MF-STGCN: Multimodal Fusion Sparse Spatio-Temporal Graph Networks with Physics-Aware Gating for Ocean Eddy Forecasting

Yanling Du[1], Zhenli Qian[1], Zhihong Huang[1], Ke Chen[2], Dongmei Huang[3], Qi He[1,*]

[1]*College of Information Technology, Shanghai Ocean University, Shanghai 201306, China*

[2]*East China Sea Forecasting and Disaster Reduction Center, Ministry of Natural Resources, Shanghai 201306, China*

[3]*College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 201306, China*

yldu@shou.edu.cn, m240751950@st.shou.edu.cn, 2116492721h@gmail.com,
chenke@ecs.mnr.gov.cn, dmhuang@shou.edu.cn, qihe@shou.edu.cn

***Abstract*—Accurate prediction of the trajectories of mesoscale ocean eddies is imperative for comprehending oceanic dynamic processes and ensuring offshore engineering safety. However, extant methodologies generally treat eddies as isolated systems, thereby constraining their capacity to capture the intricate and sparse interactions among multiple coexisting eddies. Moreover, unimodal trajectory data frequently prove inadequate in fully characterizing the nonlinear evolutionary patterns of eddies, which are subject to modulation by their intrinsic physical features as well as the surrounding environmental context. To address these challenges, this paper proposes the multimodal fusion sparse spatio-temporal graph convolutional network (MF-STGCN). Initially, the model designs a physical encoder and an adaptive gating fusion mechanism to dynamically adjust the contribution weights of physical features and environmental information. This approach effectively addresses the issues of multivariate heterogeneous data alignment and feature enhancement. Secondly, a sparse graph learning strategy is proposed based on asymmetric convolution and zero-softmax normalization. It enables the automatic mining of directed interactions with physical significance while effectively eliminating redundant noise. Finally, spatio-temporal features are aggregated in parallel through a dual-stream spatio-temporal graph convolutional network (STGCN) and a temporal convolutional network (TCN) to achieve trajectory prediction. Evaluated on the 1993–2022 South China Sea (SCS) mesoscale eddy dataset, MF-STGCN outperforms Transformer, ETPNet, and existing GNN baselines across multiple metrics. Visualization analysis further confirms the model's remarkable robustness and superiority in handling abrupt trajectory shifts and long-term predictions.**

***Index Terms*—Mesoscale eddies, spatio-temporal trajectory prediction, graph convolutional networks, sparse interaction, and multimodal fusion.**

I. Introduction

Oceanic mesoscale eddies carry approximately 90% of the kinetic energy in the ocean, serving as the core vehicle for global energy cascading and material transport [1], [2]. Their dynamic behavior directly governs the ocean's heat transport to polar regions, influencing the global heat balance pattern [3]. Simultaneously, through vertical pumping and horizontal transport, eddies regulate the distribution of nutrients and plankton, playing a crucial role in fisheries resource management and global carbon sink assessments [4]. Furthermore, the intense flow fields associated with eddies pose significant operational safety risks to deep-sea drilling, subsea pipeline deployment, and underwater vehicle path planning [1]. Consequently, accurate prediction of mesoscale eddy trajectories is not only a core scientific issue for elucidating oceanic dynamic processes but also an urgent requirement with substantial application value in fields such as climate forecasting and ocean engineering.

With the growing availability of satellite remote sensing and reanalysis data, using deep learning to mine historical patterns for mesoscale eddy trajectory prediction has emerged as a promising research frontier. The research paradigm is evolving from early single-source data-driven approaches toward multimodal feature fusion, and further toward physics-enhanced mechanisms. Early studies focused primarily on verifying the feasibility of deep learning methods, typically treating individual eddies as isolated time series.

Specifically, Ma et al. [5] pioneered the application of ConvLSTM networks to this task, achieving indirect eddy forecasting by predicting sea level anomaly (SLA) fields combined with detection algorithms. Subsequently, Nian et al. [6] introduced a memory-augmented network and a scheduled sampling strategy to achieve precise prediction of high-order non-stationary spatiotemporal dynamics using solely SLA data. However, relying exclusively on a single SLA source often fails to capture the complex physical mechanisms underlying eddy evolution, thereby hindering further breakthroughs in predictive accuracy. According to the global observational study by Chelton et al. [2], there is a significant correlation between eddy propagation trajectories and their multi-dimensional features. Specifically, the propagation speed and direction are governed not only by geostrophic parameters but also by the modulation of the eddy's own amplitude and spatial scale. Further research by Early et al. [7] indicates that under quasi-stable states, the zonal and meridional propagation velocities are closely related to the amplitude intensity. Such highly nonlinear dependencies among multiple variables pose

* Corresponding author: Qi He, email:qihe@shou.edu.cn

severe challenges to single-trajectory modeling.

To address these challenges, the research paradigm has begun to evolve toward the multimodal feature fusion stage, aiming to enhance the model's perception by incorporating additional features and environmental information. For instance, Wang et al. [8] integrated multidimensional features—such as eddy trajectories, amplitudes, and radii—using a framework based on LSTM and Extremely Randomized Trees, achieving the collaborative prediction of both eddy properties and propagation trajectories. Another study [9] proposed the MesoGRU framework, which significantly improves prediction accuracy by fusing SLA and flow field data. Simultaneously, related research incorporating spatio-temporal attention mechanisms [10], [11] has further deepened the capability to extract critical information from multi-source data. Building on these advancements, the research focus over the past two years has shifted toward enhancing the model's physical consistency and generalization. For instance, Ge et al. [12] proposed ETPNet, which embeds physical laws of eddy motion as constraints within the network architecture to achieve precise trajectory prediction. Meanwhile, Zhang et al. [13] introduced EddyTPNet, exploring physics-knowledge fusion and transfer learning strategies to address prediction challenges across diverse oceanic regions.

Despite significant progress in current physics-enhanced methods, most remain confined to a single-eddy prediction framework, treating eddies as isolated individuals. However, physical oceanography research indicates that eddy motion is not an isolated event. As illustrated in Fig. 1, in real-world oceanic environments, eddy trajectories are collectively modulated by the background flow field, the β-effect, and nonlinear interactions among neighboring eddies [14]–[19]. Such interactions often exhibit pronounced asymmetry and sparsity. Studies on dipole evolution by Long and Tang et al. [20] demonstrate that the interaction between cyclonic and anticyclonic eddies is often accompanied by the significant intensification of one while the other weakens or even dissipates, reflecting the asymmetric forces between strong and weak eddies. Furthermore, while eddy clusters exist in localized regions such as the Luzon Strait, eddies are typically distributed discretely across the vast open ocean. For an eddy located in the northern South China Sea, its motion is primarily governed by the local flow field and neighboring eddies.

Regarding the multi-eddy interaction problem, the Eddy-SPIN model recently proposed by Tang et al. [21] represents a pioneering effort. This work introduced a novel spatial projection method to map 1D trajectories of multiple eddies into a 2D image space, providing the first verification that modeling eddy-eddy interactions significantly improves prediction accuracy. However, this method relies on regular grid convolutions, which struggle to effectively adapt to the sparse interaction and non-uniform distribution characteristics of eddies. Furthermore, the image projection approach cannot explicitly characterize the dynamically evolving relationships between individual eddies. In contrast, graph neural networks (GNNs) can directly model interactions between discrete objects in non-Euclidean topological spaces. Their success in fields such as traffic flow [22] and pedestrian trajectory [23] prediction suggests that graph structures are an ideal paradigm for modeling such sparse interactions among discrete individuals.

To address the two core challenges of complex multivariate temporal dependencies and the difficulty in quantifying sparse inter-eddy interactions, we propose a multimodal fusion-based sparse spatio-temporal graph convolutional network (MF-STGCN). The model first employs a physical encoder to achieve a unified representation of eddy trajectory sequences and multidimensional physical features. An adaptive gating mechanism is then introduced to dynamically adjust the contribution weights of different modalities according to the eddy's current evolutionary state, thereby capturing the deep coupling between morphology and motion. Subsequently, to model the sparse and asymmetric interactions among multiple eddies, MF-STGCN adopts a sparse graph learning strategy that combines asymmetric convolutions with self-attention to construct directed interaction graphs. Finally, zero-softmax normalization filters redundant noise connections, enabling more precise quantitative modeling of inter-eddy relationships. The primary contributions of this work are summarized as follows:

- We develop a unified end-to-end MF-STGCN framework that synergistically models trajectory sequences, physical features, and environmental information, thereby overcoming the limitations of traditional unimodal prediction methods.
- We design a physical encoder and an adaptive gated network to resolve the challenges of fusing heterogeneous multivariate temporal features and physical features, enhancing the model's perception of complex temporal dependencies within eddies.
- We propose a sparse graph learning strategy based on asymmetric convolutions that eliminates redundant noise connections, achieving the precise quantitative modeling of sparse and asymmetric interactions among multiple eddies.
- Extensive experiments on the SCS dataset demonstrate that MF-STGCN outperforms state-of-the-art methods across multiple evaluation metrics and exhibits superior robustness in both long-term forecasting and abrupt-turn scenarios.

II. DATASET AND PREPROCESSING

This section details the data sources and preprocessing methodologies employed in this study. We integrate multi-source datasets derived from the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO) database and the Copernicus Marine Environment Monitoring Service (CMEMS), with a specific focus on the SCS. As the largest marginal sea in the Northwest Pacific [24], the SCS provides an ideal testbed for eddy dynamics due to its complex bathymetry and intense eddy activity [25]. By incorporating daily-resolution data from 1993 to 2022, the model captures

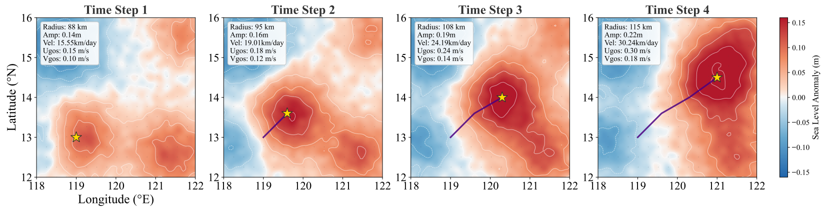


Fig. 1: Spatiotemporal evolution and multivariate dynamics of a mesoscale eddy. Its trajectory is modulated by both intrinsic physical properties (e.g., radius, amplitude, and velocity) and the surrounding environmental field (e.g., geostrophic flow). The annotated variables highlight the dynamic interplay between the eddy's morphology and motion throughout its lifecycle.

eddy behaviors across diverse timescales, including intra-seasonal and inter-annual variations.

A. Data Sources and Research Area

To enhance the representativeness of mesoscale eddy features, this study integrates multiple datasets from AVISO and CMEMS (1993–2022). Additionally, the geographic scope of the SCS was expanded to cover 0°–25°N and 100°–125°E to mitigate overfitting in trajectory prediction. The SCS is characterized by complex bathymetry and high spatiotemporal variability [26] in eddy generation and dissipation.

- **Mesoscale Eddy Trajectory Atlas (MET):** Eight core features (e.g., longitude, latitude, amplitude, radius, and velocity) are extracted from the AVISO database, which is identified based on the closed contours of sea-surface height (SSH) fields. [27].
- **Environmental Fields (SLA):** The background oceanic context is characterized by gridded sea level anomaly (SLA) and absolute geostrophic velocities at $0.25°$ resolution, with the latter decomposed into zonal (U_{gos}) and meridional (V_{gos}) components.

B. Spatiotemporal Alignment and Preprocessing

To ensure data consistency, we implemented a systematic pipeline, as shown in Fig. 2, to align discrete trajectories with continuous gridded fields. For each eddy center $(\lambda_{\text{MET}}, \phi_{\text{MET}})$, a nearest-neighbor search within a neighborhood radius of $0.25°$ was performed to extract the corresponding SLA and geostrophic velocity $(U_{\text{gos}}, V_{\text{gos}})$ values. In cases where multiple grid points matched an eddy center, the nearest grid point was selected. To mitigate scale variance across heterogeneous variables, Z-score normalization is applied to all features to facilitate model convergence. Furthermore, trajectories with a life cycle shorter than 30 days were excluded to filter transient noise and ensure dynamical stability.

III. METHODOLOGY

A. Problem Definition

We formulate the mesoscale eddy trajectory prediction as a multivariate spatio-temporal forecasting problem on a dynamic graph. The objective of mesoscale eddy trajectory prediction is to forecast future spatial positions. Given an observation sequence $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$ of N eddies within a historical time window T, where $\mathbf{X}_t \in \mathbb{R}^{N \times D}$

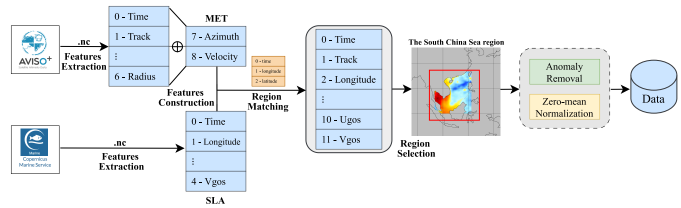


Fig. 2: Preprocessing pipeline for multi-source mesoscale eddy trajectory data. The workflow illustrates the integration of heterogeneous data from AVISO and CMEMS.

includes the longitude and latitude coordinates, physical features, and environmental information for each eddy across D feature dimensions. The prediction goal is a trajectory coordinate sequence for the next L_{pred} time steps, denoted as $\mathcal{Y} = \{\mathbf{Y}_{T+1}, \mathbf{Y}_{T+2}, \dots, \mathbf{Y}_{T+L_{pred}}\}$, where $\mathbf{Y}_t \in \mathbb{R}^{N \times 2}$ represents the longitudinal and latitudinal positions of all eddies at time t. The interaction is modeled as a sparse graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where the adjacency matrix $\mathbf{A}$ is dynamically learned to characterize the directed and sparse relationships between eddies.

B. Overall Framework

The proposed MF-STGCN framework is shown in Fig. 3, Initially, to address the alignment and fusion of multimodal features, the model employs a physical encoder to map physical features and environmental information into a latent space, subsequently utilizing an adaptive gated network to dynamically regulate the contribution of these physical semantics to the motion features; this process generates spatio-temporal representations enriched with physical semantics, providing a robust foundation for subsequent modeling. Building upon these representations, a sparse graph learning mechanism is introduced to quantify dynamic eddy interactions, where sparse adjacency matrices in both spatial and temporal dimensions are learned via self-attention and asymmetric convolutional networks to explicitly filter redundant noise connections. These learned structures enable a dual-stream spatio-temporal graph convolutional network to aggregate features in parallel, capturing spatial interaction patterns and temporal evolution laws within complex oceanic flow fields. Finally, the aggregated features are fed into a TCN, where dilated convolutions are leveraged to expand the receptive field for

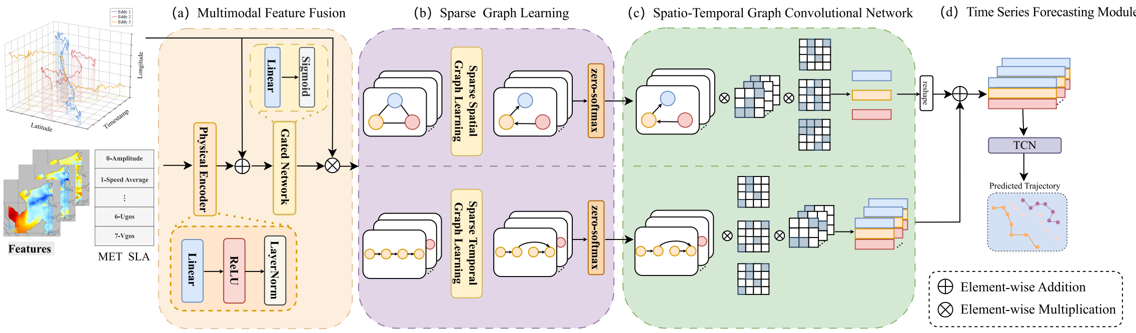


Fig. 3: The comprehensive architecture of the proposed MF-STGCN.

the efficient extraction of long-range temporal dependencies, thereby completing the trajectory prediction task.

C. Multimodal feature Fusion

To resolve complex multivariate dependencies, we design a multimodal encoding and fusion mechanism. Input features are categorized into motion modality and physical modality.First, the motion modality comprises the spatio-temporal position of an eddy. Each eddy is uniquely identified by an eddy ID, and its position at time t is defined by its longitude and latitude coordinates: $p_i^t = [\text{lon}_i^t, \text{lat}_i^t] \in \mathbb{R}^2$. The historical trajectory for eddy i over a time window T is then denoted as $P_i = \{p_i^1, \ldots, p_i^T\}$. Second, the physical modality encompasses the physical features and environmental information of the eddy. The physical features and environmental data of the eddy i at time t are denoted as $x_i^t \in \mathbb{R}^8$, which include features such as amplitude, effective radius, velocity, etc. The historical sequence is represented as $X_i = \{x_i^1, \ldots, x_i^T\}$.

Due to the significant differences in scale and semantics between the physical and motion modalities, direct concatenation would lead to an imbalanced feature distribution. Consequently, the model first performs a non-linear mapping of the physical features through a physical encoder:

$$H_{phys}^t = \text{LayerNorm}(\text{ReLU}(w_p X_{phys}^t + b_p)) \tag{1}$$

Where w_p and b_p are the learnable weight matrix and bias vector of the physical encoder, responsible for projecting raw features into a unified latent space.

In strong current regions, the environmental flow field plays a dominant role, whereas in weak current regions, internal inertia prevails. To dynamically adjust the contributions of physical features and environmental information across different motion stages, this paper introduces an adaptive gated network:

$$G^t = \sigma(W_g[X_{motion}^t \parallel H_{phys}^t] + b_g) \tag{2}$$

$$F_{fused}^t = X_{motion}^t + (H_{phys}^t \odot G^t) \tag{3}$$

where σ represents the sigmoid function, $\parallel$ denotes the concatenation operation, and $\odot$ denotes element-wise multiplication. W_g and b_g denote the trainable parameters of the adaptive gated network. The gating coefficient $G^t \in (0, 1)$ adaptively controls the injection of physical features and environmental information. This ensures that key physical driving constraints are embedded while simultaneously preserving the original trajectory trends.

D. Sparse Graph Learning

Mesoscale eddies in the ocean exhibit significant spatial locality and directionality, with substantive physical responses occurring only between a limited number of eddies. Traditional fully connected graph methods introduce substantial noise, making it difficult to quantify these interaction relationships. To address this, we propose a sparse graph learning module, as illustrated in Fig. 4, which consists of two branches: sparse spatial graph learning and sparse temporal graph learning.

*a) **Sparse Spatial Graph Construction** :* This branch aims to quantify the asymmetric sparse relationships between eddies, identifying those that exhibit significant interactions. First, a multi-head self-attention mechanism is employed to calculate the potential correlations between eddies within a global scope. For notation simplicity, we denote $Z = F_{\text{fused}}$.The query matrix Q and key matrix K are computed as $Q = ZW_Q$ and $K = ZW_K$, respectively. The dense interaction scores are then computed as follows:

$$E_{\text{dense}} = \frac{(ZW_Q)(ZW_K)^T}{\sqrt{d_k}} \tag{4}$$

where d_k is the key dimension serving as the scaling factor. This step captures the pairwise dependencies between all eddies; however, it introduces significant noise, specifically from eddy pairs that lack actual physical connections.

Next, to characterize the anisotropy of interactions between eddies and filter out noise, the model introduces an asymmetric convolutional network. As illustrated in Fig. 4, convolution kernels of size 3×1 and 1×3 are employed to extract directional features from the dense interaction map E_{dense}:

$$M_{\text{spa}} = \sigma(\text{Conv}_{1\times3}(E_{\text{dense}}) + \text{Conv}_{3\times1}(E_{\text{dense}})) \tag{5}$$

where σ is the sigmoid function. This asymmetric design captures the anisotropic interaction between eddies, reflecting the fact that the mutual interaction forces are often inherently non-symmetric. By extracting features along source and target node dimensions that correspond to meridional and zonal directions,

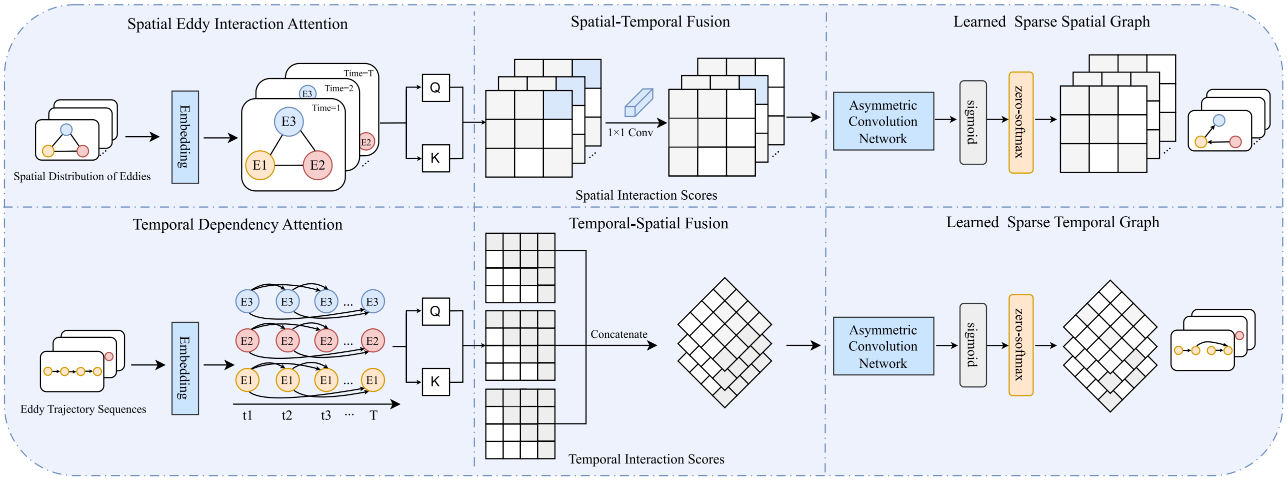


Fig. 4: Schematic of the proposed Sparse Graph Learning architecture.

the model yields a more consistent spatial interaction intensity map M_{spa}.

To precisely characterize the asymmetric sparse relationships between eddies and filter out those that genuinely interact, the model adopts a threshold truncation strategy combined with zero-softmax normalization:

$$M_{\text{mask}} = \mathbb{I}(M_{\text{spa}} > \xi) \tag{6}$$

$$\hat{A} = \text{ZeroSoftmax}(E_{\text{dense}} \odot M_{\text{mask}}) = \frac{(e^{x}-1)^2}{\sum_j (e^{x_j}-1)^2 + \epsilon} \tag{7}$$

Specifically, where x and x_j denote the elements of the masked interaction matri,$(E_{\text{dense}} \odot M_{\text{mask}})$, and ϵ is a small constant to ensure numerical stability. M_{mask} is a binary adjacency matrix obtained by thresholding M_{spa} at ξ (set to 0.5 in this study). Interaction weights below ξ are set to 0, thereby quantifying sparse effective interactions while filtering out redundant noise.

*b) **Sparse Temporal Graph Construction**:* This branch aims to quantify the motion dependencies of a single eddy across the temporal dimension, addressing the influence of historical states on the current moment. A self-attention mechanism is utilized to capture long-range dependencies, while features are aggregated through spatio-temporal fusion. Similarly, by leveraging asymmetric convolution and the zero-softmax mechanism, a learnable sparse temporal graph is generated. This graph structure explicitly filters out irrelevant historical interference and precisely quantifies the effective motion trends of the eddies.

E. Spatio-Temporal Feature Aggregation

With the sparse graphs established, the model aggregates spatio-temporal information via a dual-stream architecture. Here, F_{fused} denotes the fused node features, while $\hat{A}_{\text{spa}}$ and $\hat{A}_{\text{tmp}}$ represent the sparse spatial and temporal adjacency matrices, respectively. The framework first extracts neighborhood features through the spatial graph, followed by historical dependency aggregation via the temporal graph:

$$H_{ST} = \text{GCN}_{\text{tmp}}(\text{GCN}_{\text{spa}}(F_{\text{fused}}, \hat{A}_{\text{spa}}), \hat{A}_{\text{tmp}}) \tag{8}$$

In the opposite sequence, the model prioritizes capturing the inertial evolution in the temporal dimension:

$$H_{TS} = \text{GCN}_{\text{spa}}(\text{GCN}_{\text{tmp}}(F_{\text{fused}}, \hat{A}_{\text{tmp}}), \hat{A}_{\text{spa}}) \tag{9}$$

Finally, a 1×1 convolution is employed to fuse the outputs from both streams. This process yields a high-level representation that integrates rich physical semantics with spatio-temporal topologies.

F. Temporal Convolutional Prediction and Loss Function

*a) **Temporal Convolutional Network**:* To capture long-range temporal dependencies while avoiding gradient vanishing, the model employs a TCN. By stacking dilated convolutional layers, the TCN exponentially expands the receptive field. The operation is defined as follows:

$$O^t = \sum_{k=0}^{K-1} W_k \cdot H^{(t-k\cdot d)} + b \tag{10}$$

Where O^t denotes the output feature at time t, H is the input feature from the preceding layer, and W_k and b represent the learnable filter weights and bias, respectively. The hyperparameters d and K signify the dilation factor and kernel size. This multi-scale structure enables the model to simultaneously capture short-term fluctuations and long-term inertial trends in eddy motion.

To account for the stochastic nature of eddy motion, MF-STGCN performs joint regression on future longitudes and latitudes. To stabilize training and capture two-dimensional spatial correlations, the output layer employs a Gaussian regression head that parameterizes a bivariate Gaussian distribution. The output at time t is a five-dimensional vector:

$$Y^t_{\text{pred}} = [\mu^t_{\text{lon}}, \mu^t_{\text{lat}}, \sigma^t_{\text{lon}}, \sigma^t_{\text{lat}}, \rho^t] \tag{11}$$

where $\mu^t = (\mu^t_{\text{lon}}, \mu^t_{\text{lat}})$ and $\sigma^t = (\sigma^t_{\text{lon}}, \sigma^t_{\text{lat}})$ represent the predicted mean positions and standard deviations, respectively, and ρ^t is the correlation coefficient. During the inference stage, the mean μ^t is adopted as the final deterministic prediction coordinate, representing the most probable location within the learned distribution.

b) ***Loss Function:*** To achieve stable joint regression of longitude and latitude and capture their spatial coupling, MF-STGCN adopts a Gaussian-form negative log-likelihood (NLL) loss. Instead of relying on a single distance metric, the formulation leverages second-order spatial statistics, providing the model with the capacity to dynamically adjust to prediction errors in complex environments.

$$\mathcal{L}(\Theta) = \sum_{i=1}^{N} \sum_{\Delta t=1}^{L_{pred}} -\log P(y_i^{t+\Delta t} \mid \mu_i^{t+\Delta t}, \Sigma_i^{t+\Delta t}) \quad (12)$$

where Θ represents the set of all learnable parameters in the model, and the covariance matrix Σ is constructed from the predicted $\sigma_{\text{lon}}, \sigma_{\text{lat}}$, and ρ to characterize the anisotropic spatial uncertainty. During inference, only the predicted mean μ_t is utilized as the final deterministic prediction, enabling robust structured training while preserving a direct trajectory output.

IV. EXPERIMENTS

A. Experimental Setup and Evaluation Metrics

We compare MF-STGCN to three types of cutting-edge baselines: graph-based models ASTGNN [22] and SGCN [22]; eddy trajectory models ETPNet [12] and Eddy-SPIN [13]; and classical temporal models Transformer [28], Crossformer [29], DLinear [30], and TimeXer [31].

The experiments utilize a mesoscale eddy trajectory dataset covering the SCS spanning from 1993 to 2022, comprising 12,847 sequences. The dataset is split into training, validation, and test sets with a ratio of 70%, 15%, and 15%, respectively. MF-STGCN uses an embedding size of 64, 7 asymmetric convolution layers, and 5 TCN layers. All models are implemented in PyTorch 2.0 and trained on an RTX 4090 GPU with a learning rate of 0.001 and a batch size of 32. To ensure fairness, all baselines are re-implemented and tuned under the same settings.

We evaluate MF-STGCN using mean squared error (MSE) and mean absolute error (MAE) for deterministic coordinate deviations. Geospatial accuracy is quantified by the mean geodesic distance (MGD) via the Haversine formula, and long-term stability by the final geodesic distance (FGD) at the terminal prediction step. Additionally, we define the prediction success rate as the proportion of trajectories with an MGD below thresholds of 10 km, 20 km, and 30 km to evaluate the model's practical utility across varying precision requirements.

B. Comparative Results and Analysis

To evaluate predictive performance, we compare MF-STGCN against baselines using input horizons of 10, 15, and 20 days. These experiments assess how varying observation windows impact 7-day eddy trajectory forecasting accuracy. The results for these three scenarios are detailed in Table I, Table II, and Table III, respectively. MF-STGCN consistently outperforms all baselines. By integrating physical flow field information, MF-STGCN effectively reduces positional bias during non-linear drifts compared to models like Transformer and TimeXer. Unlike Eddy-SPIN, which relies on a spatial projection method, our model utilizes asymmetric convolution and the zero-softmax mechanism to construct sparse graphs. This approach filters redundant noise and captures genuine interactions, maintaining superior accuracy even with longer input sequences.

TABLE I: Performance Comparison ($L = 10$)

Model	MAE	MSE	FGD (km)	MGD (km)	Success Rate (%)		
					10km	20km	30km
Transformer (2017)	0.1888	0.0661	37.48	32.36	3.90	24.00	49.50
Crossformer (2023)	0.1425	0.0411	37.94	24.68	6.20	45.50	73.70
DLinear (2023)	0.1098	0.0299	35.04	19.18	24.70	63.80	83.40
TimeXer (2024)	0.0982	0.0251	31.93	16.85	33.50	68.10	83.20
ETPNet (2023)	0.1964	0.0733	48.69	34.04	3.83	14.09	31.05
Eddy-SPIN (2025)	0.1239	0.0740	30.16	29.88	11.44	38.33	57.04
ASTGNN (2021)	0.1488	0.0489	43.27	25.58	3.56	17.40	36.01
SGCN (2021)	0.1313	0.0725	27.44	17.23	38.69	68.82	84.61
MF-STGCN (Ours)	**0.0784**	**0.0144**	**15.89**	**13.57**	**40.54**	**70.21**	**84.65**

TABLE II: Performance Comparison ($L = 15$)

Model	MAE	MSE	FGD (km)	MGD (km)	Success Rate (%)		
					10km	20km	30km
Transformer (2017)	0.1750	0.0662	33.18	30.06	4.80	27.20	51.90
Crossformer (2023)	0.1608	0.0491	37.97	27.55	0.90	35.10	67.30
DLinear (2023)	0.1293	0.0378	38.34	22.45	17.30	54.00	76.60
TimeXer (2024)	0.1435	0.0441	38.76	24.68	15.60	48.30	71.40
ETPNet (2023)	0.2040	0.0796	46.63	35.26	4.76	25.78	48.11
Eddy-SPIN (2025)	0.1546	0.1056	34.36	30.62	7.04	29.56	51.91
ASTGNN (2021)	0.1401	0.0441	42.60	24.02	2.92	16.93	35.98
SGCN (2021)	0.1288	0.0698	27.01	16.87	41.13	70.53	85.16
MF-STGCN (Ours)	**0.0797**	**0.0159**	**17.29**	**13.81**	**41.36**	**71.05**	**85.19**

TABLE III: Performance Comparison ($L = 20$)

Model	MAE	MSE	FGD (km)	MGD (km)	Success Rate (%)		
					10km	20km	30km
Transformer (2017)	0.1418	0.0532	27.68	24.35	5.10	27.80	57.60
Crossformer (2023)	0.1237	0.0353	35.59	21.53	19.60	58.50	79.10
DLinear (2023)	0.1212	0.0341	36.52	21.08	19.50	58.00	80.00
TimeXer (2024)	0.1017	0.0264	32.45	17.48	31.10	68.90	79.40
ETPNet (2023)	0.1904	0.0680	43.34	33.26	6.67	27.85	50.91
Eddy-SPIN (2025)	0.1409	0.0713	31.97	28.22	10.64	31.56	52.85
ASTGNN (2021)	0.1421	0.0443	40.66	24.44	3.15	18.51	38.89
SGCN (2021)	0.1234	0.0651	25.92	16.14	41.75	70.64	84.61
MF-STGCN (Ours)	**0.0869**	**0.0195**	**19.86**	**15.08**	**43.62**	**72.96**	**86.48**

The 7-day forecast performance of MF-STGCN is evaluated in Fig. 5. While errors naturally accumulate over time, MF-STGCN exhibits superior robustness, with MGD increasing only from 5.07 km to 13.57 km. Notably, it reduces the Day 7 error by approximately 58% compared to the divergent Transformer model. This stability confirms that the sparse graph effectively suppresses error accumulation by filtering spatial noise.

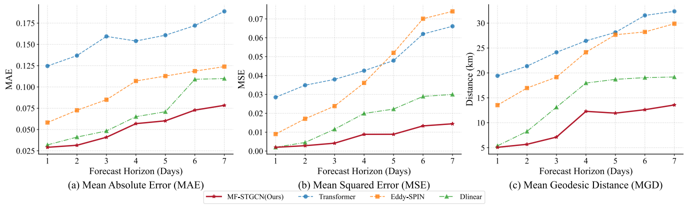


Fig. 5: Comparison of predictive performance across different forecast horizons (1–7 days) with a fixed input length of 10 days.

C. Ablation Study

To verify the contribution of each core component, we design four distinct variants of MF-STGCN in Table IV. Specifically, w/o Gate denotes the removal of the gated network, while w/o Sparse utilizes a dense adjacency matrix instead of the sparse directed interaction modeling. Furthermore, w/o Bidirectional implements only unidirectional graph convolutions, and w/o PhysEnc signifies the exclusion of the physical encoder, relying solely on single-modal movement data.

TABLE IV: Ablation of MF-STGCN Components

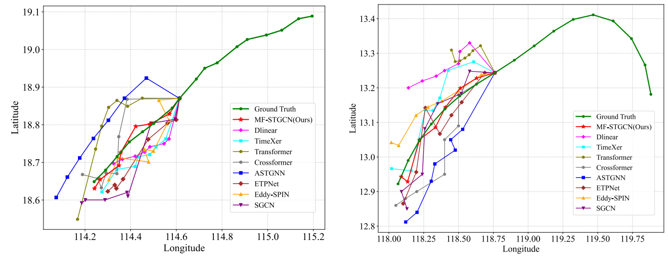
Variants	**MAE**	**MSE**	**FGD (km)**	**MGD (km)**	**Success Rate (%)**		
					10km	**20km**	**30km**
w/o Gate	0.0832	0.0166	16.32	14.42	38.16	68.34	83.49
w/o Sparse	0.0870	0.0167	17.26	15.08	30.59	62.11	79.46
w/o Bidirectional	0.0794	0.0149	16.20	13.77	38.31	68.60	83.68
w/o PhysEnc	0.0842	0.0159	16.85	14.58	33.13	63.94	80.73
MF-STGCN (Ours)	**0.0784**	**0.0144**	**15.89**	**13.57**	**38.69**	**68.82**	**84.65**

The resulting performance metrics quantify these contributions. Notably, the removal of the sparse graph learning mechanism causes the most significant degradation, with the MGD increasing to 15.08 km. This confirms that dense connectivity introduces redundant noise, whereas sparse graphs ensure focus on genuine eddy interactions. Removing the physical encoder leads to a 6.0% increase in FGD error, validating that intrinsic features provide critical dynamical priors for trajectory accuracy. Furthermore, the exclusion of the gated network results in higher errors than simple concatenation. This indicates that the relative importance of motion and physical modalities evolves over time. By dynamically modulating relevant information, the adaptive gating network achieves seamless integration of multimodal data across different evolutionary stages.

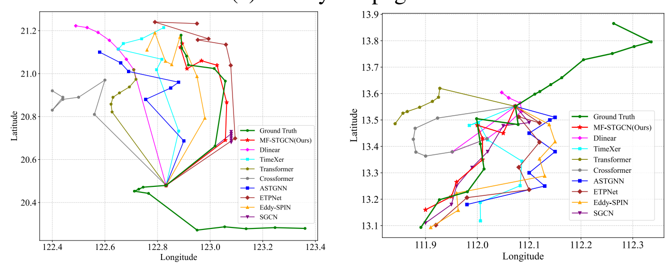
D. Case Study and Trajectory Visualization

To evaluate the predictive performance of MF-STGCN in complex marine environments, we selected two representative scenarios for visualization: steady propagation and abrupt trajectory deflection.

1) **Steady Propagation:** As shown in Fig. 6(a), while most baseline models suffer from trajectory divergence during steady propagation, MF-STGCN maintains close proximity to the ground truth due to its sparse graph learning mechanism that filters background noise.



(a) Steady Propagation



(b) Abrupt Trajectory Deflection

Fig. 6: Comparison of trajectory predictions in complex scenarios.

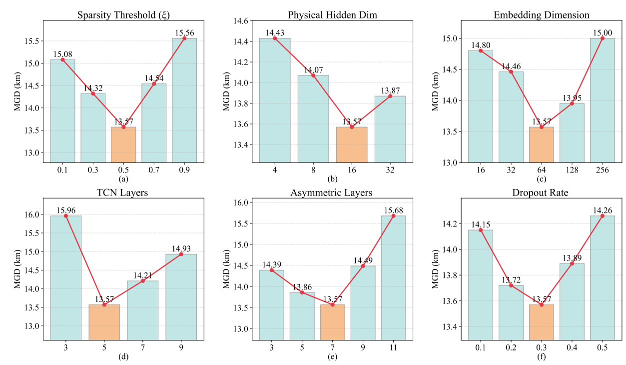


Fig. 7: Parameter sensitivity analysis using MGD (km) across six key hyperparameters: (a) Sparsity Threshold ξ, (b) Embedding Dimension, (c) Physical Hidden Dim, (d) TCN Layers, (e) Asymmetric Conv Layers, and (f) Dropout Rate. The optimal configurations are highlighted in orange.

2) **Abrupt Trajectory Deflection:** Fig. 6(b) illustrates a scenario with a sharp trajectory shift. While DLinear tends to project the trajectory linearly based on historical momentum, MF-STGCN accurately captures this sudden shift. When positional data alone is insufficient to determine the future path, the model dynamically incorporates physical features and environmental information such as geostrophic velocity and amplitude to perceive changes in environmental driving forces.

E. Parameter Sensitivity Analysis

We conduct a sensitivity analysis over six key hyperparameters to assess the robustness of MF-STGCN, as shown in Fig. 7. The embedding dimension determines feature capacity; MGD error reaches its minimum at 64, while larger dimensions trigger overfitting. For sparsity threshold ξ, peak performance occurs at 0.5; lower values introduce background noise, while higher ones eliminate essential interactions. Regarding multimodal fusion, 16-dimensional encoding avoids both the information bottleneck of 4D and the redundancy of 32D features. For temporal modeling, five TCN layers best capture eddy evolution cycles, as fewer layers restrict the receptive field. Asymmetric convolutions reach optimal performance at seven layers, beyond which over-smoothing homogenizes node features. Finally, the MGD reaches its minimum at a dropout rate of 0.3, as this optimal value balances noise reduction with the preservation of essential spatiotemporal features. MF-STGCN achieves the best balance between accuracy and efficiency under the configurations highlighted in orange.

V. Conclusion

This study addresses the challenges of complex multivariate temporal dependencies and sparse interactions in mesoscale eddy trajectory prediction by proposing the MF-STGCN. The model integrates a gated multimodal fusion mechanism with an asymmetric sparse graph structure, effectively quantifying the sparse interactions among multiple eddies. Experiments based on publicly available ocean datasets demonstrate that MF-STGCN significantly outperforms mainstream methods in terms of long-term prediction accuracy and robustness. Future research will integrate additional environmental variables to construct a more comprehensive multi-physics coupled model to further enhance forecasting precision.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 42376194.

References

[1] Y. Zhang, N. Wang, Z. Lin, K. Liu, and H. Wang, "The surface and three-dimensional characteristics of mesoscale eddies: a review," *Advances in Earth Science*, vol. 35, no. 6, pp. 568–580, 2020.

[2] D. B. Chelton, M. G. Schlax, and R. M. Samelson, "Global observations of nonlinear mesoscale eddies," *Progress in oceanography*, vol. 91, no. 2, pp. 167–216, 2011.

[3] Z. Zhang, W. Wang, and B. Qiu, "Oceanic mass transport by mesoscale eddies," *Science*, vol. 345, no. 6194, pp. 322–324, 2014.

[4] D. J. McGillicuddy Jr, "Mechanisms of physical-biological-biogeochemical interaction at the oceanic mesoscale," *Annual Review of Marine Science*, vol. 8, no. 1, pp. 125–159, 2016.

[5] C. Ma, S. Li, A. Wang, J. Yang, and G. Chen, "Altimeter observation-based eddy nowcasting using an improved conv-lstm network," *Remote Sensing*, vol. 11, no. 7, p. 783, 2019.

[6] R. Nian, Y. Cai, Z. Zhang, H. He, J. Wu, Q. Yuan, X. Geng, Y. Qian, H. Yang, and B. He, "The identification and prediction of mesoscale eddy variation via memory in memory with scheduled sampling for sea level anomaly," *Frontiers in Marine Science*, vol. 8, p. 753942, 2021.

[7] J. J. Early, R. Samelson, and D. B. Chelton, "The evolution and propagation of quasigeostrophic ocean eddies," *Journal of Physical Oceanography*, vol. 41, no. 8, pp. 1535–1555, 2011.

[8] X. Wang, H. Wang, D. Liu, and W. Wang, "The prediction of oceanic mesoscale eddy properties and propagation trajectories based on machine learning," *Water*, vol. 12, no. 9, p. 2521, 2020.

[9] X. Wang, X. Wang, M. Yu, C. Li, D. Song, P. Ren, and J. Wu, "Mesogru: Deep learning framework for mesoscale eddy trajectory prediction," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[10] X. Wang, C. Li, X. Wang, L. Tan, and J. Wu, "Spatio–temporal attention-based deep learning framework for mesoscale eddy trajectory prediction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3853–3867, 2022.

[11] H. Tang, J. Lin, and D. Ma, "Direct prediction for oceanic mesoscale eddy geospatial distribution through prior statistical deep learning," *Expert Systems with Applications*, vol. 249, p. 123737, 2024.

[12] L. Ge, B. Huang, X. Chen, and G. Chen, "Medium-range trajectory prediction network compliant to physical constraint for oceanic eddy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[13] X. Zhang, B. Huang, G. Chen, L. Ge, M. Radenkovic, and G. Hou, "Global oceanic mesoscale eddies trajectories prediction with knowledge-fused neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[14] X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang, and F. Wang, "Deep-learning-based information mining from ocean remote-sensing imagery," *National Science Review*, vol. 7, no. 10, pp. 1584–1605, 2020.

[15] G. Chen, Y. Hou, and X. Chu, "Mesoscale eddies in the south china sea: Mean properties, spatiotemporal variability, and impact on thermohaline structure," *Journal of Geophysical Research: Oceans*, vol. 116, no. C6, 2011.

[16] Z. Zhang, J. Tian, B. Qiu, W. Zhao, P. Chang, D. Wu, and X. Wan, "Observed 3d structure, generation, and dissipation of oceanic mesoscale eddies in the south china sea," *Scientific reports*, vol. 6, no. 1, p. 24349, 2016.

[17] C. Qiu, Z. Yi, Y. Qi, D. Su, D. Shao, and H. Mao, "Evolutions and dynamics of mesoscale eddies in south china sea," 2025.

[18] R. Abernathey and G. Haller, "Transport by lagrangian vortices in the eastern pacific," *Journal of Physical Oceanography*, vol. 48, no. 3, pp. 667–685, 2018.

[19] H. Wang, B. Qiu, H. Liu, and Z. Zhang, "Doubling of surface oceanic meridional heat transport by non-symmetry of mesoscale eddies," *Nature Communications*, vol. 14, no. 1, p. 5460, 2023.

[20] S. Long, F. Tian, J. Tang, F. Yu, F. Zhang, W. Ma, X. Zhang, and G. Chen, "Integrated observation of an asymmetric eddy dipole in the south china sea," *Earth System Science Data Discussions*, vol. 2025, pp. 1–24, 2025.

[21] H. Tang, Y. Liu, Z. Ma, S. Tabeta, J. Lin, and D. Ma, "Eddy-spin: A spatial projection inference network for oceanic mesoscale eddy trajectory prediction," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[22] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong, "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5415–5428, 2021.

[23] L. Shi, L. Wang, C. Long, S. Zhou, M. Zhou, Z. Niu, and G. Hua, "Sgcn: Sparse graph convolution network for pedestrian trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8994–9003.

[24] Y. Liu, L. Yu, and G. Chen, "Characterization of sea surface temperature and air-sea heat flux anomalies associated with mesoscale eddies in the south china sea," *Journal of Geophysical Research: Oceans*, vol. 125, no. 4, p. e2019JC015470, 2020.

[25] J. Ji, C. Dong, B. Zhang, Y. Liu, B. Zou, G. P. King, G. Xu, and D. Chen, "Oceanic eddy characteristics and generation mechanisms in the kuroshio extension region," *Journal of Geophysical Research: Oceans*, vol. 123, no. 11, pp. 8548–8567, 2018.

[26] G. Chen, Y. Hou, and X. Chu, "Mesoscale eddies in the south china sea: Mean properties, spatiotemporal variability, and impact on thermohaline structure," *Journal of Geophysical Research: Oceans*, vol. 116, no. C6, 2011.

[27] W. Meng, Z. Yanwei, L. Zhifei, and W. Jiawang, "Temporal and spatial characteristics of mesoscale eddies in the northern south china sea: statistics analysis based on altimeter data," *Advances in Earth Science*, vol. 34, no. 10, p. 1069, 2019.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[29] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *The eleventh international conference on learning representations*, 2023.

[30] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.

[31] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, and M. Long, "Timexer: Empowering transformers for time series forecasting with exogenous variables," *Advances in Neural Information Processing Systems*, vol. 37, pp. 469–498, 2024.