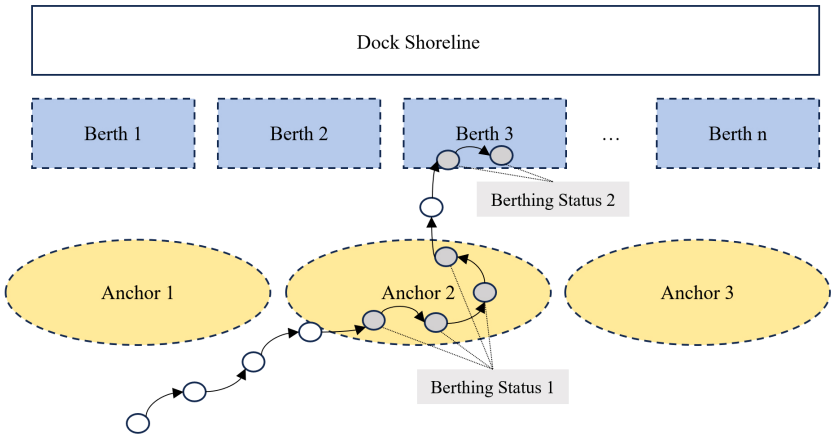


FIGURE 5
Methods for mining berthing statuses through AIS trajectory analysis.

total of 45,648 entries of vessel entry and departure data that conformed to the study's criteria were successfully filtered.

The berth statistics dataset was collected based on the utilization of berths at Shanghai Port and the surrounding port areas throughout 2021, with the detailed scope of berth data collection depicted in Figure 6. The Yangtze River Delta region, recognized as one of China's most critical economic zones, features an extensive array of berths within a dynamic maritime transportation environment. Selecting this region for the study offers a more comprehensive insight into the complexities and distinctive challenges of berth management and vessel transportation within major port clusters.

The statistics compiled include the service duration and the vessels served at each berth. However, during the data collection process, instances were noted where multiple ships were docked at the same berth simultaneously. This could occur either because the berth is sufficiently large to accommodate several smaller ships at once, or due to larger vessels at adjacent berths overlapping into the berth under study, thus influencing the recorded docking data. To accurately assess berth usage while preserving the diversity of data across different ship types, this study implemented a specialized statistical method for berth data. Initially, each berth was treated as an individual research unit, with its operational duration calculated based on the arrival and departure times of all ships that docked there. During this period, the docking of each type of ship was recorded separately to explore the docking characteristics specific to the berth. This approach resulted in a comprehensive dataset comprising 151,682 ship berthing records.

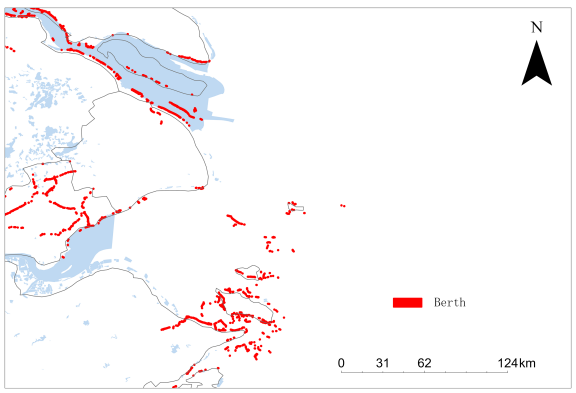


FIGURE 6
The scope of the berth statistics dataset, with red marked points indicating the locations of the berths being analyzed.

In the process of forming triplets, the connections between vessels and infrastructure are precisely articulated utilizing two entities (for instance, a vessel and a berth) and a relation (such as "docks at"). For the Tianjin Port vessel entry and departure statistics dataset, a triadic relationship among the ship's MMSI, navigation channel, anchorage, and berth was established based on the vessel's entry and departure data. A total of 21,374 triadic data entries were compiled, including relationships such as (MMSI, channel, channel number), (MMSI, berth, berth number), and (MMSI, anchorage, anchorage number). For the Shanghai Port berth statistics dataset, a triadic relationship was also delineated between different berths and their territorial belonging, as well as the characteristics of the vessels serviced by these berths. A total of 12,885 triadic data entries were constructed, covering relationship types such as (berth ID, affiliation, port), (berth ID, affiliation, port area), (berth ID, type of docked ship, ship type), (berth ID, affiliated, berth office), (berth ID, purpose of berth, purpose), and (berth ID, tonnage of berth ship, tonnage type). The dataset categorizes ship's tonnage into four types: small-medium (0–5,000 t), medium (5,000–20,000 t), large (20,000–50,000 t), and extra-large (over 50,000 t).

Tables 1 and 2 provide a comprehensive enumeration of different types of vessels included in the port-related dataset and the specific data types utilized in this study, respectively.

The triplets, including those associated with vessels and berths, are identified and listed in Table 3. A knowledge graph was constructed utilizing these identified triplets.

Figure 7 illustrates a schematic of the maritime knowledge graph, constructed using triadic relationships derived from the Tianjin Port entry and exit data, along with the Shanghai Port

TABLE 1 Number of ships in each category in the port-related dataset.

Ship Category	Quantity
Dry bulk carrier	2610
Product oil tanker	600
Container vessel	365
Fishing boat	510
Roll-on-roll-off ship	115
LNG/LPG carrier	105
Total	4305

TABLE 3 Triplet extraction in this paper.

Type of Triplets	Triplet Expression
Vessels	*<Vessel_i, Moor at, Berth_i>*
	<Vessel_i, Anchor at, Anchorage_i>
	<Vessel_i, Sail, Channel_i>
Berths	*<Berth_j, Usage, Berth Usage>*
	<Berth i, Port_Subjection, Port i>
	<Berth i, Harbor_Subjection, Harbor District i>
	<Berth i, Berth Office_Subjection, Berth Office i>
	<Berth i, Berthed, Ship Type>
	<Berth i, Berthed Ship's Tonnage, Tonnage Type >

berth statistics data. This knowledge graph's triadic construction method effectively captures the relational features between ships and port maritime facilities. By extracting analyzable relational semantics, it lays a solid foundation for subsequent tasks, such as maritime shipping feature analysis, enabling a deeper understanding of the dynamics and interactions within port operations.

In this study, Figures 7A–C, depict homogeneous relationship graphs representing ships at anchorages, berths, and in navigation channels, respectively. These figures highlight the relationship characteristics between ships and specific categories of maritime facilities. By analyzing these relational and connective patterns among the maritime entities, the extracted features can be leveraged for various downstream analytical tasks. Figures 7D, E illustrate the relationship graphs of individual ships and berths with associated maritime entities, respectively. Through these focused entity relationship graphs, features of particular entities are more effectively extracted and modeled, enhancing the precision of the analysis. Figures 7F, G are heterogeneous maritime information knowledge graphs, constructed using the datasets from ship entries and exits at Tianjin Port and berth statistics from Shanghai Port, respectively. These graphs vividly display the connections and interactions among a range of maritime entities. In these heterogeneous graphs, not only are relationships between entities described, but concepts such as meta-paths are also employed. These meta-paths facilitate the consideration of structural and semantic connections between entities, thus enabling a more profound understanding of the complex relationships inherent in

TABLE 2 Data types included in this study's dataset.

Data	Data Type	Data Description
AIS Data	MMSI	A unique nine-digit numerical code used to identify vessels, maritime mobile communication satellite service stations, and other radio stations within the maritime mobile communication system.
	Tunnel	Designated routes in waterways for vessel navigation, typically marked to ensure safe and efficient passage.
	Anchorage	Designated areas in the waters near ports, provided for vessels to temporarily anchor while waiting or to conduct other activities.
	Berth	Specific areas near ports or coastlines, designated for vessels to dock and for loading or unloading cargo or passengers.
Maritime Shipping Data	Port	A waterfront facility that provides services for docking, loading and unloading cargo, and embarking or disembarking passengers.
	Port Area	A larger administrative or management area that includes one or more ports along with their surrounding land facilities and related infrastructure.
	Berth Office	An institution or organization responsible for managing and maintaining a specific berth or a group of berths, tasked with arranging vessel docking, maintaining berth facilities, and ensuring the safety and efficiency of berth operations.
Vessel Data	Vessel Type	Different types of vessels are crucial for understanding aspects such as the function of the ship, shipping routes, types of cargo transportation, and the adaptability of port facilities.
	Vessel Usage	Determines the design and operational mode of a vessel, the required port facilities, and its compatibility with specific types of cargo and routes.
	Vessel Tonnage	Affects the cargo carrying capacity of a vessel, suitable shipping routes, required berth depth and port facilities, as well as its decisive impact on shipping costs and efficiency.

10.3389/fmars.2024.1390931

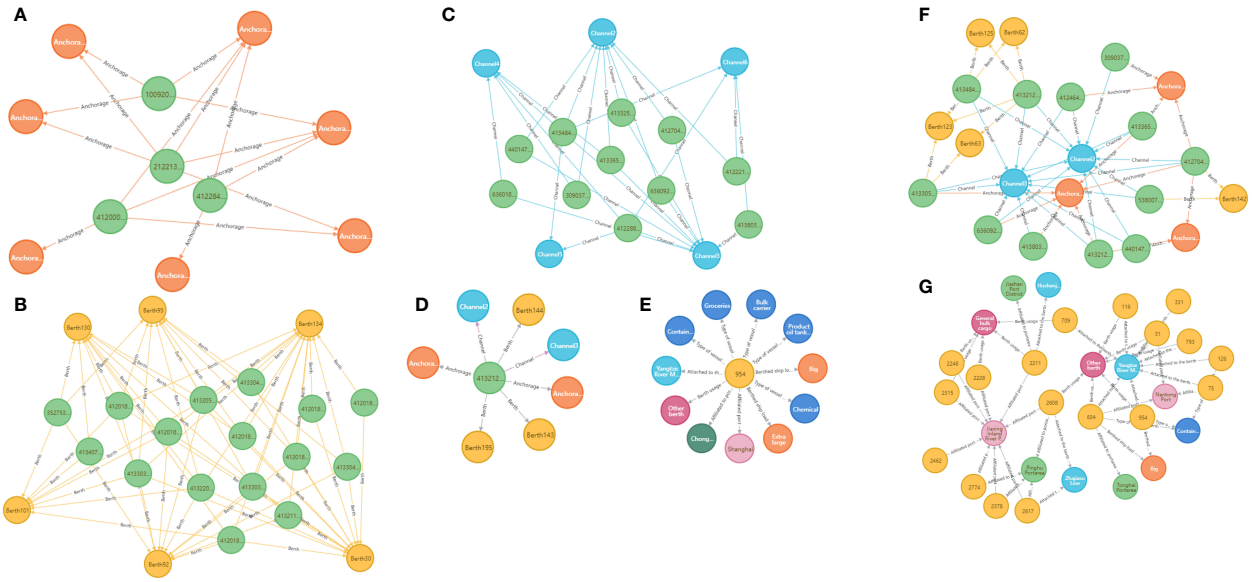


FIGURE 7
(A–G) Examples of graph visualization results.

the data. This comprehensive approach supports enhanced data mining and analytical tasks by providing a deeper insight into the multifaceted features of maritime data.

3.3 Introduction to application scenarios of the framework

Scenario 1 - Ship Type Prediction: Ship type prediction stands as a pivotal technology within the intelligent maritime domain, essential for enhancing ship traffic monitoring, port management, cargo tracking, and coastal patrol and border security measures. This study leverages knowledge graph embedding models to distill features from multiple dimensions, generating comprehensive feature vectors. Through subsequent dimensionality reduction and classification processes, ships under test are accurately classified into their respective predefined types, showcasing the application's potential in operational optimization. As illustrated in Figure 8, from a micro-perspective, ships routinely exhibit behaviors such as anchoring, sailing, and berthing while navigating within harbors. Their trajectories intersect with specific geographic spaces like fairways, anchorages, and berths. Rather than

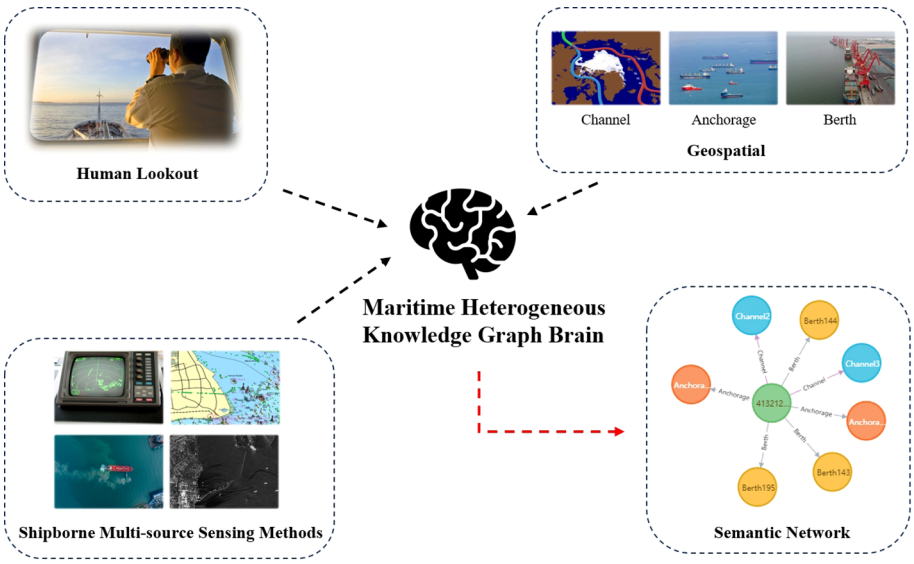


FIGURE 8
The scenario of ship type prediction.

relying solely on numerical data represented by latitude and longitude, ship captains focus more on the semantic information concerning "where the ship is navigating", "where it is anchoring", and "where it is berthing" during their voyage. Upon receiving lookout information (such as passing a lighthouse or traversing a waterway) and multisource perception data, captains also abstract these ship behaviors into semantic networks and engage in contemplation. This process results in the formation of a "small network" that has the characteristic features of ship navigation.

Scenario 2 - Similar Berth Recommendation: Similar berth recommendation represents a practical application within the intelligent maritime domain, utilizing advanced information technology to recommend optimal docking locations for ships. This scenario facilitates port resource optimization, shipping company route planning, and the efficiency of cargo handling processes. By establishing a berth knowledge graph and applying similarity calculations and recommendations, the study identifies and suggests berths with similar characteristics, enhancing operational efficiency and decision-making processes. As shown in Figure 9, berths demonstrate specific berthing preferences for different vessels, influenced by their infrastructure attributes, such as ship size and capacity, as well as commercial attributes like port ownership and port area. These factors necessitate their integration into the autonomous berthing decision-making process. To accommodate this, a multidimensional berth similarity model has been developed, which effectively recommends berths by comprehensively considering these diverse factors.

4 Experimental results and analysis

This section presents a comprehensive evaluation of the proposed model across various datasets and application scenarios, focusing on downstream tasks including similar berth recommendation and ship type prediction. Utilizing a custom-built port-related dataset, the study embarks on a multi-faceted examination: Initially, the semantic construction of ships and port infrastructure through the knowledge graph is outlined, setting the stage for in-depth analysis. Evaluation metrics specific to each dataset are introduced, providing a benchmark for assessing model performance. The efficacy of individual graph embedding models is scrutinized across different datasets, highlighting their capabilities and limitations. The performance of combined models on the datasets is evaluated, with ablation experiments and internal analyses conducted to ascertain the contribution of each framework module.

4.1 Ship type prediction

4.1.1 Evaluation metrics

For ship type prediction, a supervised classification approach was employed, necessitating the use of conventional evaluation metrics to determine the effectiveness of the classification models. These metrics include:

Accuracy: Measures the ratio of correctly classified ship samples within the test set.

Precision: Indicates the ratio of correctly identified ship samples to all samples classified by the model.

Recall: Represents the ratio of correctly identified samples within each ship type category of the test set.

F1 Score: Calculates the harmonic mean between Precision and Recall, providing a balance between the two metrics.

Macro Average: Computes the arithmetic mean of the metrics for each ship type category, treating each category with equal importance. This metric, however, may be affected by categories with fewer samples.

Weighted Average: Determines the weighted mean of each ship type category's metrics, according to the proportion of samples in each category relative to the total sample set. This approach accounts for sample imbalances across categories.

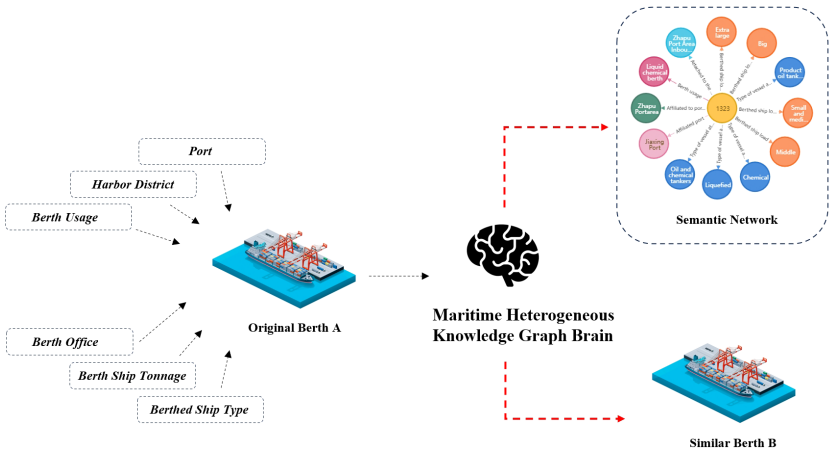


FIGURE 9
Schematic diagram of the similar berth recommendation scenario.

4.1.2 Experimental setup

The experiment utilized a series of relational triplets from the port-related dataset, derived from ships' entry and exit records. Introduce the OpenKE graph embedding model framework (Han et al., 2018) using PyTorch and Sci-kit Learn libraries. Configure embedding vectors for entities within the graph to dimensions of 50, 100, and 200 using graph embedding and supervised ML classification models. Comparative analysis across several classification models, including KNN, decision tree, random forest, SVM, and Gaussian naive Bayes, aimed to categorize ships into six distinct classes: dry bulk carriers, product oil tankers, container vessels, fishing boats, roll-on-roll-off ships, and LNG/LPG carriers. The dataset underwent a random split, allocating 80% for training and 20% for testing, with this division repeated five times to compute the average metric scores.

4.1.3 Experimental results

To evaluate the experimental effectiveness of the framework proposed in this study for ship classification tasks using port-related datasets, and to determine the optimal model combination, various graph embedding models and embedding dimensions were explored in the classification experiments. Figures 10A–C display the Precision, Recall, and F1 Scores for various graph embedding models and dimensions within the dataset, incorporating all dimensionality reduction techniques and classification methods in the analysis. Significantly, the TransR model demonstrated superior average performance across all evaluated metrics with an embedding dimension of 50. It achieved precision, recall, and F1 scores of 0.744, 0.749, and 0.737, respectively. Meanwhile, the ComplEx and DistMult models also showed commendable performance across various metrics and dimensions, highlighting the effectiveness of these embedding models in classifying ship types.

To ascertain the most effective dimensionality reduction technique and optimal embedding dimensions, we focused on the TransR, ComplEx, and DistMult models due to their superior average performance in preliminary experiments. Figures 11–13 illustrate the average Precision (Figures 11A, 12A, and 13A), Recall (Figures 11B, 12B, and 13B), and F1 Scores (Figures 11C, 12C, and 13C) for the ship classification tasks within the port-related dataset across different embedding dimensions. They highlight results across various dimensionality reduction methods and embedding dimensions, utilizing all classification approaches within our framework. The TransR model, with an embedding dimension of 50 and utilizing t-SNE for dimensionality reduction, emerged as the most effective configuration for ship classification. Notably, its performance remained superior when reduced to either 2 or 3 dimensions, outperforming other model combinations. Specifically, when reduced to 2 dimensions, the model achieved an accuracy of 0.811, a recall of 0.812, and an F1 score of 0.808. Similarly, when the classification was reduced to 3 dimensions, it recorded an accuracy of 0.800, a recall of 0.802, and an F1 score of 0.798.

Building upon the identified optimal combination of the embedding model and dimensionality reduction method—specifically, the TransR model at an embedding dimension of 50 with t-SNE—further investigations were conducted to explore the effects of different ML classification methods on ship classification results. Figures 14 and 15 visualize the classification results for each ship in the test set, employing the TransR model reduced to either two or three dimensions via t-SNE, across various classification methods. Figures 14A–E and 15A–E present the classification results when using KNN, decision tree, random forest, SVM, and Gaussian naive Bayes methods, respectively, with reductions to two and three dimensions.

The visualization results highlight that various classification models are effective at distinguishing different categories of ships when all ship vectors are projected onto the same coordinate system. Particularly in regions where features are prominent and distinguishable, most models excel in this differentiation. Notably, the TransR model, at an embedding dimension of 50 and using t-

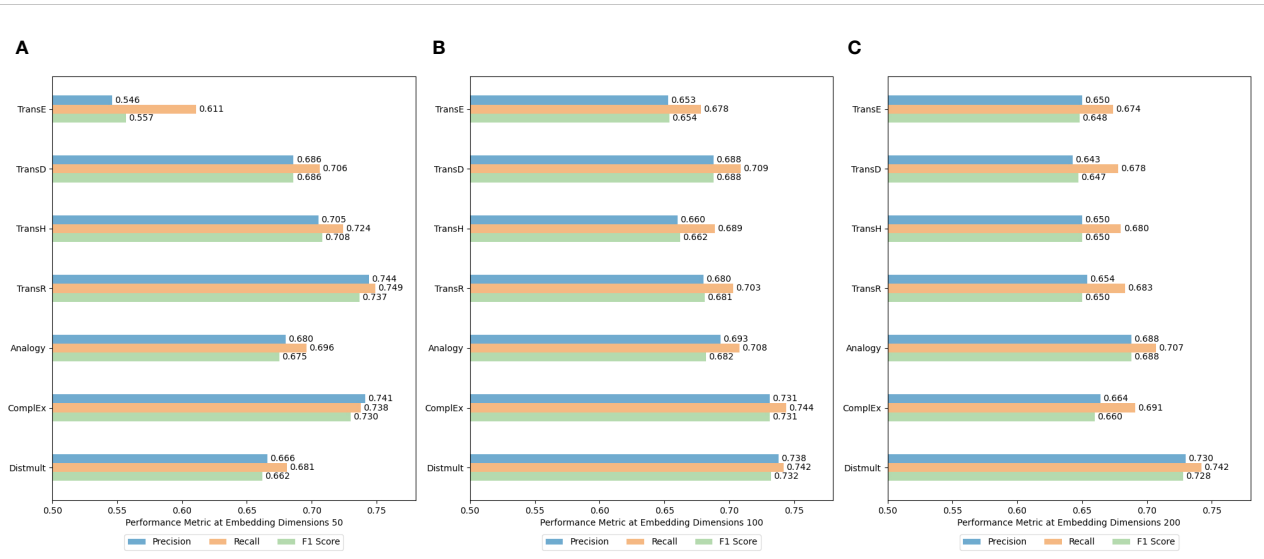


FIGURE 10
(A–C) Average classification results for the port-related dataset classification task using different embedding models.

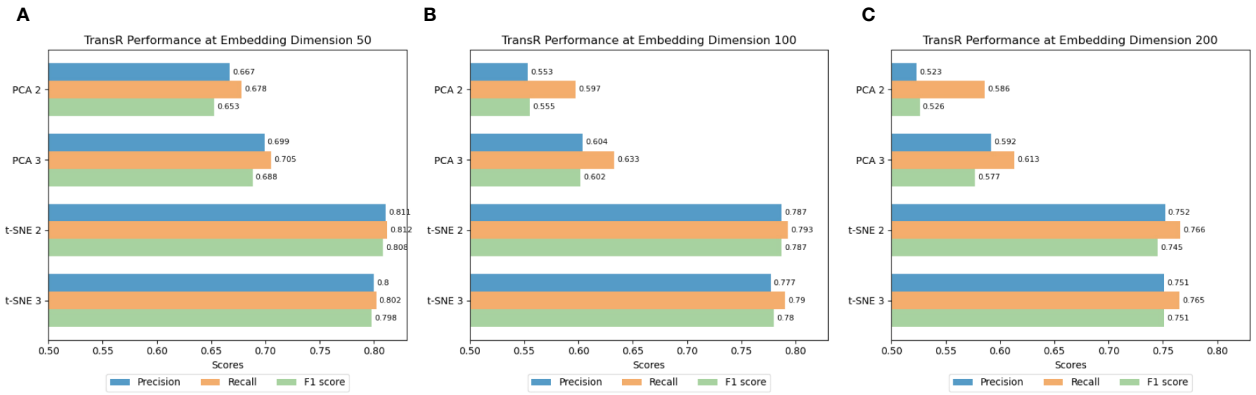


FIGURE 11
(A–C) Classification results of TransR using different dimensionality reduction methods and dimensions.

SNE for reducing dimensionality to 2D, demonstrated relatively optimal classification results when paired with classification methods such as KNN or Random Forest. These combinations achieved accuracies of 0.825 and 0.823, respectively, underscoring their effectiveness in accurately classifying ship types.

To enhance the accuracy of experimental results, the dataset was partitioned randomly into the same ratio five times, with each division undergoing t-SNE dimensionality reduction to 2D and 3D for classification experiments. Figures 16A–C illustrate the average Precision, Recall, and F1 Scores for various ship categories under 2D dimensionality reduction. Conversely, Figures 16D–F display these metrics under 3D dimensionality reduction. Figures 17A, B provide a comparison of Macro Average results for overall test data Precision, Recall, and F1 Scores across different levels of dimensionality reduction, while Figures 17C, D contrast the weighted average results for these metrics. The experimental findings reveal that classification outcomes with 2D dimensionality reduction generally surpass those with 3D reduction. Notably, bulk cargo ships exhibit the best classification results, achieving an accuracy of 0.890, a recall of 0.905, and an F1 score of 0.891. The substantial variation in classification results across different ship categories can primarily be attributed to the differing volumes of data in the dataset. The dataset contains the smallest amount of data for LNG/LPG ships, which significantly impacts the extraction of semantic information from their embedding vectors, leading to lower classification accuracy. In contrast, bulk cargo ships, which are represented with a larger volume of data, show higher classification accuracy.

Figure 18 examines classification accuracy across varying dimensionality reductions. The findings suggest that employing the TransR model with a 50-dimensional embedding and reducing it to 2D using t-SNE, in conjunction with the KNN model, yields a more accurate classification performance for most ship categories in the task of ship classification on a port-related dataset. The superior performance of this experimental combination could be attributed to the dataset's particular compatibility with the TransR model's ability to capture relationships and hierarchies, where the 50-dimensional embeddings retain adequate semantic information without leading to overfitting. The integration of 2D t-SNE reduction with the KNN classification model capitalizes on t-SNE's strength in preserving local data features, with the 2D space enhancing KNN's effectiveness in applying its distance-based classification principle.

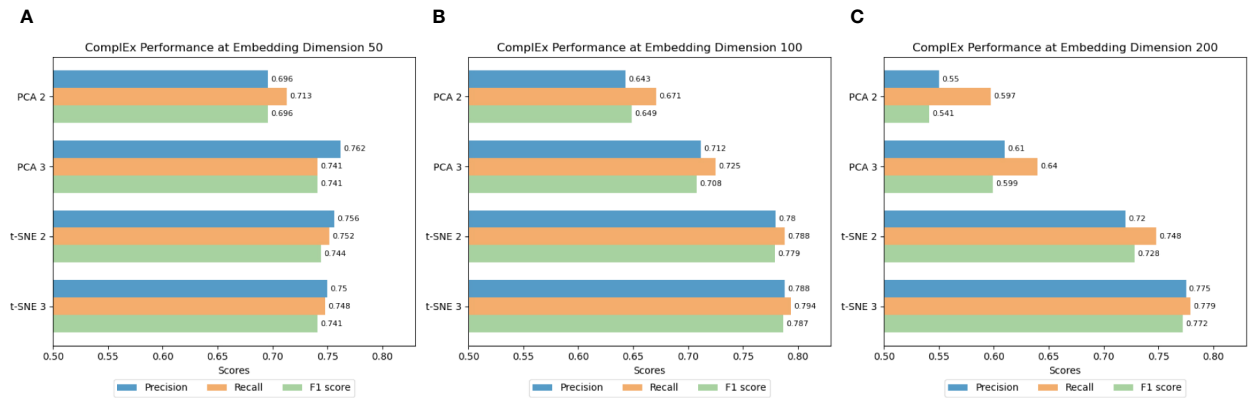


FIGURE 12
(A–C) Classification results of ComplEx using different dimensionality reduction methods and dimensions.

10.3389/fmars.2024.1390931

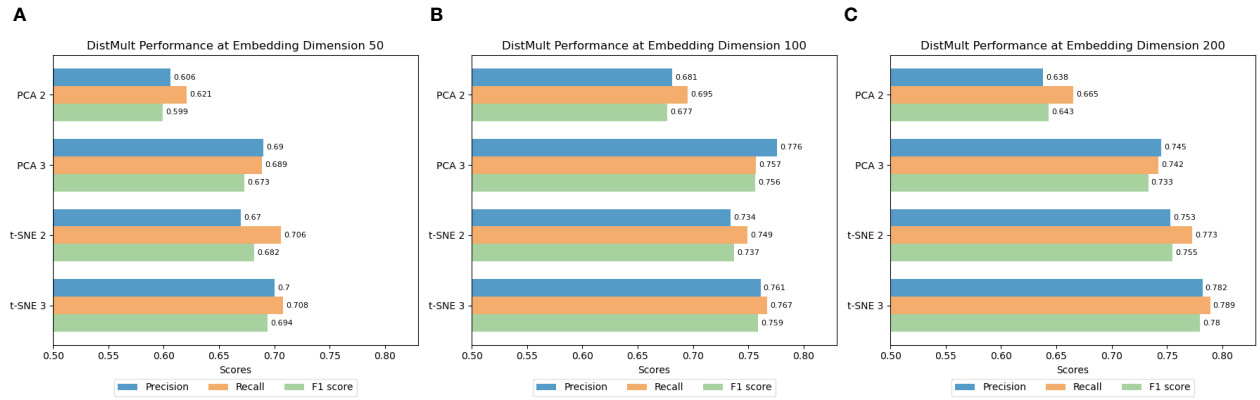


FIGURE 13
(A–C) Classification results of DistMult using different dimensionality reduction methods and dimensions.

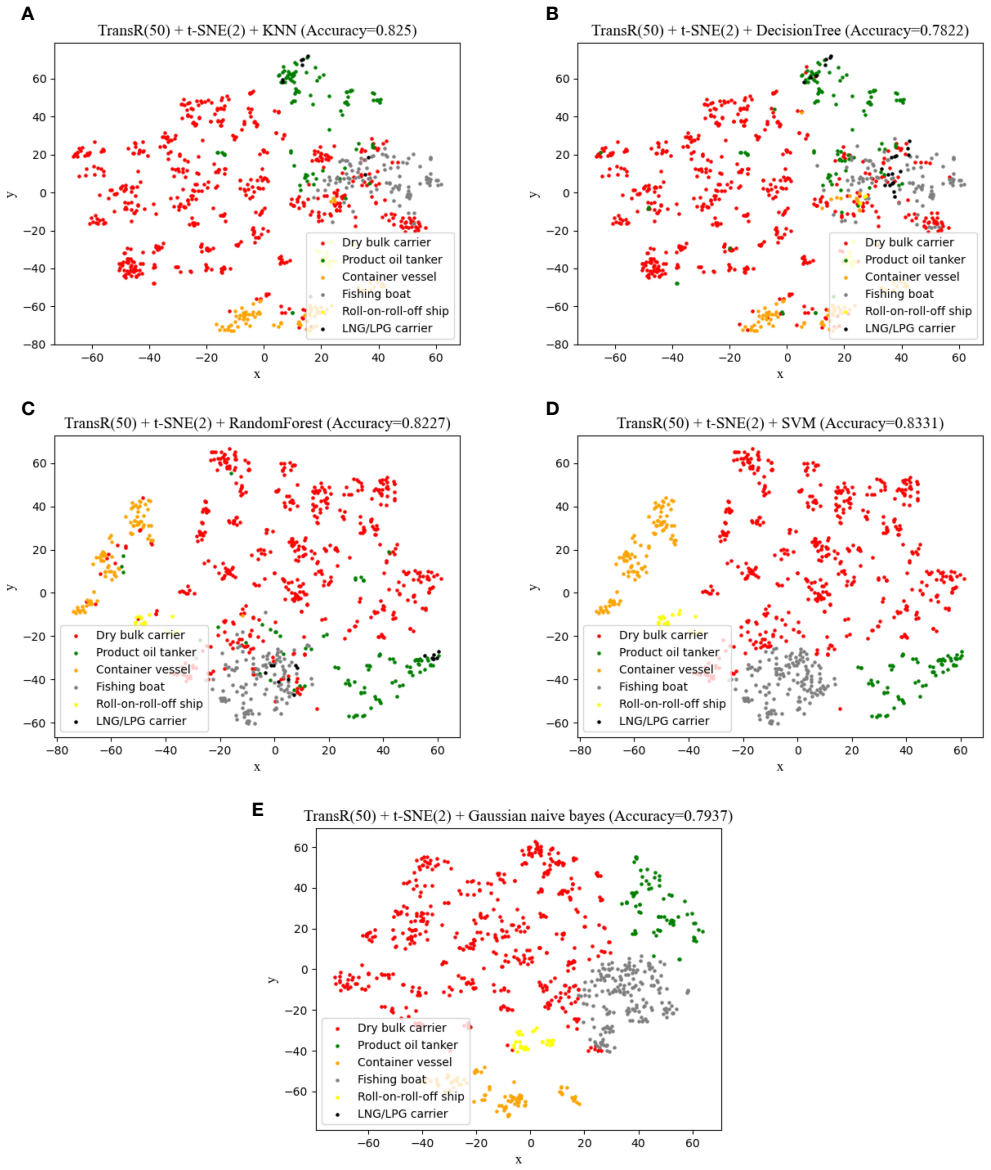


FIGURE 14
(A–E) Classification results using different methods with 50-dimensional TransR model embedded and t-SNE reduced to two dimensions.

10.3389/fmars.2024.1390931

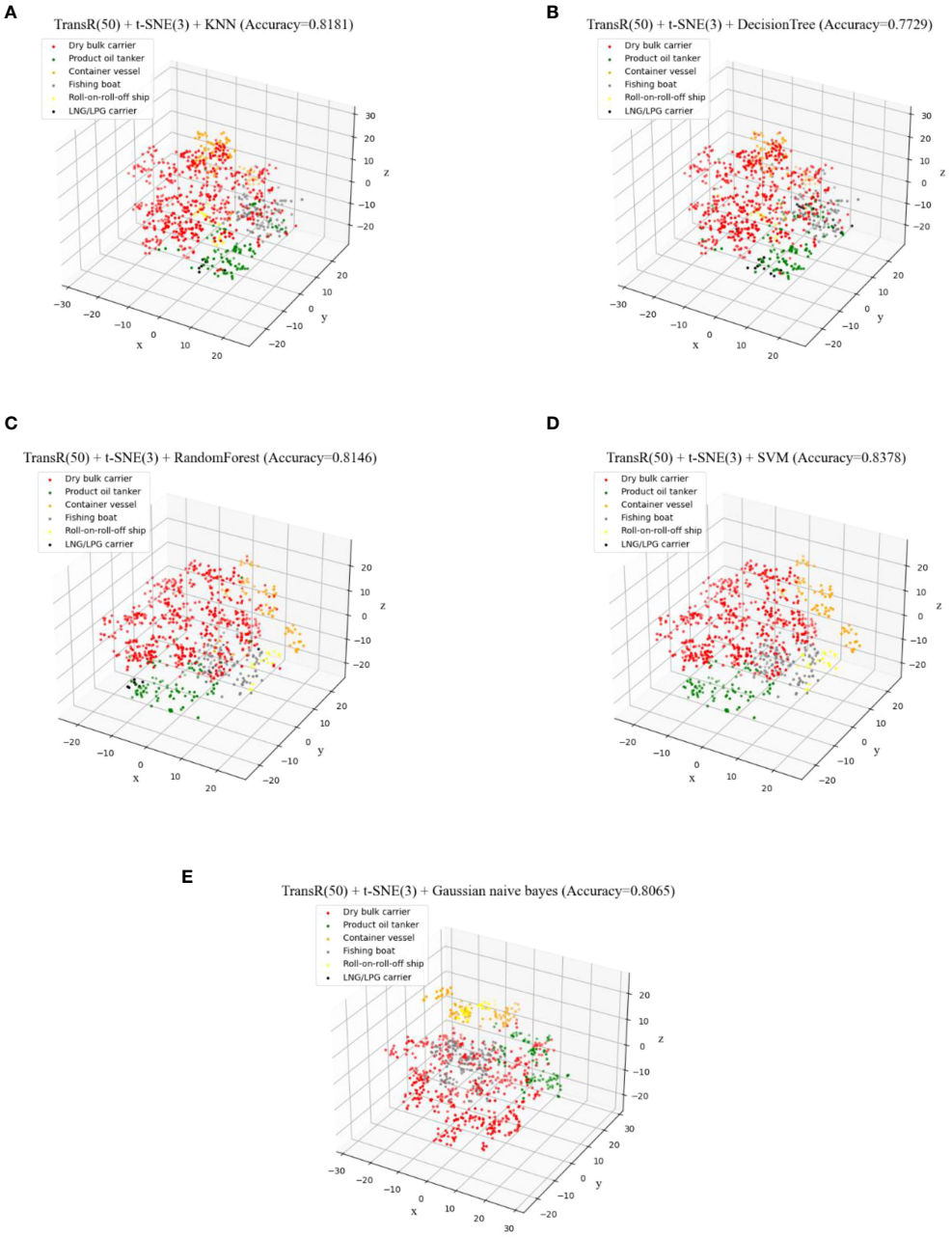


FIGURE 15
(A–E) Classification results using different methods with 50-dimensional TransR model embedded and t-SNE reduced to three dimensions.

4.1.4 Ablation study

An ablation study was conducted to discern the impact of each component within our framework on the classification task, specifically examining the effect of omitting the dimensionality reduction module (DRM). Model variants tested included configurations with and without DRM, focusing on the classification of ships using 50-dimensional embeddings from the TransR model paired with the KNN method.

Figures 19A–C present the Precision, Recall, and F1 Scores for ship classification within the port-related dataset, comparing performances with and without t-SNE reduction to 2D. Compared to the classification results prior to dimensionality reduction, the performance in categories with a larger sample size improved following the application of t-SNE reduction. The Precision, Recall, and F1 Scores across the board also enhanced post t-SNE reduction, underscoring the effectiveness of the dimensionality reduction module in ship classification endeavors. For instance, the classification accuracy for bulk carriers improved from 0.856 to 0.877. The application of dimensionality reduction techniques plays a crucial role in eliminating noise and primarily preserving key semantic features, which aids the classification model in focusing on distinctive attributes, thus enhancing classification accuracy. Additionally, dimensionality reduction often results in a more uniform distribution of vectors in the

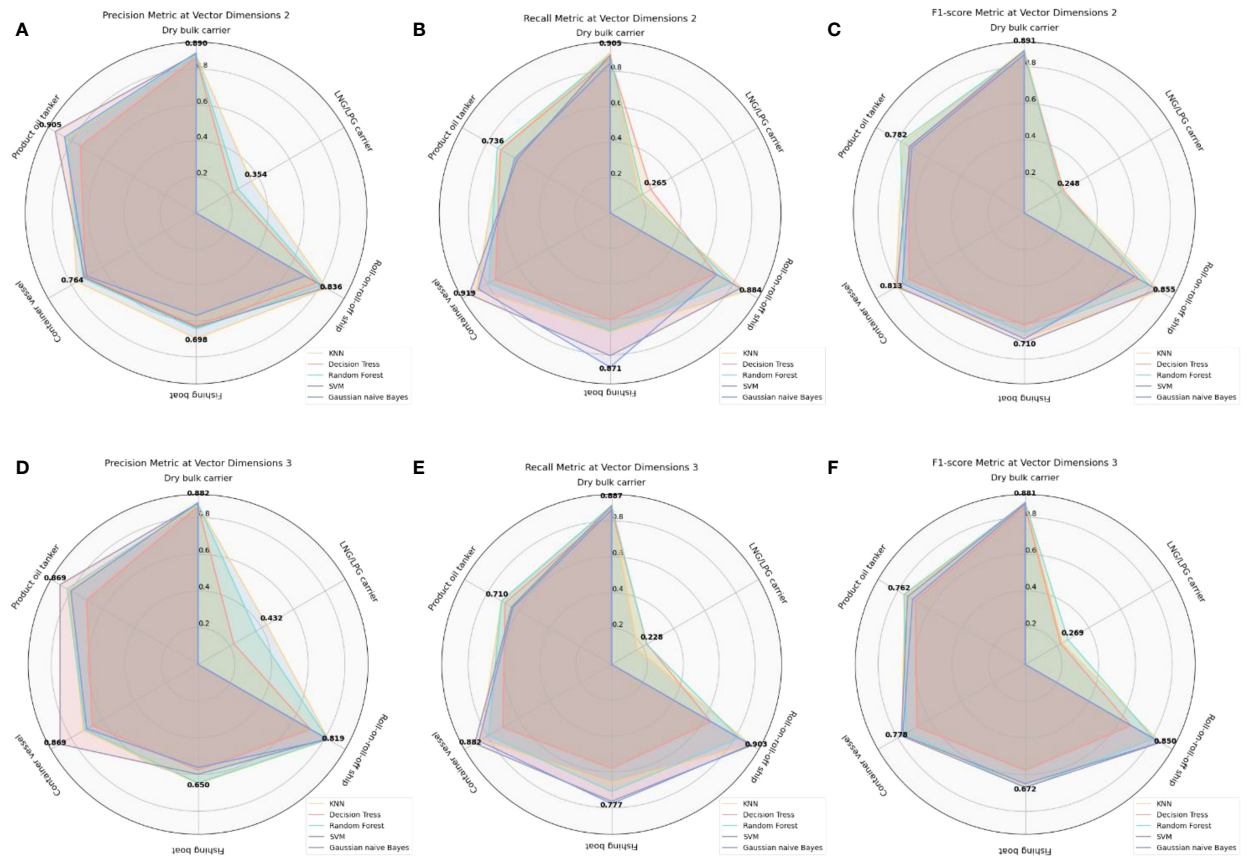


FIGURE 16
(A–F) Classification results for each ship category using different dimensionality reduction dimensions and classification methods with 50-dimensional TransR model combined with t-SNE reduction.

space, with similar vectors becoming more clustered. This clustering is advantageous for distance-based classifiers such as KNN, as it supports more precise classification results.

4.2 Similar berth recommendation

4.2.1 Evaluation metrics

For the development of a similar berth recommendation system, six key dimensions were identified for analysis: affiliated port, belonging harbor district, affiliated berth office, berth usage, type of ships berthed at the berth, and tonnage of ships berthed at the berth. These dimensions serve as the basis for evaluating and determining the similarity between two berths. The evaluation accuracy p_a for a given berth a within this system can be mathematically defined as Equation (2):

$$p_a = r_a t \quad (0 \le r \le 6) \tag{2}$$

where t denotes the total number of similar berths identified for a given berth under comparison, r_a represents the count of identical relationships between each suggested similar berth and the berth being compared a. When extending this calculation across the entire dataset, the model's overall evaluation accuracy, or the total count of identical relationships P_{Full}, can be expressed as Equation (3):

$$P_{Full} = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} r_i t \quad (0 \le r_i \le 6) \tag{3}$$

where n is the total number of berths within the dataset. The model's performance is better indicated by the higher number of identical relationships found between similar berths across the entire dataset and the corresponding berths they are compared with.

Moreover, to further assess the model's recommendation capabilities, we introduced two additional metrics: the average count of identical relationships for top t recommended berths P_{Topt} and the average count of identical relationships per recommended berth P_{single}. The calculation methods are shown in Equation (5) and Equation (6) respectively. These metrics aim to capture the model's general recommendation effectiveness for a single compared berth and the precision of the model's recommendations for each similar berth, respectively:

$$P_{Topt} = \frac{P_{Full}}{n} = \frac{\sum_{i=1}^{n} p_i}{n} = \frac{\sum_{i=1}^{n} r_i t}{n} \quad (0 \le r_i \le 6) \tag{4}$$

$$P_{Single} = \frac{P_{Topt}}{t} = \frac{P_{Full}}{nt} = \frac{\sum_{i=1}^{n} p_i}{nt} = \frac{\sum_{i=1}^{n} r_i}{n} \quad (0 \le r_i \le 6) \tag{5}$$

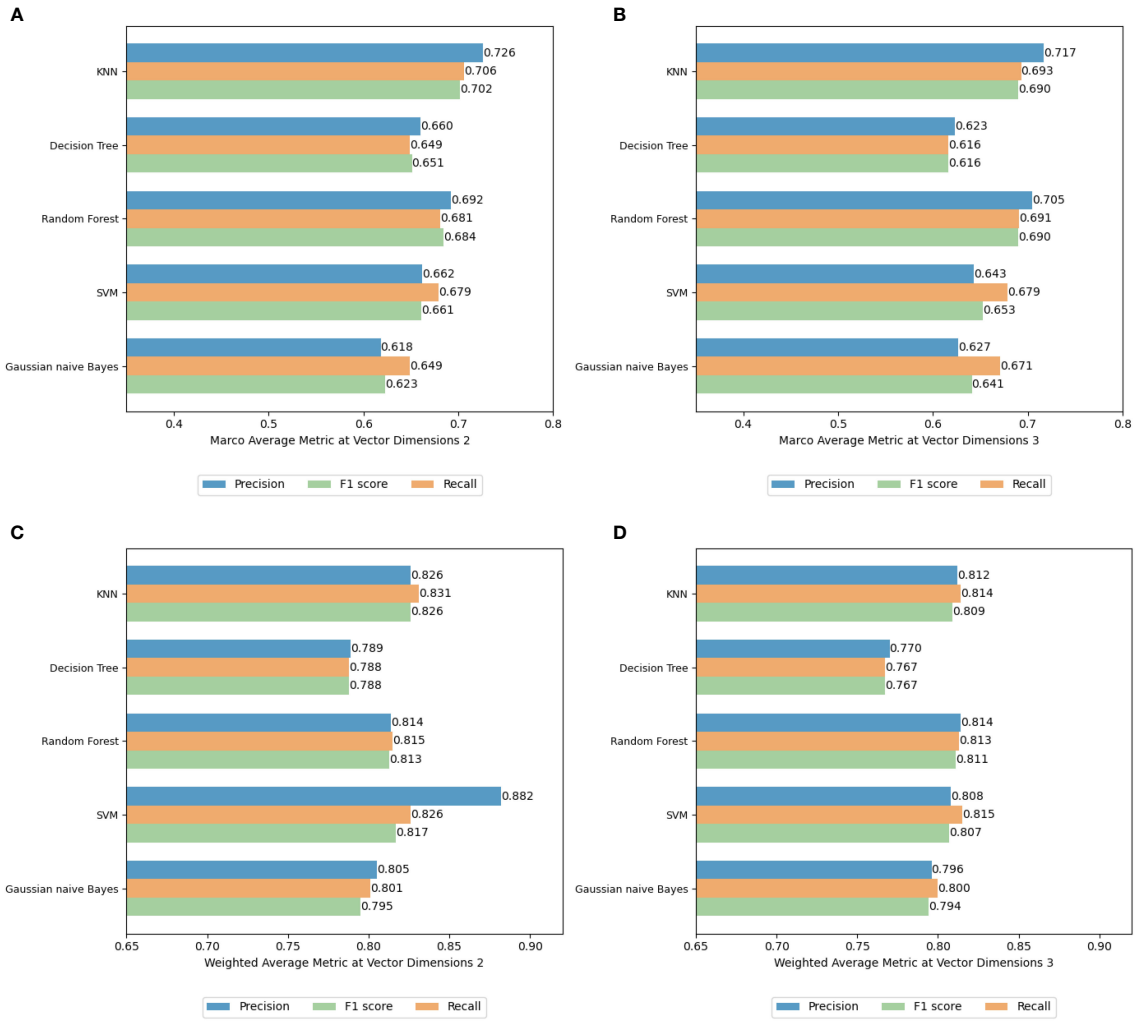


FIGURE 17
(A–D) Macro average and weighted average results using different dimensionality reduction dimensions and classification methods with 50-dimensional TransR model combined with t-SNE reduction.

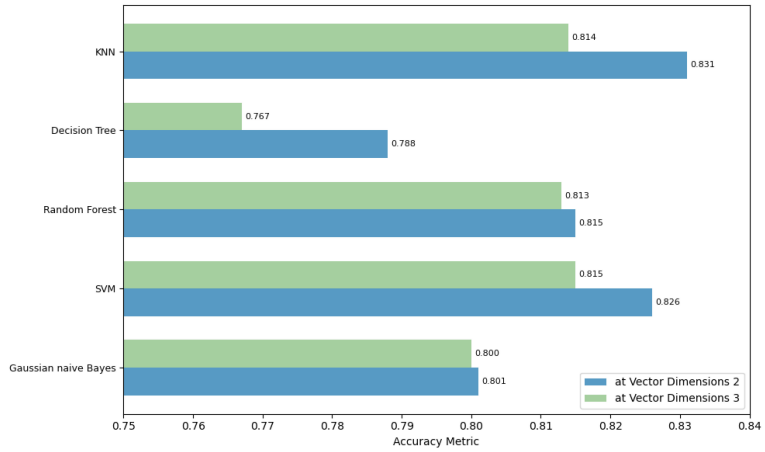


FIGURE 18
Accuracy using different dimensionality reduction dimensions and classification methods with 50-dimensional TransR model combined with t-SNE reduction.

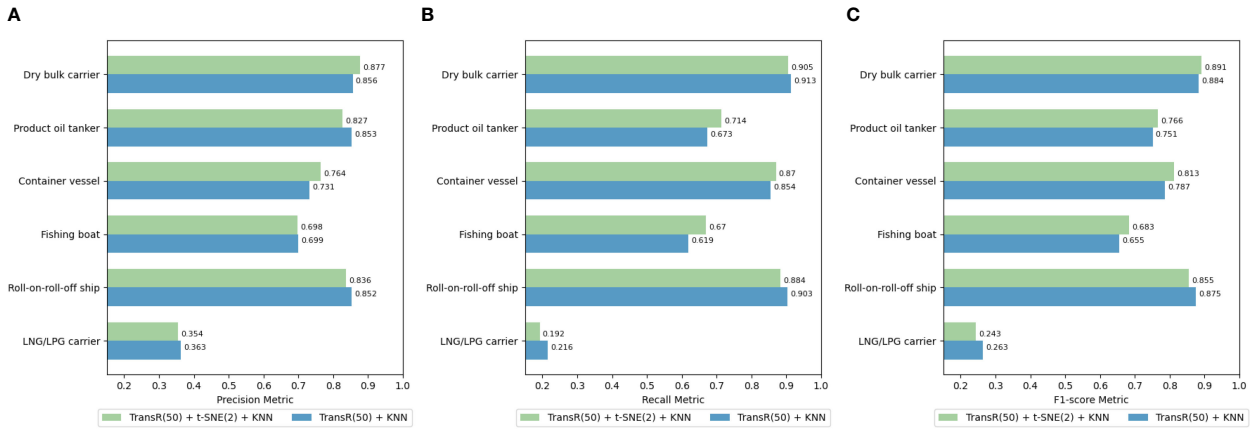


FIGURE 19
(A–C) Classification results before and after using t-SNE reduction with 50-dimensional TransR model combined with KNN classification method.

4.2.2 Experimental setup

This experiment employed the port-related dataset, where relational triplets were pre-constructed for each berth comparison. The top 20 berths with the highest similarity were recommended for analysis. Utilizing the PyTorch framework, graph embedding and berth similarity analysis models were implemented, with embedding vector dimensions set at 50, 100, and 200.

4.2.3 Experimental results

The proposed framework's efficacy was evaluated on the port-related dataset, particularly examining the influence of dimensionality reduction on similar berth recommendations. Figures 20A–C display the total count of identical relationships, the average count of identical relationships for the top 20 recommended berths, and the average count of identical relationships per recommended berth using different dimensions of the TransE model after various dimensionality reduction methods. The findings suggest that omitting dimensionality reduction results in the most accurate berth recommendations. The average count of identical relationships per recommended berth for embedding vectors with dimensions of 50, 100, and 200 reached 2.420, 3.157, and 3.324, respectively, all surpassing performances in scenarios where dimensionality reduction was applied. This indicates that the complex relationship dimensions within the similar berth knowledge graph, which are more numerous than those in ship classification, contribute to a richer set of embedding vector features. Reducing these dimensions could potentially remove critical features, thereby decreasing the accuracy of the recommendations.

Figures 21A–C present the total count of identical relationships, the average count of identical relationships for the top 20 recommended berths, and the average count of identical relationships per recommended berth, respectively, for tasks involving similar berth recommendations within the port-related dataset, utilizing various embedding techniques in our framework. The findings indicate that the Analogy method outperforms others in terms of recommendation precision with an embedding dimension of 50. Given that the port-related dataset encompasses

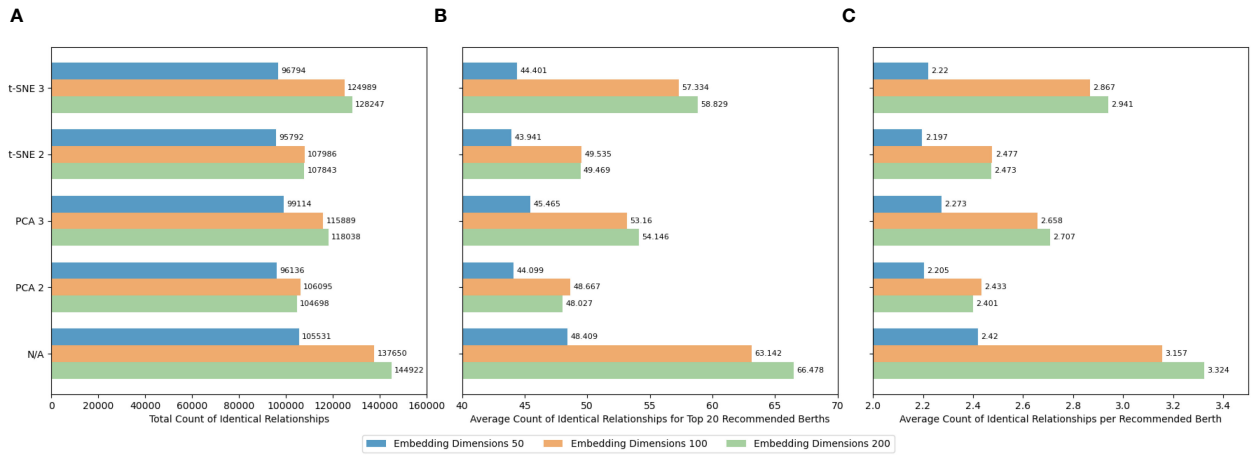


FIGURE 20
(A–C) Recommendation accuracy of different dimensionality reduction methods in the port-related dataset.

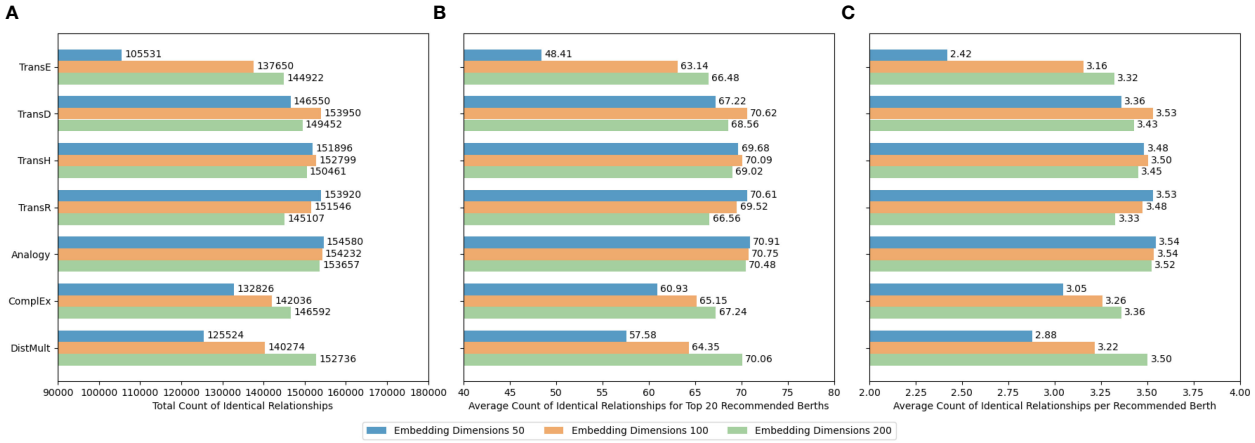


FIGURE 21
(A–C) Recommendation accuracy of different embedding models in the port-related dataset.

six types of triplets, this result implies that the berth knowledge graph encompasses a wide array of relationship types. The Analogy method's superior performance, particularly at lower dimensions, suggests its proficiency in capturing this diversity and effectively representing these relationships without leading to overfitting. The Analogy model's distinct advantage lies in its ability to handle analogical relationships, which, within the context of the port-related dataset, could be interpreted as the resemblance in characteristics or functions among different berths.

Based on the foundation of single base embedding models, this study further explored a combined model strategy for enhancing the accuracy of similar berth recommendations. Figures 22A–C present the comparative results of the total count of identical relationships, the average count of identical relationships for the top 20 recommended berths, and the average count of identical relationships per recommended berth in the port-related dataset for the task of recommending similar berths when the similarity score weights of the two sub-models in the combined model are set at a 1:1 ratio. Based on the integrated model approach, the combination of any two models exhibits superior recommendation accuracy compared to the performance of either model independently. This indicates that the combined model possesses enhanced adaptability and efficacy in the task of recommending similar berths.

Particularly, the synergistic combination of TransD, Analogy, and DistMult models outperforms other model pairings in the task of similar berth recommendation. The TransD and Analogy models, when combined at an embedding dimension of 100, showcase the highest efficacy, achieving an average count of identical relationships per recommended berth of 3.748. This enhancement can be attributed to TransD's capability to encapsulate complex semantic relationship features through distinct mapping matrices for each entity-relation, coupled with the Analogy model's proficiency in handling symmetry and anti-symmetry in entity relationships. When TransD, known for its dynamic mapping capabilities, is integrated with Analogy, which excels in capturing symmetry, the combined model potentially enhances its ability to understand and represent berth similarities

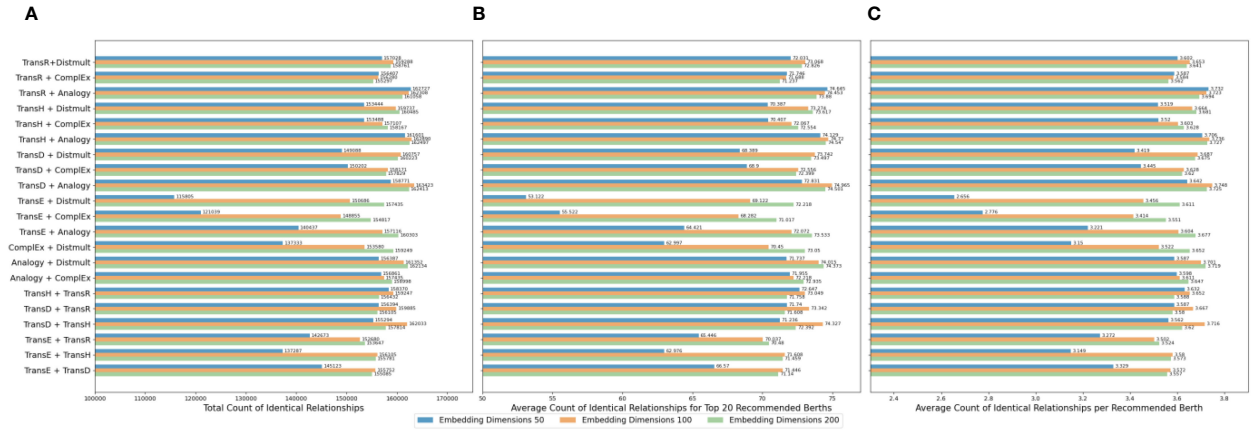


FIGURE 22
(A–C) Recommendation accuracy of various combined models in the port-related dataset.

more comprehensively. The synergy between TransD's flexibility in handling complex relationships and Analogy's efficiency in recognizing symmetric patterns could offer a more nuanced representation of berth similarities.

To delve deeper into how weight distribution affects the similarity results, this study introduces a weight parameter n to the combined model consisting of TransD, Analogy, and DistMult. This exploration aims to discern the relative influence of each model within the trio on the combined model's results. The Equation (6) to assess the impact of merging models A and B with a specific weight value n is conceptualized to quantify their combined effect on similarity assessments:

$$\boldsymbol{Similarity_{union} = nSimilarity_A + (1-n)Similarity_B} \quad (6)$$

Merging the similarity calculation values from two models allows the combined model to adopt the recommendation tendencies and strengths of both constituent models. Figures 23–25 illustrate the variations in the average count of identical relationships per recommended berth for the TransD, Analogy, and DistMult models across different embedding dimensions, utilizing the port-related dataset for the recommendation of similar berths. Specifically, Figures 23A–C detail the trends for the TransD model at embedding dimensions of 50, 100, and 200, respectively. Figures 24A–C present the trends for the Analogy model at these same dimensions, while Figures 25A–C depict the trends for the DistMult model.

The data reveal that the introduction of the combined model markedly improves the performance in recommending similar berths, with the optimal effects varying at different dimensions according to the weight value n. The synergy between the TransD and Analogy models is most effective at $n = 0.4$, where it achieves an average count of identical relationships per recommended berth of 3.755. The Analogy model exhibits a high level of proficiency in identifying similarities between berths for the recommendation task within the dataset used in this study. Although the weight assigned to TransD is marginally lower, its contribution to the overall similarity calculation remains substantial. The structured relationship recognition facilitated by TransD's dynamic mapping mechanism may complement the Analogy model's capabilities, addressing aspects of similarity that the Analogy model alone might overlook.

Using the optimal combination obtained from this experiment, the actual effect of recommending similar berths was tested, taking three berths in the Yangshan Port area of Shanghai Port as an example. The selected comparable berths and the recommended berths with high similarity are shown in Figure 26, where the red indicates the selected comparable berths, and the blue indicates the recommended candidate berths with high similarity.

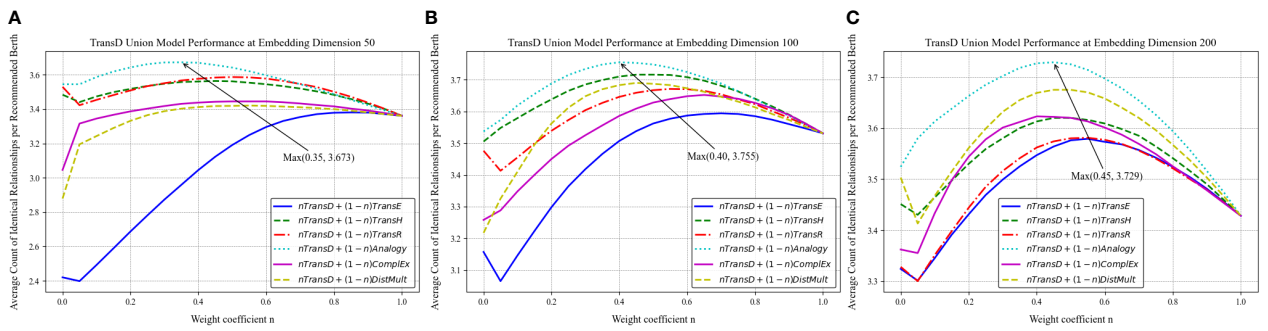


FIGURE 23
(A–C) Trend of similar berth recommendation results by the TransD model at different dimensions with changes in weight value n.

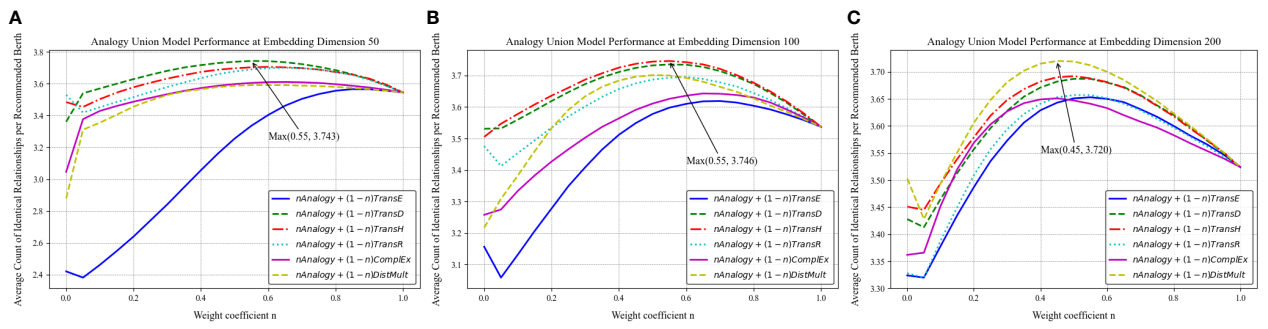


FIGURE 24
(A–C) Trend of similar berth recommendation results by the Analogy model at different dimensions with changes in weight value n.

10.3389/fmars.2024.1390931

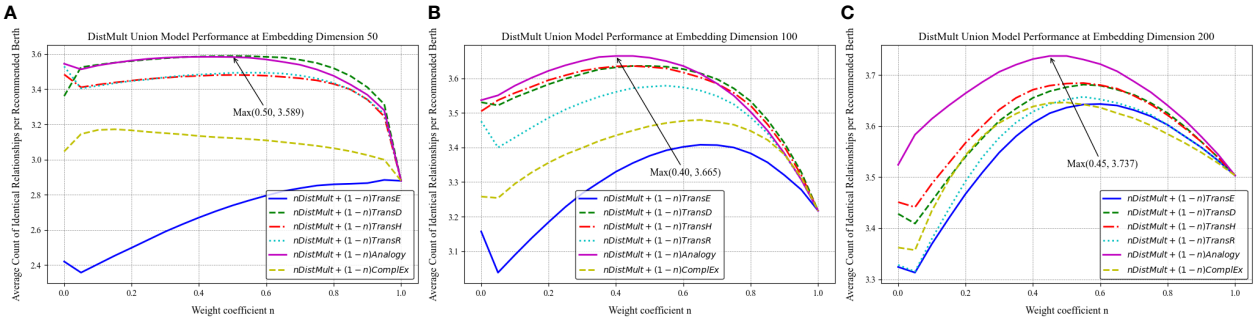


FIGURE 25
(A–C) Trend of similar berth recommendation results by the DistMult model at different dimensions with changes in weight value *n*.

5 Discussion

The framework introduced in this study utilizes knowledge graphs and graph embedding technologies to adeptly navigate the complexities of multi-source data fusion and its applications within the maritime sector. Through empirical validation in two distinct application scenarios—ship type prediction and similar berth recommendation—this research underscores the framework's capacity to amalgamate AIS data with static maritime information, thereby offering sophisticated intelligent recommendation and prediction capabilities in the maritime domain. Nonetheless, the research framework encounters limitations, notably the static nature of the knowledge graphs and graph embedding models used. This limitation constrains the models' ability to comprehend and incorporate temporal dynamics, which are critical for capturing more nuanced semantic relationships. Integrating temporal dimensions into the framework could significantly enrich its semantic capacity, thereby enhancing the precision of its recommendations and predictions. By constructing maritime semantic networks through knowledge graphs, we can effectively profile various entities in the maritime domain, such as vessels and shore-based facilities. This approach enables a more intuitive representation of the relationships and structures among these entities and allows for detailed descriptions and classifications, presenting the complex systems and multidimensional characteristics of the maritime domain from various perspectives and levels. For example, at the application level, taking vessel traffic services (VTS) as a case study, integrating this framework can equip vessels with more intelligent and efficient navigation, traffic monitoring, and decision-making services.

Looking forward to future research based on this study, a key challenge is how to achieve deeper integration with other multi-source data, such as remote sensing. Furthermore, with the continuous emergence and development of large language models like GPT (Brown et al., 2020) and Gemini (Team G, 2023), there is an opportunity to gain a deeper understanding of the contextual and semantic information within maritime knowledge graphs. Leveraging the maritime and shipping expertise contained in these graphs can significantly enhance the logical reasoning and problem-solving capabilities of large language models in the

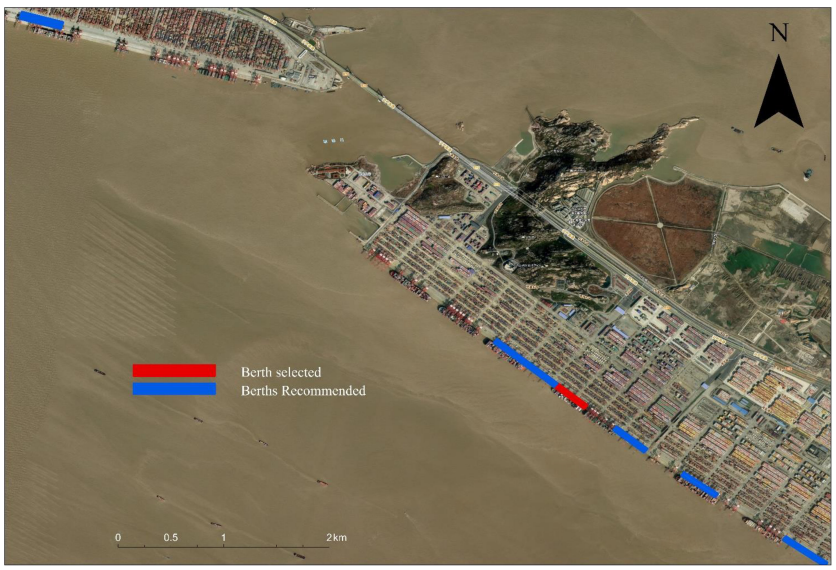


FIGURE 26
Example of similar berth recommendation effectiveness using the optimal model combination of this framework.

maritime domain. This advancement can lead to the development of more specialized maritime large language models, offering highly specialized intelligent analysis and decision-support. Such tools are applicable to improving navigation management, vessel monitoring, environmental monitoring, and enhancing maritime safety.

6 Conclusions and future work

To conclude and propose directions for future research, this study has established a semantically comprehensible framework utilizing knowledge graph technology to elucidate ship behaviors and berth profiles, presenting a pioneering intelligent maritime application framework. This framework capitalizes on knowledge graphs and graph embedding techniques to address key challenges in the maritime industry, such as ship type prediction and berth recommendation. By constructing an intricate knowledge graph that encompasses extensive information on ships and infrastructure, and by exploring various entity relationships through triplets, this research lays a robust foundation for translating complex relational data into actionable vector representations. Through exhaustive comparative analyses, the efficacy of different graph embedding models within maritime contexts was evaluated, investigating how variations in embedding dimensions and model weight distribution influence overall performance. The results of this investigation not only validate the utility of knowledge graphs and graph embedding technologies in maritime applications but also contribute significant empirical evidence and insights for future research endeavors. The developed intelligent transportation systems can aid unmanned devices in making precise decisions based on this model, thereby ensuring the safety of jurisdictional waters and dock efficiency.

While this study has made significant contributions, there are several areas that warrant further exploration. Future research could focus on advancing model fusion techniques to better integrate diverse data sources and improve predictive accuracy. Optimizing real-time data integration and dynamic updates is another critical area that could enhance the responsiveness and accuracy of the system in live environments. Additionally, developing and applying cross-domain knowledge graphs could provide deeper insights by linking data across different sectors, enhancing the comprehensiveness of analyses. Enhancing the scalability and generalization capabilities of the framework is also crucial, as it would allow the system to handle larger datasets and apply learned insights across various maritime contexts. Furthermore, exploring the integration of this framework with large model domains, such as advanced machine learning and artificial intelligence platforms, could open up new avenues for more sophisticated analytical tools and decision-support systems.

Data availability statement

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

Author contributions

YL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. XL: Writing – review & editing. ZW: Software, Writing – review & editing. QM: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. WX: Software, Writing – review & editing. YY: Writing – review & editing. PW: Investigation, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China: [Grant Number 52372316]; the Natural Science Foundation of Fujian Province: [Grant Number 2021J01821; 2023J01804].

Conflict of interest

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

Publisher's note

References

Ahmed, U., Srivastava, G., Djenouri, Y., and Lin, J. C. W. (2022). Knowledge graph-based trajectory outlier detection in sustainable smart cities. *Sustain. Cities. Soc.* 78, 103580. doi: 10.1016/j.scs.2021.103580

Bertram, N., Dunkel, J., and Hermoso, R. (2023). I am all EARS: Using open data and knowledge graph embeddings for music recommendations. *Expert Syst. Appl.* 229, 120347. doi: 10.1016/j.eswa.2023.120347

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* 26. doi: 10.5555/2999792.2999923

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901. doi: 10.18653/v1/2021.mrl-1.1

Chen, X., Liu, Y., Achuthan, K., and Zhang, X. (2020). A ship movement classification based on Automatic Identification System (AIS) data using Convolutional Neural Network. *Ocean. Eng.* 218, 108182. doi: 10.1016/j.oceaneng.2020.108182

Chen, L., Liu, D., Yang, J., Jiang, M., Liu, S., and Wang, Y. (2022). Construction and application of COVID-19 infectors activity information knowledge graph. *Comput. Biol. Med.* 148, 105908. doi: 10.1016/j.compbiomed.2022.105908

Chen, X., Wang, M., Ling, J., Wu, H., Wu, B., Li, C., et al. (2024). Ship imaging trajectory extraction via an aggregated you only look once (YOLO) model. *Eng. Appl. Artif. Intell.* 130, 107742. doi: 10.1016/j.engappai.2023.107742

Deng, C., Wang, S., Liu, J., Li, H., Chu, B., Zhu, J., et al. (2023). Graph Signal Variation Detection: A novel approach for identifying and reconstructing ship AIS tangled trajectories. *Ocean. Eng.* 286, 115452. doi: 10.1016/j.oceaneng.2023.115452

Escorcia-Gutierrez, J., Gamarra, M., Beleño, K., Soto, C., and Mansour, R. F. (2022). Intelligent deep learning-enabled autonomous small ship detection and classification model. *Comput. Electrical. Eng.* 100, 107871. doi: 10.1016/j.compeleceng.2022.107871

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., et al. (2020). "Introduction: what is a knowledge graph?," In: *Knowledge Graphs* (Cham: Springer), 1–10. doi: 10.1007/978-3-030-37439-6_1

Gan, L., Ye, B., Huang, Z., Xu, Y., Chen, Q., and Shu, Y. (2023). Knowledge graph construction based on ship collision accident reports to improve maritime traffic safety. *Ocean. Coast. Manage.* 240, 106660. doi: 10.1016/j.ocecoaman.2023.106660

Gao, M., and Shi, G. Y. (2020). Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms. *Ocean. Eng.* 205, 106919. doi: 10.1016/j.oceaneng.2020.106919

Guo, S., Zhang, H., and Guo, Y. (2023). Toward multimodal vessel trajectory prediction by modeling the distribution of modes. *Ocean. Eng.* 282, 115020. doi: 10.1016/j.oceaneng.2023.115020

Han, X., Cao, S., Lv, X., Lin, T., Liu, Z., Sun, M., et al. (2018). "Openke: An open toolkit for knowledge embedding," in *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations* (Brussels, Belgium: Association for Computational Linguistics), 139–144. doi: 10.18653/v1/D18-2

Ji, G., He, S., Xu, L., Liu, K., and Zhao, J. (2015). "Knowledge graph embedding via dynamic mapping matrix," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Beijing, China: Long Papers), Vol. Volume 1, 687–696.

Jia, C., Ma, J., Yang, X., and Lv, X. (2023). RAGAN: A Generative Adversarial Network for risk-aware trajectory prediction in multi-ship encounter situations. *Ocean. Eng.* 289, 116188. doi: 10.1016/j.oceaneng.2023.116188

Jiang, D., Wang, R., Xue, L., and Yang, J. (2024). Multisource hierarchical neural network for knowledge graph embedding. *Expert Syst. Appl.* 237, 121446. doi: 10.1016/j.eswa.2023.121446

Li, W., Liu, X., Tao, W., Zhang, L., Zou, J., Pan, Y., et al. (2024). Location and time embedded feature representation for spatiotemporal traffic prediction. *Expert Syst. Appl.* 239, 122449. doi: 10.1016/j.eswa.2023.122449

Li, G., Liu, M., Zhang, X., Wang, C., Lai, K.-H., Qian, W., et al. (2022). Semantic recognition of ship motion patterns entering and leaving port based on topic model. *J. Mar. Sci. Eng.* 10, 2012. doi: 10.3390/jmse10122012

Li, J., Xiang, J., and Cheng, J. (2023a). EARR: using rules to enhance the embedding of knowledge graph. *Expert Syst. Appl.* 10, 120831. doi: 10.1016/j.eswa.2023.120831

Li, Y., Xie, W., Yang, Y., Mei, Q., Wang, Z., Li, Z., et al. (2023b). Research on the carbon emissions traceability inventory and multi-horizon prediction of ship carbon emissions: a case study of Tianjin Port. *Front. Mar. Sci.* 10, 1174411. doi: 10.3389/fmars.2023.1174411

Liang, M., Zhan, Y., and Liu, R. W. (2021). MVFFNet: multi-view feature fusion network for imbalanced ship classification. *Pattern Recognit. Lett.* 151, 26–32. doi: 10.1016/j.patrec.2021.07.024

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. *Proc. AAAI. Conf. Artif. Intell.* 29, p2181–2187. doi: 10.1609/aaai.v29i1.9491

Liu, H., Jia, Z., Li, B., et al. The model of vessel trajectory abnormal behavior detection based on graph attention prediction and reconstruction network[J].. *Ocean Engineering*, 2023, 290, 116316.

Liu, H., Jia, Z., Li, B., Liu, Y., and Qi, Z. (2023b). The model of vessel trajectory abnormal behavior detection based on graph attention prediction and reconstruction network. *Ocean. Eng.* 290, 116316. doi: 10.1016/j.oceaneng.2023.116316

Liu, H., Wu, Y., and Yang, Y. (2017). "Analogical inference for multi-relational embeddings," in *International Conference on Machine Learning*. Sydney NSW Australia, 2168–2178 (PMLR).

Liu, C., Zhang, X., Xu, Y., Xiang, B., Gan, L., and Shu, Y. (2023a). Knowledge graph for maritime pollution regulations based on deep learning methods. *Ocean. Coast. Manage.* 242, 106679. doi: 10.1016/j.ocecoaman.2023.106679

Liu, X., Zhang, Y., Zou, H., Wang, F., Cheng, X., and Wu, W. (2023c). Multi-source knowledge graph reasoning for ocean oil spill detection from satellite SAR images. *Int. J. Appl. Earth Observation. Geoinformation.* 116, 103153. doi: 10.1016/j.jag.2022.103153

Ma, J., Jia, C., Shu, Y., Liu, K., Zhang, Y., and Hu, Y. (2021). Intent prediction of vessels in intersection waterway based on learning vessel motion patterns with early observations. *Ocean. Eng.* 232, 109154. doi: 10.1016/j.oceaneng.2021.109154

Murray, B., and Perera, L. P. (2021). Proactive collision avoidance for autonomous ships: Leveraging machine learning to emulate situation awareness. *IFAC-PapersOnLine* 54, 16–23. doi: 10.1016/j.ifacol.2021.10.067

Nickel, M., Tresp, V., and Kriegel, H. P. (2011). A three-way model for collective learning on multi-relational data. *Proc. ICML. (International. Conf. Mach. Learning).* 11, 3104482–3104584. doi: 10.5555/3104482.3104584

Shin, Y., Kim, N., Lee, H., Young, S., Hansen, M., Yoon, Y., et al. (2024). Deep learning framework for vessel trajectory prediction using auxiliary tasks and convolutional networks. *Eng. Appl. Artif. Intell.* 132, 107936. doi: 10.1016/j.engappai.2024.107936

Team G, Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J. (2023). Gemini: a family of highly capable multimodal models. *arXiv. preprint. arXiv:2312.11805*. doi: 10.48550/arXiv.2312.11805

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. (2016). "Complex embeddings for simple link prediction," in *International Conference on Machine Learning*. New York NY USA, 2071–2080 (PMLR).

Wang, S., Li, Y., Xing, H., and Zhang, Z. (2024). Vessel trajectory prediction based on spatio-temporal graph convolutional network for complex and crowded sea areas. *Ocean. Eng.* 298, 117232. doi: 10.1016/j.oceaneng.2024.117232

Wang, S., Li, Y., Zhang, Z., and Xing, H. (2023a). Big data driven vessel trajectory prediction based on sparse multi-graph convolutional hybrid network with spatio-temporal awareness. *Ocean. Eng.* 287, 115695. doi: 10.1016/j.oceaneng.2023.115695

Wang, Y., Liu, J., Liu, R. W., Wu, W., and Liu, Y. (2023b). Interval prediction of vessel trajectory based on lower and upper bound estimation and attention-modified LSTM with bayesian optimization. *Physica. A.: Stat. Mechanics. Its. Appl.* 630, 129275. doi: 10.1016/j.physa.2023.129275

Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowledge. Data Eng.* 29, 2724–2743. doi: 10.1109/TKDE.2017.2754499

Wang, Z., Zhang, J., Feng, J., and Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. *Proc. AAAI. Conf. Artif. Intell.* 28, 1112–1119. doi: 10.1609/aaai.v28i1.8870

Wang, H., Zhong, H., Chen, A., Li, K., He, H., and Qi, Z. (2023c). A knowledge graph for standard carbonate microfacies and its application in the automatical reconstruction of the relative sea-level curve. *Geosci. Front.* 14, 101535. doi: 10.1016/j.gsf.2023.101535

Xiao, Y., Li, X., Yin, J., Liang, W., and Hu, Y. (2023). Adaptive multi-source data fusion vessel trajectory prediction model for intelligent maritime traffic. *Knowledge-Based. Syst.* 277, 110799. doi: 10.1016/j.knosys.2023.110799

Xie, W., Li, Y., Yang, Y., Peng, W., Zhishan, W., Zhaoxuan, L., et al. (2023). Maritime greenhouse gas emission estimation and forecasting through AIS data analytics: a case study of Tianjin port in the context of sustainable development. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1308981

Yang, P., Wang, H., Huang, Y., Yang, S., Zhang, Y., Huang, L., et al. (2024). LMKG: A large-scale and multi-source medical knowledge graph for intelligent medicine applications. *Knowledge-Based. Syst.* 284, 111323. doi: 10.1016/j.knosys.2023.111323

Yang, B., Yih, W. T., He, X., Gao, J., and Deng, L. (2014). Embedding entities and relations for learning and inference in knowledge bases. *arXiv. preprint. arXiv:1412.6575*. doi: 10.48550/arXiv.1412.6575

Zhang, C., Bin, J., and Liu, Z. (2024). TrajBERT-DSSM: Deep bidirectional transformers for vessel trajectory understanding and destination prediction. *Ocean. Eng.* 297, 117147. doi: 10.1016/j.oceaneng.2024.117147

Zhang, C., Liu, S., Guo, M., and Liu, Y. (2023). A novel ship trajectory clustering analysis and anomaly detection method based on AIS data. *Ocean. Eng.* 288, 116082. doi: 10.1016/j.oceaneng.2023.116082

Zheng, K., Zhang, X., Wang, C., Cui, H., and Wang, L. (2022). "Adaptive Path Planning for Autonomous Ships Based on Deep Reinforcement Learning Combined with Images," in *International Conference on Autonomous Unmanned Systems* (Springer Nature Singapore, Singapore), 1706–1715.

Zhengyu, Z., Qinhao, W., and Zhiqiang, Z. (2022). Knowledge graph construction and intelligent application based on enterprise-level big data of nuclear power industry. *Proc. Comput. Sci.* 214, 1334–1338. doi: 10.1016/j.procs.2022.11.313

Zhou, Y., Daamen, W., Vellinga, T., and Hoogendoorn, S. P. (2019). Ship classification based on ship behavior clustering from AIS data. *Ocean. Eng.* 175, 176–187. doi: 10.1016/j.oceaneng.2019.02.005

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 26 June 2024
DOI 10.3389/fmars.2024.1406381

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Fausto Ferreira,
University of Zagreb, Croatia
Zhen Kang,
Tianjin University, China

*CORRESPONDENCE
Noriko Tada
norikot@jamstec.go.jp

RECEIVED 25 March 2024
ACCEPTED 11 June 2024
PUBLISHED 26 June 2024

CITATION
Tada N, Nagano A, Tanaka S, Ichihara H, Suetsugu D, Matsuoka D, Ito M, Kuwatani T and Honda MC (2024) Challenge for multifaceted data acquisition around active volcanoes using uncrewed surface vessel. *Front. Mar. Sci.* 11:1406381. doi: 10.3389/fmars.2024.1406381

Challenge for multifaceted data acquisition around active volcanoes using uncrewed surface vessel

Noriko Tada[1*], Akira Nagano[2], Satoru Tanaka[1], Hiroshi Ichihara[1,3], Daisuke Suetsugu[1], Daisuke Matsuoka[4], Makoto Ito[2], Tatsu Kuwatani[1] and Makio C. Honda[2]

[1]Research Institute for Marine Geodynamics, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa, Japan, [2]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa, Japan, [3]Earthquake and Volcano Research Center, Graduate School of Environmental Studies, Nagoya University, Nagoya, Aichi, Japan, [4]Research Institute for Value-Added-Information Generation, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan

Monitoring of volcanic eruptions, the atmosphere, and the ocean, along with their scientific understanding, can be achieved through multifaceted observations that include camera images, topographic deformations, elastic waves, geology, and the chemical constituents and temperatures of air and water. However, regions of increased volcanic activity and/or shallow waters are difficult to access by crewed ship due to danger or grounding risk. To overcome these difficulties, we used an uncrewed surface vessel (USV), Bluebottle, to operate multiple observation around oceanic volcanoes in the Bonin Island arc in May 2023. Even under adverse sea conditions, the USV successfully reached a remote observation site, Nishinoshima volcano, which is about 130 km away from Chichijima Island where the USV started out from. The USV conducted five days of observations at shallower than 500 m water depth around Nishinoshima Island, the first time after violent eruptions of Nishinoshima in June 2020. The USV is equipped with various sensors and data collection technologies, including a single-beam echosounder, oceanographic and meteorological sensors, an Acoustic Doppler Current Profiler, and a time-lapse camera. These tools have provided a multi-dimensional view of the underwater landscape and marine conditions near the volcano for the first time in the world. We obtained new bathymetric data, sub-bottom images, and measurements of temperature, salinity, and pH. This study leverages advanced technologies and innovative methodologies to enhance our understanding of marine and geological phenomena.

KEYWORDS

USV, volcanic islands, Nishinoshima, bathymetric survey, marine geological survey, ocean monitoring, ocean surveillance

1 Introduction

Volcanic activities, many of which occur in volcanic islands and submarine volcanoes in the Earth, can potentially cause severe damage to human activities, as the 2022 eruption of the Hunga Tonga-Hunga Ha'aoai volcano generated a huge amount of volcanic ash that reached the mesosphere (e.g., Matoza et al., 2022) and a global tsunami (e.g., Proud et al., 2022). Monitoring of volcanic eruptions and their scientific understanding can be achieved by multifaceted observations, including camera images, topographic deformations, elastic waves, geology, and chemical constituents and temperatures of air and water. Recently, the understanding and monitoring of inland volcanoes have progressed based on integrated land-based observations (e.g., Iguchi et al., 2019). However, studies of oceanic volcanoes have lagged far behind those of inland volcanoes due to fundamental inaccessibility and difficulty for the scientists to deploy instruments. In addition, observations by crewed ship are very costly. Moreover, areas inaccessible to crewed vessels, such as regions of increased volcanic activity and/or shallow waters prone to grounding risks, become blind spots for observations even in coastal waters.

Sea surface currents vary on wide ranges from basin-scale to mesoscale (~10–100 km) and smaller scale (i.e., submesoscale) and decadal to daily scales. Recently, mesoscale and submesoscale eddies have been highlighted because they may contribute to the transport and redistribution of heat and materials (Ando et al., 2021). The kinetic energies of mesoscale and submesoscale eddies are limited to western boundary currents such as the Kuroshio and the Gulf Stream and around topographies such as islands and seamounts. Mesoscale and submesoscale eddies in the Kuroshio Extension region have been intensively studies (e.g., Inoue et al., 2016; Nagano et al., 2016). Meanwhile, observations of mesoscale and submesoscale features generated by topographies are quite limited except for areas performed in specific projects (Johnston et al., 2019). In addition, the atmosphere is considered to be modified by the temperature variations associated with topography-induced mesoscale and submesoscale ocean features through the air–sea interaction but the modification has not been revealed yet.

The advancements in automation and miniaturization of observing instruments have enabled the efficient and cost-effective execution of observations through the use of uncrewed surface vessels (USVs) (Patterson et al., 2022). Besides, such small platforms do not disturb the atmosphere and ocean as much as shipboard observations (Nagano et al., 2022). Thus, uncrewed observations provide a remarkable means realizing data acquisition in previously inaccessible oceanic regions. In particular, USV's observations of oceanic volcanoes have great potential in term of 1) the ability to observe in the near-shore region can fill observation gap the above-mentioned, 2) multiple USVs can be deployed around the volcano to build a cost-effective observation network, and 3) USVs can obtain atmospheric and oceanic parameters related to volcanic activities and the topography-generated eddies (however, note that chemical investigations often require sampling and analysis of water from volcanic lakes on land). On the other hand, USV observations still face challenges such as navigation permit issues and lack of fuel, power, and communication capacity.

Nishinoshima Island and Fukutoku-Oka-no-Ba submarine volcano (hereafter, FOB) in the Bonin Island arc (Figure 1) are recently erupted volcanoes with the above-mentioned observational difficulties. Nishinoshima Island is an uninhabited island located about 1000 km south of Tokyo and about 130 km west of the nearest inhabited island, Chichijima Island. Nishinoshima volcano resumed volcanic activity in 2013 for the first time in about 40 years and has been actively erupting ever since, significantly enlarging the portion of the island above sea level (Maeno et al., 2021). In June 2020, due to navigation warnings, it became not possible to go within 1.5 km of the summit crater of Nishinoshima, resulting in blank areas on the latest topographical and nautical charts.

FOB is situated about 335 km south of Nishinoshima Island. On August 13, 2021, FOB experienced a major eruption, producing a vast amount of pumice and an eruption column up to 16,000 m (e.g., Maeno et al., 2022; Yoshida et al., 2022). An island temporarily emerged but was eroded by waves and disappeared below the sea surface about three months later. FOB is expected to have very shallow waters, making observation by crewed vessels risky due to the high probability of grounding. After significant eruptions at both volcanoes, attempts were made to acquire new bathymetric data to compare with pre-eruption maps. We do not have any bathymetric data, i.e., depth data, following eruption, so it is unclear whether the changes in depth are gradual or if features like side volcanoes or chimneys have formed, suddenly making the water shallower. Additionally, discolored water and bubbles, presumably caused by volcanic activity, have been observed around Nishinoshima Island and crater area of FOB, and the potential impacts on crewed ships and their occupants are not yet understood. Therefore, in the 2022 crewed voyage, we adopted a precautionary approach by not approaching areas shallower than 500 meters with crewed ships.

In this paper, we present a considerably challenging observation conducted using a single engineless USV targeting the volcanoes Nishinoshima and FOB, which are located a significant distance from any inhabited islands. The study includes multiple

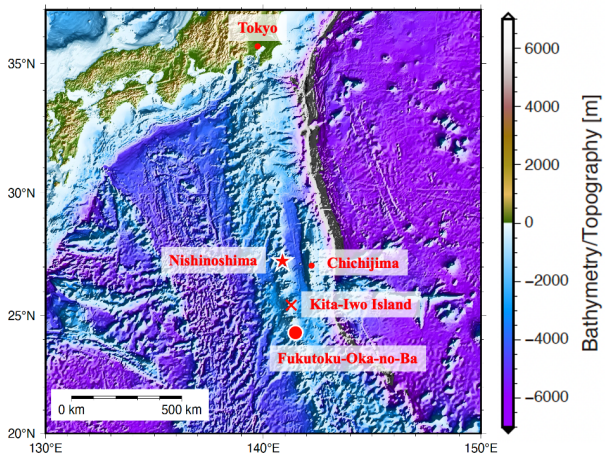


FIGURE 1
Location of Chichijima Island, Nishinoshima Island, Kita-Iwo Island, and Fukutoku-Oka-no-Ba submarine volcano (FOB). Elevation data are from the ETOPO1 global relief model (NOAA, 2009).

observational objectives deploying USVs in remote volcanic regions for the first time in the world. First, the USV and the observation instruments mounted on the USV are described. Then an overview of the operations from departure to observation and return is given. Next, the obtained data are described. Based on these results, the usefulness and problems of these observations, including operations, are discussed.

2 Methods

2.1 USV

We utilized a USV Bluebottle (Supplementary Table S1; Figure 2), developed and owned by the Australian private company OCIUS (https://ocius.com.au/). The Bluebottle boasts a total length of 6.8 m and, when its sail is deployed, reaches a height of approximately 5.0 m above the waterline. The vessel is designed with a payload capacity of up to 350 kg. Figures 2A, B illustrate the Bluebottle with its sail raised and lowered, respectively, showcasing the vessel's versatility in adapting to various marine research and observation conditions. The Bluebottle is propelled primarily by wind power and supplemented by wave and solar power (photovoltaics) and does not use an engine. Remote control, facilitated through the use of Iridium, adjusts the sail angle for wind-driven navigation from a land-based station. The current location of the Bluebottle, along with essential navigation information such as wind direction and speed, and the view ahead, can be monitored nearly in real-time via a web interface (Figure 3). The vessel sails at speed of 2–4 knots under the solar-powered mode and 0.5–1.5 knots wave-powered mode. The nominal maximum speed can reach 6.0 knots. The Bluebottle is equipped with an Automatic Identification System (AIS), a 360-degree time-lapse camera, radar, lights, and a horn. OCIUS monitors Bluebottle's navigation around the clock using AIS, radar, and the 360-degree camera. While the vessel lacks automated obstacle avoidance capabilities, potential collisions or accidents are mitigated through remote control operation.

2.2 Data acquisition instruments

Prior to the Bluebottle survey commencement, all instruments were configured to meet project specifications (Supplementary Table S2). Heading, rolling, and positional data was collected

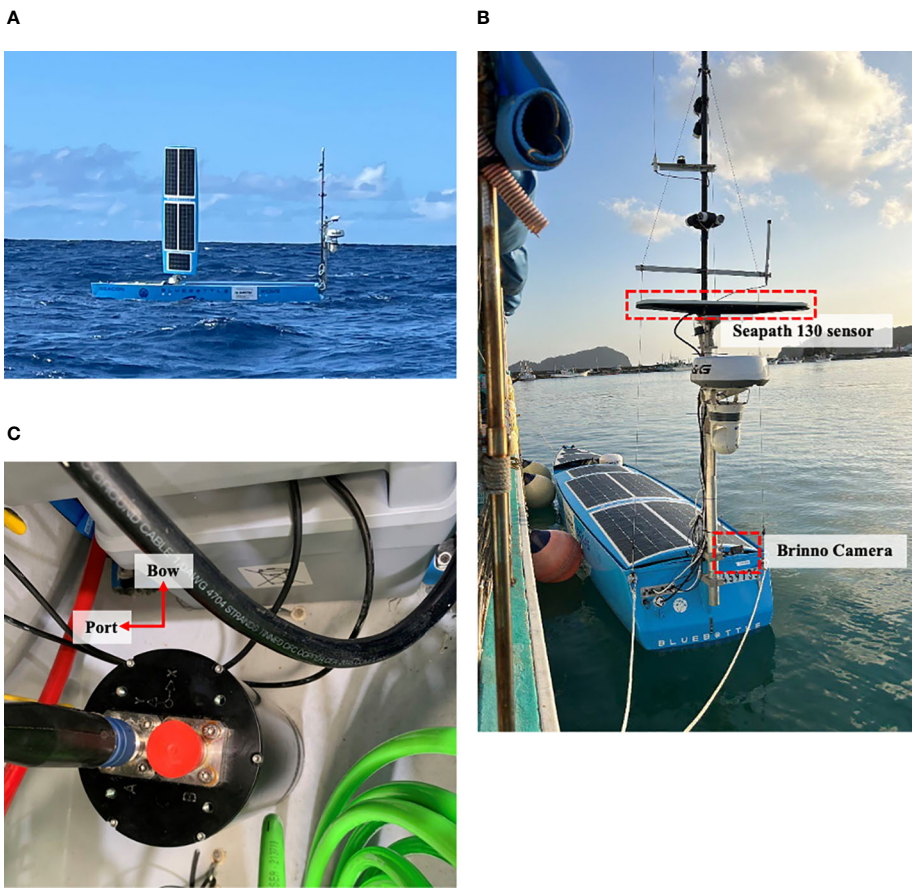


FIGURE 2
The Bluebottle **(A)** with its sail raised and **(B)** lowered. The locations of Seapath 130 sensor, Brinno Timelapse camera. **(C)** Spider cable connected to a fixed motion reference unit.

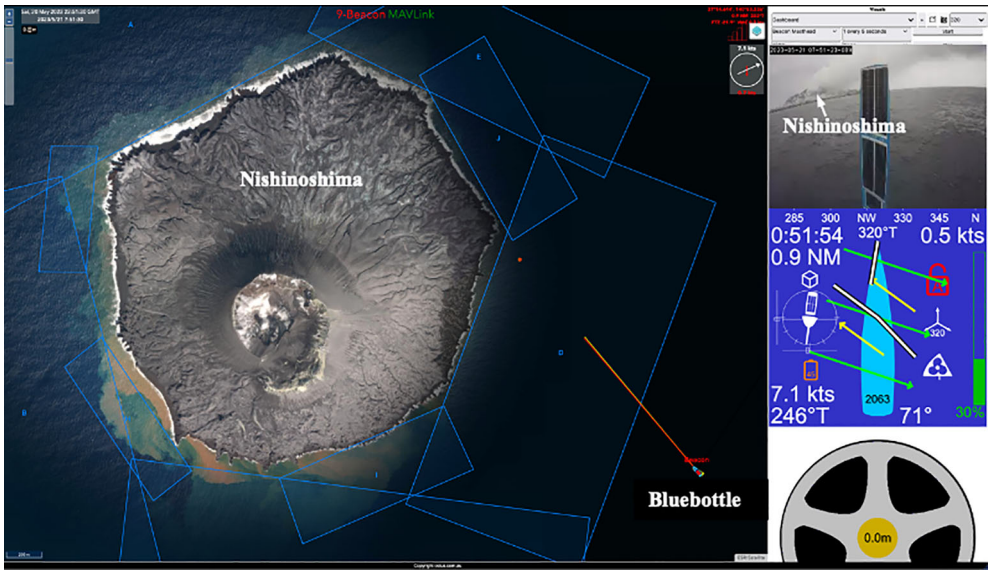


FIGURE 3
An example of a web interface showing the current location of the Bluebottle with essential navigation information.

using a Seapath 130 sensor settled on top of the Bluebottle (Figure 2B). The Seapath 130 sensor contains integrated Global Navigation Satellite System (GNSS) antennas and receivers and is connected to a fixed motion reference unit (MRU), which is housed within an underwater enclosure, using a spider cable (Figure 2C). The MRU orientation was installed based on guidelines and configured within the Seapath 130 software to ensure accurate Inertial Navigation System (INS) and positional data.

Acquiring accurate topographical and geological data is paramount for understanding the dynamic changes of oceanic island volcanoes and submarine volcanoes. Depth data, essential for this purpose, can be comprehensively obtained through sub-bottom profiling, enabling the determination of sediment layer thickness and, in areas of volcanic activity, the depth of volcanic ash layers. Additionally, side-scan sonar data facilitate the examination of reflectance intensity, serving as a critical tool for discerning the composition of the seabed, particularly for identifying exposed lava formations in regions not obscured by volcanic ash. To achieve this integrated data acquisition, the Kongsberg EA640/EA440 All-in-one system was deployed on the Bluebottle, as illustrated in Figure 4. The advanced system incorporates a 200 kHz side-scan sonar, a 15 kHz sub-bottom profiler/transducer, and a 200kHz transducer, all connected through the EA440 Wide Band Transceiver (WBT) operating with a sampling interval of 0.1 seconds. The bathymetry survey utilized both 15 kHz and 200 kHz frequencies to maximize data accuracy and resolution. Installation of the single-beam echosounder, side-scan sonar, and sub-bottom profiler at the aft

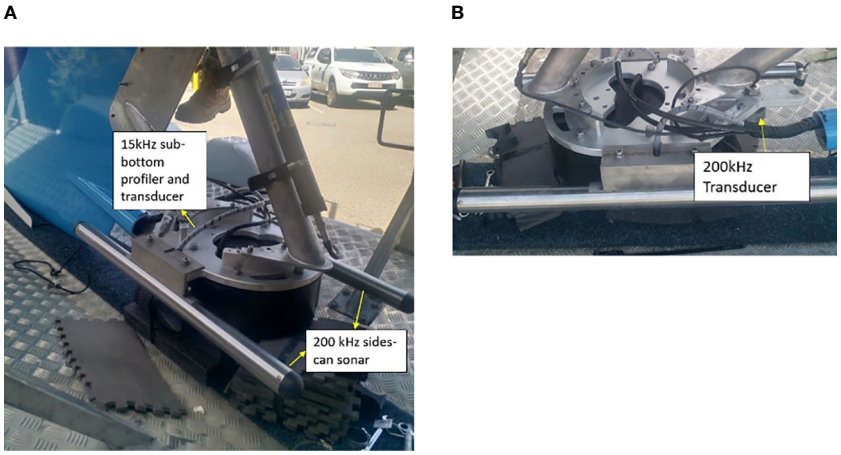


FIGURE 4
The location of the EA640/440 All-in-one system.

end of the keel ensures optimal performance and data collection efficiency. The entire suite of instruments is meticulously configured and managed via the EA440 software interface, ensuring seamless operation for the collection of submarine topography and geology data.

We used a suit of meteorological sensors, an Airmar WS-200WX IPx7, mounted on the Bluebottle to obtain variations in atmospheric parameters (air temperature, atmospheric pressure, and wind speed/direction). The device was installed to the highest point of the communications mast (the rear mast) of the vessel to ensure minimal interference with wind speed and direction readings for collection of wind speed/direction, air temperature, and atmospheric pressure data (Figure 5A). The height of the device from the sea level was approximately 3 m. A YDHS-01 humidity sensor was used to measure relative humidity on this mission (Figure 5A). However, the humidity sensor and a cable connection between the sensor and data logging system was submerged by heavy rainfall in the first trial. Because the connector could not be completed the repair until the deployment on May 16, 2003, we abandoned the humidity measurement. Infrared thermometers, Heitronics CT15.10 and CT9.10WT, were installed to monitor the sea surface skin temperature. The CT15.10 was installed downward facing, and the CT9.10WT faced the sky (Figure 5A). The CT9.10WT readings were used to provide temperature compensation signal. Readings of these devices were uploaded in real-time to the OCIUS database at 1 Hz.

Changes in water temperature, salinity, dissolved oxygen concentration (DO), turbidity, phytoplankton concentration, and pH resulting from volcanic activity and ocean current variability excited by the topography were monitored using a multiparameter probe Aqua TROLL 600 (*In-Situ*, Port Colins, CO, USA) mounted on the Bluebottle. The device is anti-corrosive and has a low power consumption, rendering it highly suitable for this USV survey. The

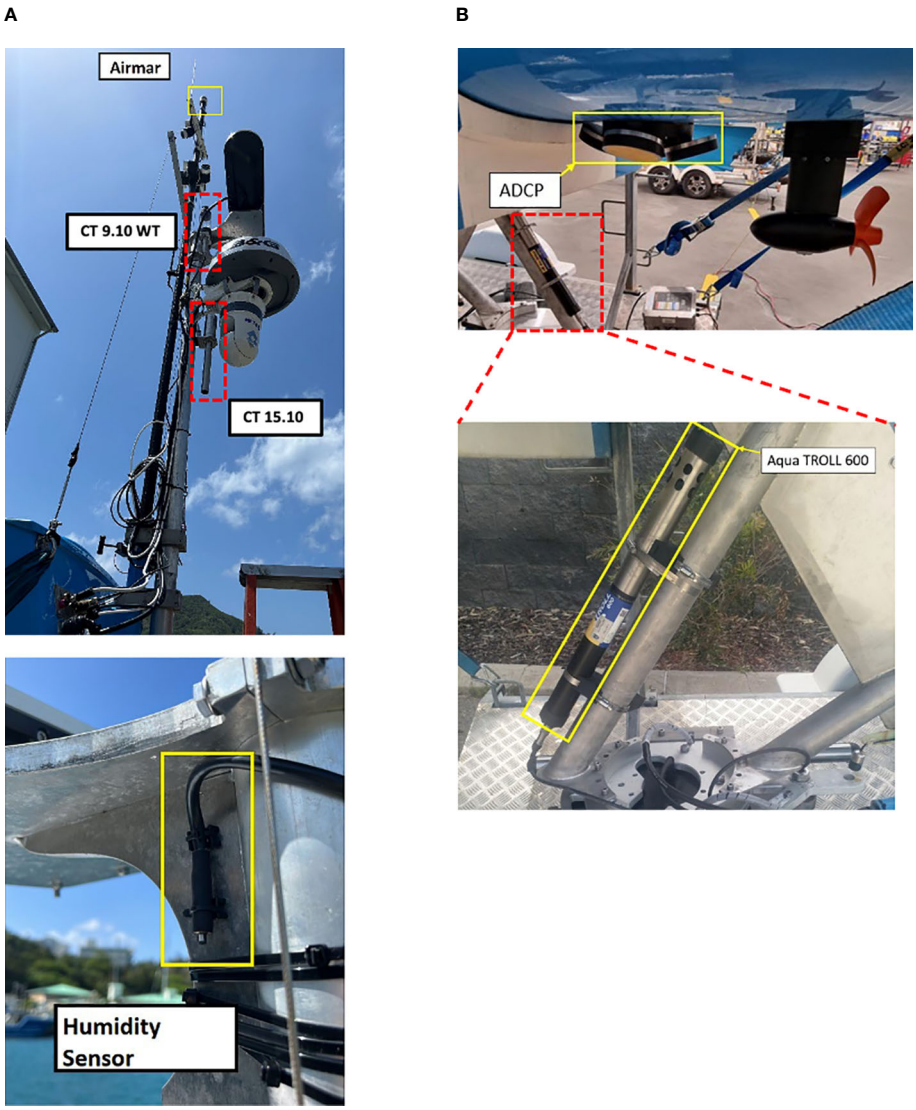


FIGURE 5
ADCP, Airmar, and Aqua TROLL 600.

probe was installed aft end of the keel of the Bluebottle at a depth of approximately 1 m with the sensor orientation pointed towards the keel to ensure minimal interference with sonar equipment (Figure 5B) and which measured pH, DO, conductivity, temperature, chlorophyll-a fluorescence, and pressure with 1-minute sampling intervals. Salinity in practical salinity unit was calculated from temperature and conductivity values. The accuracies of the instrument are 0.1°C for temperature, 0.1 for salinity, 0.1 for pH, and 0.1 mg/L for DO.

For more detailed analysis of the hydrographic structures around the island and volcano, temperature and salinity were also measured each 10 min by using another conductivity-temperature-depth sensor JES10mini (Offshore Technologies, Yokohama, Japan) installed near the Aqua TROLL 600 probe. The temperature and conductivity accuracies of the JES10mini are 0.005°C and 0.005 S/m, respectively. The accuracy of computed salinity is better than 0.04.

Moreover, a backscatter meter with fluorometer (FLBBSB, Seabird Science, USA) was installed next to JES10mini to observe the spatio-temporal variabilities in underwater biogenic and lithogenic particles.

Additionally, an Acoustic Doppler Current Profiler (ADCP) (Nortek Signature100 ADCP, Norway) was installed downward to the bottom end of the rear keel (Figure 5B) to measure the vertical structure of current velocity, enhancing our understanding of underwater current patterns and temporal variations. The ADCP is a long-range current profiler and collects a current profile at depths from 0 to 500 m every 1 minute. In this survey, the instrument was set to record current profiles after averaging 10 minutes. Because this instrument does not detect the seabed, we removed data deeper than the water depth based on data collected by the 15 kHz sub-bottom profiler. In regions of water depths greater than 500 m, where are located far from the island, no data was obtained below 500 m depth.

To monitor the color of the seawater and floating objects in the area where the Bluebottle is sailing, we took time-lapse photos of the ocean surface. Specifically, understanding how marine debris, including ocean plastics, is distributed in the open sea is crucial for grasping the overall situation of the marine plastic issue, which has become a significant problem in recent years. The Brinno TLC2020-C Timelapse Camera (Figure 2B) was utilized to capture downward facing images of the water's surface during the mission. The camera's picture frequency was chosen to maximize the framerate of the images while ensuring a sufficient level of charge was maintained. Information from the device was recovered at the end of the mission.

3 Results and discussion

3.1 Observation cruises

The Bluebottle was transported from Fremantle (Australia) to Chichijima Island (Japan), which is an inhabited island of Japan relatively close to Nishinoshima and FOB, to conduct observation cruises in the Bonin region. As the Bluebottle is a USV owned by OCIUS, an Australian company, transporting and operating it in a territory of Japan through challenging procedures related to import and adherence to Japanese laws. For instance, any floating object exceeding 3.3 meters in length requires ship registration and inspection on site before its voyages. That forced us to have an additional preparation period and costs for the USV surveillance.

The Leg 1 cruise of the Bluebottle started on April 24, 2023, from Futami port of Chichijima Island toward FOB (Figure 1). The Bluebottle had advanced near the southern vicinity of Kita-Iwo Jima. However, due to difficulties encountered in data acquisition, it was decided to return to Chichijima Island on April 27. The Bluebottle arrived back at Futami port of Chichijima Island on May 1, marking the conclusion of Leg 1 of the cruise. Unfortunately, topographical, geological, meteorological, and oceanographic data were not acquired around FOB this time.

The ship track of each day in the Leg 2 journey is illustrated in Figure 6A. The Leg 2 cruise of the Bluebottle commenced on May 16 from Futami port, heading towards Nishinoshima Island. Upon arrival in the observation area on May 19, the Bluebottle began operations of the EA640/440 All-in-one system to acquire topographical and geological datasets. Initially, the Bluebottle navigated along planned survey lines running northeast to southwest on the northwest side of Nishinoshima Island, spaced 50 m apart, to collect data. The Bluebottle complete two round trips mostly along planed survey lines. However, the interaction between the vessel's heading, tidal currents, and wind direction and power prevented it from sailing linearly along the survey lines at all times. Figure 6B shows the fluctuation in the battery's remaining charge during this phase. From Figure 6B, it is evident that during May 19 to May 21, 2023, the solar-powered battery was hardly able to charge, and the battery percentage consistently decreased. This suggests that the sunlight was insufficient to recover the battery's remaining charge, and just five days after departure, the battery's remaining charge had already halved. Consequently, to acquire accurate bathymetric data over a wider area while minimizing power consumption, we decided to navigate the Bluebottle in a way that requires minimal power, adjusting the navigation according to the prevailing wind directions and tidal currents at the eastern side of Nishinoshima Island. Moreover, interested in the variations in water depth depending on the direction from the crater, we also conducted circumnavigational observations around the island. With the approach of the powerful Typhoon Mawar on May 20 (minimum pressure 900 hPa and maximum wind speed 60 m/s; Japan Meteorological Agency, 2023), the Bluebottle was forced to depart from the observation area around Nishinoshima Island on May 23, and arrived back at Chichijima Island on May 28, thereby concluding the Leg2 cruise. Even under adverse sea conditions, the Bluebottle, which operates solely on natural energy without a combustion engine, successfully reached a remote observation site and conducted five days of observations. However, the inability to allocate a sufficient observation period remains a challenge for future endeavor.

The currents around Nishinoshima Island were much stronger

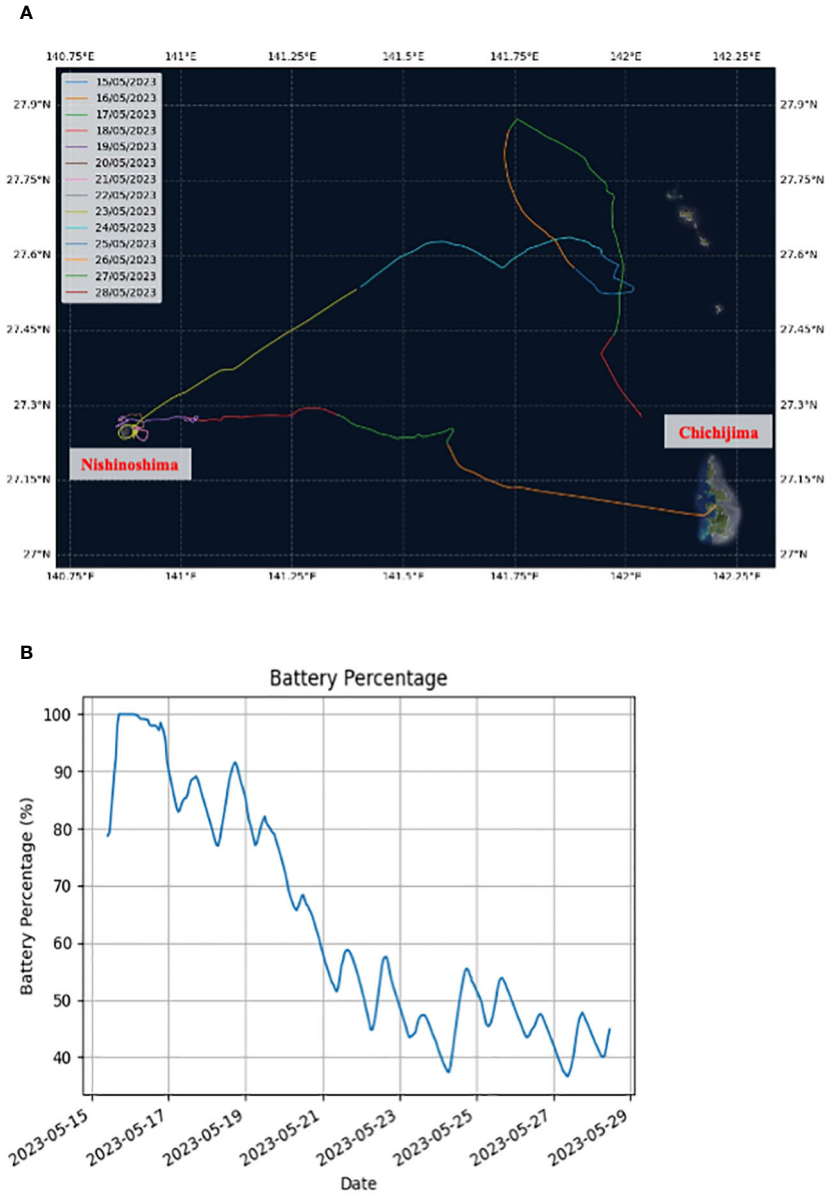


FIGURE 6
(A) Mission track of the Leg 2 cruise and **(B)** the battery charge remaining during the Leg 2 cruise.

than previously anticipated. In this condition, the Bluebottle needed to use a significant amount of throttle to achieve the desired transects. Because the Bluebottle had to save throttle usages for sustainable operation, the Bluebottle had to loiter in the lee of the currents. In the western North Pacific, typhoons frequently occurs and the Kuroshio current, other adjacent current systems, and mesoscale eddies are associated with quite strong currents. To conduct observations in the severe sea conditions in the western North Pacific, USVs, including Bluebottle, should have so strong body as to stay under such severe conditions and sufficient mobility to escape typhoons and other atmospheric disturbances. Additionally, in the extratropical regions, cloudy and rainy days are relatively more frequent particularly in winter and early summer (the rainy season in Japan). More development of the solar-powered battery management system is considered to be essential to perform long-term observations regardless of the season.

3.2 Observed data

3.2.1 Topographical and geological data

In the Leg 2 cruise, single-beam bathymetric data were collected from May 19 to May 23 around Nishinoshima Island. Over this five-day survey period, a total of 27,717 raw bathymetric files were logged. The data from the single-beam echosounder, covering depths of 0–1200 m at 15 kHz and 0–150 m at 200 kHz were processed into a grid with a resolution of 10 meters and are

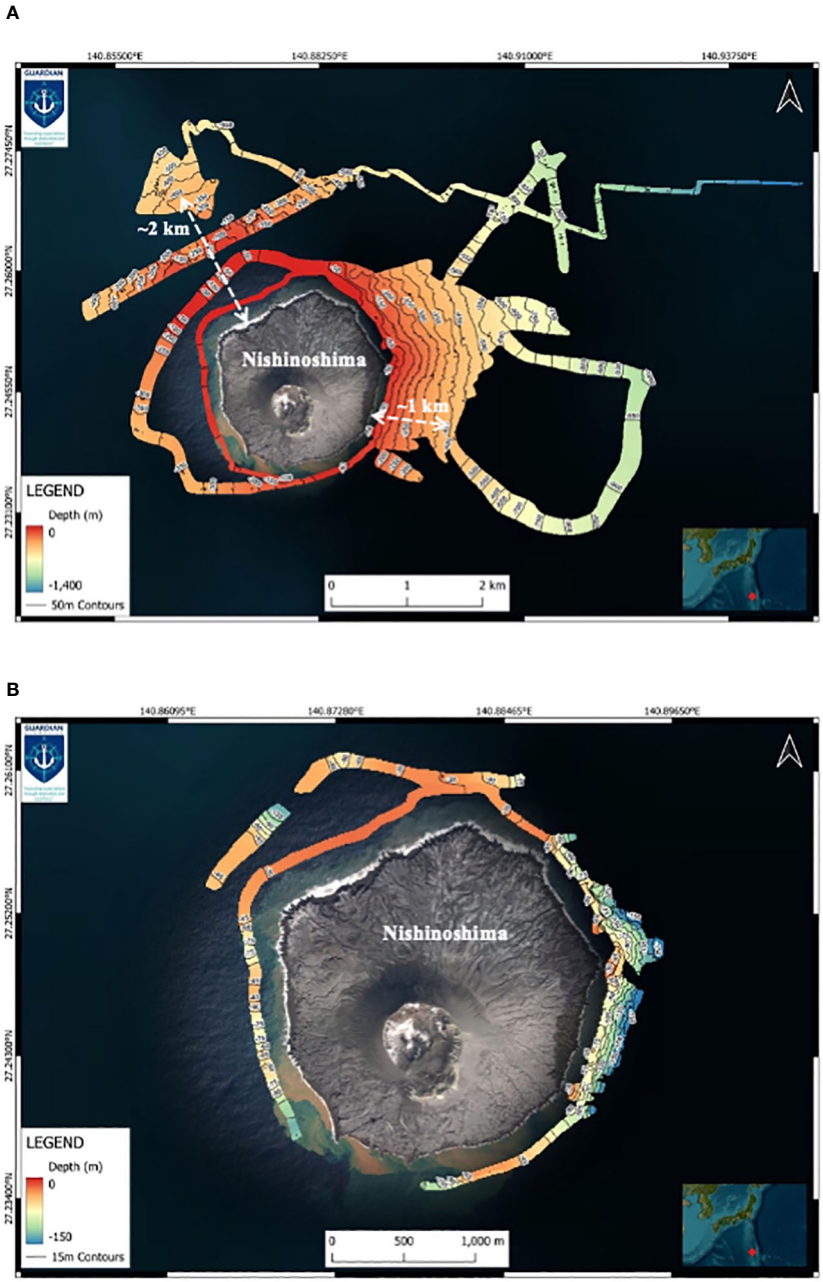


FIGURE 7
Seafloor topography data sampled **(A)** at 15 kHz and **(B)** at 200 kHz. The white double arrows and numbers indicate the distances from the shoreline to a water depth of 400 meters.

presented in Figures 7A, B, respectively. Figure 7A reveals that although the data acquisition is uneven, it clearly delineates the contrasting seabed topographies on the northwestern and eastern sides of the island. The depth reaches 400 meters at about 3 km from the crater and approximately 2 km from the shoreline on the northwestern side, whereas on the eastern side, it occurs less than 2 km from the crater and about 1 km from the shoreline. This suggests a more gradual slope on the northwestern side compared to the eastern side. The same trend is observable in Figure 7B, which has higher resolution in shallower areas. The bathymetric data, which nearly encircles the island, indicate that the northern to northwestern sides have shallower depths compared to other directions. The bathymetric data reveal a steep slope of approximately 40 degrees on the northeast to eastern flank of the volcano, where water depths exceed 100 m.

Sub-bottom profiler data collected at 15 kHz to the northeast and to the southeast of Nishinoshima Island are shown in Figures 8A, B, respectively. Both sub-bottom profiler datasets exhibit continuous high reflection layers. Figure 8A displays two such layers, with the lower high reflection layer appearing similar to the high reflection layer in Figure 8B, suggesting that these layers represent the seabed foundation. In contrast, the upper high reflection layer in Figure 8A shows lower

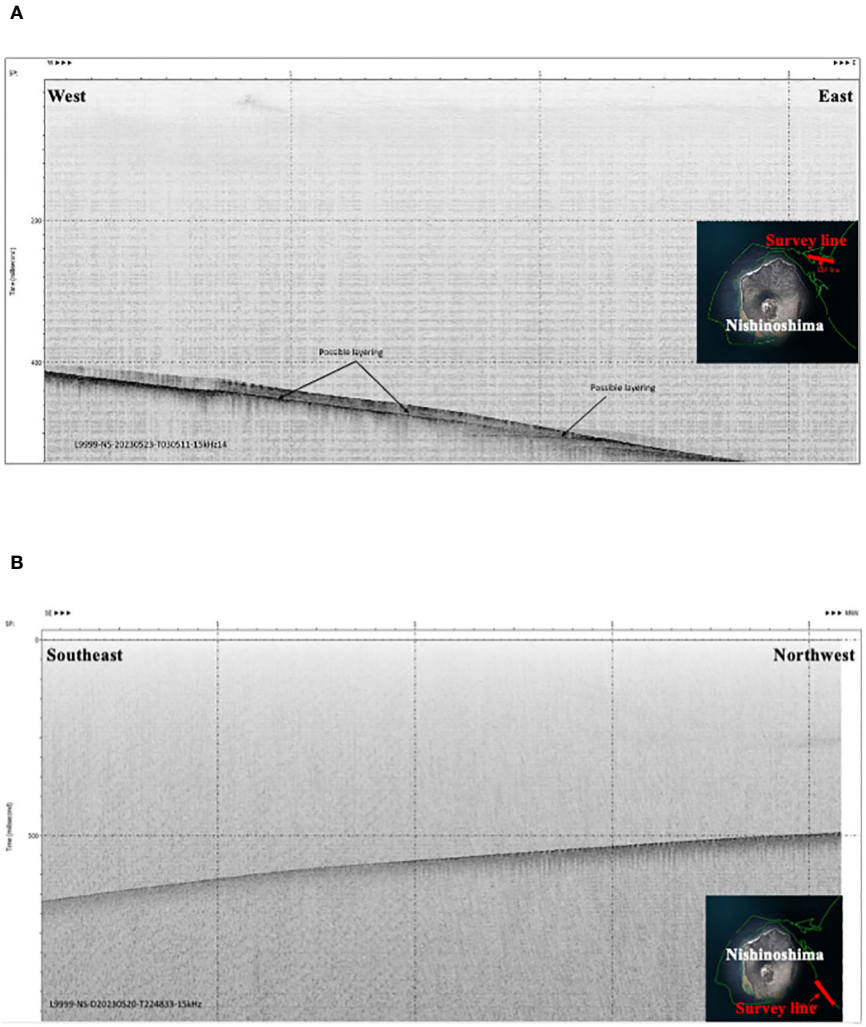


FIGURE 8
Sub-bottom profiles acquired **(A)** in the northeast and **(B)** in the southeast of Nishinoshima Island.

reflectivity on the west side, and the layer thickness between the high reflection layers seems to thin towards the east. The thinning of layers away from Nishinoshima Island indicates that volcanic ash erupted from the volcano has accumulated on top of the foundation, forming layers. The weak reflection area on the west side may indicate that the volcanic ash has not yet solidified sufficiently. Moving forward, we aim to extract as much information as possible about the horizontal extent and thickness variation of this presumed ash layer from the data we got in this observation, to assess the amount of volcanic ash deposited on the seabed.

The attempt to obtain meaningful acoustic image of the seabed using side-scan sonar was unsuccessful. The side-scan sonar used in this study operated at a high frequency of 200 kHz, which is highly effective in shallow water. Given that the shallowest observed water depth was around 30 meters (Figure 7B), it is unlikely that no data were captured at all. Therefore, future efforts will involve verifying the side-scan sonar data in conjunction with sub-bottom profiler data and water depth data. This will allow us to thoroughly review whether any data were indeed collected by the side-scan sonar and to revalidate the data accordingly.

3.2.2 Meteorological data

Immediately after the passage of a low-pressure system, the Bluebottle departed at the Chichijima Island and observed the northwesterly wind (Figure 9A). Throughout the observation, two low-pressure systems passed during the observation. Associated with the atmospheric disturbances, the wind was observed to rotate and to be equivalent to or exceeds approximately 10 m/s during the most of the observation. Because of cloudiness during the passages of the low-pressure systems, the Bluebottle battery was concerned not to be sufficiently charged and not to continue the observation.

Due to the cold skin effect (e.g., Saunders, 1967), skin temperature similarly varied with temperature at 1 m depth, i.e., SST, but was always lower than SST (Figures 9B, C). The low skin temperature is considered to be closely related to the heat exchange between the atmosphere and ocean. The low-skin-temperature may

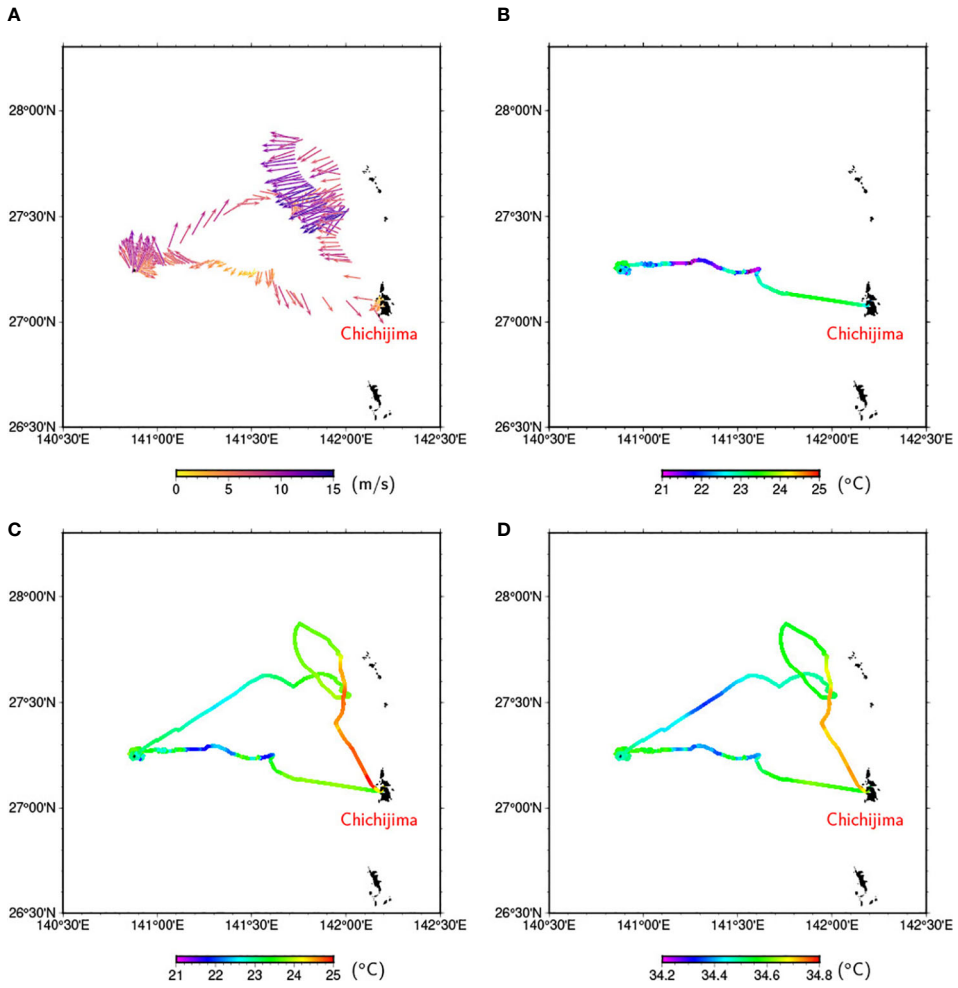


FIGURE 9
(A) Wind vectors in m/s on the Bluebottle track. Wind speed is indicated by color bar at the bottom of the panel. **(B)** Skin temperature in °C on the Bluebottle track. **(C)** Temperature and **(D)** salinity on the Bluebottle track observed by the CTD JES10mini.

suggest that the turbulent plus long-wave heat flux was always upward.

3.2.3 Water properties

Figures 9C, D show temperature and salinity observed by the CTD JES10mini along the Bluebottle track, which are referred in this paper as sea surface temperature (SST) and sea surface salinity (SSS), respectively. SST and SSS ranged in 21–25°C and 34.2–34.8, respectively, which are consistent with climatology in this region. Notably, the SST and SSS are found to covary on spatial scales of about 10–50 km, which is within the mesoscale range. For example, the Bluebottle observed cold and fresh water on the westward track from 141.6°E to 141.1°E and warm and saline water on the southward track from 27.6°N to 27.0°N. Furthermore, focusing on the variations in temperature and salinity around Nishinoshima Island, the SST and SSS varied on time scales shorter than approximately 6 hours (Supplementary Figure S1). During the observation, a mesoscale cyclonic eddy was detected to the northeast of Nishinoshima Island by the satellite sea surface height observation. The high-frequency SST and SSS variabilities may be attributed to submesoscale (< 10 km) hydrographic structures generated around the volcano in a background southeastward flow generated by the mesoscale eddy.

Eastward currents over 0.5 m/s were observed to be along the contours of the topography to the northwest of the island by the ADCP (for example, current velocity vectors at 7 m depth are shown in Figure 10). The current speed is much larger than the estimated velocity from altimetric sea surface height data (~0.2 m/s). The strong current may be driven by the cyclonic eddy and modified by the topography. Remarkably, the ADCP observed currents directed to the west on the leeward side of the island or east of the island. Such current normal to topographic contours must not be in the geostrophic balance and be highly variable in space and time, as observed in SST and SSS (Figures 9C, D).

On the way to Nishinoshima Island, DO varied from 6.3 to 6.6 mg/l (197 to 206 micro mol/l) (Figure 11A). These values are typical values observed in late spring in the western Pacific subtropical region (~ 200 micro mol/l; e.g., Honda, 2020). East of Nishinoshima

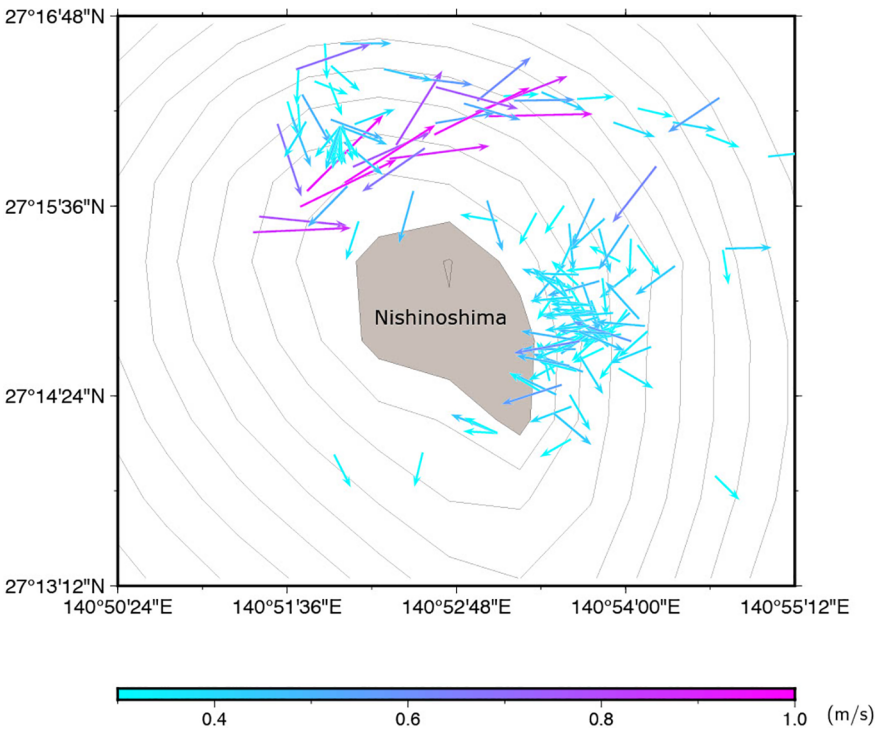


FIGURE 10
ADCP velocity vectors in m/s at the uppermost layer (7 m depth). Current speed is represented by the color bar at the bottom. Vectors with speed greater than 0.3 m/s were plotted. Bottom topography based on ETOPO 2022 was shown by contours at intervals of 100 m. The area shaded in gray represents Nishinoshima Island.

Island, DO decreased suddenly although data were limited (Figure 11B). Essentially, DO correlated well with SST (Supplementary Figure S2), while DO did not correlate with Chl-a (not shown). It is indicative of that DO concentration were determined by, not primary productivity, but solubility. However, lower DO deviated from SST-DO linier regression line. Thus, the water east of Nishinoshima Island was likely affected by outflow from the island. This outflow water might be reductive due to volcanic gas such as CO_2 and SO_2 (e.g., de Ronde and Stucker, 2015).

As the Bluebottle was approaching to Nishinoshima Island, pH values decreased and the minimum was observed at the east of the island (Figure 12). Lower pH near volcanically-active Nishinoshima Island was expected before this investigation as reported previously (Sato et al., 2018). However, absolute value ranged from 8.6 and 9.7 and these values are never observed in the surface water of western North pacific subtropical area (usually ~ 8.1; Honda et al., 2017). Based on pH measurement of discolored seawater sampled by using multicopter, low pH of ~6.5 was observed around Nishinoshima (Takai, 2024). In addition, pH sensor values monotonically decreased during deployment and did not have any significant correlations with other properties; unfortunately, the observed change might be artifact.

Backscatter of light of 700 nm wave length ($bbp_{(700)}$) values were lower than 0.01 m^{-1} or under detection limit on the way to and north of Nishinoshima (Figure 13A). This is typical value in the open ocean in the western Pacific subtropical area (e.g., Fujiki, 2019; Taketani, 2021; Kitamura, 2022). However, backscatter increased largely east side of Nishinoshima (Figure 13B). The $bbp_{(700)}$ > 0.05 m^{-1} (upper detection limit) has not been observed in the open ocean of the western Pacific previously. This unusual high backscatter coincides with the fact that the surface seawater around Nishinoshima Island discolored shown in pictures taken by time-lapse camera on the Bluebottle (Figure 14). Based on chemical analysis, Takai (2024) reported that this discoloring (red-brown color) was attributed to suspended particles rich in iron.

Chl-a also increased slightly east of Nishinoshima Island (Figure 15). Weak but significant positive correlation was observed between Chl-a and backscatter, especially near Nishinoshima Island (solid black circles in Supplementary Figure S3A). Thus, although high backscatter was mainly attributed to iron-rich suspended particles as discussed in the previous paragraph, particles detected as backscatter might be partly phytoplankton. In addition, Chl-a observed at night near Nishinoshima Island correlated negatively with temperature (solid blue crosses in Supplementary Figure S3B). However, the maximum of Chl-a was very low (at most 0.15 mg/m^3) and, thus, it is hard to say that, due to supply of micro (iron)/macro (e.g., nitrogen and phosphate) nutrient, volcanic activity or the outflow from Nishinoshima Island enhanced phytoplankton.

3.2.4 Surface photo images

From May 16, 2023, to May 28, 2023, the Brinno Timelapse Camera recorded data, capturing photos every five minutes,

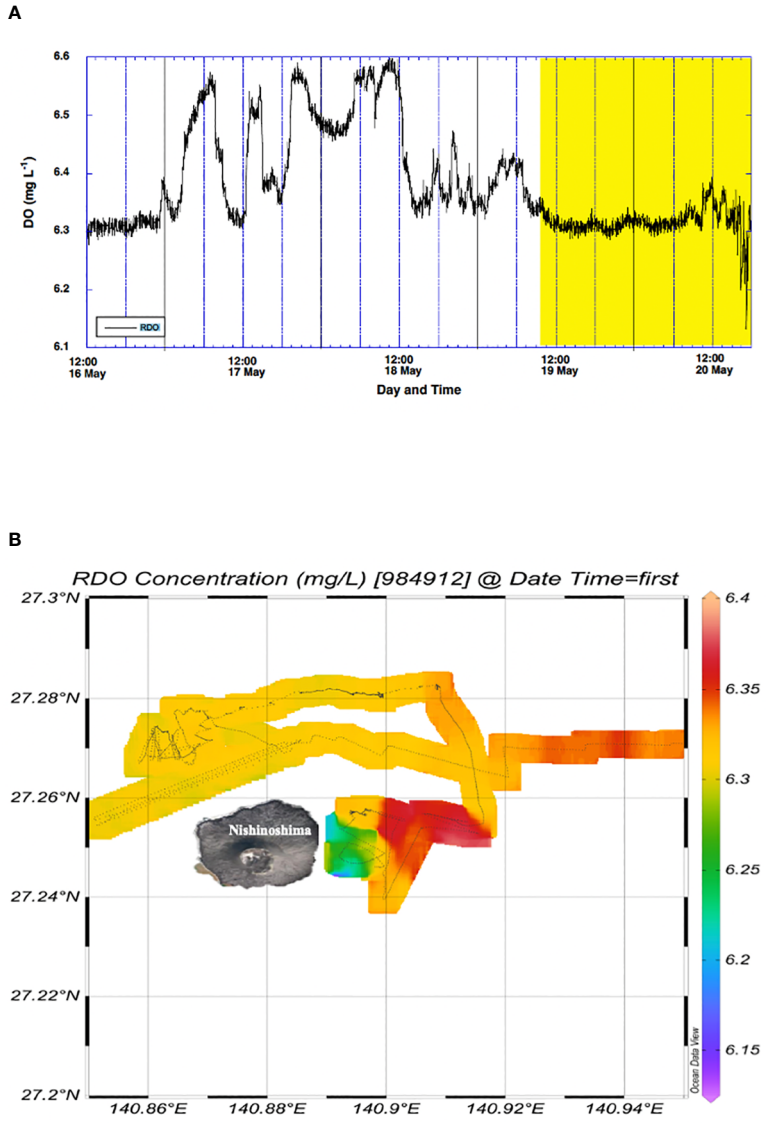


FIGURE 11
DO data **(A)** time-series variability west of 142°E. Yellow hatched area denotes the period when the USV-BB stayed near Nishinoshima. **(B)** horizontal distribution near Nishinoshima.

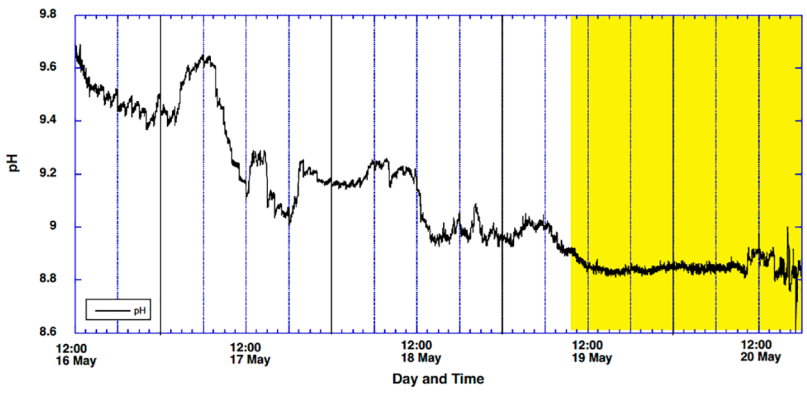


FIGURE 12
Time-series variability in pH west of 142°E. Yellow hatched area denotes the period when the USV-BB stayed near Nishinoshima.

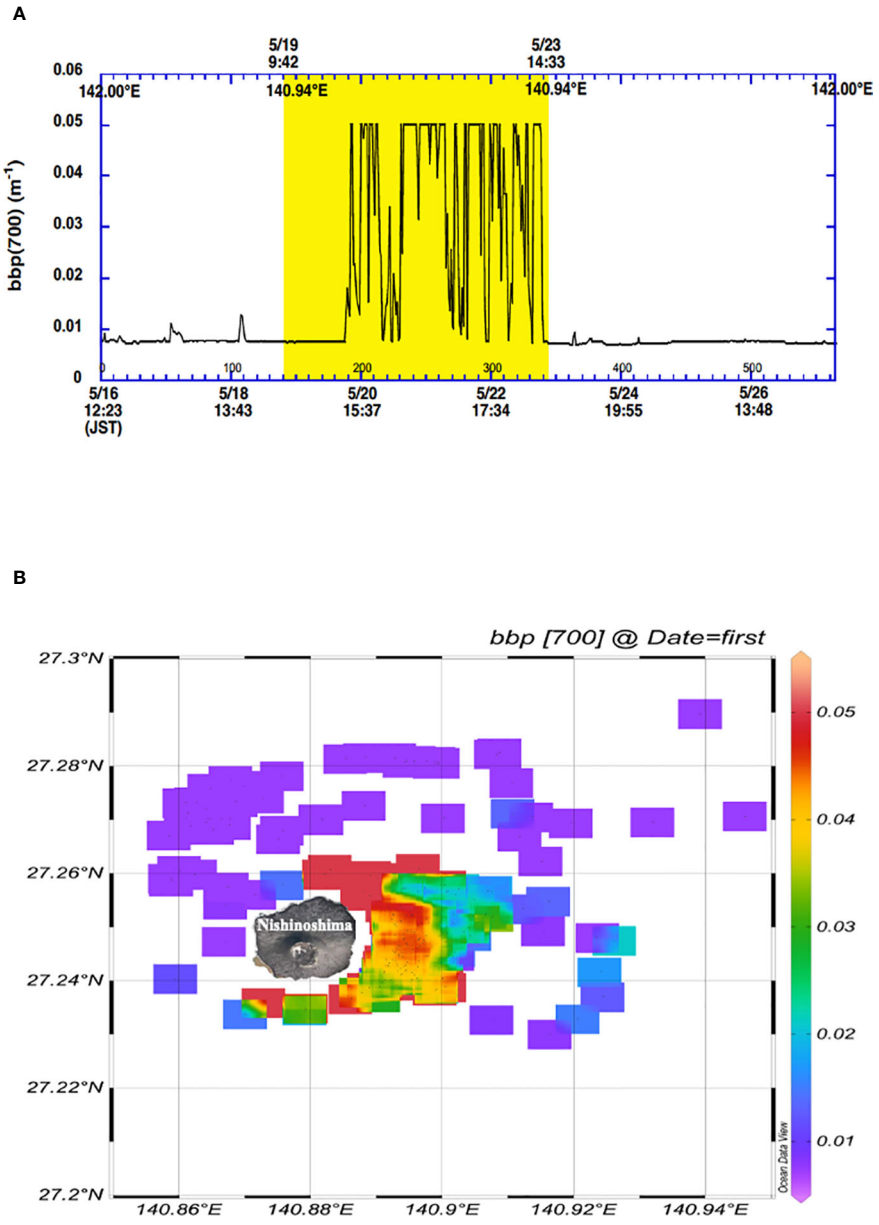


FIGURE 13
Backscatter data [$bbp_{(700)}$ (m^{-1})] **(A)** time-series variability west of 142°E. Yellow hatched area denotes the period when the USV-BB stayed near Nishinoshima shown in **(B)**. **(B)** horizontal distribution near Nishinoshima.

resulting in a total of 3506 photos throughout the mission. These images, with timestamps in Japan Standard Time (JST, UTC+9), offer a detailed visual log of the marine environment during this period. Figures 14A-D shows varying oceanic conditions: Figures 14A, B reveal possibly marine debris (plastic bottle) and leaves floating on the sea surface, respectively. Figures 14C, D both show the effect of time on the images. Figure 14C, taken at 8:31 a.m. after dawn, shows sunlight reflecting off the surface of the water, reducing visibility. Figure 14D, taken around 4:00 a.m. before dawn, is mostly a noise-like image.

Figures 14E, F, derived from Bluebottle's 360-degree camera, further illustrates these observations. Figure 14E shows Nishinoshima Island and its fumarolic gases, with the surrounding seawater appearing in its normal blue state. In contrast, Figure 14F, taken closer to the island, clearly shows the Bluebottle navigating through discolored waters, reinforcing the impact of volcanic activity on water coloration. This photographic evidence, cross-referenced with vessel data including location, heading, and gyroscopic measurements, provides a comprehensive view of the environmental changes encountered. The variation in water color, from the typical oceanic blue to a murky appearance near Nishinoshima Island, underscores the significant influence of volcanic activity on marine conditions.

The utilization of the Brinno Timelapse Camera and Bluebottle's 360-degree camera system provided unprecedented visual data on marine conditions, enabling a detailed analysis of

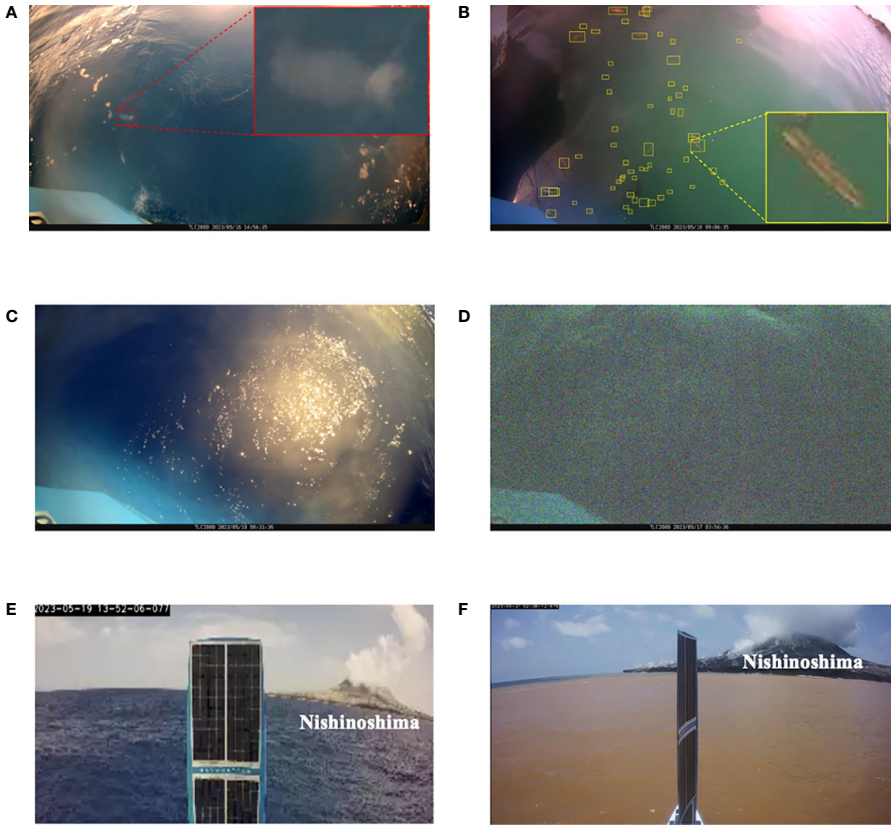


FIGURE 14
Downward facing camera and Bluebottle's 360-degree camera images.

changes over time. The timelapse photography allowed for continuous monitoring of the marine environment, capturing dynamic processes and transient events that traditional observation methods might miss.

3.2.5 Implications for marine volcano observation with USV

The USV conducted five days of observations at shallower than 500 m water depth around Nishinoshima Island, the first time after the big eruption of Nishinoshima in June 2020. The single-beam bathymetric data around Nishinoshima Island reveal the presence of a steep slope with an inclination of approximately 40 degrees on the eastern flank (Figure 7). This is significantly larger than the inclinations observed on the north side and is consistent with the major growth of Nishinoshima Island on the north side of the crater during the large-scale eruptions from December 2019 to June 2020 (Maeno et al., 2021). The clear reflective surfaces and their depth changes in the sub-bottom profile data (Figure 8) are likely indicative of the distribution of volcanic sediments. Although these volcanological interpretations will be left for subsequent papers, these results hold important implications for understanding the formation of Nishinoshima volcano. Additionally, the continuous day-time observations of fumarolic activity obtained through 360-degree time-lapse photography (Figure 14) serve as evidence that our observations have realized real-time monitoring of Nishinoshima volcano.

pH anomalies particularly on the east side near Nishinoshima Island (Figure 12) and the color anomalies in seawater observed through camera images (Figure 14) suggest seawater abnormalities associated with the activity of Nishinoshima volcano. Besides, the observed submesoscale variations in water temperature, salinity, and current around Nishinoshima Island may be influenced by the presence of this massive volcano, which extends from 4000 m below the sea floor to above the sea surface, and potentially by volcanic activities. The submesoscale ocean variations can have impacts on the atmospheric boundary layer. Clarifying phenomena that involve interactions among the solid Earth, oceans, and atmosphere, as observed in this comprehensive USV-based observation, hold great significance.

4 Concluding remarks

Our preliminary investigation around the volcanic island evinced that the employment of USV presents a significant advantage in marine research, particularly in hazardous or inaccessible areas, although we have to overcome some operational problems, such as increasing of electrical capacity and driving power of USV, to realize sustainable observations. Its ability

10.3389/fmars.2024.1406381

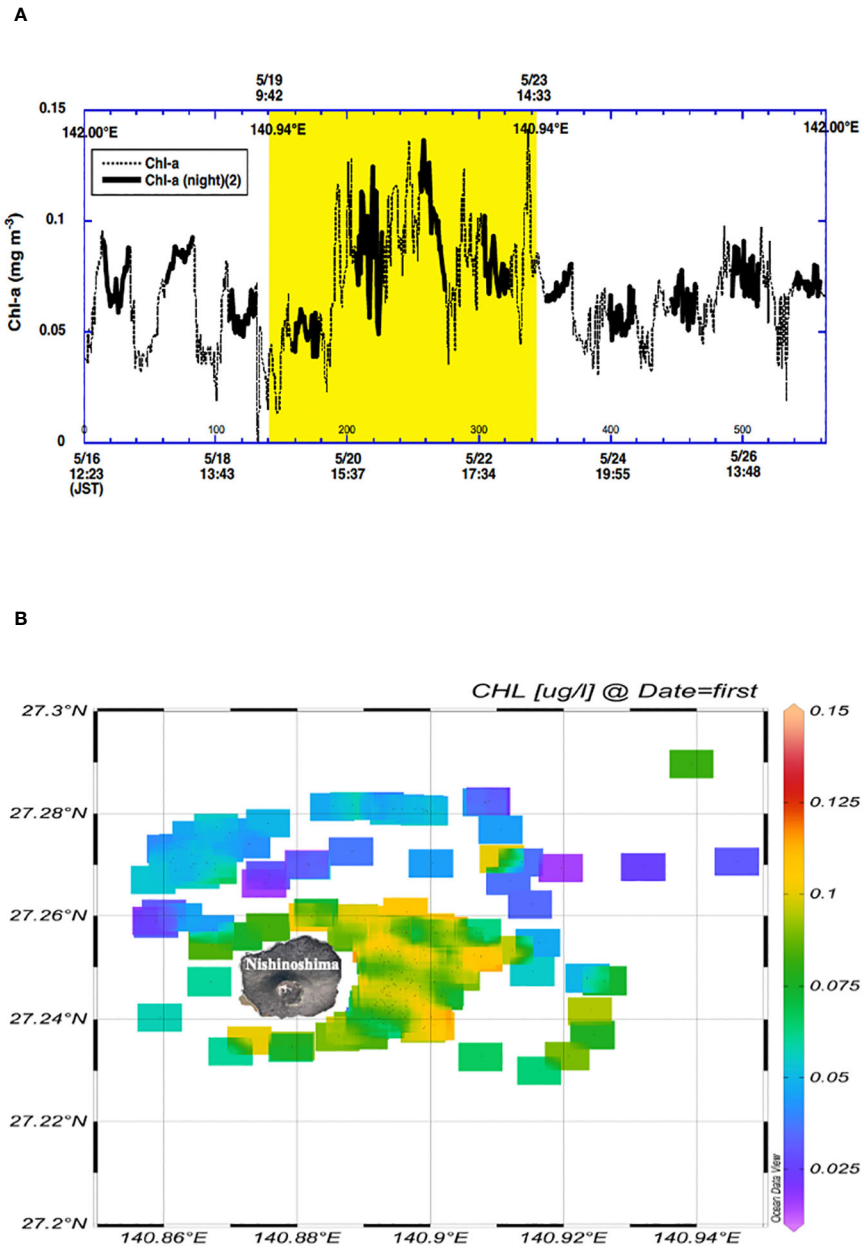


FIGURE 15
Chl-a data **(A)** time-series variability west of 142°E. Yellow hatched area denotes the period when the USV-BB stayed near Nishinoshima shown in **(B)**. Bold line shows Chl-a at night (19:00–04:00) without occurrence of quenching at day time. **(B)** horizontal distribution near Nishinoshima.

to navigate autonomously and gather multiple data in real-time under challenging conditions, such as near volcanic activity zones or in turbulent weather, minimizes risks to human researchers and reduces operational costs. Additionally, the integration of various sensors and data collection technologies, such as the single-beam echosounder for bathymetric data, and the oceanographic and meteorological sensors, despite their limitations, has provided a multi-dimensional view of the underwater landscape and marine conditions. In this study, we discussed the data acquired individually; however, moving forward, we plan to advance the understanding of various phenomena around Nishinoshima Island by integrating and analyzing multiple datasets in conjunction.

Data availability statement

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

Author contributions

NT: Writing – original draft, Writing – review & editing, Project administration, Validation. AN: Writing – original draft, Writing – review & editing, Validation, Visualization. ST: Writing – review & editing, Project administration. HI: Writing – original draft,

Writing – review & editing, Validation. DS: Writing – review & editing, Project administration. DM: Writing – original draft, Writing – review & editing, Validation. MI: Writing – review & editing, Project administration. TK: Writing – review & editing. MH: Writing – original draft, Writing – review & editing, Project administration, Validation, Visualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (Grant Nos. JP20H02236).

Acknowledgments

We extend our sincere gratitude to the OCIUS for their technology and acceptance of our challenging mission, and the Chichijima Fisheries Cooperative, Ogasawara Village Office, Tokyo Metropolitan Ogasawara Island Branch Office, OCIUS, the Ministry of Land, Infrastructure, Transport and Tourism, the Japan Coast Guard, and people of Chichijima Island for their invaluable support and assistance to conduct the Bluebottle observation. We would like to express our sincere gratitude to Dr. Toshiya Fujiwara for his invaluable advices and discussions for this study. This study was supported by ERI JURP 2024-W-03 in Earthquake Research Institute, the University of Tokyo. We are also grateful to the two reviewers for their insightful comments and suggestions, which significantly improved 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1406381/full#supplementary-material

References

Ando, K., Lin, X., Villanoy, C., Danchenkov, M., Lee, J.-H., He, H.-J., et al. (2021). Half century of scientific advancements since the cooperative study of the kuroshio and adjacent regions (CSK) program –need for a new kuroshio program. *Prog. Oceanogr* 193, 102513. doi: 10.1016/j.pocean.2021.102513

de Ronde, C. E. J., and Stucker, V. K. (2015). "Seafloor Hydrothermal Venting at Volcanic Arcs and 81 Backarcs," in The Encyclopedia of Volcanoes, *Chapter 47*, (Netherlands: Elsevier) 823–849.

Fujiki, T. (2019). *MR19-02 Preliminary Cruise Report* (Yokosuka, Japan: JAMSTEC). doi: 10.17596/0002620

Honda, M. C. (2020). Effective vertical transport of particulate organic carbon in the western north pacific subarctic region. *Front. Earth Sci.* 8. doi: 10.3389/feart.2020.00366

Honda, M. C., Wakita, M., Matsumoto, K., Fujiki, T., Siswanto, E., Sasaoka, K., et al. (2017). Comparison of carbon cycle between the western Pacific subarctic and subtropical time-series stations: highlights of the K2S1 project. *J. Oceanography*. 73:647–667. doi: 10.1007/s10872-017-0423-3

Iguchi, M., Nakamichi, H., Tanaka, H., Ohta, Y., Shimizu, A., and Miki, D. (2019). Integrated monitoring of volcanic ash and forecasting at sakurajima volcano, Japan. *J. Disaster Res.* 14, 798–809.

Inoue, R., Suga, T., Kouketsu, S., Kita, T., Hosoda, S., Kobayashi, T., et al. (2016). Western North Pacific Integrated Physical-Biogeochemical Ocean Observation Experiment (INBOX): Part 1. Specifications and chronology of the S1-INBOX floats. *J. Mar. Res.* 74, 43–69.

Japan Meteorological Agency (2023). Available online at: https://www.data.jma.go.jp/yoho/data/typhoon/T2302.pdf.

Johnston, T. M. S., Schönau, M. C., Paluszkiewicz, T., MacKinnon, J. A., Arbic, B. K., Colin, P. L., et al. (2019). Flow Encountering Abrupt Topography (FLEAT): A multiscale observational and modeling program to understand how topography affects flows in the western North Pacific. *Oceanography* 32, 10–21. doi: 10.5670/oceanog.2019.407

Kitamura, M. (2022). *MR21-06 Preliminary Cruise Report* (Yokosuka, Japan: JAMSTEC). doi: 10.17596/0002532

Maeno, F., Kaneko, T., Ichihara, M., Suzuki, Y. J., Yasuda, A., Nishida, K., et al. (2022). Seawater-magma interactions sustained the high column during the 2021 phreatomagmatic eruption of Fukutoku-Oka-no-Ba. *Commun. Earth Environ.* 3, 260. doi: 10.1038/s43247-022-00594-4

Maeno, F., Yasuda, A., Hokanishi, N., Kaneko, T., Tamura, Y., Yoshimoto, M., et al. (2021). Intermittent growth of a newly-born volcanic island and its feeding system revealed by geological and geochemical monitoring 2013-2020, nishinoshima, ogasawara, Japan. *Front. Earth Sci.* 9. doi: 10.3389/feart.2021.773819

Matoza, R. S., Fee, D., Assink, J. D., Iezzi, A. M., Green, D. N., Kim, K., et al. (2022). Atmospheric waves and global seismoacoustic observations of the January 2022 Hunga eruption, Tonga. *Science* 377, 95–100. doi: 10.1126/science.abo7063

Nagano, A., Hasegawa, T., Ariyoshi, K., Iinuma, T., Fukuda, T., Fujii, N., et al. (2022). USV-Observed Turbulent Heat Flux Induced by Late Spring Cold Dry Air Incursion over Sub-Mesoscale Warm Regions off Sanriku, Japan. *Sensors* 22, 9695. doi: 10.3390/s22249695

Nagano, A., Suga, T., Kawai, Y., Wakita, M., Uehara, K., and Taniguchi, K. (2016). Ventilation revealed by the observation of dissolved oxygen concentration south of the Kuroshio Extension during 2012–2013. *J. Oceanogr.* 72, 837–850. doi: 10.1007/s10872-016-0386-9

NOAA NGDC. (2009). ETOPO1 1 Arc-minute global relief model. doi: 10.7289/V5C8276M

Patterson, R. G., Lawson, E., Udyawer, V., Brassington, G. B., Groom, R. A., and Campbell, H. A. (2022). Uncrewed surface vessel technological diffusion depends on cross-sectoral investment in open-ocean archetypes: A systematic review of USV applications and drivers. *Front. Mar. Sci.* 8:736984. doi: 10.3389/fmars.2021.736984

Proud, S. R., Prata, A. T., and Schmauss, S. (2022). The January 2022 eruption of Hunga Tonga-Hunga Ha'apai volcano reached the mesosphere. *Science* 378, 554–557. doi: 10.1126/science.abo4076

Sato, I., Ono, T., Morishita, T., Hamasaki, S., Takahashi, H., and Nogami, K. (2018). Temporal change in chemical composition of seawater around Nishinoshima volcano according to volcanic activity. *Rep. Hydrographic Oceanographic Res.* 55, 101–109.

Saunders, P. M. (1967). The temperature at the ocean-air interface. *J. Atmos. Sci.* 24, 269–273. doi: 10.1175/1520-0469(1967)024<0269:TTATOA>2.0.CO;2

Takai, S. (2024). Geochemical study of seawater and hydrothermal fluids obtained around volcanic edifices in the Japan waters. Tokyo University of Marine Science and Technology, 85.

Taketani, F. (2021). *Preliminary Cruise Report MR21-01* (Yokosuka, Japan: JAMSTEC). doi: 10.17596/0002527

Yoshida, K., Tamura, Y., Sato, T., Hanyu, T., Usui, Y., Chang, Q., et al. (2022). Variety of the drift pumice clasts from the 2021 Fukutoku-Oka-no-Ba eruption, Japan. *Island Arc* 31, e12441. doi: 10.1111/iar.12441

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 22 August 2024
DOI 10.3389/fmars.2024.1386267

OPEN ACCESS

EDITED BY
Xinyu Zhang,
Dalian Maritime University, China

REVIEWED BY
Xin He,
Naval Aviation University, China
Xinyi Xie,
Zhejiang University, China

*CORRESPONDENCE
Christopher J. Peck
cjpeck04@gmail.com

RECEIVED 14 February 2024
ACCEPTED 08 July 2024
PUBLISHED 22 August 2024

CITATION
Peck CJ, Langedock K, Boone W, Fourie F, Moulaert I, Semeraro A, Sterckx T, Geldhof R, Groenendaal B and Ponsoni L (2024) The use of autonomous underwater vehicles for monitoring aquaculture setups in a high-energy shallow water environment: case study Belgian North Sea.
Front. Mar. Sci. 11:1386267.
doi: 10.3389/fmars.2024.1386267

The use of autonomous underwater vehicles for monitoring aquaculture setups in a high-energy shallow water environment: case study Belgian North Sea

Christopher J. Peck[1]*, Kobus Langedock[1], Wieter Boone[1], Fred Fourie[1], Ine Moulaert[1], Alexia Semeraro[2], Tomas Sterckx[3], Ruben Geldhof[4], Bert Groenendaal[5] and Leandro Ponsoni[1]

[1]Flanders Marine Institute (VLIZ), Ostend, Belgium, [2]Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Ostend, Belgium, [3]DEME Group, Zwijndrecht, Belgium, [4]Jan De Nul, Aalst, Belgium, [5]Sioen Industries NV, Ardooie, Belgium

Effective and frequent inspections are crucial for understanding the ecological and structural health of aquaculture setups. Monitoring in turbid, shallow, and dynamic environments can be time-intensive, expensive, and with a certain level of risk. The use of monitoring techniques based on autonomous vehicles is an attractive alternative approach because these vehicles are becoming easier to use, cheaper and more apt to carry different sensors. In this study, we used an Autonomous Underwater Vehicle (AUV) equipped with interferometric side scan sonar to observe an aquaculture setup in the Belgain North Sea. The surveys provided information on the longlines and indicated that the mussel dropper lines touched the seabed, implying that mussel growth weighed the longlines down. The side scan imagery also captured significant scouring around the longline anchors and localized debris on the seabed, which is important information to ensure the long-term sustainability of the setup and impact on the seabed. The results show that observing mussel longlines in a turbid, shallow, and high-energy environment using an AUV is a viable technique that can provide valuable information. Thus, the present study provides key insights into the application of innovative uncrewed monitoring techniques and forms an important step towards efficient and sustainable management of offshore aquaculture setups.

KEYWORDS

autonomous underwater vehicles, side scan sonar, mussel aquaculture setup, shallow high-energy environment, Belgian North Sea

1 Introduction

Aquaculture setups for mussels cultivation are utilized globally, accounting for 94% of the world's mussel production (Avdelas et al., 2021). In addition to their role in food production, and within a context of increasing threat of coastal zones by climate change-related processes, mussel aquaculture setups have also been explored as a strategy to kickstart and sustain mussel reefs on the seafloor (Goedefroo et al., 2022; Boulenger et al., 2024), offering a nature-based solution (Krull et al., 2015; Seddon et al., 2020; Van der Meulen et al., 2023) for coastal protection (Murray et al., 2002; Borsje et al., 2011; Temmerman et al., 2013; Walles et al., 2016). The rationale is that mussel reefs at the seafloor show potential in providing coastal protection by trapping and stabilizing sediments, diminishing wave impact, and alleviating the effects of sea level rise (Goedefroo et al., 2022; Ells and Muarry 2012; Koch et al., 2009; Boulenger et al., 2024). In both scenarios, whether deployed as a food production system or a nature-based solution, these setups necessitate regular monitoring to assess their structural integrity, the health status of mussels, and their environmental impact.

Traditionally, mussel aquaculture monitoring has been predominantly conducted manually (Bao et al., 2020), often involving divers (Lowry et al., 2014; Hicks et al., 2015; Ali et al., 2022), which is labor intensive and carries an amount of risk. Diving efforts are also restricted by environmental conditions such as visibility, strong currents and waves, leading to limited coverage. With increasing investments and automation in the food industry, aquaculture has become one of the fastest-growing sectors of food production globally (Allison, 2011) and consequently, monitoring techniques are continuously evolving and adapting to the specific characteristics of individual aquaculture setups (e.g., species, covered area, depth, proximity to the coast, accessibility) and the environmental conditions in which they operate. In this direction, efforts to minimize human intervention in sampling approaches for mussel aquaculture setups and seafloor reefs have increased, aided by aerial (e.g., Barbosa et al., 2022) and marine uncrewed vehicles (Bao et al., 2020).

Commonly, aerial-based remote sampling relies on optical techniques (Massarelli et al., 2021), rendering them ineffective in turbid waters due to reduced visibility caused by suspended sediment and other particulate matter which interferes with optical signals (Zhao et al., 2018). In contrast, *in situ* measurements through marine uncrewed vehicles enable detailed monitoring of the aquaculture setups and underlying seabed. The use of marine robots with application to aquaculture monitoring has increased in the latest years through different approaches and employed different types of vehicles (see Ubina and Cheng (2022) for a review), such as Uncrewed Surface Vehicle (USV; e.g., Sousa et al., 2019), Remotely Operated Vehicle (ROV; e.g., Amundsen et al., 2021), or even through USV-ROV interactions (Osen et al., 2018). Although most of the use of autonomous vehicles in aquaculture was focused on the pisciculture industry, operations by USVs and ROVs would still be adaptable and applicable for mussel aquaculture and, therefore, eliminate the need for divers and associated risks. However, the spatial coverage and environmental conditions might still be an issue even if to a lesser extent. USV operations might be hampered by waves and currents and restricted at near-surface inspections. ROV dives have reduced spatial coverage and present the risk of entanglement between the umbilical cable and the mussels' long- and dropper lines. Autonomous Underwater Vehicles (AUVs) do not have these issues as there is no requirement for an umbilical cord and they are able to repeat pre-programmed missions (Wynn et al., 2014). AUVs have also been tested to autonomously inspect fish farm cages and the surrounding water quality (Karimanzira et al., 2014).

While optical methods encounter limitations across various sampling platforms (e.g., divers, aerial uncrewed vehicles, ROVs, and SUVs), AUVs are also well-suited for employing acoustic methods, such as side scan sonar (Wynn et al., 2014; McGeady et al., 2023). This technology finds extensive application in hydrographic (Mitchell and Somers, 1989; Ryant, 1975) and marine geological surveys (Johnson and Helferty, 1990; Greene et al., 1999), and benthic habitat monitoring (Brehmer et al., 2003; Marsden et al., 2023; Greene et al., 2018; Ali et al., 2022). Essentially, side scan sonar detects seabed objects and discerns sediment types, while also able to inspect water column structures like piers and bridge supports for signs of damage (Clausner and Pope, 1988; Murphy et al., 2011; Bryant 1975; Hou et al., 2022). A hull-mounted side scan sonar system was previously employed to monitor installations of a mussel aquaculture setup in the French Mediterranean, although clarity for detecting mussel dropper lines was occasionally limited (Brehmer et al., 2006). While often operated from vessels, side scan sonar surveys are increasingly being conducted by AUVs (Wynn et al., 2014). AUVs offer advantages over vessel-mounted or towed side scan sonar, as they can fly relatively close to the targets (seabed and mussel lines), enabling the collection of more tailored datasets.

However, the use of AUVs in shallow and turbid, high-energy environments poses challenges. The strong currents and waves in such environments can cause the AUVs to roll, significantly impacting the quality of the side scan sonar data. Furthermore, local current velocities may exceed the vehicle's speed (Wynn et al., 2014) and, therefore, considerably impact the sampling strategy.

This study aimed to investigate the hypothesis that *"AUVs equipped with interferometric side scan sonar can effectively monitor mussel aquaculture installations, including long- and dropper lines, anchoring systems, and the seabed beneath, in high-energy and turbid environments"*. To validate this hypothesis, the following questions will be addressed: 1) Can an AUV safely conduct surveys of aquaculture infrastructure in shallow and turbid, high-energy environments? 2) If so, which side scan sonar settings yield the highest quality data? 3) What relevant information can be ascertained about the aquaculture setup? 4) What pertinent information can be unraveled about the seabed surrounding the aquaculture setup?

This paper is organized as follows: Section 2 introduces the aquaculture setup and the study area (Section 2.1), along with the AUV survey strategy (Section 2.2) and methods applied on the data collection, processing, and analysis (Section 2.3). Section 3 presents the results and corresponding discussions, including side scan sonar settings (Section 3.1), outcomes from the mussel aquaculture setup

(Section 3.2), and seafloor inspections (Section 3.3). Lastly, Section 4 provides objective answers to the research questions and offers concluding remarks regarding the raised hypothesis.

2 Methods

2.1 Aquaculture setup and study area

As part of the Coastbusters 2.0 project, three mussel longlines were deployed in two different areas in the Belgian North Sea (Figure 1A), near the municipality of De Panne (Figure 1B). The site closer to the shore was considered sheltered, as it is adjacent to sand banks on the offshore side. The site further offshore is on the other side of the sand banks and is considered exposed (Figure 1B).

The longlines were spaced between 30 and 40 m apart at the sheltered site (Figure 1C) and between 20 and 50 m at the exposed site. Each longline was approximately 150 m long and consisted of a (near)surface line secured on the seabed by an anchor on either extremity. Two types of anchors were used at each site: a screw anchor, which was drilled directly into the seabed, and a block anchor with a chain attached to the end of the longline (Figure 2). Along the longlines, there were 36 three-meter-long mussel dropper lines, where mussel larvae could attach, spaced 1.5 m from each other. At the end of each dropper line, there was a concrete block weighing the lines down vertically in the water column. Buoys were attached to the longline in order to keep the longline occupied by dropper lines at the surface and not resting on the seabed. The mussel larvae were expected to settle on the dropper lines and metamorphose into the juvenile phase, allowing mussels to grow before dropping onto the seabed (Figure 2).

The seabed in the region is mainly composed of sand with fine to medium grain size (Degraer et al., 2000), with an area of muddy sand to the south of the sheltered site. At both sites, the mussel longlines were deployed at a depth of 5 m LAT (lowest astronomical tide). The tide in the area is semi-diurnal with a range of 3.5 m during the neap tides and more than 5 m during spring tides. The large tidal range is associated with strong tidal currents, with peaks exceeding 1 m s^{-1} in the nearshore area (Haerens et al., 2012).

Wind speed data from the Westinder weather and wave height data from the Trapegeer wave buoy (both from the Agency for Maritime Services and Coast, Flemish Government, 2023) were collected for the two survey years and months (Figure 3). The average prevailing wind and wave direction is southwesterly. However, because of the angle of the coast, the highest waves are generated when the wind is northwesterly (Fettweis et al., 2012).

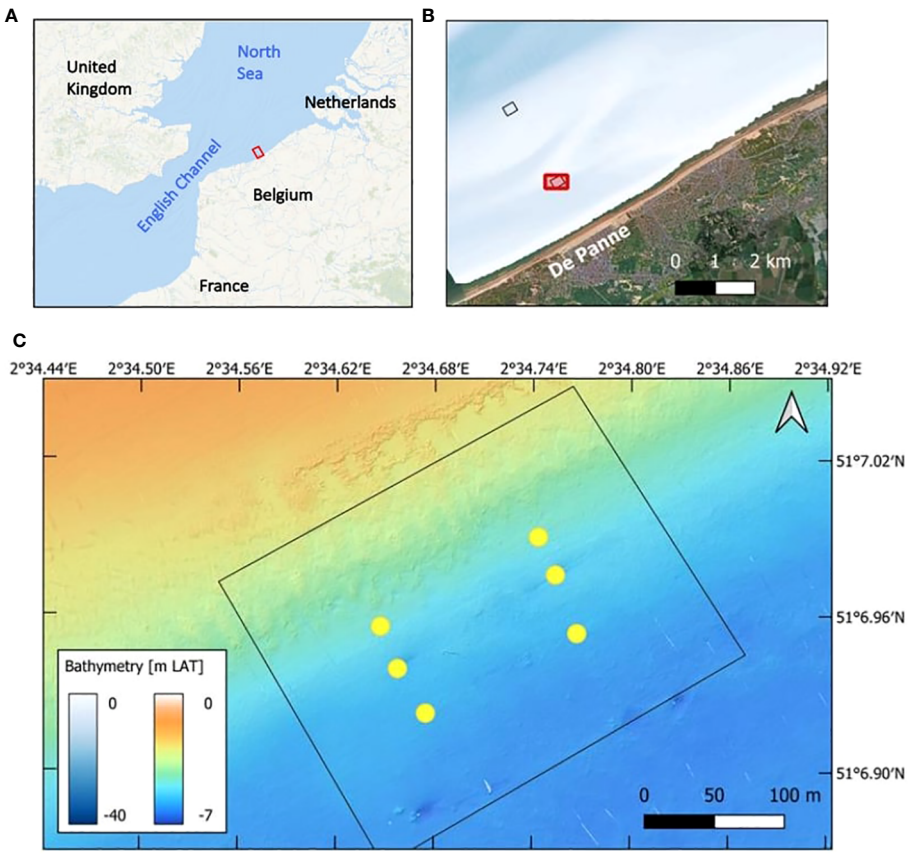


FIGURE 1
Location of aquaculture setup deployment sites off the coast of Belgium in the North Sea **(A)**, near the city of De Panne **(B)**. The longlines were deployed at an exposed site (black rectangle) and a sheltered site (red rectangle). Each side is composed of three long-lines **(C)** with 36 dropper lines, as represented in Figure 2. Base map: ESRI Ocean.

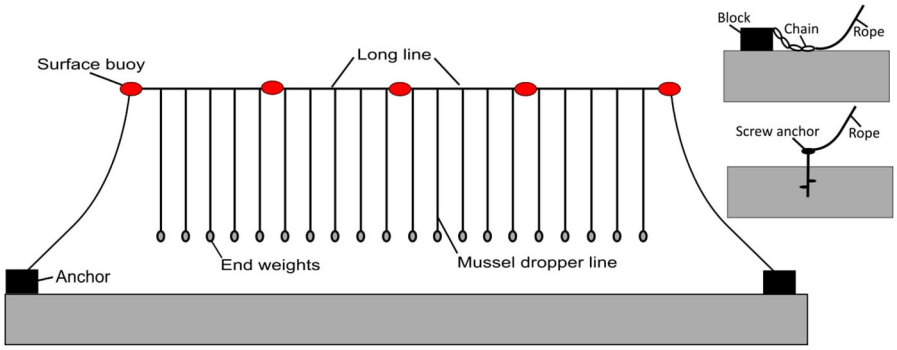


FIGURE 2
Sketch of the deployed structures in the context of the Coastbusters 2.0 project displaying longlines, surface buoys, dropper lines, end-weights, and anchors. Note that the sketch is not to scale; not all 36 dropper lines are represented in the figure. Three of these structures were deployed at each site.

The winds were similar between the two months with the majority of the wind coming from the southwest. However, October has stronger winds in all directions and has periods of stronger winds and higher waves (Figures 3C–F). More northwesterly winds occur during October (Figures 3C, E) while August has periods of easterly winds (Figure 3A).

2.2 AUV surveys

The AUV used in this research (AUV Barabas) is a commercial-off-the-shelf Teledyne Gavia modular vehicle (Figure 4A). For the surveys, it had a configuration of 2.87 m long, with a nose cone, a single battery pack module, and modules for navigation, control, surveying and propulsion. The AUV Barabas was equipped with an iXBlue Compact C5 inertial navigation system (INS) aided by a Pathfinder Doppler Velocity Log (DVL), European Geostationary Navigation Overlay Service-capable (EGNOS) Differential Global Positioning System (DGPS), and Keller pressure sensor (± 0.005 accuracy) (Figure 4A). The INS houses three fiber-optic gyros and three accelerometers. The data from the INS and the available aiding sensors are fed into a Kalman filter, which calculates a best guess of the position, speed, and altitude of the vehicle in all three dimensions, as well as their respective error estimates (Table 1). After starting the vehicle and prior to deployment, a continuous DGPS input is needed to align the INS. When the AUV Barabas is submerged, the combination of the Compact C5 INS and the Pathfinder DVL limits the increase in the position error to 0.02% of the distance travelled (CEP 50). Each time the AUV Barabas surfaces, the DGPS antenna will be able to obtain a new position fix, and the positioning error will decrease.

The speed of the AUV is vital to collect high quality side scan imagery, particularly in an environment with a strong current. The AUV's speed can be set either as a fixed revolutions per minute (RPM) for the propeller or a fixed speed over ground (SOG). With a set RPM, the AUV's SOG fluctuates with current speed, impacting data consistency. Alternatively, setting a fixed SOG allows real-time adjustment of propeller RPM for more consistent data, but affects maneuverability. Low RPM in current reduces maneuverability, while high RPM risks aggressive behavior and mission aborts. A minimum RPM of 500 and maximum of 1000 maintained stability, with an optimal SOG of 1.7 m s^{-1} in background currents up to 0.5 m s^{-1}. For background currents with speed between 0.5 and 0.8 m s^{-1}, the AUV started to respond more erratically to disturbing factors. For background flow between 0.8 and 1 m s^{-1}, the AUV was still able to track a sampling line, but occasionally aborted the mission in the turn between lines, since it was unable to generate enough speed for drastic maneuvers. Above 1 m s^{-1} currents, the increased risk of a mission abort and the reduced data made us decide not to deploy the AUV. Based on this empirical approach, the AUV Barabas sailed at a fixed speed over ground of 1.7 m s^{-1} for this study, ensuring optimal flying mode and fixed sampling resolution to facilitate post-processing of the side scan sonar mosaic (see Section 2.3). At the very rear of the AUV, a nozzle contains the single propeller and four individually controlled fins. The fins thus control roll (actively maintaining 0° roll), pitch (used for depth control) and heading (used for track keeping). The AUV is not equipped with actuators to move laterally.

The AUV Barabas was successfully deployed in three campaigns (August and October 2021 and October 2022) equipped with a Klein 3500 dual-frequency interferometric side scan sonar (see details in Section 2.3) by the research vessel Simon Stevin. Several surveys of varying lengths were conducted for each campaign. Since the AUV Barabas was operating in-between and around the mussel longlines in a high-energy environment, a few considerations needed to be taken for each survey. The surveys were conducted with the AUV Barabas flying aligned either against or with the prevailing tidal current to prevent sideways movements of the vehicle, since this would lead to poor side scan sonar imagery. As mentioned above, navigation with AUVs is known to be challenging as vehicles lose GPS signals when underwater, and the error of INS increases the longer a vehicle is underwater (Wynn et al., 2014;

10.3389/fmars.2024.1386267

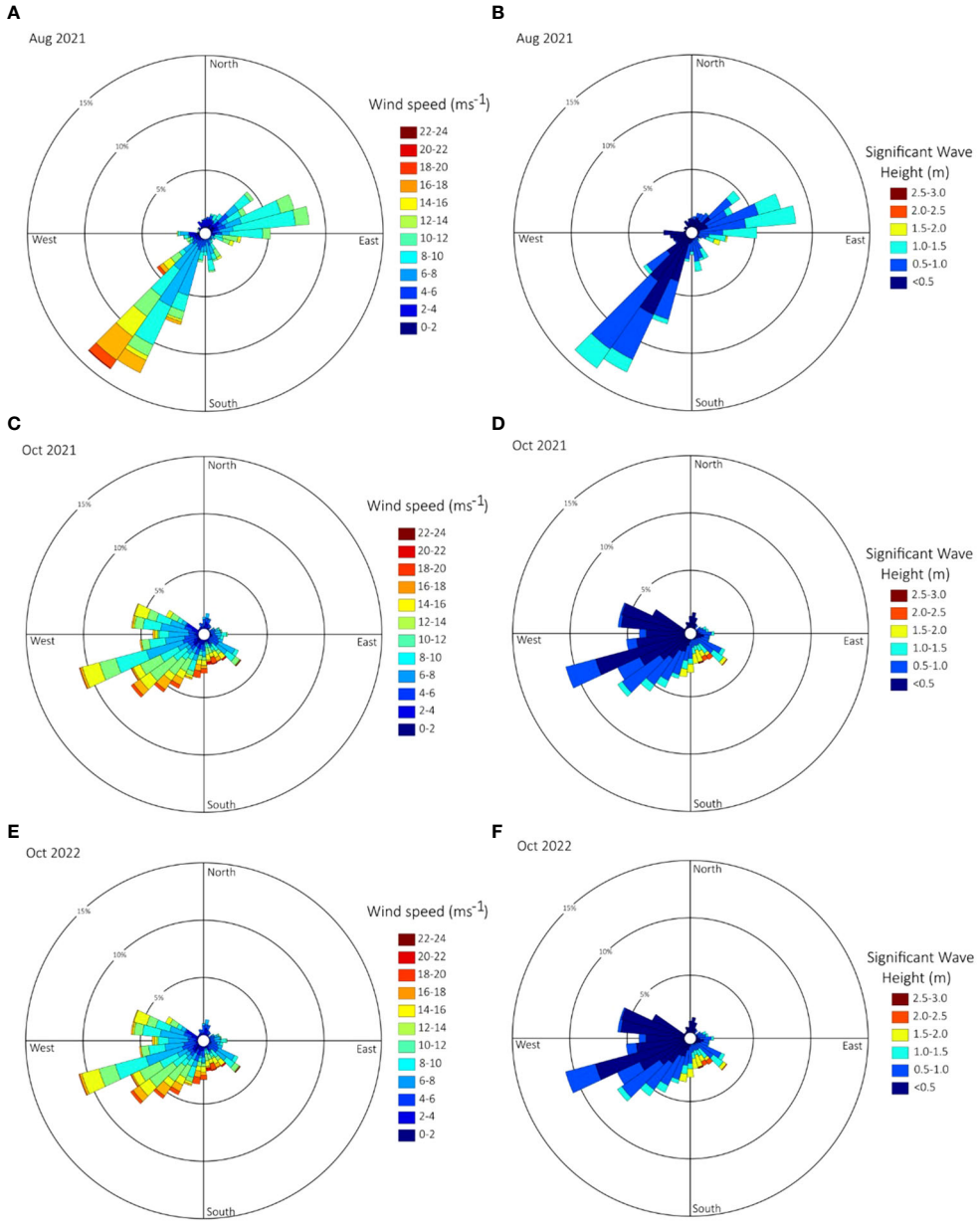


FIGURE 3
Wind speed and direction **(A, C, E)** from the Westinder weather station and significant wave height and direction **(B, D, F)** from the Trapegeer buoy for the year of 2021. Data was acquired from the Agency for Maritime Services and Coast, Flemish Government (2023).

Paull et al., 2018). To avoid collision with the longlines, the survey lines that passed closest to the obstacles were executed first when the position uncertainty was minimal. The mussel lines were installed parallel to the main current flow, but as the current turns with the tides, the mussel lines would curve between the anchors (either towards or away from the coast). This behavior was anticipated and taken into account when planning the survey tracks (Figure 4B). The track plan was relatively close to the longlines to ensure that they were captured in the side scan sonar swath (Figure 4C). However, the distance between the AUV Barabas and the longline would vary with the state of the tide and the ambient current.

2.3 Data collection, processing, and analyses

Side scan sonar operates by emitting a signal, commonly referred to as a "beam", which travels through the water column and reaches the seafloor or structures located on either side of the sampling platform. This platform is typically a towfish or vessel (Blondel, 2009). In our study, however, the side scan sonar was mounted on an AUV (e.g., Wynn et al., 2014). The AUV Barabas was equipped with a Klein 3500 dual-frequency interferometric side scan sonar. Figure 4A shows the positioning of the side scan sonar

10.3389/fmars.2024.1386267

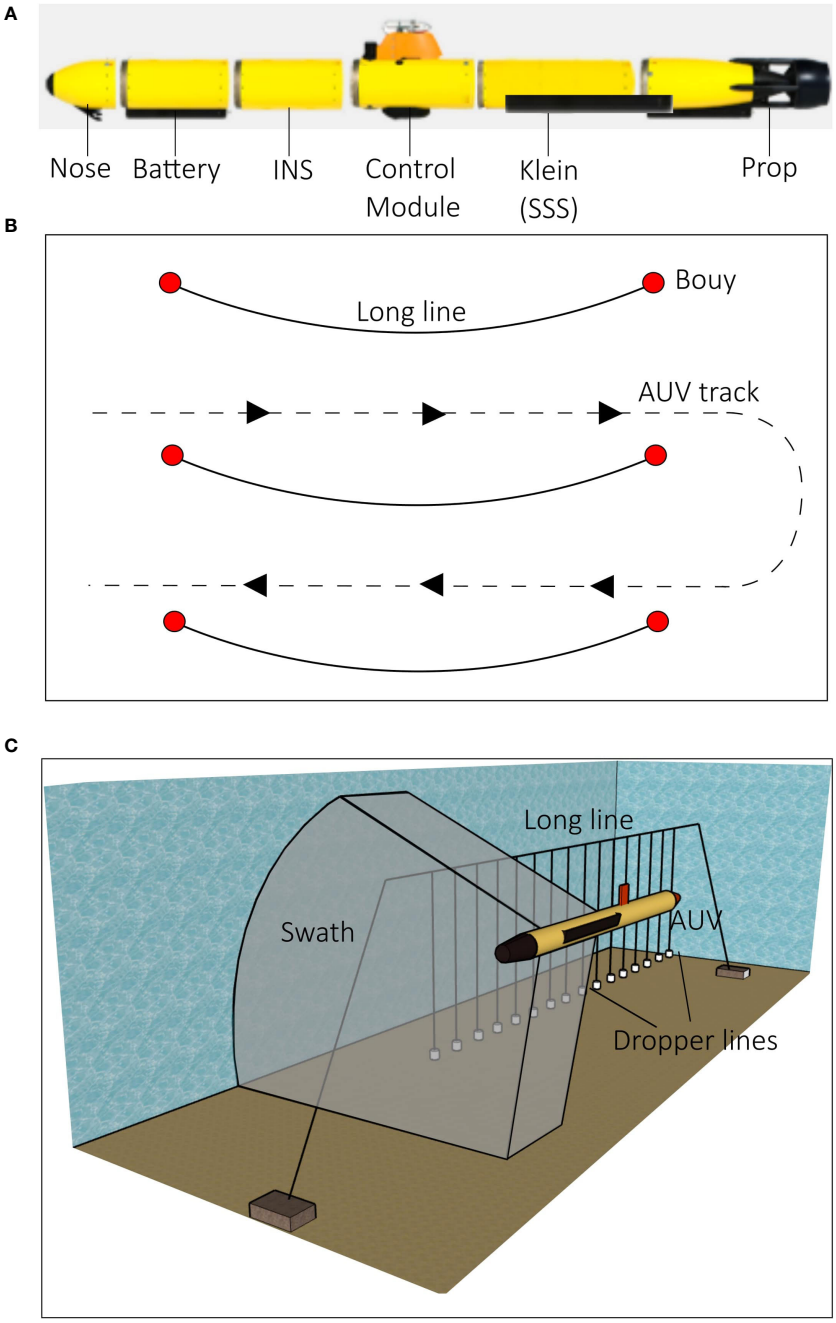


FIGURE 4
(A) Modules of the AUV Barabas used for the surveys, including the nose, battery, internal navigation system (INS), control module, Klein side scan sonar (SSS) module, and the prop. Typical survey pattern conducted by the AUV **(B)** and how it would fly next to the mussel longlines so that the side scan sonar could capture it **(C)**. The image is a stylized representation and not to scale.

system on the vehicle. Once the data was collected, it was displayed and processed using two different methods: waterfall and mosaics. The raw data was plotted as waterfall images using the SonarPro software (version 14.1, Klein Marine Systems, Inc). Waterfall images depict the intensity of the sonar signal return. While various methods exist for presenting waterfall images, in this study, distances are plotted along the x-axis, with increasing distance from the center of the figure. Additionally, successive data samples are represented along the y-axis, which corresponds to both time and distance, considering the sampling rate and the vehicle's cruising speed. Waterfall images enable the identification and measurement of features and contacts, as well as the acquisition of individual positions. However, due to the extensive number of samples, only a fraction of the data can be displayed. Additionally, these images cannot be integrated with other data sources due to the absence of georeferencing.

On the other hand, the mosaics are georeferenced images and were created with SonarWiz software (SonarWiz 7, Chesapeake

TABLE 1 Details and accuracy of AUV Barbara navigation equipment.

Navigation sensor/ input	Model	Standard deviation and Kahlman settings
INS	iX Blue Phins Compact C5	Heading accuracy calculated realtime. SD <0.1° during survey Roll and pitch accuracy calculated realtime SD < 0.002° during survey
DGPS	Septentrio AsteRx4	Position accuracy estimated realtime. During this survey approximately 5 m after surfacing, decreasing to minimum 0.36 m after continuous fixing (SD)
DVL	RDI 600 kHz Pathfinder	Fixed input value. SD of linear speed 0.05 m/s in every direction
Pressure sensor	Kalman PA-30X	Fixed input value. SD of depth 1 m (± 0.005 accuracy)

Technology, Inc), where movements of the AUV Barabas (pitch, roll, and yaw) are also compensated for. Depending on the quality and quantity of the images, building a mosaic can be labor intensive, requiring specialized software and some resolution can be lost. A mosaic can be exported in a wide range of data formats, allowing the data to be overlaid with other georeferenced data.

As a dual-frequency system, the sonar transducer emitted acoustic signals at two distinct frequencies: a relatively high frequency of 900 kHz and a relatively low frequency of 455 kHz. The high-frequency (900 kHz) beam has a horizontal opening angle of 0.34°, and the range (single side) can be set between 15 m and 75 m. The low-frequency (455 kHz) beam had a horizontal opening angle of 0.48°, and the range could be set between 30 m and 200 m. The optimal altitude above the seafloor is 10–15% of the sonar range, meaning that the operational range of the sonar will be limited by the local depth. The along-track sampling is inversely proportional to the AUV's sailing speed and sonar range, whereas the cross-track resolution of the side scan sonar's Compressed High Intensity Radar Pulse (CHIRP) relies on the sampling frequency. A variety of side scan sonar settings were tested during each survey to define which settings obtained the most useful data. The altitudes tested were 3 m and 5 m, with corresponding ranges of 30 m and 50 m. Both high-frequency and low-frequency data were collected simultaneously. Finally, pulse lengths of 2 ms and 1 ms were tested, with a respective sampling frequency of 31250 Hz and 62500 Hz, providing an across track resolution of 4.8 cm and 2.4 cm.

3 Results and discussion

3.1 Optimizing AUV surveys and side scan sonar settings

For this research, high and low-frequency transducer settings were used and analyzed both in a waterfall display and as mosaic images. The high frequency was used for the analysis as it provides the highest resolution of the data. The low-frequency setting was run at the same time and was used for redundancy.

The survey area is affected by wind and wave action, which impact the AUV Barabas and the quality of side scan sonar data when the vehicle is close to the sea surface. Therefore, the altitude of the AUV Barabas combined with the tidal height is critical. There were two main survey heights: 3 and 5 m from the seabed. At 5 m, roll artefacts were seen in both the waterfall and in the mosaic data (Figure 5A). The mosaic looked 'wavy' and had shading in places where there is no object to cast an acoustic shadow (Figure 5A). It is clear that the AUV Barabas was too close to the surface and was being excessively impacted by the surface waves. When the survey height was set at 3 m from the seabed, the roll artefacts were significantly reduced (Figure 5B). With the survey site being relatively shallow, even small surface waves (<0.5 m) will have more of an effect on the AUV Barabas, reducing the operational window in which the largest coverage and highest quality of data can be obtained. The altitude of the AUV Barabas affects the maximum distance that can be imaged by the side scan sonar, which is also known as the range. In general, a lower range provides a higher image resolution but covers a smaller area of the seabed. A range of 10 times the altitude was found to cover the largest area while providing the highest resolution (Flemming, 1976). In the three campaigns, two ranges were used: 30 m and 50 m. Since 3 m altitude produced better quality data, it became the predominant survey altitude and, therefore, 30 m became the predominant acoustic range used.

Throughout the campaigns, different surveys used a pulse length of either 2 ms (31250 Hz sampling frequency) or 1 ms (62500 Hz sampling frequency), as shown in Figure 6. We found the higher across track resolution of the 1 ms pulse length to provide a clearer image when analyzing the objects on the seabed in the waterfall, such as an anchor (Figure 6A, B). However, when using the mosaic format, we found that the higher resolution provided little added value, and the contrast between the seabed and objects on the seabed decreased (Figure 6C). Therefore, a pulse length of 2 ms is better suited for mosaics, as it creates a clearer image of the seabed and provides better contrast between objects (Figure 6D). This could be clarified by the stretching of pixels when georeferencing: at a sonar range of 30 m and sailing at 1.7 m s^{-1}, the along-track sampling resolution is 6.8 cm, which is closer to the across-track resolution of the 2 ms pulse (4.8 cm) than to that of the 1 ms pulse (2.4 cm). When using side scan sonar mosaics and waterfall together, the choice of pulse length depends on the specific mission objectives.

3.2 Aquaculture setup inspection

The most data was collected from the sheltered site and is only data analyzed in the following section. The images captured by the side scan sonar on the AUV Barabas clearly show the mussel longlines in both the waterfall and the mosaics, even when the longlines are above the transducer and sometimes at the surface of the water column. When compiling the survey images together into a mosaic, it is possible to see all three longlines (Figures 7, 8). Using

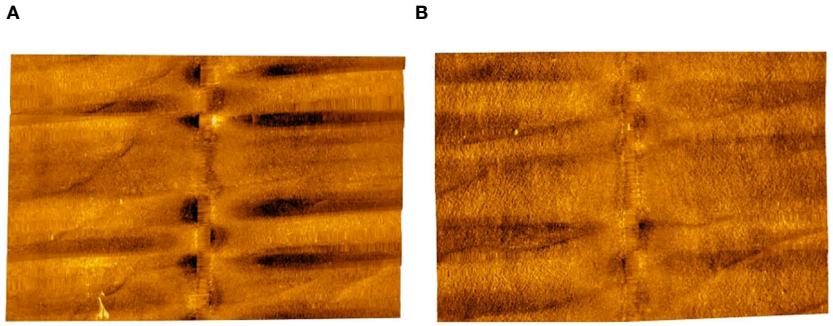


FIGURE 5
Images showing the difference in the quality of side scan sonar data at a relatively high altitude of 5 m **(A)** and low altitude of 3 m from the seabed **(B)**. The side scan sonar data settings were as follows: frequency 950 kHz, altitude 5 m **(A)** and 3 m **(B)**, range 50 m **(A)** and 30 m **(B)**, pulse length 2 ms. Both images have had the same degree of processing applied.

this technique, a surveyor can confirm that the lines and their anchors are still intact and assess how the lines are being impacted by the local currents. An example of this is the distinctive curve of the longlines as they are pushed in the direction of the tide, curving towards the shore during the flood current or away from the shore during the ebb current (Figures 7, 8).

In addition to the mosaics, the waterfall images show the individual dropper lines as the AUV Barabas flies parallel to the

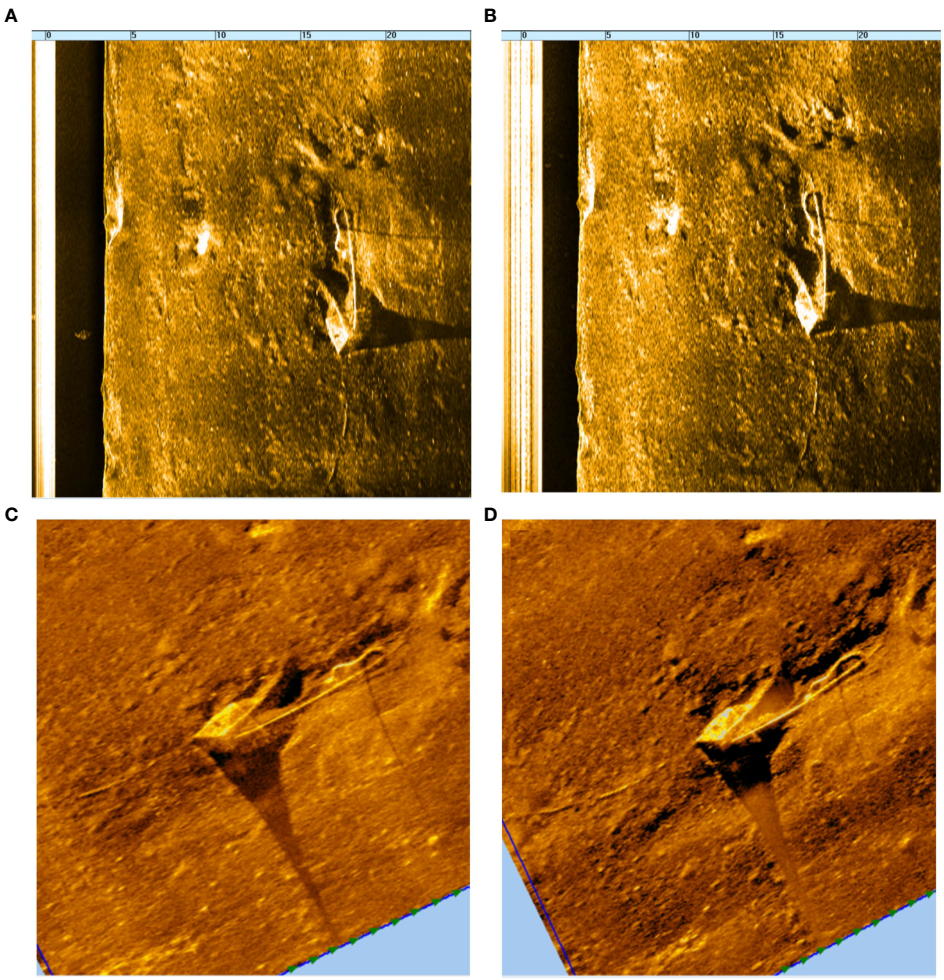


FIGURE 6
Waterfall and mosaic images of an anchor on the seafloor comparing the two pulse rates used 1 ms **(A, C)** and 2 ms **(B, D)**. The other side scan sonar data settings were as followed: Frequency 950 kHz, Altitude 5 m, Range 30m.

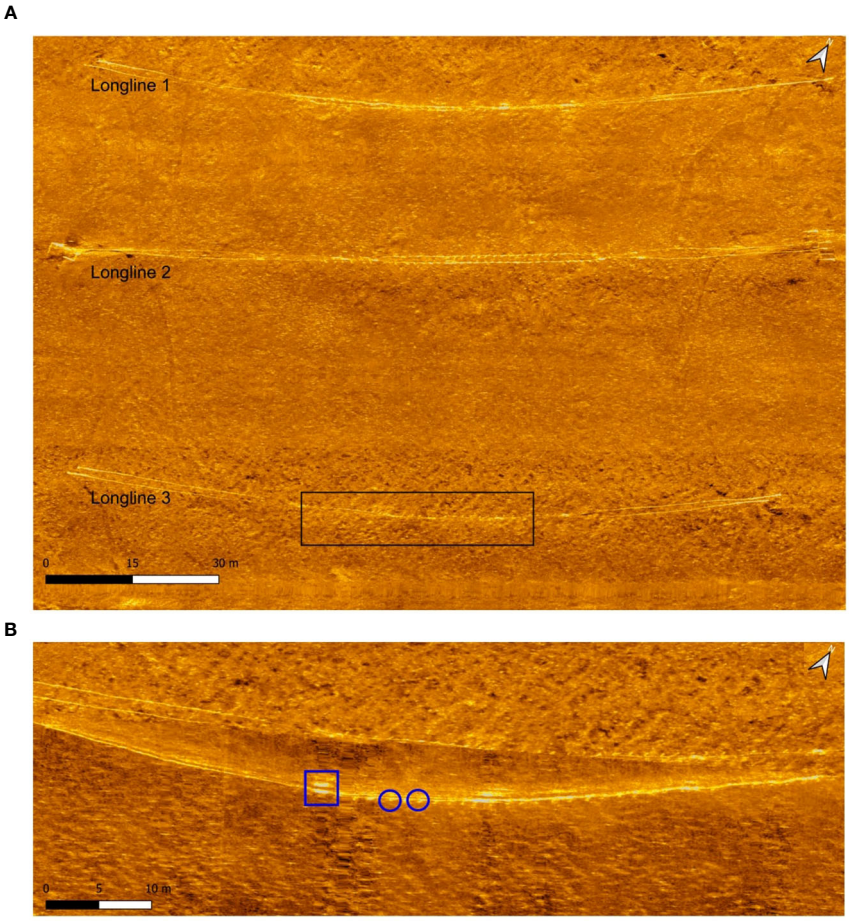


FIGURE 7
Side scan sonar mosaic of the sheltered site conducted in August 2021, with the three mussel longlines clearly visible **(A)** and the individual mussel dropper lines **(B)**. Some of the individual dropper lines are highlighted in blue, with the square indicating a surface buoy on the longline. The side scan sonar data settings were as follows: frequency 950 kHz, altitude 3 m, range 30 m, pulse length 2 ms.

lines, and if the dropper lines are not visible, the acoustic shadows often appear (Figure 9). The shadows not only show that the mussel dropper lines are intact, but by using SonarPro and measuring the distance between the start of the shadow and the dropper line, one can determine their height above the seabed (Figure 9). From analyzing the waterfall image (Figure 9), there are three places where the shadow of the longline or the dropper lines touch the respective reflections, indicating that the objects are on the seabed. Looking in further detail at the mussel dropper lines at the top of the image, there are nine on the seabed. The dropper lines then quickly rise off the seabed with an increase of 1.2 m between dropper lines 2 and 3 (Figure 9).

Using the AUV Barabas to conduct repeated surveys provides insight into how the longlines might change over time. The August 2021 survey (Figure 7B) was able to pick out the buoys on the line and the dropper lines. When the entire dropper line is at approximately the same distance from the transducer, the dropper lines are displayed as dots, highlighted in blue, similar to a top-side view. The same survey was conducted in October 2021 (Figure 8B), and in this survey, some dropper lines were visibly resting on the seabed. The two surveys were conducted at similar states of tide (Table 2), so this could indicate that the mussel dropper lines were heavier in October than in August, suggesting there had been mussel growth compatible with the relatively higher mussel growth and ventral thickness of the shells expected at this time of the year (Nagarajan et al., 2006). The dropper lines on longline 3 were clearly dragged along the seabed, as this longline does not have a uniform curve like lines 1 and 2 (Figure 8A). With the mussel dropper lines resting on the seabed, they are at risk of entanglement (Figure 8B). They add extra stress on the longline and will also impact the surrounding seabed as the dropper lines are dragged back and forth with the tide. From an aquaculture perspective, the dropper lines should avoid touching the seabed, as the quality of the mussels will be impacted. It is well known that mussels suspended in the water column produce a higher yield, as they are able to feed constantly and require less cleaning to remove sand and grit before consumption (Cheong and Lee, 1984). From a reef building perspective (as is the case for the Coastbusters 2.0 project), dropper lines on the seafloor could hamper the creation of

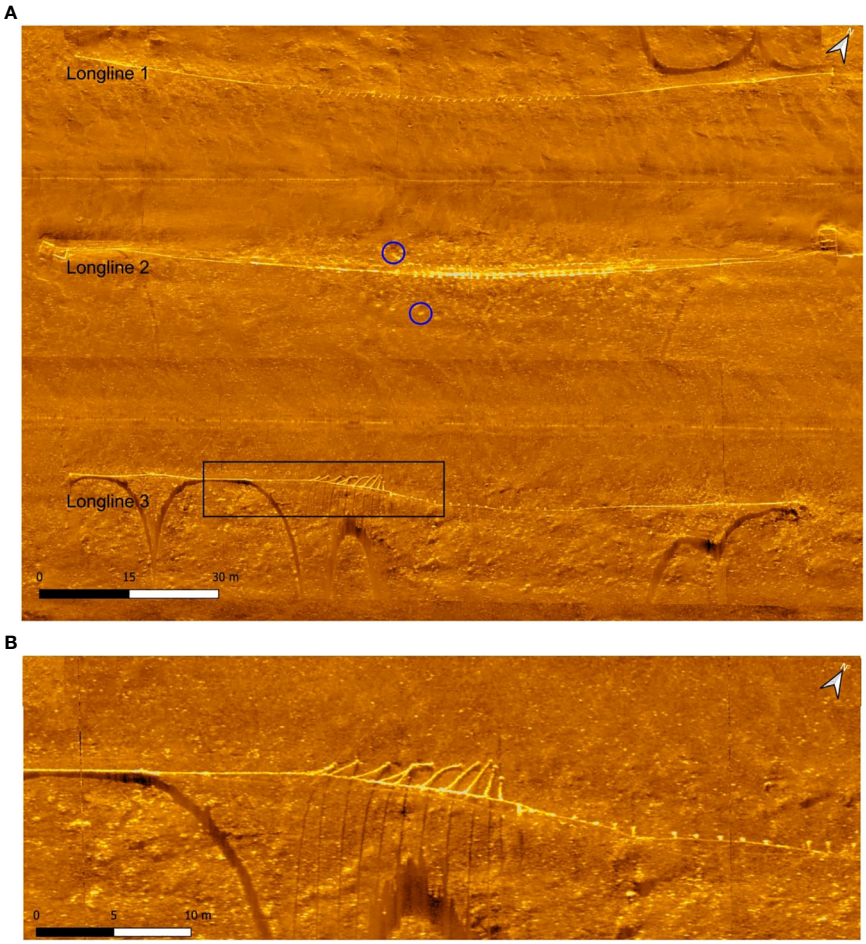


FIGURE 8
Side scan sonar mosaic of the sheltered site conducted in October 2021, with the three mussel longlines clearly visible. Potential contacts of mussels on the seafloor are highlighted in blue **(A)**. Individual mussel dropper lines resting on the seabed **(B)**. The side scan sonar data settings were as follows: frequency 950 kHz, altitude 3 m, range 30 m, pulse length 2 ms.

a reef. As the dropper lines are dragged along the seabed with tidal currents, the concrete weights at the end of the line can induce significant scour and disturb any organisms on the seabed.

3.3 Seabed inspection and environmental impact

Using side scan sonar to detect objects, such as reefs or shipwrecks, and monitor scouring on the seabed is a common practice (Johnson and Helferty, 1990; Penrose et al., 2005). Our surveys indicated both mussels on the seabed and seabed scouring from the aquaculture setup. The surveys conducted in October 2021 detected a large number of reflections that appeared as bright spots surrounding longline 2 (Figure 8A). The reflections are similar to how mussels appear on side scan sonar in previous studies (Powers et al., 2015) and mussels were found beneath long lines by diving surveys in summer and winter of 2021, but no surveys were conducted when the data for Figure 8A was collected (Islam et al., 2024).

The AUV itself will have very little impact on the environment and the seafloor as it needs to be kept at least 3 m off the seabed in order to collect high quality side scan data as discussed in section 2.2. The Gavia AUV is also battery powered and contributes to the decarbonization of marine fleets. However, the mussel long lines can negatively impact the seabed. Although longlines are deployed in high-energy environments where the sediment is routinely disturbed, scouring can still have an impact on benthic habitats (Broad et al., 2020) and can potentially release carbon stored in the seafloor, which may then enter the atmosphere (Atwood et al., 2024). From analyzing the side scan imagery, the scouring is most significant around the block anchors for the mussel longlines (Figure 10). Scouring is important to monitor, especially in shallow high-energy environments, as it is well known to affect the stability of structures on the seabed (Sumer and Fredsøe, 2002). For the two types of anchors (screw and block anchors, as described in Section 2.1), there are different degrees of scouring. The screw anchor shows no evidence of scouring, while there is significant scouring around the block and chain anchor (Figure 10). There is a

10.3389/fmars.2024.1386267

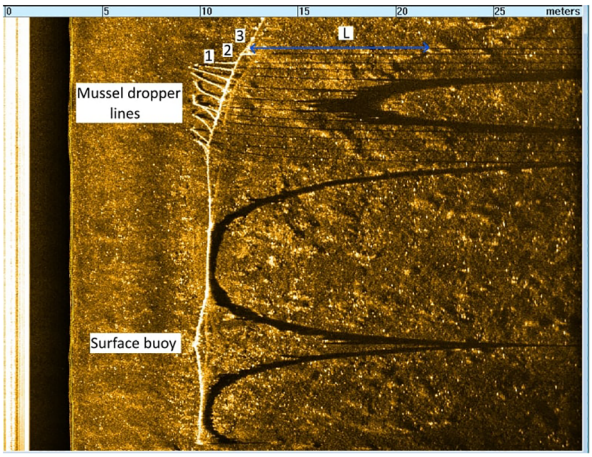


FIGURE 9
Waterfall images displaying of the longline and mussel dropper lines. L is the measured distance between the acoustic shadow and the object, used to calculate the height of bottom of surface buoy = 2.1 m, dropper line 1 = 0 m, dropper line 2 = 0.36m and dropper line 3 = 1.5m. The side scan sonar data settings were as follows: frequency 950 kHz, altitude 3 m, range 30 m, pulse length 2 ms.

large pit around the block and under the chain, which is a result of the tidal current pushing the chain back and forth. Scouring was more apparent in waterfall images than in mosaics, where the processing and combining of survey tracks reduced the appearance of anchor pits. Anchors and their chains are well known to create pits and displace sediments. These sediments usually support a variety of marine life, including polychaetes, crustaceans, mollusks, and sponges, all of which are vulnerable to anchors and their associated scour (Sorokin et al., 2005; Pitcher et al., 2009). Given that there is no scouring present when using a screw anchor, it would be the preferable method of the two anchors used here for mussel longline deployment to prevent the unintentional destruction of a potential mussel reef.

4 Summary and conclusions

The objective of this study was to test the following hypothesis: *"AUVs equipped with interferometric side scan sonar can effectively monitor mussel aquaculture installations, including long- and dropper lines, anchoring systems, and the seabed beneath, in high-energy and turbid environments"*. To address this hypothesis the following four questions were raised. First, can an AUV safely conduct surveys of aquaculture infrastructure in shallow and turbid, high-energy environments? Three campaigns and several surveys were successfully conducted in and around the mussel lines in the Belgian North Sea, an area with current speed exceeding 1 m s^{-1}. This study shows that side scan sonar mounted on an AUV is a viable technique in support of management of aquaculture and coastal protection setups. Apart from mapping the seabed, structures in the water column can be detected and analyzed. Using an AUV in a shallow, high-energy environment is not without challenges. The strong current and waves associated with the high-energy environment can cause roll, which has the greatest impact on the quality of the side scan sonar data (Figure 4). To combat the strong currents, AUV Barabas was aligned either against or with the prevailing tidal current to prevent the instrument from moving sideways. The speed of the AUV was also set a fixed SOG of 1.7 m s^{-1} as that was found to be the best for the current conditions and provided the best quality of data. The impact of surface waves was addressed by flying the AUV Barabas at a lower altitude; however, the AUV Barabas still needed to be high enough to detect the mussel longlines on the surface. The close proximity to submerged obstacles called for a high navigational accuracy, which was ensured by planning short missions with intermittent GPS fixes.

TABLE 2 Tidal information during AUV Barabas surveys displayed in Figures 5 and 6.

Survey date	Survey time	Low tide	High tide
12/08/2021	11:44	10:57	16:36
08/10/2021	10:19	9:30	15:07

The second question was: which side scan sonar settings yield the highest quality data? The altitude of the AUV impacts coverage of the side scan sonar. With the AUV Barabas at a lower altitude (3 m instead of 5 m from the seabed) to avoid impact from surface waves, the range is lower and covers less of the seabed. However, the resolution is higher and the AUV will pitch and roll less, providing more detailed images. As expected the high frequency setting (900 kHz) produced the highest data resolution. The quality of the data produced by the two pulse lengths used depended upon the format in which the data was displayed. In the mosaic, a pulse length of 2 ms provided better data, whereas a pulse length of 1 ms provided more detail when displayed in the waterfall.

The third question was: what relevant information could be ascertained about the aquaculture setup? The surveys provided

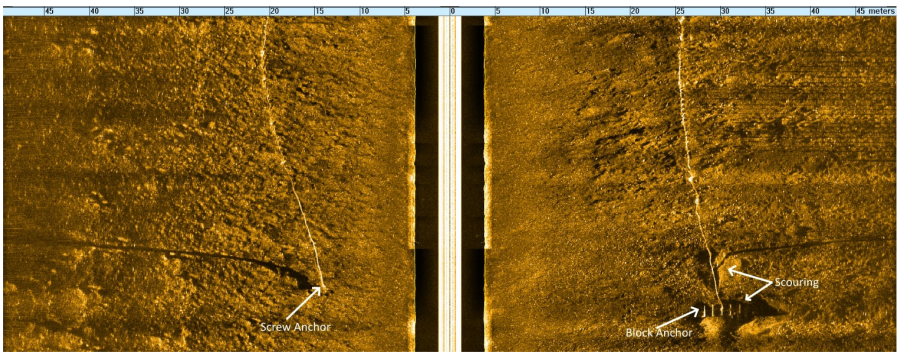


FIGURE 10
Waterfall images of significant scouring around the two different anchors of the mussel longlines, a screw anchor (left) and block anchor (right). The side scan sonar data settings were as follows: frequency 950 kHz, altitude 5 m, range 50 m, pulse length 2 m.

details on not only the longlines and their anchors but also the individual dropper lines in the water column. The surveys show how far the dropper lines are above the seabed and when they rest upon it. This is relevant for both aquaculture setups, as it provides insight into the stressors placed on longlines in high-energy environments and the impact of the setup on the seabed. If the survey shows the dropper lines on the seabed, the entire setup can be properly adjusted to ensure that the mussel dropper lines stay above the seabed. This is critical for both the quality of the mussels harvested in aquaculture and the development of a mussel reef for coastal protection.

Finally, we addressed the fourth question to investigate what pertinent information can be unraveled about the seabed surrounding the aquaculture setup and, therefore, investigated the potential impact on the seabed. The side scan sonar surveys also collected data on the surrounding environment and show that the main impact of the aquaculture setup on the seabed is scouring, particularly around the anchors. A traditional block and chain anchor induces significantly more scouring than a screw anchor. Scouring from the anchor could affect not only organisms in the surface substrate but also any mussels that settle on the seabed from the dropper lines.

The results demonstrate that employing an Autonomous Underwater Vehicle (AUV) equipped with a side scan sonar is a viable and innovative technique for monitoring mussel longlines. As these techniques continue to evolve, they could offer a more cost-effective and simpler alternative to the traditionally challenging logistics of conventional monitoring methods. This advancement underscores the potential for significant enhancements in aquaculture management, particularly in optimizing operational efficiencies and reducing risks. However, monitoring with an AUV is not without drawbacks. Side scan sonar does not provide direct information on the health of the mussels setup such as occurrence of predators or fouling. The positioning error from underwater navigation increases the longer the AUV is under the water. Therefore, shorter and more frequent dives with surfacing to regain accurate positioning from satellite navigation systems might be required.

Another limitation is the size of the AUV itself. With a length of almost 3 m, the used AUV Barabas has a relatively large turning radius, which is not ideal for a lawnmower pattern flown between two mussel longlines. The size and weight also require a support vessel for safe launch and recovery. Since most of our findings are not specific to the AUV Barabas used, the issues could be overcome by using smaller AUVs. Smaller, cheaper AUVs are emerging on the market, and while they are unable to match the sonar range, depth rating, and navigation precision of a survey class AUV, these factors might be less of a concern in very shallow water. The data quality could be sufficient for operational monitoring, with much lower acquisition costs and a minimal logistical footprint.

The approach used in this study shows that it is an efficient and safe alternative to combine AUVs and side scan sonar in the detailed monitoring of an aquaculture setup. A similar approach could be implemented in the monitoring of offshore mussel farms at a large production scale, allowing for intense monitoring in short periods. This monitoring strategy could also be suitable for other aquaculture setups. For example to monitor structural integrity of pisciculture cages, oyster cages/beds, and more particularly their impact on the seafloor. The AUV platform allows for multiple sensors to be collecting data at once so environmental data could be collected along with side scan sonar data, providing information on temperature, salinity and water pH. Using an AUV also provides a low risk alternative to site inspection to either identify suitable areas for aquaculture installations or after disaster, such as structural damage after a storm. Monitoring aquaculture setups with an AUV could also be used in tandem with other marine robotic platforms such an ROV. An AUV survey would provide information about a site as a whole, then an ROV could perform detailed visual inspection on areas of interest and look at specific problems for example biofouling on the mussels.

Data availability statement

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

Author contributions

CP: Formal analysis, Visualization, Writing – original draft, Writing – review & editing. KL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. FF: Data curation, Methodology, Writing – review & editing, Conceptualization, Investigation. LP: Writing – original draft, Writing – review & editing. IM: Writing – review & editing. AS: Writing – review & editing. TS: Writing – review & editing. RG: Writing – review & editing. BG: Writing – review & editing. WB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was part of the Blue Cluster Project Coastbusters 2.0 and funded by Flanders Innovation & Entrepreneurship (VLAIO) through the funding HBC.2019.0037. Blue Cluster was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

We would like to acknowledge Flanders Innovation & Entrepreneurship (VLAIO) and Blue Cluster for funding the Coastbusters 2.0 Project. Furthermore, we gratefully acknowledge the crew of RV Simon Stevin and DAB Vloot for their continuous support in performing the research at sea. Finally, we thank the Marine Robotics Center (MRC) of the Flanders Marine Institute (VLIZ) for the use of AUV Barabas in our research.

Conflict of interest

Author TS was employed by company DEME Group. Author RG was employed by company Jan De Nul. Author BG was employed by company Sioen Industries NV.

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

Agency for Maritime Services and Coast, Flemish Government. (2023). *Flemish banks monitoring network*. Available at: https://meetnetvlaamsebanken.be/.

Ali, A., Abdullah, M. R., Safuan, C. D. M., Afiq-Firdaus, A. M., Bachok, Z., Akhir, M. F. M., et al. (2022). Side-scan sonar coupled with scuba diving observation for enhanced monitoring of benthic artificial reefs along the coast of Terengganu, Peninsular Malaysia. *J. Mar. Sci. Eng.* 10, 1309. doi: 10.3390/jmse10091309

Allison, E. H. (2011). *Aquaculture, fisheries, poverty and food security* (Penang: The WorldFish Center), 62. Working Paper 2011-65.

Amundsen, H. B., Caharija, W., and Pettersen, K. Y. (2021). Autonomous ROV inspections of aquaculture net pens using DVL. *IEEE J. Oceanic Eng.* 47, 1–19. doi: 10.1109/JOE.2021.3105285

Atwood, T. B., Romanou, A., DeVries, T., Lerner, P. E., Mayorga, J. S., Bradley, D., et al. (2024). Atmospheric CO2 emissions and ocean acidification from bottom-trawling. *Front. Mar. Sci.* 10, 1125137. doi: 10.3389/fmars.2023.1125137

Avdelas, L., Avdic-Mravlje, E., Borges Marques, A. C., Cano, S., Capelle, J. J., Carvalho, N., et al. (2021). The decline of mussel aquaculture in the European Union: causes, economic impacts and opportunities. *Rev. Aquaculture* 13, 91–118. doi: 10.1111/raq.12465

Bao, J., Li, D., Qiao, X., and Rauschenbach, T. (2020). Integrated navigation for autonomous underwater vehicles in aquaculture: A review. *Inf. Process. Agric.* 7, pp.139–pp.151. doi: 10.1016/j.inpa.2019.04.003

Barbosa, R. V., Jaud, M., Bacher, C., Kerjean, Y., Jean, F., Ammann, J., et al. (2022). High-resolution drone images show that the distribution of mussels depends on microhabitat features of intertidal rocky shores. *Remote Sens.* 14, 5441. doi: 10.3390/rs14215441

Blondel, P. (2009). "The Handbook of Sidescan Sonar." in *Springer Praxis Books* (Berlin, Heidelberg: Springer). doi: 10.1007/978-3-540-49886-5_4

Borsje, B. W., van Wesenbeeck, B. K., Dekker, F., Paalvast, P., Bouma, T. J., van Katwijk, M. M., et al. (2011). How ecological engineering can serve in coastal protection. *Ecol. Eng.* 37, 113–122. doi: 10.1016/j.ecoleng.2010.11.027

Boulenger, A., Lanza-Arroyo, P., Langedock, K., Semeraro, A., and Van Hoey, G. (2024). Nature-based solutions for coastal protection in sheltered and exposed coastal waters: integrated monitoring program for baseline ecological structure and functioning assessment. *Environ. Monit. Assess.* 196, 316. doi: 10.1007/s10661-024-12480-x

Brehmer, P., Gerlotto, F., Guillard, J., Sanguinède, F., Guénnegan, Y., and Buestel, D. (2003). New applications of hydroacoustic methods for monitoring shallow water aquatic ecosystems: the case of mussel culture grounds. *Aquat. Living Resour.* 16, 333–338. doi: 10.1016/S0990-7440(03)00042-1

Brehmer, P., Vercelli, C., Gerlotto, F., Sanguinède, F., Pichot, Y., Guennégan, Y., et al. (2006). Multibeam sonar detection of suspended mussel culture grounds in the open sea: Direct observation methods for management purposes. *Aquaculture* 252, 234–241. doi: 10.1016/j.aquaculture.2005.06.035

Broad, A., Rees, M. J., and Davis, A. R. (2020). Anchor and chain scour as disturbance agents in benthic environments: trends in the literature and charting a course to more sustainable boating and shipping. *Mar. pollut. Bull.* 161, 111683. doi: 10.1016/j.marpolbul.2020.111683

Bryant, R. S. (1975). "Side Scan Sonar for Hydrography - An Evaluation by the Canadian Hydrographic Service", *The International Hydrographic Review*. 52 (1).

Cheong, L., and Lee, H. B. (1984). *Mussel farming* (Bangkok, Thailand: Secretariat, Southeast Asian Fisheries Development Center).

Clausner, J. E., and Pope, J. (1988). May. "Application of side-scan sonar for inspection of coastal structures." in *Offshore Technology Conference*. (OTC), OTC–5781. doi: 10.4043/5781-MS

Degraer, S., Van Lancker, V., Moerkerke, G., Van Hoey, G., Vincx, M., Jacobs, P., et al. (2000). Intensive evaluation of the evolution of a protected benthic habitat. *HABITAT RUG Section Mar. Biology-Sedimentary geology Eng. geology*, 1–16.

Ells, K., and Murray, A. B. (2012). Long-term, non-local coastline responses to local shoreline stabilization. *Geophysical Res. Lett.* 39. doi: 10.1029/2012GL052627

Fettweis, M., Baeye, M., Lee, B. J., Chen, P., and Yu, J. C. (2012). Hydro-meteorological influences and multimodal suspended particle size distributions in the Belgian nearshore area (southern North Sea). *Geo-Marine Lett.* 32, 123–137. doi: 10.1007/SÛ0367-011-0266-7

Flemming, B. W. (1976). Side-scan sonar: a practical guide. *Int. hydrographic Rev.* 53, 65–92.

Goedefroo, N., Benham, P., Debusschere, E., Deneudt, K., Mascart, T., Semeraro, A., et al. (2022). Nature-based solutions in a sandy foreshore: A biological assessment of a longline mussel aquaculture technique to establish subtidal reefs. *Ecol. Eng.* 185, 106807. doi: 10.1016/j.ecoleng.2022.106807

Greene, A., Rahman, A. F., Kline, R., and Rahman, M. S. (2018). Side scan sonar: A cost-efficient alternative method for measuring seagrass cover in shallow environments. *Estuarine Coast. Shelf Sci.* 207, 250–258. doi: 10.1016/j.ecss.2018.04.017

Greene, H. G., Yoklavich, M. M., Starr, R. M., O'Connell, V. M., Wakefield, W. W., Sullivan, D. E., et al. (1999). A classification scheme for deep seafloor habitats. *Oceanol. Acta* 22, 663–678. doi: 10.1016/S0399-1784(00)88957-4

Haerens, P., Bolle, A., Trouw, K., and Houthuys, R. (2012). Definition of storm thresholds for significant morphological change of the sandy beaches along the Belgian coastline. *Geomorphology* 143, 104–117. doi: 10.1016/j.geomorph.2011.09.015

Hicks, D., Cintra-Buenrostro, C., Kline, R., Shively, D., and Shipley-Lozano, B. (2015). "Artificial reef fish survey methods: counts vs. Log-categories yield different diversity estimates," in *Proceedings of the 68th Gulf and Caribbean Fisheries Institute*, 74–79. Available online at: https://www.REEF.org.

Hou, S., Jiao, D., Dong, B., Wang, H., and Wu, G. (2022). Underwater inspection of bridge substructures using sonar and deep convolutional network. *Advanced Eng. Inf.* 52, 101545. doi: 10.1016/j.aei.2022.101545

Islam, M., Semeraro, A., Langedock, K., Moulaert, I., Stratigaki, V., Sterckx, T., et al. (2024). Inducing mussel beds, based on an aquaculture long-line system, as nature-based solutions: effects on seabed dynamics and benthic communities. *Nature-Based Solutions* 6, 100142. doi: 10.1016/j.nbsj.2024.100142

Johnson, H. P., and Helferty, M. (1990). The geological interpretation of side scan sonar. *Rev. Geophys.* 28, 357. doi: 10.1029/RG028i004p00357

Karimanzira, D., Jacobi, M., Pfützenreuter, T., Rauschenbach, T., Eichhorn, M., Taubert, R., et al. (2014). First testing of an AUV mission planning and guidance system for water quality monitoring and fish behavior observation in net cage fish farming. *Inf. Process. Agric.* 1, 131–140. doi: 10.1016/j.inpa.2014.12.001

Koch, E. W., Barbier, E. B., Silliman, B. R., Reed, D. J., Perillo, G. M., Hacker, S. D., et al. (2009). Non-linearity in ecosystem services: temporal and spatial variability in coastal protection. *Front. Ecol. Environ.* 7, 29–37. doi: 10.1890/080126

Krull, W., Berry, P., Bauduceau, N., Cecchi, C., Elmqvist, T., Fernandez, M., et al. (2015). Towards an EU research and innovation policy agenda for nature-based solutions & re-naturing cities. *Final report of the Horizon 2020 expert group on nature-based solutions and re-naturing cities*., 1–70. doi: 10.2777/479582

Lowry, M. B., Glasby, T. M., Boys, C. A., Folpp, H., Suthers, I., and Gregson, M. (2014). Response of fish communities to the deployment of estuarine artificial reefs for fisheries enhancement. *Fisheries Manage. Ecol.* 21, 42–56. doi: 10.1111/fme.12048

Marsden, J. E., Marcy-Quay, B., Dingledine, N., Berndt, A., and Adams, J. (2023). Physical and biological evolution of constructed reefs–long-term assessment and lessons learned. *J. Great Lakes Res.* 49, 276–287. doi: 10.1016/j.jglr.2022.10.008

Massarelli, C., Galeone, C., Savino, I., Campanale, C., and Uricchio, V. F. (2021). Towards sustainable management of mussel farming through high-resolution images and open source software—The taranto case study. *Remote Sens.* 13, 2985. doi: 10.3390/rs13152985

McGeady, R., Runya, R. M., Dooley, J. S., Howe, J. A., Fox, C. J., Wheeler, A. J., et al. (2023). A review of new and existing non-extractive techniques for monitoring marine protected areas. *Front. Mar. Sci.* 10, 1126301. doi: 10.3389/fmars.2023.1126301

Mitchell, N. C., and Somers, M. L. (1989). Quantitative backscatter measurements with a long-range side-scan sonar. *IEEE J. Oceanic Eng.* 14, pp.368–pp.374. doi: 10.1109/48.35987

Murphy, R. R., Steimle, E., Hall, M., Lindemuth, M., Trejo, D., Hurlebaus, S., et al. (2011). Robot-assisted bridge inspection. *J. Intelligent Robotic Syst.* 64, 77–95. doi: 10.1007/s10846-010-9514-8

Murray, J. M. H., Meadows, A., and Meadows, P. S. (2002). Biogeomorphological implications of microscale interactions between sediment geotechnics and marine benthos: a review. *Geomorphology* 47, 15–30. doi: 10.1016/S0169-555X(02)00138-1

Nagarajan, R., Lea, S. E., and Goss-Custard, J. D. (2006). Seasonal variations in mussel, Mytilus edulis L. shell thickness and strength and their ecological implications. *J. Exp. Mar. Biol. Ecol.* 339, 241–250. doi: 10.1016/j.jembe.2006.08.001

Osen, O. L., Leinan, P. M., Blom, M., Bakken, C., Heggen, M., and Zhang, H. (2018). October. A novel sea farm inspection platform for norwegian aquaculture application. *OCEANS 2018 MTS/IEEE Charleston*, 1–8. doi: 10.1109/OCEANS.2018.8604648

Paull, L., Seto, M., Leonard, J. J., and Li, H. (2018). Probabilistic cooperative mobile robot area coverage and its application to autonomous seabed mapping. *Int. J. Robotics Res.* 37, 21–45. doi: 10.1177/0278364917741969

Penrose, J. D., Siwabessy, P. J. W., Gavrilov, A., Parnum, I., Hamilton, L., Bickers, A., et al. (2005). Acoustic techniques for seabed classification. *Cooperative Research Centre for Coastal Zone Estuary and Waterway Management, Technical Report*. 32 (11).

Pitcher, C. R., Burridge, C. Y., Wassenberg, T. J., Hill, B. J., and Poiner, I. R. (2009). A large scale BACI experiment to test the effects of prawn trawling on seabed biota in a closed area of the Great Barrier Reef Marine Park, Australia. *Fisheries Res.* 99, 168–183. doi: 10.1016/j.fishres.2009.05.017

Powers, J., Brewer, S. K., Long, J. M., and Campbell, T. (2015). Evaluating the use of side-scan sonar for detecting freshwater mussel beds in turbid river environments. *Hydrobiologia* 743, 127–137. doi: 10.1007/s10750-014-2017-z

Seddon, N., Chausson, A., Berry, P., Girardin, C. A., Smith, A., and Turner, B. (2020). Understanding the value and limits of nature-based solutions to climate change and other global challenges. *Philos. Trans. R. Soc. B* 375, 20190120. doi: 10.1098/rstb.2019.0120

Sorokin, S. J., Currie, D. R., and Ward, T. M. (2005). Sponges from the great Australian bight marine park (Benthic protection zone). *Report to the Wildlife Conservation Fund, Department of Environment and Heritage South Australia*. SARDI Aquatic Sciences Centre, Adelaide.

Sousa, D., Hernandez, D., Oliveira, F., Luís, M., and Sargento, S. (2019). A platform of unmanned surface vehicle swarms for real time monitoring in aquaculture environments. *Sensors* 19, 4695. doi: 10.3390/s19214695

Sumer, B. M., and Fredsøe, J. (2002). The mechanics of scour in the marine environment. *World Sci.* 17. doi: 10.1142/ASOE

Temmerman, S., Meire, P., Bouma, T. J., Herman, P. M. J., Ysebaert, T., and De Vriend, H. J. (2013). Ecosystem-based coastal defence in the face of global change. *Nature* 504, 79–83. doi: 10.1038/nature12859

Ubina, N. A., and Cheng, S. C. (2022). A review of unmanned system technologies with its application to aquaculture farm monitoring and management. *Drones* 6, 12. doi: 10.3390/drones6010012

Van der Meulen, F., IJff, S., and van Zetten, R. (2023). Nature-based solutions for coastal adaptation management, concepts and scope, an overview. *Nordic J. Bot.* 2023, e03290. doi: 10.1111/njb.03290

Walles, B., Troost, K., van den Ende, D., Nieuwhof, S., Smaal, A. C., and Ysebaert, T. (2016). From artificial structures to self-sustaining oyster reefs. *J. Sea Res.* 108, 1–9. doi: 10.1016/j.seares.2015.11.007

Wynn, R. B., Huvenne, V. A., Le Bas, T. P., Murton, B. J., Connelly, D. P., Bett, B. J., et al. (2014). Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience. *Mar. geology* 352, 451–468. doi: 10.1016/j.margeo.2014.03.012

Zhao, H., Zhang, H., Miao, C., Ye, X., and Min, M. (2018). Linking heat source–sink landscape patterns with analysis of urban heat islands: Study on the fast-growing Zhengzhou City in Central China. *Remote Sens.* 10, 1268. doi: 10.3390/rs1008126

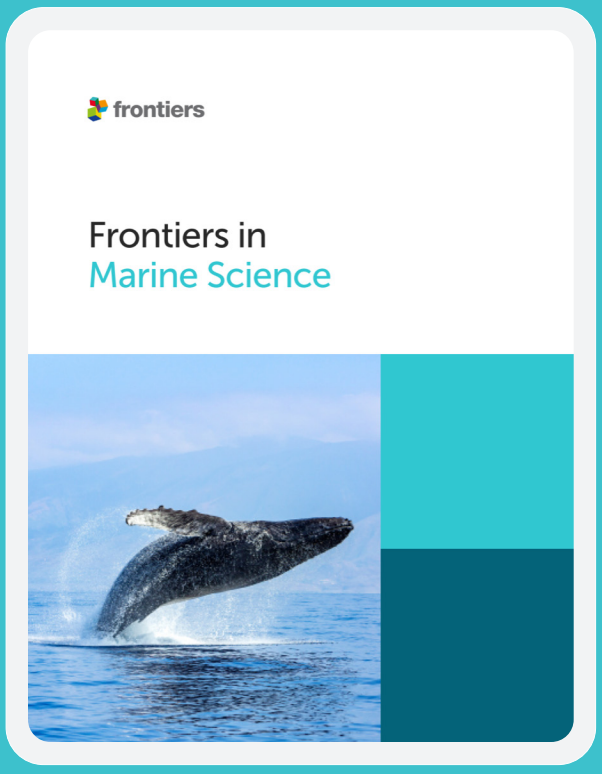
Frontiers in Marine Science

Explores ocean-based solutions for emerging global challenges

The third most-cited marine and freshwater biology journal, advancing our understanding of marine systems and addressing global challenges including overfishing, pollution, and climate change.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact