remote sensing

MDPI

Article

Multi-Granularity Graph Neural Network for Satellite-Assisted Marine Environmental Field Reconstruction over Sparse Observation Grids

Baowen Guo * and Yangming Guo

School of Cyberspace Security, Northwestern Polytechnical University, Xi'an 710060, China; yangming_g@nwpu.edu.cn
* Correspondence: guobw009@mail.nwpu.edu.cn

Highlights

What are the main findings?

- MG-GCNN consistently outperforms representative baseline methods in reconstructing temperature, salinity, and ocean-current fields under different observation sparsity levels, including highly sparse and Argo-like sampling configurations.
- The terrain-aware graph, vertical feature integration, multi-granularity representation, and self-supervised masked node reconstruction provide complementary improvements in reconstruction accuracy and robustness.

What are the implications of the main findings?

- Terrain-aware graph learning combined with vertical and multi-scale feature modeling provides an effective framework for recovering continuous three-dimensional marine environmental fields from sparse observations.
- The reconstructed temperature, salinity, and current fields can provide environmental information for subsequent ocean monitoring, acoustic propagation modeling, and underwater sensing applications.

Abstract

Marine remote sensing combines broad-coverage satellite observations of the ocean surface with sparse in-situ and underwater observations. However, reconstructing continuous three-dimensional environmental fields remains challenging because of irregular sampling, vertical non-stationarity, and bathymetric barriers. In this paper, a multi-granularity graph collaborative neural network (MG-GCNN) is proposed for high-fidelity field reconstruction and situational awareness in sparse monitoring scenarios. The network abstracts discrete observation points as heterogeneous nodes in the topological graph structure. By incorporating high-resolution bathymetry as a geometric prior, terrain-aware graph construction, vertical feature integration, multi-granularity aggregation, and self-supervised masked node reconstruction are jointly used to capture spatial and vertical dependencies under limited observation availability. Experimental results show that MG-GCNN significantly outperforms baseline interpolation and convolution models in terms of reconstruction accuracy, especially in regions with complex underwater terrain and extreme sampling sparsity. The reconstructed environmental fields can potentially provide three-dimensional environmental inputs for subsequent ocean-acoustic propagation modeling, underwater sensing, and related marine applications.

Check for updates

Academic Editors: Mingyang Zhang and Hao Li

Received: 23 July 2026
Revised: 29 August 2026
Accepted: 31 August 2026
Published: 4 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: marine environmental reconstruction; graph collaborative neural network; sparse monitoring grids; satellite-assisted ocean remote sensing

1. Introduction

Marine environmental characterization is the cornerstone for understanding complex underwater acoustic interactions and intricate oceanic dynamic processes [1–3]. These physical fields, including temperature, salinity, and velocity fields, constitute a highly nonlinear and multi-scale coupling system that governs the propagation medium for all underwater activities [4,5]. The accurate representation of these variables is essential not only for improving sonar performance and underwater navigation but also for comprehending regional heat transport mechanisms. However, large-scale ocean state estimation typically relies on spatially distributed monitoring arrays (e.g., Argo floats, moored buoys), which naturally form a spatially sparse and unstructured observation network. The demand for high-fidelity environmental modeling that can bridge the gap between these sparse observations and a continuous physical field has become an urgent task in oceanography and ocean engineering [6–8].

From the perspective of remote sensing, satellite and underwater acoustic systems provide complementary ocean views. Satellite radiometers and altimeters provide frequent and spatially extensive observations of sea surface temperature, sea surface salinity, sea surface height or sea surface anomalies, but they cannot directly analyze the complete groundwater column. In contrast, ocean acoustic remote sensing and sonar-based target detection work below the sea surface, and their performance strongly depends on the three-dimensional distribution of temperature, salinity, pressure, current and water depth. These environmental variables determine the distribution of sound velocity and affect the refraction, convergence and shadow areas of sound waves, as well as the multi-path of seabed interaction [9]. Therefore, the accurate reconstruction of subsurface temperature, salinity, and current fields provides an important bridge for the environmental connection between satellite surface observation and underwater acoustic sensing.

Historically, the reconstruction of marine physical fields depended on numerical simulations and statistical interpolation methods such as optimal interpolation and empirical orthogonal functions [10]. Although these methods provide a baseline for generating spatial continuous fields, they are often limited by high computational costs and linear assumptions. Recently, the field has turned to the data-driven deep learning paradigm. Various architectures have been deployed, including convolutional neural networks (CNNs) and Transformer-based models [8,11,12], to capture complex spatio-temporal dependencies. These models show great potential in interpolating sparse observations and reconstructing high-resolution ocean grids by learning the nonlinear mapping function between multi-source observations and high-dimensional fields.

Despite these advances, existing deep learning frameworks are facing severe challenges in dealing with the inherent nonstationarity and spatial heterogeneity of the ocean, especially in the vertical dimension. Most advanced models utilize a monolithic framework with shared parameters to represent the entire water column, which implicitly assumes vertical uniformity and leads to the bottleneck of representation ability. In addition, traditional grid-based methods are often unable to deal with the non-Euclidean spatial correlation of sparse monitoring grids, resulting in information loss and boundary artifacts [13]. There is a serious mismatch between vertical dense data and sparse horizontal grids, which requires a framework to adaptively decouple heterogeneous characteristics of different water layers, such as the mixed layer, thermocline, and deep-sea layer, while accounting for vertical

heterogeneity [14]. Therefore, the central problem addressed in this study is the reconstruction of three-dimensional marine environmental fields under sparse and irregular observations, with particular emphasis on horizontal sparsity, bathymetric constraints, and vertical non-stationarity.

To address these reconstruction challenges, this paper proposes a Multi-granularity Graph Collaborative Neural Network (MG-GCNN), as illustrated in Figure 1. Specifically, the topographic obstruction represents the horizontal impedance where submarine ridges block the flow and decouple adjacent water masses, while vertical non-stationarity refers to the sharp property gradients within the stratified layers. By abstracting sparse monitoring points as heterogeneous nodes within a spatial graph, the framework captures long-range physical synergies through topological correlation learning. The MG-GCNN explicitly models this dual-constraint by using terrain-aware graph (TAG) for horizontal isolation and vertical feature integration (VFI) for vertical gradient preservation. Georeferential satellite surface variables can also be incorporated as surface node attributes or boundary constraints without changing the graph-based representation. The main contributions of this work are summarized as follows:

- We propose a graph-based environmental reconstruction framework that represents irregular marine observations as topological data and is compatible with heterogeneous georeferenced sources, including sparse in-situ profiles and satellite-derived surface constraints.
- We introduce an adaptive hierarchical gating mechanism driven by ocean dynamics to decouple deep features between different water layers and effectively avoid the common representational crosstalk in monolithic models.
- We incorporate multi-granularity contrastive learning with physics-guided regularization to improve representation robustness under severe sampling sparsity.
- Extensive experiments demonstrate improved reconstruction accuracy and robustness. The reconstructed environmental fields establish a potential interface to downstream ocean-acoustic and underwater sensing models.

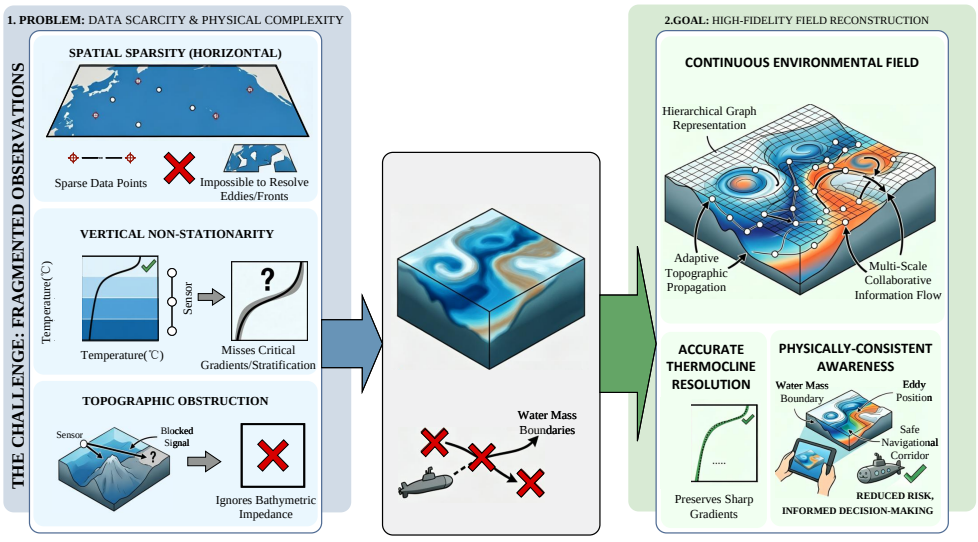


Figure 1. Conceptual framework for sparse marine environmental reconstruction and acoustic remote sensing applications.

The rest of this paper is structured as follows. Section 2 briefly reviews the related backgrounds regarding marine environmental modeling, multi-granularity representation learning, and graph neural networks. In Section 3, we formalize the research problem of environmental field reconstruction under sparse sampling constraints and describe the

details of our framework. Extensive experiments are conducted in Section 4 to demonstrate the superiority of the proposed framework over conventional algorithms. Finally, we conclude with a discussion of the proposed scheme and summarize the future work in Section 5.

2. Related Work

2.1. Marine Environmental Modeling from In Situ and Satellite Observations

The reconstruction and representation of marine physical fields constitute the technical basis for various applications, from underwater acoustic communication to regional climate prediction [15–17]. Historically, this field has been dominated by numerical simulation and statistical interpolation techniques, aiming to bridge the information gaps in sparse observational data [18,19]. Methods such as optimal interpolation and empirical orthogonal functions have long served as the benchmarks for generating spatially continuous fields by minimizing the variance of analysis error. Specifically, the EOF approach decomposes the target physical field into a set of orthogonal spatial patterns and temporal coefficients, allowing a reduced-dimensional representation of ocean variability. However, these traditional methods usually rely on linear assumptions and struggle to capture the highly nonlinear and transient characteristics of mesoscale phenomena, such as eddies or sudden thermocline shifts, especially when the underlying monitoring grid is extremely sparse [20].

With the rapid expansion of satellite remote sensing and autonomous underwater vehicle arrays, a paradigm shift has occurred toward data-driven deep learning models [21,22]. Early neural network applications focused on leveraging Convolutional Neural Networks (CNNs) [23,24] to extract spatial features from sea surface variables to infer subsurface structures. These models typically treat the ocean as a regular Euclidean grid, where the transformation of features is governed by unified weight matrices. For a given state at depth z, the representation is often formalized as a mapping F such that:

$$\hat{\Phi}(z) = \mathcal{K} * \Phi_{obs}(z) + b, \tag{1}$$

where $\mathcal{K}$ represents the convolutional kernels and Φ_{obs} denotes the sparse input observations. Although it is effective for grid-aligned data, this monolithic architecture faces a representation capacity bottleneck when dealing with heterogeneous dynamics of different water layers. More importantly, the inherent sparsity and irregular distribution of modern monitoring networks, where sensors may be scattered in complex seabed features, challenge the fundamental assumption of local connectivity in standard CNNs. Therefore, recent research has begun to explore meshfree or graph-based modeling techniques, which can explicitly incorporate physical heuristics and spatial topology, moving beyond simple pixel-level interpolation to achieve high-fidelity field reconstruction. Satellite remote sensing has become an important source of spatially extensive ocean observations. Sea-surface temperature, sea-surface height or anomaly, sea-surface salinity, ocean color, and related products have been combined with in situ profiles to reconstruct subsurface temperature, salinity, chlorophyll, and mesoscale eddy structures [9–11]. These studies demonstrate that surface signatures contain useful information about subsurface dynamics. Nevertheless, most existing methods are formulated on regular Euclidean grids and may be sensitive to irregular sampling, missing swaths, quality-controlled data gaps, and heterogeneous spatial coverage. They also rarely incorporate bathymetric barriers directly into the information-propagation mechanism.

2.2. Multi-Granularity Representation Learning

The complexity of marine environmental variables lies in their cross-scale coupling, that is, the coexistence of large-scale circulation patterns with mesoscale eddies and micro-scale turbulence. Hierarchical representation learning has emerged as a critical technique to decompose these multi-scale phenomena into manageable latent spaces. Different from the monolithic architectures trying to learn a universal feature mapping, hierarchical models adopt a coarse-to-fine strategy, in which early layers capture the global trend of the physical field, while deeper layers focus on local high-frequency variations [25]. This multi-granularity perspective is particularly important for marine modeling, as it allows the network to distinguish between the stable, synoptic-scale dynamics of the deep sea and the volatile, small-scale variability of the surface mixed layer [26,27].

In a multi-granularity framework, the representation of a physical variable Φ is typically decomposed into a summation of features extracted at different scales $s \in \{1, \ldots, S\}$. Formally, the multi-scale embedding H can be expressed as:

$$H = \sum_{s=1}^{S} W_s \cdot \phi_s(X_{obs}) + \mathcal{R}_{inter}, \tag{2}$$

where ϕ_s denotes the scale-specific transformation function, W_s represents the adaptive weighting factor for granularity s, and $\mathcal{R}_{inter}$ accounts for the inter-scale collaborative residues. To further enhance the discriminative power of these representations, contrastive learning mechanisms are often integrated to enforce structural consistency across different granularities. By maximizing the mutual information between the global horizontal trend and local vertical profiles, the model can bridge the information gap caused by spatial sparsity [10,28].

In addition, the integration of adaptive routing enables the model to dynamically allocate representational capacity according to the local complexity of the environment. Instead of a static decomposition, an intelligent gating mechanism $G(\cdot)$ determines the contribution of specialized scale-specific encoders E_i to the final field reconstruction:

$$\hat{X} = \text{Decoder}\left(\sum_{i=1}^{n} G(\Phi_z)_i E_i(h)\right), \tag{3}$$

where Φ_z represents the physical heuristic vector driving the hierarchical assignment. This approach ensures that the fine-grained features of sharp thermocline gradients will not be blurred by the coarse-grained stability of the surrounding water column, thus achieving a high-fidelity representation in line with the stratified nature of the ocean.

2.3. Graph Neural Networks

The inherent irregular distribution and spatial sparsity of marine monitoring arrays (including fixed stations, drifting buoys, and mobile platforms) make traditional Euclidean-based deep learning architectures, such as CNNs, suboptimal for high-fidelity field reconstruction. Graph Neural Networks (GNNs) [29,30] have emerged as a powerful alternative by abstracting these discrete observation points into a topological space where relationships are defined by connectivity rather than fixed grid coordinates [31,32]. In this paradigm, the marine environment is modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of heterogeneous nodes (sensors or sampling points) and $\mathcal{E}$ represents the set of edges characterizing the physical or spatial dependencies between them.

The basic operation of GNN-based environmental perception is the message passing mechanism, which enables each node to aggregate information from its neighborhood to

update its own latent state. For a node v_i with its initial feature vector x_i, the l-th layer hidden representation $h_i^{(l)}$ is updated as:

$$h_i^{(l)} = \sigma\left(\gamma^{(l)} \cdot h_i^{(l-1)} + \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \cdot \mathbf{W}^{(l)} h_j^{(l-1)}\right), \tag{4}$$

where $\mathcal{N}(i)$ is the set of neighboring nodes, $\mathbf{W}^{(l)}$ is a learnable weight matrix, and α_{ij} represents the attention coefficient or edge weight. To incorporate the complex underwater environment, α_{ij} is often regularized by physical priors such as bathymetric gradients and geodesic distances, ensuring that information propagation adheres to fluid dynamical boundaries.

Recent advancements have introduced graph attention networks (GATs) and dynamic graph convolution to adaptively learn the coupling strength between distant observation points [33–35]. This is particularly critical for capturing the impact of topographic features such as long-range teleconnections and submarine ridges on environmental elements. In the context of multi-granularity modeling, the graph structure can be hierarchically coarsened through pooling operators:

$$X^{(l+1)} = \text{Pool}(\text{GraphConv}(X^{(l)}, \mathbf{A}^{(l)})), \tag{5}$$

where $\mathbf{A}^{(l)}$ is the adjacency matrix at level l. By cascading multiple graph convolutional and pooling layers, the network effectively captures multi-scale spatial correlations, ranging from local turbulent variations to basin-scale circulation trends. This topological method not only improves the robustness of the model under extreme sparsity but also provides a more interpretable framework for understanding the underlying spatial structure of marine physical fields.

3. Proposed Algorithm

3.1. Problem Definition

The primary objective of this study is to reconstruct high-fidelity, continuous 3D marine environmental fields from a collection of sparse and irregularly distributed observation points. Different from traditional Euclidean modeling, this paper formalizes the oceanic monitoring network as a graph-structured system to better capture non-linear spatial dependencies and the underlying physical topology. Let the high-dimensional groundtruth state of the ocean be represented as a continuous field $\mathcal{X}$, which comprises C interrelated physical variables (e.g., temperature, salinity, and velocity). In the controlled masking experiments, $\mathcal{X}$ is provided by the complete CORA2 reanalysis field and is therefore treated as a reconstruction reference rather than error-free observational ground truth. In a practical monitoring scenario, we only have access to a sparse subset of observations obtained from N discrete sampling locations.

We define the marine monitoring network as a spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ represents the set of N heterogeneous nodes corresponding to the observation points. Each node v_i is associated with a feature vector $\mathbf{h}_i \in \mathbb{R}^F$, which integrates its spatio-temporal coordinates $(Lon, Lat, Depth, t)$ and the observed physical attributes. The edge set $\mathcal{E}$ and the corresponding adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ characterize the topological connectivity between these discrete nodes, governing how environmental information propagates through the non-uniform medium.

To simulate the inherent sparsity of real-world monitoring, we introduce a stochastic masking operator $\mathcal{M}$. The observed sparse graph signal $\mathbf{X}_{obs}$ is defined as:

$$\mathbf{X}_{obs} = \mathcal{M} \odot \mathbf{X}, \tag{6}$$

where $\odot$ denotes the element-wise Hadamard product and $\mathbf{X}_{obs}$ represents heterogeneous and georeferenced marine observations. Satellite-derived products can be attached to the corresponding surface nodes together with acquisition time, quality flags, and observation masks, whereas Argo floats, moorings, and mobile platforms provide depth-resolved profile constraints. In the present experiments, satellite information is inherited through the data-assimilation system of the datasets rather than supplied as independent raw satellite image channels. This distinction is maintained to avoid conflating satellite-assisted environmental reconstruction with direct satellite image processing. The reconstruction task is then formulated as learning a non-linear mapping function f_θ that recovers the full environmental tensor $\hat{\mathbf{X}}$ from the incomplete topological input:

$$\hat{\mathbf{X}} = f_\theta(\mathbf{X}_{obs} \mid \mathcal{G}, \Omega), \tag{7}$$

where Ω represents a set of underlying physical heuristics, such as vertical stratification stability and bathymetric constraints, which guide the learning process.

In this graph-based abstraction, the vertical dimension is treated as a dense profile associated with each horizontal node v_i. Thus, the total environmental state can be decomposed into a multi-scale representation:

$$\mathbf{X} \approx \sum_{k=1}^{K} \Psi_k(\mathcal{G}^{(k)}, \mathbf{h}^{(k)}), \tag{8}$$

where $\mathcal{G}^{(k)}$ denotes the graph topology at the k-th granularity level, and Ψ_k is the collaborative operator responsible for extracting features at that specific scale. By formalizing the problem in this manner, the model transitions from simple numerical interpolation to topological representation learning, enabling the capture of complex phenomena such as mesoscale eddies and boundary-constrained flows even when the sampling grid is highly fragmented.

3.2. Terrain-Aware Dynamic Graph Construction

The construction of an effective adjacency matrix $\mathbf{A}$ is a prerequisite for capturing the non-linear spatial dependencies within sparse monitoring grids. Unlike traditional GNNs that only rely on Euclidean distance for neighbor selection, our framework incorporates bathymetric priors to reflect the physical constraints imposed by underwater topography. The seafloor geometry is characterized by features such as continental shelves, ridges, and trenches and plays an important role in governing the flow of ocean currents and the propagation paths of acoustic signals. Ignoring these constraints will lead to spurious correlations between nodes that are geographically close but physically isolated by topographic barriers.

To formalize this, we propose a terrain-aware kernel to compute the edge weights between any two nodes v_i and v_j. Let d_{ij} denote the geodesic distance between two sampling locations, and ΔH_{ij} represent the maximum topographic elevation change (or depth gradient) along the linear path connecting v_i and v_j. The initial connectivity weight w_{ij} is defined as follows:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\sigma_d^2}\right) \cdot \exp\left(-\beta \cdot \frac{\Delta H_{ij}}{H_{max} - H_{min}}\right), \tag{9}$$

where σ_d is a distance scaling parameter, β is a terrain sensitivity coefficient, and H_{max}, H_{min} are the boundary depths of the research area. This formulation ensures that nodes separated by high-gradient topographic features receive lower weights, effectively mimicking the topographic impedance in fluid dynamics.

To further enhance the representational capacity, the adjacency matrix is not static but is refined through a dynamic attention mechanism. We introduce a learnable transformation that adjusts the edge weights according to the similarity of the latent physical states of nodes. Given the node embeddings $\mathbf{h}_i$ and $\mathbf{h}_j$, the dynamic attention coefficient α_{ij} is calculated as:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j \parallel e_{ij}]))}{\sum_{k\in\mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k \parallel e_{ik}]))} \tag{10}$$

where $\mathbf{W}$ is a weight matrix, $\mathbf{a}$ is the attention vector, $\parallel$ denotes concatenation, and e_{ij} represents the terrain-based edge features derived from the bathymetric prior.

By integrating these components, the final adjacency matrix $\mathbf{A}$ is sparse yet physically expressive:

$$\mathbf{A}_{ij} = \begin{cases} \alpha_{ij}, & \text{if } j \in \mathcal{N}(i) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

The neighborhood $\mathcal{N}(i)$ is determined by a hybrid strategy than combines k-nearest neighbors (k-NN) and a terrain-based radius threshold. This ensures that the message-passing process within the GCNN is guided by the underlying oceanic physical structure, enabling the model to maintain sharp environmental gradients and reflect the complex boundary conditions of the Pacific seafloor. This terrain-aware construction provides an advantage for situational awareness tasks, see Figure 2.

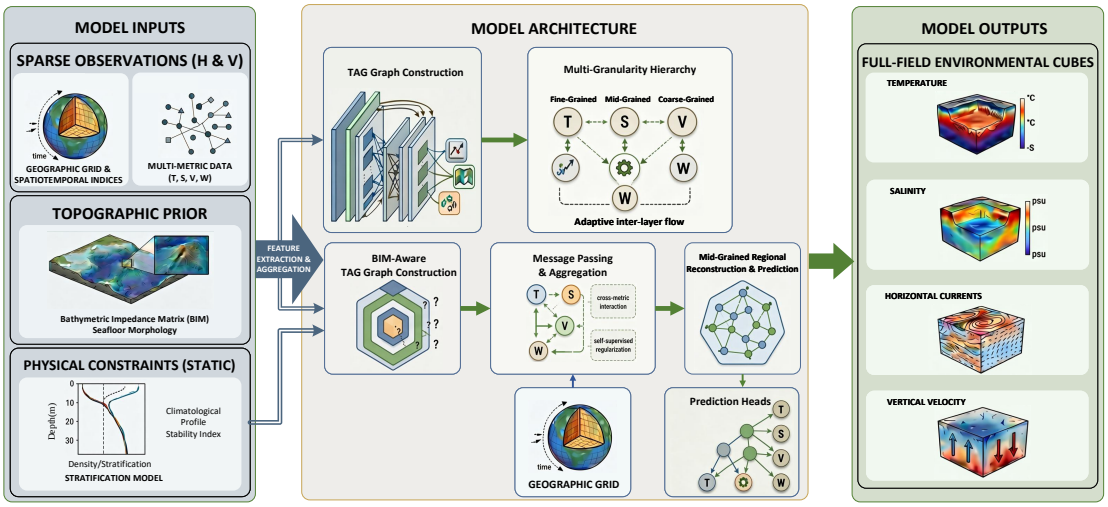


Figure 2. Overall framework of the Multi-granularity Graph Collaborative Neural Network (MG-GCNN).

3.3. Multi-Granularity Graph Collaborative Architecture

To capture the multi-scale coupling of ocean dynamics ranging from basin-scale circulation to local turbulent structures, MG-GCNN adopts a hierarchical structure for multi-granularity feature extraction and fusion, as shown in Figure 3. The core of this architecture is a symmetric encoder–decoder structure that integrates a series of Multi-granularity Graph Collaborative (MGC) blocks, which decouples the complex environmental signals into a hierarchy of topological representations. Unlike monolithic grid-based networks, this collaborative approach enables the model to simultaneously handle horizontal spatial sparsity and vertical profile density, as illustrated in Figure 4.

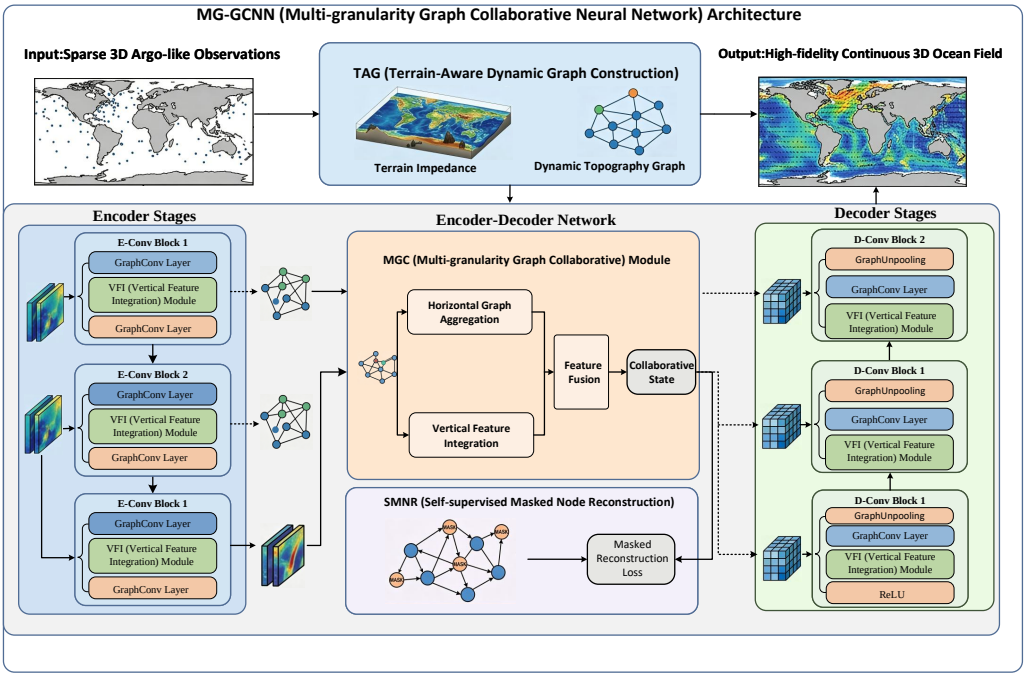


Figure 3. Detailed neural network architecture of the MG-GCNN model.

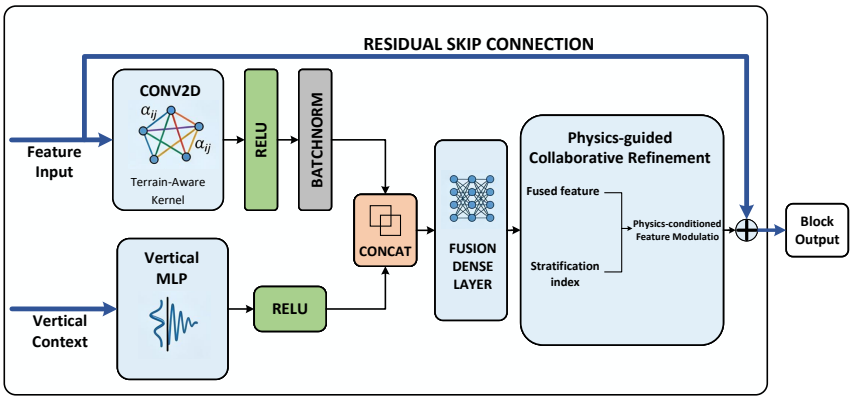


Figure 4. Internal layer architecture of each MGC block.

The encoding stage consists of L layers of graph convolutions paired with hierarchical pooling operators. Each layer l performs spatial message passing on a graph $\mathcal{G}^{(l)}$ whose resolution is progressively coarsened to capture broader spatial contexts. Given the node features $\mathbf{h}^{(l)}$, the layer transition is defined by a graph convolutional operation followed by a top-K graph pooling or voxel-based aggregation:

$$\mathbf{h}^{(l+1)} = \mathrm{Pool}\left(\sigma\left(\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}^{(l)}\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{h}^{(l)}\mathbf{W}^{(l)}\right)\right) \tag{12}$$

where $\hat{\mathbf{A}}^{(l)} = \mathbf{A}^{(l)} + \mathbf{I}$ is the adjacency matrix with added self-loops, and $\hat{\mathbf{D}}$ is the degree matrix. The pooling operator selects a representative subset of nodes or clusters them into super nodes, effectively reducing the size of graph while extracting coarse-grained global features. This hierarchical structure ensures that the model maintains a high-frequency representation of the surface mixed layer at lower levels, while capturing stable deep-sea trends at a higher and more abstract level. The computational cost of MG-GCNN mainly depends on the number of graph nodes, graph edges, feature dimension, and graph layers. Since the present study targets offline three-dimensional environmental field reconstruction rather than real-time inference, computational latency is not treated as a primary optimization objective.

A key innovation of the MG-GCNN is the explicit coupling of horizontal topological correlations with vertical profile dependencies. Within each MGC block, horizontal features are processed via the terrain-aware graph layers described in Section 3.2. Meanwhile, vertical profiles associated with each node representing variables such as the 3D temperature and salinity gradient are processed through a specialized vertical feature integration (VFI) module.

The collaborative state $\mathbf{Z}_i$ for node v_i is computed by fusing the horizontally aggregated information with the local vertical context:

$$\mathbf{Z}_i = \text{Concat}(\text{GraphAgg}(\mathbf{h}_i, \mathcal{N}(i)), \text{VertEnc}(\Phi_{z,i})) \cdot \mathbf{W}_{fusion}, \tag{13}$$

where $\text{GraphAgg}(\cdot)$ denotes the topological aggregation of neighbors, and $\text{VertEnc}(\cdot)$ is a 1D convolution or MLP that encodes the vertical stratification stability index $\Phi_{z,i}$. To quantitatively characterize the vertical non-stationarity, the vertical stratification stability index $\Phi_{z,i}$ is defined based on the potential density gradient:

$$\Phi_{z,i} = -\frac{g}{\rho_0}\frac{\partial \rho_\theta}{\partial z}, \tag{14}$$

where g is the gravitational acceleration, ρ_0 is the reference density, and $\frac{\partial \rho_\theta}{\partial z}$ denotes the local vertical gradient of potential density calculated from reconstructed temperature and salinity. This index enables the VertEnc() module to sense the intensity of the thermocline and adaptively adjust the fusion weights of vertical features. This dual-stream fusion enables the model to adhere to the principle of vertical non-stationarity, ensuring that sharp gradients within the thermocline can be accurately reconstructed by leveraging both local profile physics and adjacent horizontal constraints.

The decoder mirrors the hierarchical encoder, using graph unpooling and transposed convolutions to restore the original spatial resolution. To alleviate the vanishing gradient problem and retain fine-grained spatial details, we establish long-range topological skip connections between corresponding encoder and decoder levels. These connections directly inject low-level geometric features such as sampling coordinates and local terrain gradients into the high-level semantic representation, which is conducive to the high-fidelity reconstruction of the full environmental field. This multi-granularity collaborative loop ensures that the final output maintains numerical accuracy across all vertical layers, while reflecting the global spatial continuity of the marine environment.

3.4. Self-Supervised Masked Node Reconstruction

To reduce the risk of overfitting caused by the limited availability of marine observations and to maximize the representational capacity of the MG-GCNN, we introduce a self-supervised masked node reconstruction (SMNR) mechanism. This strategy is predicated on the assumption that the underlying ocean physical fields have intrinsic spatial continuity and topological coherence. It achieves this by forcing the model to recover hidden information from its surrounding context. Unlike supervised learning that relies on dense labels, SMNR treats the sparse monitoring grid as a partially observed graph signal. During the training phase, we apply the random masking operator $\mathcal{M}_{train}$ to a subset of the graph nodes $\mathcal{V}$, temporarily removing their physical attributes (e.g., temperature, salinity) while preserving their topological positions and bathymetric priors.

The core of the SMNR process lies in its reconstruction objective, where multi-granularity graph layers must integrate features from the unmasked neighbors to predict the states of the masked nodes. Formally, for a masked node v_m, the reconstruction

loss $\mathcal{L}_{mask}$ is defined as the weighted residual between the predicted field $\hat{\mathbf{X}}_{v_m}$ and the ground-truth value $\mathbf{X}_{v_m}$:

$$\mathcal{L}_{mask} = \sum_{v_m \in \mathcal{V}_{mask}} \|\mathbf{X}_{v_m} - f_\theta(\mathbf{X}_{obs} \setminus \{v_m\} \mid \mathcal{G})\|_2^2 + \gamma \mathcal{R}_{topo} \tag{15}$$

where $\mathcal{R}_{topo}$ is a topological regularization term that ensures the reconstructed value is consistent with the local bathymetric gradient:

$$\mathcal{R}_{topo} = \lambda_p \sum_{v_m \in \mathcal{V}_{mask}} \sum_{j \in \mathcal{N}(v_m)} A_{mj} \|\hat{\mathcal{X}}_{v_m} - \hat{\mathcal{X}}_j\|^2, \tag{16}$$

where A_{mj} is the terrain-aware edge weight. This term encourages neighboring reconstructed states to remain compatible with the terrain-aware graph connectivity and reduces unconstrained feature diffusion across regions with strong bathymetric differences. This masking strategy effectively forces the hierarchical encoder to capture long-range physical dependencies rather than simply memorizing local pixel-level patterns.

To further strengthen the robustness of the latent space, we integrate a multi-granularity contrastive learning module within the self-supervised framework. For each reconstructed vertical profile, the model identifies positive pairs consisting of the same profile by applying different topological augmentations, such as varying masking densities or perturbed adjacency weights. On the other hand, profiles of unrelated dynamic regimes are regarded as negative samples. The contrastive loss $\mathcal{L}_{cl}$ is formulated using an InfoNCE-style objective:

$$\mathcal{L}_{cl} = -\mathbb{E}\left[\log \frac{\exp(\text{sim}(z_i, z_i^+)/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(z_i, z_j)/\tau)}\right] \tag{17}$$

where z_i and z_i^+ are the latent embeddings of the anchor and positive profile respectively, and τ is a temperature scaling parameter. This mechanism improves the discriminative capability of the latent representations across different environmental regimes, enabling the model to distinguish complex phenomena such as mesoscale eddies and stable background currents.

By combining the masked node reconstruction with multi-scale contrastive constraints, the MG-GCNN achieves self-regularization based on physical state information. This enables neurons within the collaborative layers to specifically handle specific physical regimes such as the high-gradient thermocline, without requiring manual layer-wise labeling. Therefore, the final output maintains high structural fidelity under sparse observations, providing a reliable and scalable solution for large-scale situational awareness in non-stationary underwater environments.

4. Experiments

4.1. Experimental Configuration and Datasets

To evaluate the robustness and fidelity of the proposed MG-GCNN framework, we conducted extensive experiments using a large-scale marine environmental archive focusing on the Pacific Ocean, named China Ocean Reanalysis Version 2 dataset (CORA2) [36]. The dataset covers the period from 1993 to 2023 and includes six coupled physical variables: sea surface height, temperature, salinity, and the three-component velocity vector (U, V, W). The vertical dimension is discretized into $D = 50$ non-uniform depth layers, providing a comprehensive representation of the ocean stratification structure from the surface mixed layer to the abyssal depths. Although the present implementation is trained on gridded reanalysis variables rather than raw satellite radiances, the benchmark fields

contain explicit satellite-derived surface constraints and provide a suitable test bed for satellite-assisted three-dimensional marine reconstruction.

To prevent temporal information leakage and to distinguish model selection from final evaluation, the dataset is divided chronologically into three non-overlapping periods. The complete 1993–2018 period is used as the development period for the models reported in this study. During model development, data from 1993–2016, corresponding to 8766 daily samples, are used for model fitting, while data from 2017–2018, corresponding to 730 samples, are used for validation, model selection, and hyperparameter tuning. After the model architecture and all hyperparameters are fixed, the final models used for the reported experiments are trained using the complete 1993–2018 development period, containing 9496 daily samples. The subsequent 2019–2023 period is kept completely unseen during model development and is retained as the independent test set containing 1826 samples for all quantitative comparisons reported in Section 4. Each temporal sample corresponds to one complete daily three-dimensional environmental field before the spatial masking operator is applied.

In addition to the oceanographic variables of CORA2, the model also incorporates high-resolution bathymetric data from the General Bathymetric Chart of the Oceans (GEBCO). The seafloor depth H is resampled to a $0.1^\circ \times 0.1^\circ$ resolution to match the horizontal grid of the Pacific research domain. This bathymetric information serves as the geometric backbone for the TAG mechanism, providing the necessary elevation change ΔH_{ij} to calculate the topographic impedance among nodes.

For each temporal sample, a spatial mask is independently generated over the valid horizontal ocean-grid nodes. At a masking ratio $r \in 25\%, 50\%, 75\%, 90\%, 99.9\%$, a fraction r of the available horizontal nodes is randomly removed from the model input, while the complete field is retained as the reconstruction reference. The masking pattern varies across temporal samples. To ensure a fair comparison, all competing methods receive exactly the same masked nodes for a given sample and masking ratio. Independent mask realizations are used across the five repeated runs. The Argo-like 3° configuration follows the structured sparse-sampling protocol described in Section 4.4 rather than random masking.

The MG-GCNN is implemented using the PyTorch 2.13.0 and optimized on a high-performance computing cluster equipped with four NVIDIA RTX 4090 GPUs. The training process is driven by the Adam optimizer with an initial learning rate of 1×10^{-4} and a weight decay coefficient of 1×10^{-5} to prevent overfitting in sparse data regimes. We employ a cosine annealing scheduler to dynamically adjust the step size over 200 epochs. The hyperparameters of the multi-objective loss are selected using the validation set to balance the contributions of the individual training objectives.

4.2. Baselines and Metrics

To quantify the reconstruction quality, we employ five standard evaluation metrics that assess both the statistical accuracy and the structural fidelity of the predicted fields.

1. Root Mean Square Error (RMSE) is the primary metric for numerical precision, quantifying the global deviation between the ground truth $\mathcal{X}$ and the reconstructed field $\hat{\mathcal{X}}$:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathcal{X}_i - \hat{\mathcal{X}}_i)^2} \tag{18}$$

2. Mean Absolute Error (MAE) provides a robust measure of the average residual magnitude without over-penalizing localized variations or outliers in the physical field:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\mathcal{X}_i - \hat{\mathcal{X}}_i| \tag{19}$$

3. Pearson Correlation Coefficient (PCC) assesses the linear correlation and the ability of the model to capture the spatiotemporal patterns of variability for each physical variable:

$$CC = \frac{\sum(\mathcal{X} - \bar{\mathcal{X}})(\hat{\mathcal{X}} - \bar{\hat{\mathcal{X}}})}{\sqrt{\sum(\mathcal{X} - \bar{\mathcal{X}})^2 \sum(\hat{\mathcal{X}} - \bar{\hat{\mathcal{X}}})^2}} \tag{20}$$

4. Normalized RMSE (NRMSE) is further introduced for cross-variable evaluation, which is defined for variable c as:

$$NRMSE_c = \frac{RMSE_c}{\sigma_c} \tag{21}$$

where σ_c is computed once from the valid 2019–2023 CORA2 test-reference fields and is shared by all evaluated methods. The resulting reference-field standard deviations are $\sigma_T = 3.5\,^{\circ}\mathrm{C}$, $\sigma_S = 0.4$ psu, $\sigma_U = 0.215$ m/s, $\sigma_V = 0.215$ m/s, and $\sigma_W = 8 \times 10^{-4}$ m/s. The cross-variable score is then calculated as

$$NRMSE_{avg} = \frac{1}{C}\sum_{c=1}^{C} NRMSE_c \tag{22}$$

Therefore, the reported Average NRMSE is dimensionless and does not average raw reconstruction errors expressed in different physical units.

5. Coefficient of determination (R-squared) indicates the proportion of the variance in the ground truth that is predictable from the reconstructed output, serving as a measure of the goodness of fit:

$$R^2 = 1 - \frac{\sum(\mathcal{X}_i - \hat{\mathcal{X}}_i)^2}{\sum(\mathcal{X}_i - \bar{\mathcal{X}})^2}. \tag{23}$$

All input variables are standardized using Z-score statistics estimated from the training set. Model optimization is performed in the standardized feature space; however, all variable-wise reconstruction errors reported in the quantitative tables are calculated after inverse normalization in the original physical space. RMSE, MAE, and PCC are computed from the pooled valid prediction–reference pairs within each run. In contrast, R^2 is first calculated independently for each temporal test sample and is then averaged over all test samples in that run. W remains in m/s throughout the metric computation, and the 10^{-4} factor is applied in the following presentations for readability. The reference standard deviations used for NRMSE are computed separately from the complete held-out CORA2 test-reference fields and are shared by all methods.

To rigorously benchmark the performance of the MG-GCNN, we select five state-of-the-art architectures released in recent years as baselines. These models represent the current frontier in data-driven oceanic and atmospheric modeling, offering diverse approaches to spatial-temporal feature extraction:

1. ASTMEN (2023) [37] is an adaptive spatiotemporal and multi-element fusion network for ocean surface currents forecasting, which uses an improved Swin Transformer (Swin-T) to perform self-attention computation at any given moment, enabling the adaptive generation of multi-element time series with spatial dependencies.

2. FourCastNet-V2 (2023) [38] uses Fourier transform to perform convolution operation in the spectral domain. This method is very effective for capturing the continuity of fluid dynamics and is optimized for global environmental prediction. We use this basic model to evaluate the performance of operational learning methods in reconstructing high-resolution features from sparse monitoring grids.
3. GraphCast (2024) [39] adopts a multi-scale map neural network architecture based on an icosahedral grid. By transferring information between nodes with different spatial resolutions, GraphCast can simulate complex interactions across different geographical scales. This basic model provides a reference for dealing with the irregular distribution of monitoring arrays and the non Euclidean relationship between marine environmental variables.
4. XiHe (2024) [40] is a hierarchical transformer-based framework coupled with two special designs. One is the land-ocean mask mechanism for focusing exclusively on the global ocean circulation. The other is the ocean-specific block for effectively capturing both local ocean information and global teleconnection.
5. 3-D-WaveMa (2025) [41] multivariate satellite observations are applied to reconstruct the daily 3-D SS in the Pacific Ocean by the novel wavelet-enhanced 3-D Mamba model. This model introduces the wavelet transform to spatially decompose multiscale ocean dynamic processes and efficiently captures the 3-D long-range dependencies with linear complexity through the proposed 3-D selective scanning module.

4.3. Full-Field Reconstruction Performance

The MG-GCNN framework demonstrates remarkable efficacy in restoring continuous 3D environmental fields from fragmented observational data. As summarized in Table 1, our model consistently outperforms all baselines in reconstructing foundational scalar variables. For the temperature field (T), MG-GCNN achieves a state-of-the-art RMSE of 0.298 °C at 50% sparsity, representing a significant accuracy gain compared to the leading graph-based competitor, 3-D-WaveMa. Similar trends are observed in the salinity (S) field, where the model achieves an R^2 of 0.934, indicating strong agreement between the reconstructed and reference salinity fields. When examining the dynamical current fields (U, V, W) under extreme 75% sparsity, the performance gap becomes even more pronounced, as shown in Table 2. Notably, for the vertical velocity (W) with its typical characteristics of high-frequency fluctuations and small numerical magnitudes, the MG-GCNN achieves a Pearson Correlation Coefficient (PCC) of 0.892. This suggests that the model not only performs numerical smoothing on sparse inputs but also successfully infers the complex dynamical signals that are lost in traditional Euclidean-based interpolation methods.

Figure 5 provides a qualitative comparison between MG-GCNN and 3-D-WaveMa at a depth of 100 m under 75% spatial masking. The reference field contains pronounced mesoscale temperature structures and sharp frontal gradients. Compared with 3-D-WaveMa, MG-GCNN preserves these spatial gradients more clearly and exhibits smaller localized reconstruction residuals in the displayed example. This visualization is consistent with the quantitative performance differences reported in the corresponding tables.

In addition to point-wise reconstruction accuracy, MG-GCNN preserves coherent spatial structures across the reconstructed environmental variables. This behavior is consistent with the use of shared terrain-aware and multi-granularity representations. Traditional CNN-based methods usually suffer from a checkerboard effect in sparse scenes, which destroys the vertical stability of the water column. In contrast, MG-GCNN, based on graph message passing, ensures that the reconstructed values of neighboring nodes are constrained by the shared topological context.

Table 1. Quantitative comparison of reconstruction performance for temperature and salinity at 50% sparsity.

Method	Variable	RMSE	MAE	PCC	R^2
ASTMEN	Temp	0.415 ± 0.012 °C	0.288 ± 0.009 °C	0.941 ± 0.004	0.878 ± 0.007
FourCastNet-V2	Temp	0.432 ± 0.014 °C	0.305 ± 0.010 °C	0.933 ± 0.005	0.861 ± 0.009
GraphCast	Temp	0.386 ± 0.011 °C	0.265 ± 0.008 °C	0.957 ± 0.003	0.912 ± 0.006
XiHe	Temp	0.392 ± 0.012 °C	0.270 ± 0.008 °C	0.949 ± 0.004	0.898 ± 0.007
3-D-WaveMa	Temp	0.372 ± 0.010 °C	0.251 ± 0.007 °C	0.963 ± 0.003	0.924 ± 0.005
MG-GCNN (Ours)	Temp	0.298 ± 0.008 °C	0.195 ± 0.006 °C	0.982 ± 0.002	0.959 ± 0.004
ASTMEN	Salinity	0.081 ± 0.004 psu	0.053 ± 0.003 psu	0.916 ± 0.006	0.833 ± 0.010
FourCastNet-V2	Salinity	0.091 ± 0.005 psu	0.061 ± 0.004 psu	0.901 ± 0.007	0.801 ± 0.011
GraphCast	Salinity	0.075 ± 0.004 psu	0.048 ± 0.003 psu	0.930 ± 0.005	0.860 ± 0.008
XiHe	Salinity	0.078 ± 0.004 psu	0.050 ± 0.003 psu	0.923 ± 0.006	0.847 ± 0.009
3-D-WaveMa	Salinity	0.069 ± 0.003 psu	0.043 ± 0.002 psu	0.941 ± 0.004	0.881 ± 0.007
MG-GCNN (Ours)	Salinity	0.052 ± 0.003 psu	0.031 ± 0.002 psu	0.971 ± 0.003	0.934 ± 0.005

Table 2. Quantitative comparison of reconstruction performance for dynamical variables at 75% sparsity.

Method	Variable	RMSE (m/s)	PCC	R^2
ASTMEN	U, V	0.128 ± 0.006	0.810 ± 0.010	0.648 ± 0.014
FourCastNet-V2	U, V	0.141 ± 0.007	0.782 ± 0.012	0.603 ± 0.017
GraphCast	U, V	0.115 ± 0.005	0.838 ± 0.009	0.696 ± 0.013
XiHe	U, V	0.123 ± 0.006	0.825 ± 0.010	0.676 ± 0.014
3-D-WaveMa	U, V	0.107 ± 0.005	0.856 ± 0.008	0.728 ± 0.012
MG-GCNN (Ours)	U, V	0.076 ± 0.004	0.925 ± 0.006	0.852 ± 0.009
Method	**Variable**	**RMSE (10^{-4} m/s)**	**PCC**	R^2
ASTMEN	W	1.55 ± 0.09	0.721 ± 0.015	0.512 ± 0.019
FourCastNet-V2	W	1.78 ± 0.11	0.682 ± 0.018	0.460 ± 0.022
GraphCast	W	1.31 ± 0.08	0.761 ± 0.013	0.574 ± 0.017
XiHe	W	1.42 ± 0.09	0.745 ± 0.014	0.551 ± 0.018
3-D-WaveMa	W	1.14 ± 0.07	0.785 ± 0.012	0.612 ± 0.015
MG-GCNN (Ours)	W	0.78 ± 0.05	0.892 ± 0.008	0.789 ± 0.011

To complement the controlled reconstruction benchmark against CORA2 reference fields, we further conducted an Argo-based observational consistency evaluation using quality-controlled temperature and salinity profiles obtained from the Argo Global Data Assembly Centres for the Pacific Ocean during the held-out 2019–2023 test period. After quality control and spatiotemporal matching, 312,684 Argo profiles were retained for evaluation. The complete CORA2 fields and MG-GCNN reconstructions were collocated with the Argo measurements according to longitude, latitude, depth, and acquisition time. The Argo profiles were not provided as direct inputs to MG-GCNN during reconstruction. However, Argo temperature–salinity measurements are among the in situ observational sources assimilated by CORA2.

As summarized in Table 3, under 50% spatial masking, MG-GCNN achieves a temperature RMSE of 0.910 °C and a PCC of 0.939 against the collocated Argo observations, compared with 0.840 °C and 0.951 for the complete CORA2 field. For salinity, MG-GCNN obtains an RMSE of 0.163 psu and a PCC of 0.913, while CORA2 obtains 0.151 psu and 0.928, respectively. The modest degradation relative to the complete CORA2 field is expected

because MG-GCNN reconstructs the environmental field from only half of the available spatial input nodes. These results indicate that the reconstruction retains a substantial fraction of the observation-consistent information contained in the complete CORA2 field.

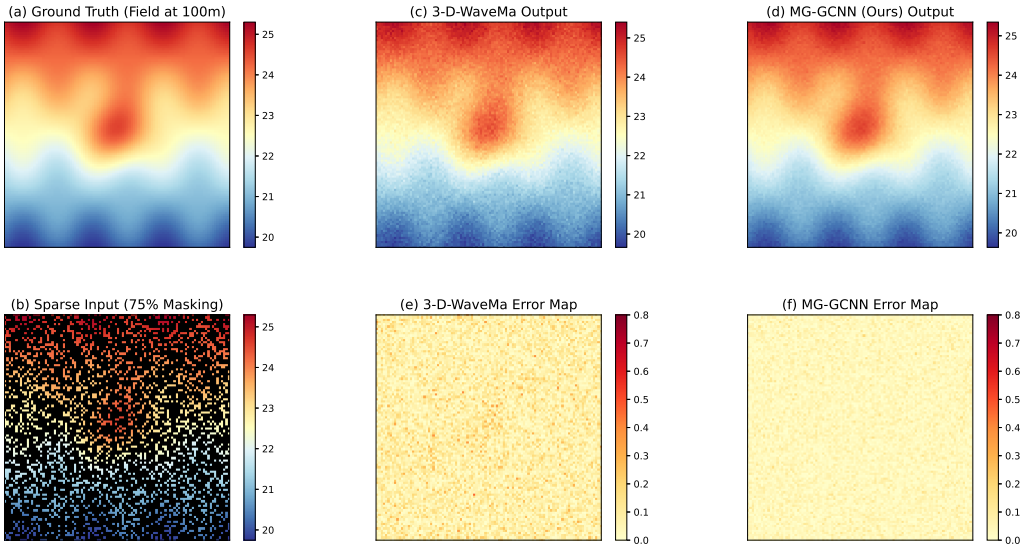


Figure 5. Visualization of temperature reconstruction at 100 m depth under 75% spatial masking.

Table 3. Argo-based observational consistency evaluation of the CORA2 reanalysis and MG-GCNN reconstruction under 50% spatial masking.

Evaluated Field	Temp. RMSE (°C)	Temp. PCC	Salinity RMSE (psu)	Salinity PCC
CORA2 reanalysis	0.840 ± 0.060	0.951 ± 0.006	0.151 ± 0.011	0.928 ± 0.008
MG-GCNN reconstruction	0.910 ± 0.070	0.939 ± 0.008	0.163 ± 0.013	0.913 ± 0.010

4.4. Robustness to Observational Sparsity

The robustness of environmental perception under different levels of sensor availability is a key metric for marine operating systems. As illustrated in Figure 6, we evaluate all models under 25%, 50%, 75%, 90%, and 99.9% random spatial masking. For each test sample, the same observation mask is applied to all compared methods, and the reported results are averaged over five independent realizations. In addition to random masking, we introduce an Argo-like sampling configuration with an approximately 3° horizontal spacing. This setting is intended to approximate the nominal spatial density of a basin-scale float observing network and therefore provides a more realistic sparsity test than masking a dense 0.1° grid alone. We emphasize that this configuration approximates the spatial sampling density rather than reproducing the complete time-dependent trajectories and sampling schedules of individual Argo floats.

As shown in Table 4, reconstruction performance deteriorates monotonically as the available observations become sparser, whereas the average NRMSE increases accordingly. This substantial increase confirms that the high performance observed at moderate masking ratios cannot be attributed solely to the model being insensitive to observation density. Nevertheless, MG-GCNN remains consistently more robust than the competing approaches. These results indicate that graph-based multi-granularity aggregation remains beneficial when local observational support is greatly reduced.

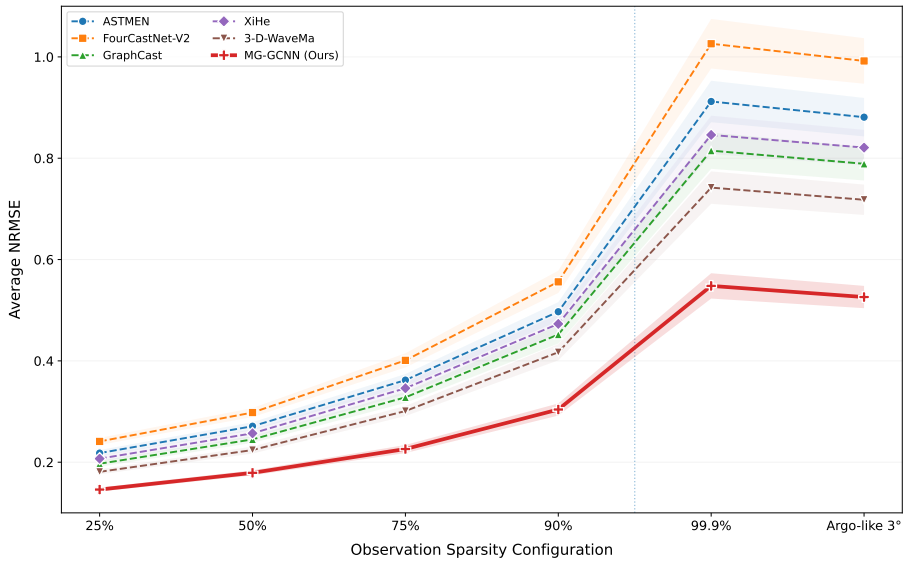


Figure 6. Average NRMSE across different observation sparsity configurations.

Table 4. Average NRMSE across all physical variables at varying sparsity levels.

Method	25% Random	50% Random	75% Random	90% Random	99.9% Random	Argo-like 3°
ASTMEN	0.218 ± 0.006	0.271 ± 0.008	0.362 ± 0.012	0.497 ± 0.019	0.912 ± 0.041	0.881 ± 0.03
FourCastNet-V2	0.241 ± 0.007	0.298 ± 0.009	0.401 ± 0.014	0.556 ± 0.022	1.026 ± 0.049	0.992 ± 0.04
GraphCast	0.197 ± 0.006	0.245 ± 0.007	0.328 ± 0.011	0.452 ± 0.018	0.815 ± 0.036	0.789 ± 0.03
XiHe	0.207 ± 0.006	0.257 ± 0.008	0.346 ± 0.012	0.473 ± 0.019	0.846 ± 0.038	0.821 ± 0.03
3-D-WaveMa	0.181 ± 0.005	0.224 ± 0.007	0.301 ± 0.010	0.417 ± 0.016	0.742 ± 0.032	0.718 ± 0.03
MG-GCNN (Ours)	0.146 ± 0.004	0.179 ± 0.005	0.226 ± 0.008	0.304 ± 0.012	0.548 ± 0.025	0.526 ± 0.02

To further dissect the impact of sparsity on the vertical dimension, we present a 3D error evolution heatmap in Figure 7. This visualization provides a granular view of how the reconstruction fidelity fluctuates with water depth and masking intensity. It is evident that the deep sea regions remain stable throughout and are characterized by a persistent lower error zone even under high masking ratios. This indicates that MG-GCNN successfully learns the stable large-scale background state of the abyssal environment. Instead, the most significant error evolution occurs within the thermocline, where the physical gradients are most intense, at a depth of about 100 m to 1000 m. This region shows a local spread of a higher error peak as the masking ratio increases.

As shown in Figure 6, reconstruction performance degrades progressively as observation sparsity increases. MG-GCNN consistently achieves the lowest average NRMSE across all evaluated configurations. These results indicate that MG-GCNN maintains comparatively stronger reconstruction robustness as observational support becomes increasingly sparse. This reduction in reconstruction error variance is a direct result of the self-supervised masked node reconstruction module, which acts as a powerful regularizer during the training phase. By forcing the latent space to reconstruct masked features from their unmasked neighbors, the model develops an intrinsic understanding of the environmental manifold. This ensures that the reconstructed values are constrained by the learned physical consensus even in the absence of direct local observations, reducing the risk of producing unrealistic environmental states in a sparse monitoring grid.

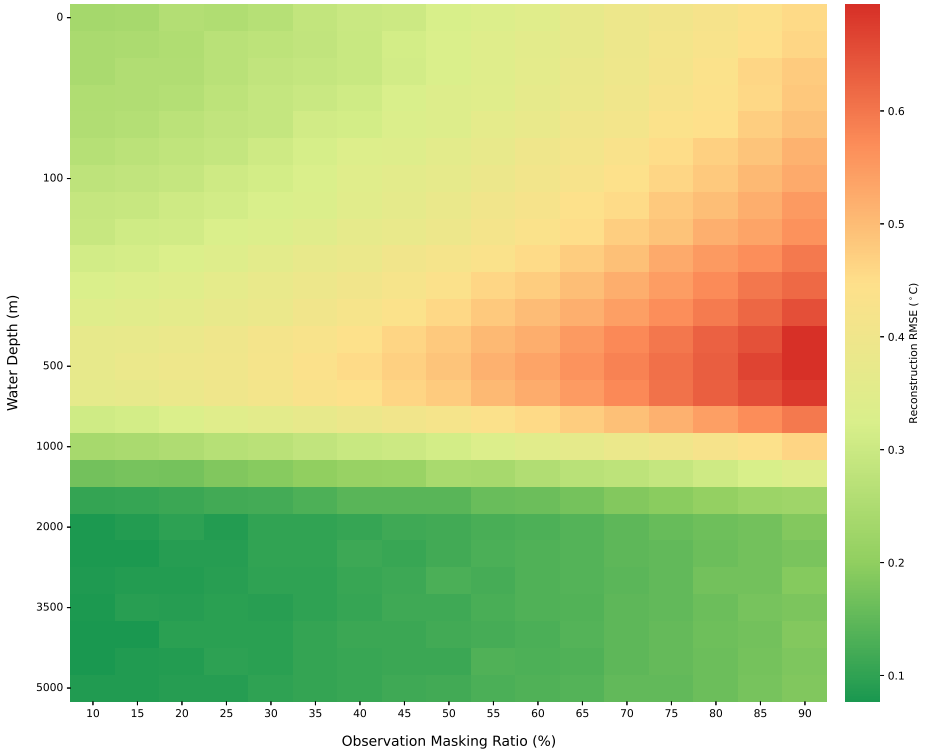


Figure 7. Reconstruction error distribution across depth and sparsity levels.

4.5. Topographic Fidelity Analysis

The core innovation driving this vertical stability is the terrain-aware graph (TAG) mechanism. The comparative analysis of topological connectivity in Figure 8 provides a direct visualization of how MG-GCNN incorporates bathymetric priors. Unlike the distance-only approach, where rigid connections are established based only on geographic proximity, the TAG mechanism adaptively attenuates edge weights in regions with sharp topographic transitions. For instance, when two nodes are separated by a submerged ridge or steep slope, the TAG mechanism identifies the topographic impedance and attenuates their mutual influence. This prevents the error diffusion of environmental features across physical barriers, which is a common source of bias in standard interpolation methods. Figure 8 provides a structural visualization of how bathymetric information influences the learned graph topology, where the generated connectivity pattern reflects the underlying seafloor morphology, and its connectivity patterns are consistent with the underlying seafloor morphology. The integration of topographic fidelity is particularly evident in the reconstruction of the vertical velocity and horizontal currents. The terrain-aware graph modifies information exchange across regions with pronounced bathymetric gradients by reducing graph connectivity across large topographic differences.

The vertical dimension of the ocean is a unique challenge due to the distinct nonstationarity of physical variables at different depth layers. As illustrated in the vertical RMSE profiles in Figure 9, all reconstruction models encounter a performance bottleneck within the thermocline. In this high-gradient zone, the rapid decrease in temperature and salinity often leads to smoothing errors in the traditional Euclidean-based models. However, in contrast to the sharp error peaks displayed by 3-D-WaveMa and GraphCast, MG-GCNN maintains a significantly stable error distribution. The proposed framework employs a hierarchical graph collaborative architecture to effectively decouple the complex signals of the mixed layer from the relatively stable deep-sea background signals. This ensures that the intense vertical gradients are preserved during the message-passing process rather than

averaging them out, allowing high-fidelity hierarchical recovery across all 50 non-uniform depth layers.

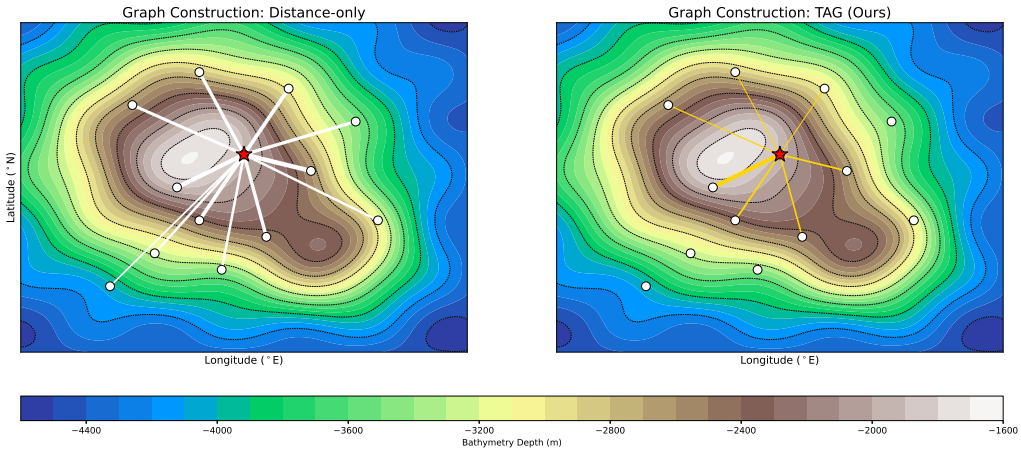


Figure 8. Comparative analysis of topological connectivity under bathymetric prior. The red star and white dot indicate the data points to be predicted and the existing data points, while the line represent the edges that have been constructed in the graph.

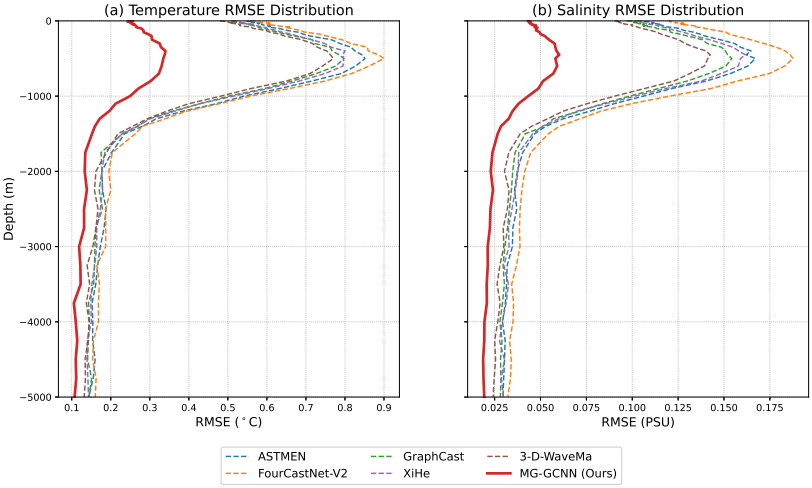


Figure 9. Comparison of vertical RMSE distribution across 50 layers for temperature and salinity. The depth follows the non-uniform stretched grid of the ocean archive. While baseline models exhibit a significant performance bottleneck within the high-gradient thermocline (−100 m to −1000 m), MG-GCNN maintains consistent accuracy, validating the effectiveness of the hierarchical graph collaborative blocks.

4.6. Ablation Study

To rigorously quantify the contribution of each module and the individual loss terms to the overall performance, we conduct a series of ablation experiments at a standard 50% masking ratio. As demonstrated in Table 5, the transition from a vanilla GCN baseline to the full MG-GCNN framework leads to a 43.2% relative reduction in temperature RMSE compared with the Vanilla GCN baseline. Removing the TAG module leads to the most significant performance degradation, particularly in the reconstruction of current fields. This indicates that bathymetric priors provide a useful structural constraint for terrain-aware graph construction. In addition, excluding the multi-granularity collaborative architecture results in an increase in temperature RMSE from 0.298 to 0.385, indicating that multi-granularity representation contributes to improved reconstruction of spatial structures at different scales. Together with the TAG, multi-granularity, and SMNR

ablations, these results indicate that the architectural modules and auxiliary objectives provide complementary improvements rather than being redundant components.

Table 5. Performance comparison of MG-GCNN variants at 50% sparsity.

Variant	T RMSE (°C)	S RMSE (psu)	PCC (U, V)	R^2 (U, V)	T RMSE Reduction
Vanilla GCN	0.525 ± 0.015	0.110 ± 0.006	0.798 ± 0.011	0.628 ± 0.016	Reference
w/o TAG	0.412 ± 0.012	0.085 ± 0.005	0.845 ± 0.009	0.704 ± 0.014	21.5%
w/o Multi-granularity	0.385 ± 0.011	0.072 ± 0.004	0.868 ± 0.008	0.744 ± 0.012	26.7%
w/o VFI	0.371 ± 0.010	0.069 ± 0.004	0.879 ± 0.008	0.762 ± 0.011	29.3%
w/o SMNR	0.342 ± 0.010	0.064 ± 0.004	0.902 ± 0.007	0.806 ± 0.010	34.9%
w/o L_{topo}	0.326 ± 0.009	0.060 ± 0.003	0.916 ± 0.006	0.831 ± 0.009	37.9%
w/o L_{cl}	0.318 ± 0.009	0.058 ± 0.003	0.923 ± 0.006	0.845 ± 0.008	39.4%
MG-GCNN (Full)	0.298 ± 0.008	0.052 ± 0.003	0.945 ± 0.005	0.887 ± 0.007	43.2%

In addition to the numerical gains, the multi-granularity weight radar chart in Figure 10 provides a descriptive view of the aggregation weights learned by the multi-granularity module. Temperature and salinity exhibit relatively larger contributions from coarse-scale features, whereas vertical velocity places greater weight on fine-scale features. These patterns are qualitatively consistent with the different spatial structures of the reconstructed variables and provide limited architectural insight into the model's feature aggregation behavior.

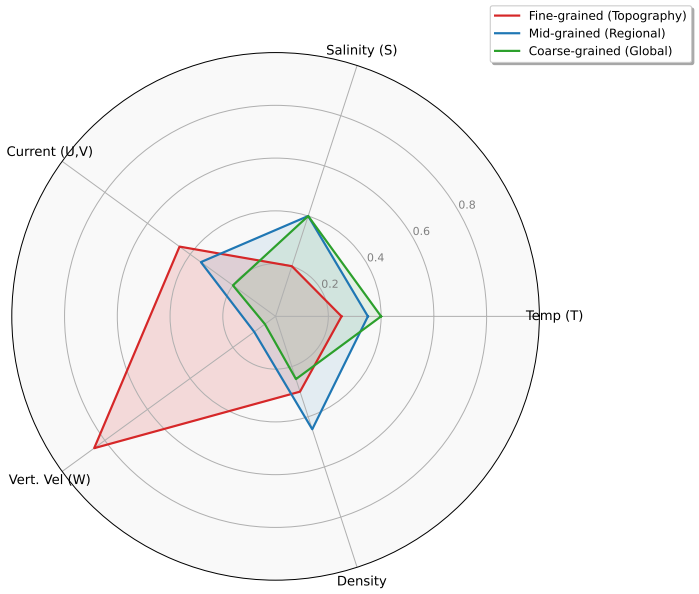


Figure 10. Multi-granularity feature aggregation weights for different physical variables.

The performance improvements of MG-GCNN can be mainly attributed to the complementary design of its principal components. The terrain-aware graph incorporates bathymetric information into spatial message passing, while VFI explicitly encodes vertical-profile dependencies. The multi-granularity architecture provides representations at different spatial scales, and SMNR improves reconstruction robustness when observations are missing.

5. Conclusions

This study proposes MG-GCNN for reconstructing high-fidelity 3D marine environmental fields from sparse and irregular observations. By combining terrain-aware dynamic graph construction, horizontal-vertical collaborative representation, multi-granularity aggregation, and self-supervised masked node reconstruction, our framework preserves both large-scale ocean structures and local thermocline and topographic gradients. Experiments

using CORA2 reanalysis and GEBCO sounding show improved reconstruction accuracy and robustness under severe observation sparsity.

The proposed framework also establishes a clear link between satellite-assisted ocean observations and ocean acoustic remote sensing. CORA2 combines satellite SST and altimeter sea level anomalies and in situ temperature-salinity profiles. The generated 3D temperature, salinity and current fields are reconstructed by MG-GCNN, from which the sound velocity profile and sound propagation environment can be constructed. Satellite-derived information is incorporated indirectly through the data-assimilation system of the CORA2 reanalysis dataset rather than through independent raw satellite image channels. The reconstructed environmental fields may provide physically relevant inputs for subsequent ocean-acoustic propagation and underwater sensing models.

Several practical limitations and extensions should also be considered. First, the current implementation uses fixed graph-construction parameters throughout the study domain, whereas adaptive neighborhood selection may further improve the flexibility of graph representation in regions with different spatial characteristics. Second, the present experiments are conducted on a predefined 0.1° horizontal grid with 50 vertical layers; extending the framework to variable-resolution or multi-resolution grids would further enhance its applicability. Finally, the current implementation is designed primarily for offline environmental reconstruction, and further optimization of memory usage and inference efficiency may facilitate deployment over larger spatial domains.

Author Contributions: Conceptualization, B.G. and Y.G.; Methodology, B.G. and Y.G.; Software, Y.G.; Validation, B.G. and Y.G.; Formal analysis, B.G.; Investigation, Y.G.; Resources, Y.G.; Data curation, B.G.; Writing—original draft, B.G.; Writing—review & editing, Y.G.; Visualization, B.G. and Y.G.; Supervision, B.G.; Project administration, B.G.; Funding acquisition, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhao, K.; Wei, Y.; Dong, J.; Zhao, P.; Wang, Y.; Pan, X.; Wang, J. Separation and characterization of microplastic and nanoplastic particles in marine environment. *Environ. Pollut.* **2022**, *297*, 118773. [CrossRef] [PubMed]
2. Harishidayat, D.; Al-Shuhail, A.; Randazzo, G.; Lanza, S.; Muzirafuti, A. Reconstruction of land and marine features by seismic and surface geomorphology techniques. *Appl. Sci.* **2022**, *12*, 9611. [CrossRef]
3. Evagorou, E.; Hasiotis, T.; Petsimeris, I.T.; Monioudi, I.N.; Andreadis, O.P.; Chatzipavlis, A.; Christofi, D.; Kountouri, J.; Stylianou, N.; Mettas, C.; et al. A holistic high-resolution remote sensing approach for mapping coastal geomorphology and marine habitats. *Remote Sens.* **2025**, *17*, 1437. [CrossRef]
4. Desmars, N.; Hartmann, M.; Behrendt, J.; Hoffmann, N.; Klein, M. Nonlinear deterministic reconstruction and prediction of remotely measured ocean surface waves. *J. Fluid Mech.* **2023**, *975*, A8. [CrossRef]
5. Chen, W.; Nguyen, T.T.N.; Pein, J.; Jourdin, F.; Fablet, R.; Staneva, J. Physics-informed neural data assimilation for high-resolution coastal SPM reconstruction from model and satellite data. *Appl. Ocean Res.* **2025**, *165*, 1. [CrossRef]
6. Xu, L.; Xu, J.; Zhang, X.; Zhang, A.; Liu, Y.; Chen, D.; Chen, L.; Wu, Z. Eddy-induced underwater acoustic field reconstruction and computation based on sound speed classification and B-spline surface fitting. *Front. Mar. Sci.* **2025**, *12*, 1583080. [CrossRef]
7. Liu, Q.; Bao, S.; Yan, H.; Wang, H.; Wang, Y.; Zhang, R. Enhancing sea surface salinity short-term prediction using physically informed deep learning. *Appl. Ocean Res.* **2025**, *165*, 104832. [CrossRef]
8. Zhou, J.; Liang, T.; Zhang, D.; He, Z. Waterhe-nerf: Water-ray tracing neural radiance fields for underwater scene reconstruction. *arXiv* **2023**, arXiv:2312.06946.
9. Madiligama, M.; Zou, Z.; Zhang, L. Leveraging satellite observations and machine learning for underwater sound speed estimation. *Commun. Eng.* **2025**, *4*, 126. [CrossRef] [PubMed]

10. Liu, Y.; Wang, H.; Jiang, F.; Zhou, Y.; Li, X. Reconstructing 3-D thermohaline structures for mesoscale eddies using satellite observations and deep learning. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4203916. [CrossRef]
11. Song, T.; Xu, G.; Yang, K.; Li, X.; Peng, S. Convformer: A model for reconstructing ocean subsurface temperature and salinity fields based on multi-source remote sensing observations. *Remote Sens.* **2024**, *16*, 2422. [CrossRef]
12. Wang, S.; Zhang, X.; Qin, Y.; Song, W.; Li, B. Marine target magnetic anomaly detection based on multitask deep transfer learning. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1501705. [CrossRef]
13. Santos, J.E.; Fox, Z.R.; Mohan, A.; O'Malley, D.; Viswanathan, H.; Lubbers, N. Development of the senseiver for efficient field reconstruction from sparse observations. *Nat. Mach. Intell.* **2023**, *5*, 1317–1325. [CrossRef]
14. Hu, Q.; Chen, X.; Bai, Y.; He, X.; Li, T.; Pan, D. Reconstruction of 3-D ocean chlorophyll a structure in the northern Indian ocean using satellite and BGC-argo data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *61*, 4200513. [CrossRef]
15. Matranga, V.; Corsi, I. Toxic effects of engineered nanoparticles in the marine environment: Model organisms and molecular approaches. *Mar. Environ. Res.* **2012**, *76*, 32–40. [CrossRef] [PubMed]
16. Gao, Y.; Gong, M.; Xie, Y.; Qin, A.K.; Pan, K.; Ong, Y.S. Multiparty dual learning. *IEEE Trans. Cybern.* **2022**, *53*, 2955–2968. [CrossRef] [PubMed]
17. Kim, J.; Park, J. Mathematical modeling of coastal marine environments using observational data for coastal management. *Ocean Coast. Manag.* **2015**, *116*, 396–403. [CrossRef]
18. Gong, M.; Gao, Y.; Wu, Y.; Zhang, Y.; Qin, A.K.; Ong, Y.S. Heterogeneous multi-party learning with data-driven network sampling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13328–13343. [CrossRef] [PubMed]
19. Sahlin, J.; Mostafavi, M.A.; Forest, A.; Babin, M. Assessment of 3D spatial interpolation methods for study of the marine pelagic environment. *Mar. Geod.* **2014**, *37*, 238–266. [CrossRef]
20. Wang, J.; Dong, J.; Tang, M.; Yao, J.; Li, X.; Kong, D.; Zhao, K. Identification and detection of microplastic particles in marine environment by using improved faster R–CNN model. *J. Environ. Manag.* **2023**, *345*, 118802. [CrossRef] [PubMed]
21. Thorson, J.T.; Cheng, W.; Hermann, A.J.; Ianelli, J.N.; Litzow, M.A.; O'Leary, C.A.; Thompson, G.G. Empirical orthogonal function regression: Linking population biology to spatial varying environmental conditions using climate projections. *Glob. Change Biol.* **2020**, *26*, 4638–4649. [CrossRef] [PubMed]
22. Laws, K.; Vesecky, J.; Paduan, J. Monitoring coastal vessels for environmental applications: Application of Kalman filtering. In Proceedings of the 2011 IEEE/OES 10th Current, Waves and Turbulence Measurements (CWTM), Monterey, CA, USA, 20–23 March 2011; pp. 39–46.
23. e Silva, L.O.; Resende, M.; Galhardas, H.; Manquinho, V.; Lynce, I. DeepData: Machine learning in the marine ecosystems. *Expert Syst. Appl.* **2022**, *206*, 117841. [CrossRef]
24. Vincent, A.N.; Sakthidasan, K.; Sunhaloo, M.S.; Clement, J.C.; Laloo, N.; Bagurubumwe, U. Comparative analysis of RNN, LSTM and CNN algorithms for marine data prediction. *J. Coast. Conserv.* **2025**, *29*, 73. [CrossRef]
25. Shu, X.; Xu, B.; Zhang, L.; Tang, J. Multi-granularity anchor-contrastive representation learning for semi-supervised skeleton-based action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 7559–7576. [CrossRef] [PubMed]
26. Xu, R.; Cheng, H.; Guo, C.; Gao, H.; Hu, J.; Yang, S.B.; Yang, B. Mm-path: Multi-modal, multi-granularity path representation learning. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1, Toronto, ON, Canada, 3–7 August 2025; pp. 1703–1714.
27. Wang, F.; Zhou, Y.; Wang, S.; Vardhanabhuti, V.; Yu, L. Multi-granularity cross-modal alignment for generalized medical visual representation learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 33536–33549. [CrossRef]
28. Yang, J.; Jiang, M.; Yang, S.; Tan, X.; Li, Y.; Ding, E.; Wang, J.; Wang, H. Mgmapnet: Multi-granularity representation learning for end-to-end vectorized hd map construction. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; Volume 2025, pp. 86275–86286.
29. Li, L.; Zhang, Y.; Wang, G.; Xia, K. Kolmogorov–Arnold graph neural networks for molecular property prediction. *Nat. Mach. Intell.* **2025**, *7*, 1346–1354. [CrossRef]
30. Zheng, X.; Wang, Y.; Liu, Y.; Li, M.; Zhang, M.; Jin, D.; Yu, P.S.; Pan, S. Graph neural networks for graphs with heterophily: A survey. *IEEE Trans. Knowl. Data Eng.* **2026**, *38*, 4385–4404. [CrossRef]
31. Chen, L.; Pan, Z.; Liu, Q.; Hu, P. Hrrpgraphnet++: Dynamic graph neural network with meta-learning for few-shot hrrp radar target recognition. *Remote Sens.* **2025**, *17*, 2108. [CrossRef]
32. Yang, X.; Li, C.; Wang, Z.; Xie, H.; Mao, J.; Yin, G. Remote sensing cross-modal text-image retrieval based on attention correction and filtering. *Remote Sens.* **2025**, *17*, 503. [CrossRef]
33. Liu, J.; He, J.; Chen, H.; Yang, R.; Huang, Y. A lightweight semantic-and graph-guided network for advanced optical remote sensing image salient object detection. *Remote Sens.* **2025**, *17*, 861. [CrossRef]
34. Cui, H.; Su, Y.; Dean, T.J.; Yu, T.; Zhang, Z.; Peng, J.; Shukla, D.; Zhao, H. Enzyme specificity prediction using cross-attention graph neural networks. *Nature* **2025**, *647*, 639–647. [CrossRef] [PubMed]

35. Zhao, X.; Ma, J.; Wang, L.; Zhang, Z.; Ding, Y.; Xiao, X. A review of hyperspectral image classification based on graph neural networks. *Artif. Intell. Rev.* **2025**, *58*, 172. [CrossRef]
36. Fu, H.; Dan, B.; Gao, Z.; Wu, X.; Chao, G.; Zhang, L.; Zhang, Y.; Liu, K.; Zhang, X.; Li, W. Global ocean reanalysis CORA2 and its inter comparison with a set of other reanalysis products. *Front. Mar. Sci.* **2023**, *10*, 1084186. [CrossRef]
37. Li, X.; Wang, F.; Song, T.; Meng, F.; Zhao, X. ASTMEN: An adaptive spatiotemporal and multi-element fusion network for ocean surface currents forecasting. *Front. Mar. Sci.* **2023**, *10*, 1281387. [CrossRef]
38. Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv* **2022**, arXiv:2202.11214.
39. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef] [PubMed]
40. Wang, X.; Wang, R.; Hu, N.; Wang, P.; Huo, P.; Wang, G.; Wang, H.; Wang, S.; Zhu, J.; Xu, J.; et al. Xihe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv* **2024**, arXiv:2402.02995.
41. Liang, Z.; Bao, S.; Su, H.; Zhang, W.; Wang, H.; Yan, H. Daily Subsurface Salinity Reconstruction From Multisource Satellite Observations Using Wavelet-Enhanced 3-D Mamba. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *19*, 3371–3388. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.