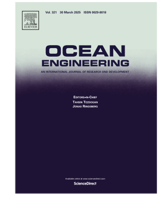
Ocean Engineering 336 (2025) 121688

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

Enhancing vessel trajectory prediction with bi-state space and squeeze excitation attention

Caiquan Xiong [a], Jiaming Li [a], Xinyun Wu [a], Donghua Liu [b,*], Qi Wang [c], Rong Gao [a], Jinjia Ruan [b]

[a] *Hubei University of Technology, No. 28, Nanli Road, Wuhan, 430068, Hubei, China*
[b] *China Waterborne Transport Research Institute, No. 8, Xitucheng Road, Beijing, 100080, Beijing, China*
[c] *Chinese Academy of Sciences, No. 52, Sanlihe Road, Beijing, 100094, Beijing, China*

ARTICLE INFO

Keywords:
Vessel trajectory prediction
Automatic identification system
Deep learning
Bidirectional state space
Squeeze excitation attention
Causal self-attention

ABSTRACT

Accurately predicting vessel trajectories is crucial for maritime traffic management and safety. Deep learning technologies have made significant progress in modeling the complex relationships of vessel trajectories. However, existing approaches suffer from two limitations: (1) inadequate capture of comprehensive features when modeling complex spatiotemporal relationships, and (2) substantial computational complexity introduced by these methods that constrains real-world deployment. To address these challenges, we propose a novel Bidirectional state Space and Squeeze Excitation attention fusion model (Bi_SSE) for vessel trajectory prediction. Specifically, we introduce a linear-complexity bidirectional state space component to capture the complex features of vessel trajectories. Additionally, we design an squeeze excitation attention with causal self-attention component dynamically adjusting feature weights to enhance feature representation. Finally, we utilize the cross-stacks method to incorporate the two components to enhance prediction performance. Experimental evaluations on the AIS dataset ct_dma demonstrate that the proposed Bi_SSE model significantly outperforms state-of-the-art methods. Furthermore, it exhibits exceptional performance in short-term (1–3 h) and long-term (10 h) prediction tasks, achieving significantly lower Haversine distances in long-term predictions. The source codes are available at: https://github.com/with45/Bi_SSE

1. Introduction

The rapid growth of the global economy has led to a significant increase in international trade volume and maritime traffic density. This surge has resulted in frequent challenges such as close-range staggered navigation between vessels and increased navigational complexity, posing substantial risks to maritime traffic safety. In response to these emerging issues, Vessel Trajectory Prediction (VTP) technology has emerged as a critical solution for modern maritime safety systems. By integrating advanced data analysis, machine learning (Li et al., 2024b; Guo et al., 2018), and artificial intelligence algorithms (Chen et al., 2023; Borkowski, 2017; Nguyen and Fablet, 2024), VTP technology demonstrates an exceptional capability to process and analyze dynamic information from vessels. The VTP technology enables accurate prediction of vessel trajectories, effectively prevents collisions, and significantly reduces the incidence of maritime traffic accidents. As an indispensable component of intelligent maritime navigation systems, VTP significantly enhances early warning mechanisms while providing robust technical support for the sustainable development of global maritime operations and has attracted widespread attention and in-depth research from academia and industry.

Existing VTP methods can be broadly classified into two categories: kinematics-based approaches and neural network-based models. The former relies on constructing state transition equations to describe motion dynamics (Sutulo et al., 2002). Perera and Soares (Perera et al., 2010) employed extended Kalman filters establishing a state transition matrix to estimate the future positions of the vessels. Zhang et al. (2019) and Guo et al. (2018) developed state transition models using hidden Markov for vessel position prediction. Liu et al. (2019a) enhanced the accuracy of the Markov-based method by integrating multiple sensor data. These kinematics-based approaches have yielded improvements in VTP. However, they are limited in modeling the continuous motion behavior of vessels and are not well-suited for predicting trajectories at consecutive time points. Concurrent with technological advancements, neural network models have emerged as a prominent research focus, leveraging data-driven capabilities to capture complex spatiotemporal relationships. Chen et al. (2025). proposed a DDQN method that utilizes convolutional neural network to capture environmental state

* Corresponding author.
E-mail address: liudonghua@wti.ac.cn (D. Liu).

https://doi.org/10.1016/j.oceaneng.2025.121688
Received 28 February 2025; Received in revised form 16 May 2025; Accepted 26 May 2025

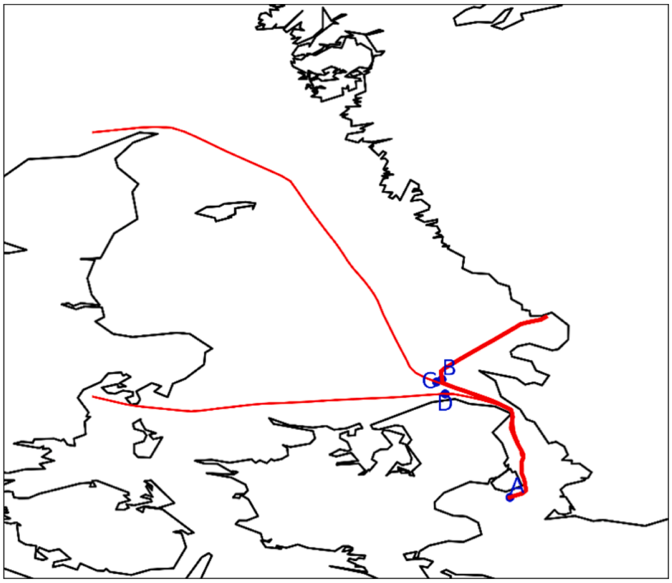


Fig. 1. Multiple destinations reachable from point A.

information features and improve the performance of ship route planning. Borkowski (2017) used Long Short Term Memory (LSTM), integrating data from the Automatic Identification System (AIS) with information from multiple sources to improve prediction accuracy. Tang et al. (2022) employed LSTM to model vessel trajectories, leveraging a data-driven approach to explore the spatiotemporal distribution of trajectories for future movement predictions. Li et al. (2019) applied clustering methods to capture trajectory features and then used LSTM, achieving sub-50-meter average prediction errors for 10-min forecasts. Suo et al. (2020) employed a Gated Recurrent Unit (GRU) to improve computational efficiency while maintaining prediction accuracy comparable to LSTM. Despite significant advancements and improved accuracy in VTP achieved by these methods, they overlook the complexity of vessel trajectories in narrow and complex traffic environments, making it difficult to capture comprehensive vessel trajectory features and subsequently affecting prediction performance.

As shown in Fig. 1, the vessel, originating at the starting point A, is capable of reaching the three destinations: B, C, and D, although substantial overlap is observed in the trajectories of vessels originating from point A and traversing points B, C and D, there exhibit notable disparities in their ultimate travel directions and subsequent trajectories. Consequently, it is imperative for the model to precisely capture the nuanced details of these trajectories and the underlying motion patterns to achieve effective predictions of future trajectories. Furthermore, when dealing with such complex dynamic trajectories, the robustness of the model to noise is relatively limited. This limitation becomes particularly evident when handling highly non-stationary data, where a significant degradation in prediction accuracy may occur.

To address these challenges, some researchers have used Transformer architecture to capture complex relationships within sequential data. In the field of traffic trajectory prediction, such as vessel (Nguyen and Fablet, 2024; Jiang et al., 2023; Takahashi et al., 2024; Jin et al., 2024; Xiong et al., 2024), vehicle, and drone trajectory forecasting (Zou et al., 2022; Chen et al., 2022; Gu et al., 2024), has witnessed Transformers' remarkable capacity to capture complex patterns and global features. Jiang et al. (2023) proposed TRFM-LS model to integrate LSTM and Transformer for VTP. LSTM is utilized to learn temporal features, while Transformer is employed to overcome the limitations of LSTM in capturing long-term sequence information. Experimental results show that TRFM-LS exhibits superior accuracy in time series trajectory prediction. Nguyen and Fablet (2024) introduced TrAISformer model to extract long-term temporal patterns from high-dimensional AIS data while mitigating heterogeneity and multi-modal issues of the data through a designed loss function. However, these Transformer-based trajectory prediction methods heavily rely on multi-head self-attention mechanism, resulting in quadratic complexity related to sequence length. This poses significant efficiency challenges in handling long vessel sequences and impedes its application in modeling long-term correlation (Wu et al., 2024). Furthermore, they struggle to encapsulate fine-grained features in trajectory data.

Recently, SSM (Kalman, 1960), originating from control system theory, has gained widespread application in natural language processing and computer vision due to their advantages in handling long sequences with linear complexity. Ma et al. (2024a) proposed the U-Mamba model based on Mamba architecture that effectively captures long-term dependencies in biomedical images, achieving state-of-the-art performance with reduced computational complexity compared to Transformer models. The MaTrRec model (Zhang et al., 2024) integrates the long-term dependency modeling capabilities of Mamba (Gu and Dao, 2023) with the global attention mechanism inherent in Transformer models, demonstrating superior performance across multiple sequential recommendation benchmarks. However, existing Mamba models exhibit vulnerability to data non-stationarity, noise, and outliers in time series modeling (Liu et al., 2022). Dao et al. proposed Mamba2 (Dao and Gu, 2024) based on Mamba. The model optimizes Mamba through the State Space Duality (SSD) mechanism and can effectively capture complex long-term dependencies. Additionally, the extension breakthrough architecture unifies State Space Model (SSM) with Transformer mechanisms achieving 2–8 times over Transformers without performance degradation, establishing a new paradigm for efficient long-term modeling.

The trajectory of a vessel exhibits strong temporal and continuous characteristics. Mamba2 can be utilized to model the vessel trajectory and capture long-term dependency features. However, the future trajectory of a vessel within a certain period is not only affected by the trajectory information in the past but also influenced by unexpected situations such as wind and waves, ocean currents, navigation plans, and sensor errors in the future period. Therefore, when modeling the trajectory of a vessel, the impact of contextual information needs to be considered. The traditional Mamba2 is usually unidirectional, and it is difficult to capture comprehensive contextual features, which may affect the prediction performance (Wu et al., 2024).

To address these issues, this paper proposes an innovative Bi_SSE model that leverages Bidirectional state Space (Bi_Space) and Squeeze Excitation attention with Causal self-attention (SECausal) to model vessel trajectory. Specifically, the Bi_Space component incorporates a bidirectional state space mechanism. The forward state space leverages the SSD of Mamba2 to extract current trajectory features, while the backward state space focuses on capturing historical trajectory features. This bidirectional approach enables more accurately capture of complex patterns and long-term dependencies in vessel trajectories. In addition, the parallel scan calculation method utilized in SSD ensures that the computational complexity is linearly, thereby effectively reducing the resource consumption of the model. The SECausal component is designed to augment the model's ability in capturing fine-grained features. This component combines squeeze excitation attention with causal self-attention. The squeeze excitation attention automatically learns the importance of features through "squeezing" and "exciting" operations, emphasizing important features while suppressing less relevant ones. The causal self-attention mechanism calculates attention weights based on the current moment and preceding information, allowing the proposed Bi_SSE model to judge the importance of features in light of historical information. The combination of them enables a more precise focus on features closely related to the current state and future trends of the vessel, disregarding irrelevant information, thereby capturing fine-grained vessel trajectory features. Subsequently, an MSEC encoder is developed by cross-stacking the Bi_Space and SECausal components. This hierarchical architecture enables the modeling of complex dependencies and the capture of comprehensive vessel trajectory features through complementary interactions between state space transformations and attention-based representations. Finally, the effectiveness of the proposed Bi_SSE model and its computational efficiency advantages are validated by comparing it with existing advanced prediction methods on publicly available datasets. Additionally, the impact of different combinations of Bi_Space and SECausal components on prediction performance is explored. The main contributions are as follows:

(1) To the best of our knowledge, this paper proposes the first bidirectional adaptive state space enhanced Bi_SSE model. This innovation enables comprehensive analysis of vessel trajectory dynamics through complementary forward-backward feature propagation and fine-grained feature extraction, significantly enhancing model robustness and interaction modeling capabilities.
(2) We constructed Bi_Space based on the Mamba2. By integrating bidirectional state space with dynamic selectivity, this approach maintains linear computational complexity while overcoming the contextual limitations of unidirectional models, significantly enhancing vessel trajectory feature capture capabilities.
(3) We designed the MSEC encoder, which integrates Bi_Space and SECausal. The former captures contextual features through bidirectional state space, while the latter strengthens critical features through attention under causal constraints, effectively modeling fine-grained features while suppressing noise. The hierarchical interaction significantly enhances feature capture capability by leveraging complementary global state propagation and adaptive attention modulation.
(4) Experimental results on public datasets show that the proposed Bi_SSE model significantly outperforms state-of-the-art models in performance while maintaining competitive computational efficiency.

2. Related work

The evolution of vessel trajectory prediction models has progressed from the initial rule-based and physical approaches to more sophisticated data-driven methodologies. In this section, we mainly review approaches that fall under traditional VTP methods and deep learning-based VTP, which are closely related to our work.

2.1. Traditional VTP methods

Many approaches have been developed to improve VTP performance. Early models estimated vessel positions over brief intervals based on current speed and heading. For example, Perera et al. (2010) introduced the Extended Kalman Filter to dynamically estimate the vessel's position, velocity, and acceleration, and combines a curved motion model with a linear position model for prediction. Diamant and Jin (2013) proposed a dead reckoning (DR) navigation method based on a single accelerometer, which classifies and projects acceleration data using an expectation-maximization algorithm to estimate vessel displacement. Skulstad et al. (2019) proposed a dead reckoning method based on recurrent neural networks, which can effectively maintain vessel position estimation and demonstrates comparable prediction accuracy to traditional Kalman filtering methods. These methods were subsequently applied to construct state transition matrices for future position predictions.

With the widespread adoption of AIS data, researchers have started using data-driven methods to improve prediction accuracy. Traditional machine learning methods, such as Support Vector Machines (SVM) and Random Forests, have been primarily employed for trajectory classification or short-term prediction tasks. Liu et al. (2019b) proposed the Support Vector Regression-based model to address the challenges of accurately modeling vessel trajectories and neural networks getting trapped in local optima. Zhang et al. (2020) developed a generalized destination prediction model based on Random Forests, leveraging the similarity between a vessel's current navigation trajectory and historical trajectories to predict its destination. These traditional data-driven approaches overcome the complexities of trajectory modeling and demonstrated significant effectiveness in prediction tasks when validated on large-scale real-world AIS datasets.

2.2. Deep learning-based VTP

Deep learning techniques, Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Graph Neural Networks (GNN) and Transformer architecture, have exhibited exceptional performance in vessel trajectory prediction. Zhu (2021) employed RNN to improve the prediction accuracy, establishing the superior performance of RNN over conventional backpropagation neural networks in AIS data processing. Tang et al. (2022) utilized LSTM to capture the motion characteristics of vessels, showcasing its advantages in handling time-series data. Zhang et al. (2023) a time-aware LSTM coupled with Generative Adversarial Network (GAN) architectures. This hybrid approach achieved optimized multi-vessel trajectory prediction through adversarial training mechanisms, yielding statistically significant improvements in both prediction accuracy and system reliability. Park et al. (2021) designed a clustering model that combined bidirectional LSTM with longest common subsequence similarity metrics, enabling more effective feature extraction for subsequent predictive modeling. Gao et al. (2021) proposed the MP-LSTM model that integrates physical constraints with LSTM to predict both future trajectory and time information over a period. Ma et al. (2024b) developed a hybrid deep learning method incorporating Seasonal-Trend Loess Decomposition, Multi-head Self-Attention, and LSTM, achieving notable improvements in ship time headway prediction robustness. Li et al. (2024a) proposed a bi-directional information fusion-driven deep network that combines bidirectional LSTM and bidirectional GRU into a three-layer information-enhanced network. This model fully utilizes the advantages of both networks, significantly improving the accuracy of prediction. These methods utilized the strength of LSTM capturing long-term dependencies and nonlinear temporal features to improve the prediction performance. However, they still face challenges including gradient decay and computational efficiency bottlenecks when processing long-term sequences, while its sensitivity to input data noise may compromise the model's generalization performance. Jiang et al. (2023) proposed the TRFM-LS model, which combines LSTM with Transformer. LSTM is used to learn temporal features, Transformer addresses the shortcomings of LSTM in capturing sequence information. Experiments have shown that TRFM-LS demonstrates higher accuracy in time series trajectory prediction. Nguyen and Fablet (2024) proposed TrAISformer, which is capable of extracting long-term temporal patterns from AIS data in high-dimensional space, while also addressing data heterogeneity and multi-modal issues through a newly designed loss function. However, Transformer attention-based models still face limitations in long-sequence tasks due to quadratic computational complexity growth (Gu and Dao, 2023; Wang et al., 2024; Sun et al., 2021).

3. Problem formulation

VTP is one of the key components in maritime traffic forecasting, aiming to predict the future state of a vessel using its historical data. In this paper, we aim to use the public AIS data ct_dma from Danish Maritime Authority (https://www.dma.dk/safety-at-sea/navigational-information/ais-data), which includes latitude(Lat), longitude(Lon), speed over ground (Sog), and course over ground (Cog) to prediction the future trajectory of vessels. The raw dataset contained approximately 712 million AIS messages from January 1, 2019, to March 31, 2019, including cargo and cruise vessels. Since raw AIS datasets often contain outliers and missing data, which can pose challenges to prediction, we preprocessed the dataset following (Jiang et al., 2023; Suo et al., 2024a). We filtered out vessel AIS records containing unrealistic speed values (Sog ≥ 30 knots), moored or at-anchor vessels, those within 1 nautical mile of the coastline, voyages with fewer than 20 data points, or durations shorter than 4 h. We merged voyages with a maximum interval of 2 h between consecutive AIS messages, split long voyages into shorter segments with a maximum sequence length of 20 h, and downsampled the data with a sampling rate of 10 min.

After preprocessing, it contains 11,888 AIS samples, which includes latitude, longitude, speed, and heading to prediction the future trajectory of vessels. $x_{0:T} = \{x_0, x_1, x_2, \dots, x_T\}$ denotes the input data of the model consists of AIS data over a period of time, where $x_t = [Lat, Lon, Sog, Cog]$ denotes the vessel's position information at time, the vessel trajectory prediction problem can be defined as:

$$\hat{y} = f(x_{T+1:T+P} \mid x_{0:T}) \tag{1}$$

where $\hat{y}$ represents the predicted vessel trajectory for the future period from $T+1$ to $T+P$, f denotes the prediction model, and $x_{0:T}$ is the historical AIS data used as input from time 0 to T.

4. The proposed bi_SSE model

The proposed Bi_SSE model mainly addresses the challenges of high computational complexity and difficulty in capturing comprehensive vessel trajectory features. In terms of improving resource consumption, we introduce Mamba2, which improved the selective SSM of Mamba through the SSD mechanism, enabling parallel scanning of data and effectively improving computational efficiency. At the same time, based on Mamba2, we designed a Bi_Space component, which uses the SSD mechanism to learn vessel trajectory features from forward and reverse directions. The forward SSD captures the features of the current time step, and the reverse SSD captures the trajectory features of a period in the past, thus achieving the learning of global context features of vessel trajectories. A SECausal component is designed to capture fine-grained vessel trajectory features. The squeeze excitation attention through "squeezing" and "exciting" operations gives more attention to important features and suppresses unimportant ones. The causal self-attention mechanism calculates the attention weight based on the current time and historical information, allowing the model to focus on current features while also judging their importance based on historical information, thereby capturing fine-grained vessel trajectory features. Finally, the MSEC module achieves the interaction between context features and fine-grained features, capturing comprehensive vessel trajectory features to improve prediction performance. As shown in Fig. 2, Bi_SSE model consists of three key components: the embedding layer, the MSEC module, and the prediction layer.

(1) *Embedding layer*: It maps the four-dimensional input AIS data x to high-dimensional data E, The embedded high-dimensional data, along with position encoding PE, serve as the input to the MSEC module.
(2) *The MSEC module*: It is composed alternately of two components: Bi_Space and and SECausal. The Bi_Space state space of Mamba2 is used to model the complex relationships between vessel trajectory sequences and capture global context features. The SECausal captures the fine-grained features of vessel sequences by merging squeeze excitation attention and causal self-attention.
(3) *Prediction layer*: The prediction layer maps the dynamic changes and spatial features captured by the MSEC module to the target space through a linear layer and activation function, enabling the prediction of the vessel's trajectory at the next step.

4.1. Embedding layer

Due to the significant differences in physical meaning and numerical scale between longitude, latitude, speed, and heading, directly embedding them into high-dimensional data may result in an imbalance in the importance weights across different feature ranges, potentially hindering the model's ability to effectively learn inter-feature relationships. The proposed Bi_SSE model employs a strategy of separate embedding for them. This approach maintains the distinct physical significance and independence of each feature, thereby facilitating the model's capacity to capture inherent feature relationships. Additionally, it can reduce training complexity and mitigate the risk of overfitting. The input four-dimensional trajectory data of the vessel is $x_{0:T} = \{x_0, x_1, x_2, \dots, x_T\}$, where $x_t = [x_{Lon}, x_{Lat}, x_{Sog}, x_{Cog}]$, $0 \le t \le T$, The x_t is embedded into high-dimensional space separately as:

$$\begin{aligned} e_{Lon} &= Embedding_{d_{Lon}}(x_{Lon}) \\ e_{Lat} &= Embedding_{d_{Lat}}(x_{Lat}) \\ e_{Sog} &= Embedding_{d_{Sog}}(x_{Sog}) \\ e_{Cog} &= Embedding_{d_{Cog}}(x_{Cog}) \end{aligned} \tag{2}$$

Then, they are concatenated to obtain $E_t = concat(e_{Lon}, e_{Lat}, e_{Sog}, e_{Cog})$, and $E = \{E_0, E_1, \dots, E_T\}$, $E \in \mathbb{R}^{(L,D)}$, where $L = T+1$, and $D = d_{Lon} + d_{Lat} + d_{Sog} + d_{Cog}$, D represents the embedding feature dimension.

4.2. The MSEC module

To capture comprehensive vessel trajectory features, we design the MSEC module, it consists two components: Bi_Space component and SECausal component, as shown in Fig. 3.

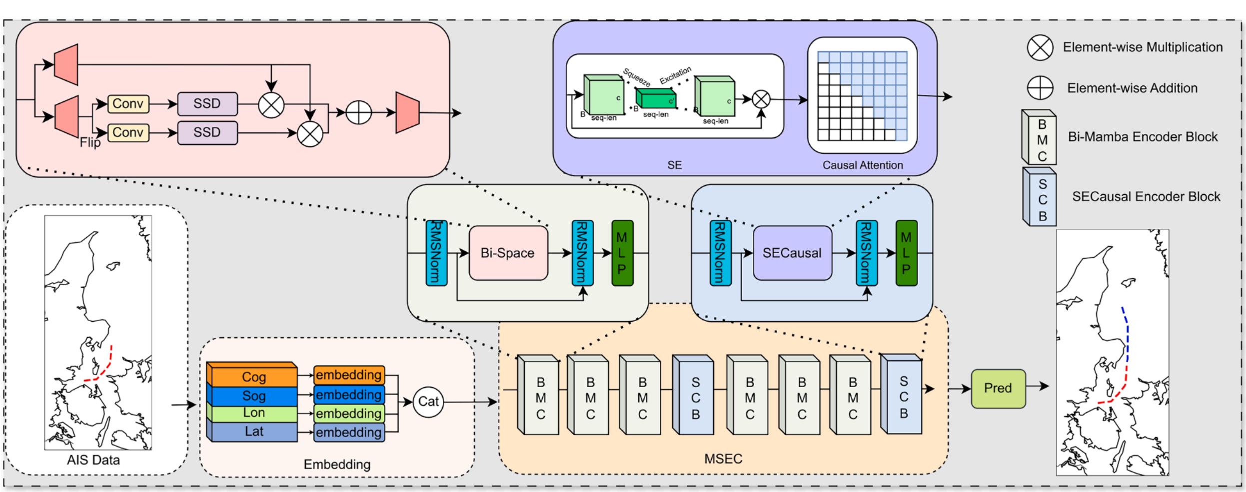


Fig. 2. Bi_SSE framework diagram.

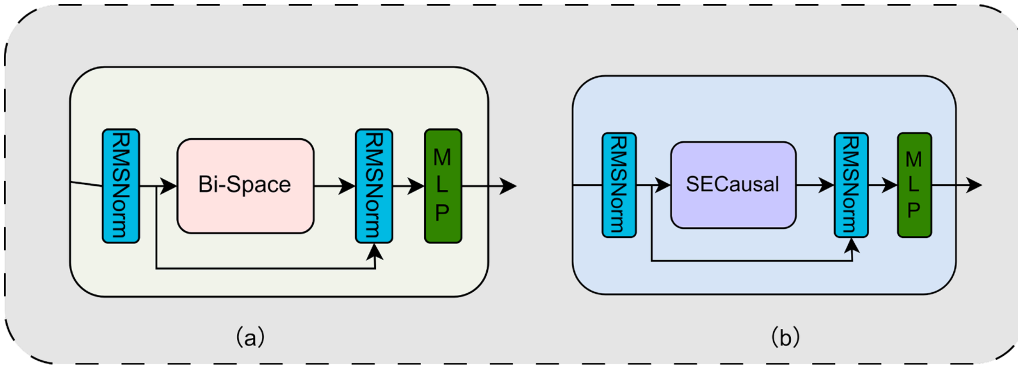


Fig. 3. The MSEC encoder: (a) the Bi_Space component and (b) the SECausal component.

(1) *Bi_Space*: The Bi_Space component is based on Mamba2 and employs a bidirectional state space to capture complex patterns of vessel trajectories and improve computational efficiency. The forward state space is utilized to derive features from the current trajectory, and the backward state space is tasked with capturing features from historical trajectories. This bidirectional strategy enhances the ability to capture complex patterns of vessel trajectories. The SSD mechanism of Mamba2 is responsible for reducing the parameter count and memory usage effectively through parallel projection and normalization mechanisms, thereby transforming the original temporal dependency modeling into linear complexity while maintaining enhanced sequence modeling capabilities.
(2) *SECausal*: The SECausal component integrates a squeeze excitation for channel-wise feature recalibration and a causal self-attention mechanism, jointly enabling dynamic feature adjustment and enhanced temporal dependency modeling of vessel trajectory.

The overall architecture employs an alternating structure, with the SECausal component inserted after every m layers of Bi_Space component. This configuration facilitates the model's ability to capture dependencies of vessel trajectories at multiple levels. The specific computational process can be represented as follows:

$$H^l = \begin{cases} \mathrm{Bi\_SS}(H^{l-1}), & \text{if } l \bmod m \neq 0 \\ \mathrm{SECausal}(H^{l-1}), & \text{if } l \bmod m = 0 \end{cases} \tag{3}$$

where l represents the layer number of the encoder, H^l is the encoded output of the lth layer, and $H^0 = E$

4.2.1. Bidirectional state space component

In the Bi_Space component, we have incorporated the core SSD of Mamba2 to enhance the computational efficiency. Specifically, the SSD encodes the internal states of vessel movement patterns into multidimensional vectors, while preserving temporal dependencies through a real-time data stream-driven state update mechanism. This approach can reduce parameters while simultaneously boosting the model's efficacy (Fig. 4)

In the VTP task, the output of the previous layer encoding, $H^{l-1} \in \mathbb{R}^{(L,D)}$, $L = T + 1$, can be viewed as the encoded result of the input data $x_{0:T}$, where $H^{l-1} = \{h_0^{l-1}, h_1^{l-1} \dots, h_T^{l-1}\}$. The spatial state equation for the vessel can be expressed as:

$$s_t = As_{t-1} + Bh_t^{l-1} \tag{4}$$

$$h_t^l = \mathrm{C}s_t + \mathrm{D}h_t^{l-1} \tag{5}$$

where $0 \le t \le T$, s_t represents the hidden state of the vessel at time, A is the state transition matrix, representing the system's state change, B denotes the input matrix, representing the impact of the input on the system's state, C is the output matrix, representing the effect of the state on the output, D represents the direct impact of the input on the output, and h_t^l represents the encoded result at time t in the current layer. The encoded result of the current encoding layer can be represented as $H^l = \{h_0^l, h_1^l, \dots, h_T^l\}$, $H^l \in \mathbb{R}^{(L,D)}$. We can derive:

$$s_0 = B_0 h_0^{l-1} \tag{6}$$

$$\begin{aligned} s_t &= A_t \cdots A_1 B_0 h_0^{l-1} + A_t \cdots A_2 B_1 h_1^{l-1} + \cdots + A_t B_{t-1} h_{t-1}^{l-1} + B_t h_t^{l-1} \\ &= \sum_{i=0}^{t} A_{t,i}^{\times} B_i h_i^{l-1} \end{aligned} \tag{7}$$

Multiplying by C_t gives the output equation:

$$h_t^l = \sum_{i=0}^{t} C_t^\top A_{t:i}^{\times} B_i h_i^{l-1} \tag{8}$$

Vectorizing for $t \in [T]$ gives the matrix form of the SSM variation:

$$H^l = \mathrm{SS}M(\mathrm{A}^l, \mathrm{B}^l, \mathrm{C}^l)(H^{l-1}) = MH^{l-1} \tag{9}$$

where $A^l \in \mathbb{R}^{(N,N)}$, $B^l \in \mathbb{R}^{(D,N)}$, $C^l \in \mathbb{R}^{(D,N)}$, N is the dimension of the hidden state in SSM, used to capture the dynamic behavior of the system, M is a lower triangular matrix, $M \in \mathbb{R}^{(T,T)}$, and when $i < j$, $M_{ij} = 0$, otherwise $M_{ij} = C_i^\top A_{i:j}^{\times} B_j = C_i^\top A_i \dots A_{j+1} B_j$.

In the Bi_Space component, we use the matrix decomposition method of Mamba2 to optimize SSM matrix computations. This process begins by partitioning the SSM matrix into specific size blocks, and then the separable properties of the matrix are used to decompose each off-diagonal

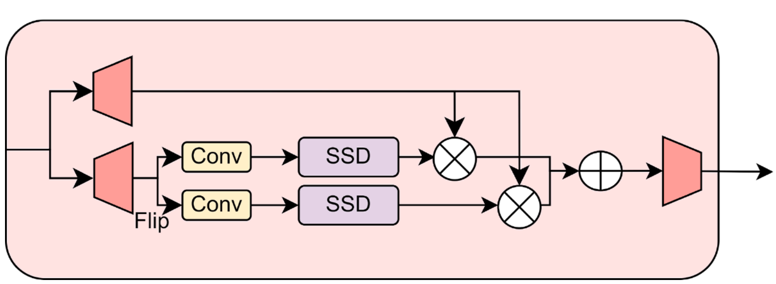


Fig. 4. Bi_Space.

block. Formula (9) can be improved as follows:

$$H^l = \mathrm{SSD}(\mathrm{A}^l,\mathrm{B}^l,\mathrm{C}^l)(H^{l-1}) = MH^{l-1} \tag{10}$$

where $A^l \in \mathbb{R}^{(N,N)}$, $B^l \in \mathbb{R}^{(D,N)}$, $C^l \in \mathbb{R}^{(D,N)}$, and $M \in \mathbb{R}^{(T,T)}$.

Compared to SSM, SSD achieves parallel computation and matrix calculations, facilitating optimal utilization of GPU resources. The Mamba2 architecture, founded on the SSD, has undergone modest refinements, with the primary alteration being the implementation of parallel SSM parameter generation, supplanting the previous sequential approach.

The Bi_Space component first uses a forward SSD to extract features from the input data, capturing the dominant features. Then, a backward SSD is used to augment the forward results, further elucidating concealed information and supplementary features within the sequential data. This bidirectional methodology enhances feature representation and more effectively models the complex relationships within sequence data. The formulation of Bi_Space is expressed as:

$$\begin{aligned} H^l &= w_1 \times SSD(A^{l_1}, B^{l_1}, C^{l_1})(H^{l-1}) \\ &\quad + w_2 \times SSD(A^{l_2}, B^{l_2}, C^{l_2})\big(\mathrm{Flip}(H^{l-1})\big) \end{aligned} \tag{11}$$

where A^{l_1}, B^{l_1}, C^{l_1} represent the matrices A, B, C used in the forward SSD, and A^{l_2}, B^{l_2}, C^{l_2} represent the matrices A,B,C used in the backward SSD. Additionally, to more flexibly combine the outputs of the forward and backward SSD, Bi_Space introduces two learnable weight parameters w_1 and w_2. These weights serve to dynamically adjust the contribution of each feature in the final feature representation by applying them to the forward and backward SSD outputs, respectively.

4.2.2. SECausal Component

We introduce a mechanism that integrates squeeze excitation attention with causal self-attention to enhance the capture of dynamic fine-grained features in trajectory data.

As shown in Fig. 5, it consists of two parts: Squeeze Excitation attention (SE) and Causal Attention. The core idea of SE is to dynamically adjust the importance of features, thereby highlighting key features and eliminating redundant information, the Causal Attention captures fine-grained features of vessel trajectories while strictly adhering to causal constraints.

SE received the hidden state $H^{l-1} \in \mathbb{R}^{T\times D}$, unlike traditional methods that use global average pooling to compress the dimension of the time step, which can cause the features of the vessel trajectory at specific moments to be influenced by adjacent time steps and to lose critical features. We directly employ a dynamic adjustment method that preserves the integrity of the time step dimension, using the parameter $\mathbf{w}$ to dynamically weight the importance of vessel trajectory features at each time step.

$$\mathbf{w} = \sigma\big(W_2 \cdot \delta(W_1 \cdot H^{l-1})\big) \tag{12}$$

where W_1 and W_2 are learnable parameter matrices, δ and σ are the ReLU and Sigmoid functions, respectively. Then, the input feature H^{l-1} is element-wise weighted along the feature dimension by the weight $\mathbf{w}$:

$$F = H^{l-1} \odot \mathbf{w} \tag{13}$$

F represents the encoded result after the SE attention.This process dynamically emphasizes the importance of different features at each time step, adapting to the non-stationarity of complex trajectory data.Then the causal self-attention block is used to capture the dynamic dependencies between time steps. The causal self-attention utilizes the masking mechanism to capture the dependencies between current and past time step information, the hidden state H^l is as follows:

$$H^l = \sum_{j\le i} \mathrm{softmax}\left(\frac{F_i \cdot F_j^\top}{\sqrt{D}}\right) \cdot F_j \tag{14}$$

w H^l represents the encoded result after SE attention at the lth layer.This process dynamically emphasizes the importance of different features at each time step, adapting to the non-stationarity of complex trajectory data. i and j represent the time steps in the sequence, with i indicating the current time step, and $j \le i$ indicating that the attention mechanism only considers information from the current and past time steps, D denotes the dimensionality of the feature vectors.

4.3. Prediction layer

The prediction layer maps the feature H output by the MSEC module to different probability spaces and uses the cross-entropy loss function (Shannon, 1948) for training to achieve the prediction of vessel trajectories. Specifically, the MSEC module encodes the input $(x_{0:T})$, including longitude, latitude, speed over ground, and course, into a high-dimensional H, which contains the spatiotemporal feature information of the next-step $(x_{1:T+1})$ trajectory. The feature $H \in \mathbb{R}^{T\times D}$ is then processed through a fully connected layer and decomposed into four subspaces, the output of each subspace can be represented as $\hat{y}_{Lat}$ $\hat{y}_{Lon}$, $\hat{y}_{Sog}$,$\hat{y}_{Cog}$, corresponding to the $x_{1:T+1}$ probability distributions of longitude, latitude, speed over ground, and course over ground, respectively. For example, for latitude, we divide its range (from 55.5° to 58.0°) into 250 probability distributions. Similarly, we perform the corresponding partitioning for longitude, speed over ground, and course to achieve more accurate trajectory predictions. We normalize the features $\hat{y}_{Lat}, \hat{y}_{Lon}, \hat{y}_{Sog}, \hat{y}_{Cog}$ through the softmax function, respectively, to obtain the probability distribution of each trajectory feature, the features of vessel are as follows:

$$\hat{y}_{Lat}, \hat{y}_{Lon}, \hat{y}_{Sog}, \hat{y}_{Cog} = f_{softmax}(split(Linear(H), d1, d2, d3, d4)) \tag{15}$$

where $\hat{y}_{Lat}, \hat{y}_{Lon}, \hat{y}_{Sog}, \hat{y}_{Cog}$ represent the $x_{1:T+1}$ four subspace features after the split. $d1$, $d2$, $d3$, $d4$ denote the number of intervals for latitude, longitude, speed over ground, and course over ground, respectively. The parameters of them are obtained by dividing the difference in their respective ranges by the interval value. The range of Lat is from 55.5° to 58.0°, with a difference of 2.5°. It is divided into intervals of 0.01°, and the number of intervals is 250, the value of $d1$ is 250. The range of Lon is from 10.3° to 13.0°, with a difference of 2.7°. It is also divided into intervals of 0.01°, resulting in 270 intervals in total, so the value of $d2$ is 270. The maximum value of the Sog is 30 knots. When divided into intervals of 1 knot each, there are 30 intervals in total, the value of

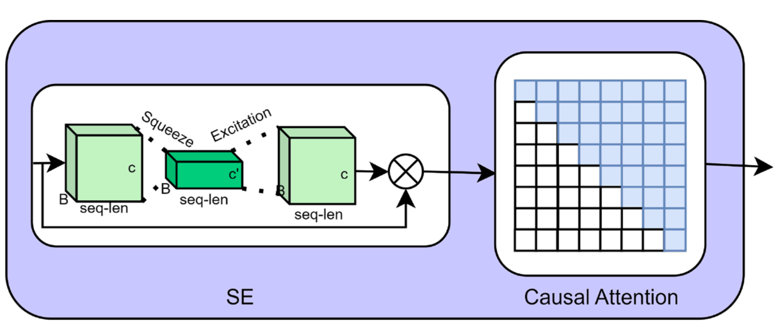


Fig. 5. The framework of SECausal component.

$d3$ is 30. The range of the Cog is from 0 to 360°, and it is divided into intervals of 5° each, there are 72 intervals in total, the value of $d4$ is 72.

We utilize the cross-entropy loss function to measure the discrepancy between the predicted probability distribution by the model and the truth values, and optimize the model parameters through backpropagation. The loss function of vessel prediction is as follows:

$$L_{Lat} = -\sum_{i=1}^{C} (y_{Lat})_i \log\left([\hat{y}_{Lat}]_i\right) \tag{16}$$

$$L_{Lon} = -\sum_{i=1}^{C} (y_{Lon})_i \log\left([\hat{y}_{Lon}]_i\right) \tag{17}$$

$$L_{Sog} = -\sum_{i=1}^{C} (y_{Sog})_i \log\left([\hat{y}_{Sog}]_i\right) \tag{18}$$

$$L_{Cog} = -\sum_{i=1}^{C} (y_{Cog})_i \log\left([\hat{y}_{Cog}]_i\right) \tag{19}$$

$$Loss = L_{Lat} + L_{Lon} + L_{Sog} + L_{Cog} \tag{20}$$

In the above expression y_{Lat}, y_{Lon}, y_{Sog}, y_{Cog} represents the $x_{1:T+1}$ true values of the four dimensions, and $\hat{y}_{Lat}, \hat{y}_{Lon}, \hat{y}_{Sog}, \hat{y}_{Cog}$ represent the probability distributions. $L_{Lat}, L_{Sog}, L_{Cog}, L_{Lat}$ are the loss function of longitude, latitude, speed, and course, and $Loss$ is the sum of these loss.

5. Training

To effectively predict trajectories, the proposed Bi_SSE model adopts a recursive approach, gradually accumulating information to infer the future trajectory. To predict the vessel's trajectory $f(X_{T+1:T+P} \mid x_{0:T})$ for the future time steps $T+1$ to $T+P$, we utilize the historical data $x_{0:T}$ from time steps 0 to T to perform P-step recursive prediction, where the data from the last time step of each prediction is concatenated with the original data, ultimately resulting in the AIS data $x_{0:T+P}$.

To ensure the predicted results closely match the actual trajectory, the Bi_SSE model is optimized through backpropagation. Specifically, the model parameters are adjusted by minimizing the error between the predicted values and the actual values. The AdamW (Loshchilov, 2017) optimizer is employed as the optimization algorithm for the model to enable end-to-end training. The AdamW optimizer is a variant of the Adam optimization algorithm, widely used in deep learning, particularly for tasks requiring high robustness and fast convergence. Compared to the standard Adam optimizer, AdamW introduces a weight decay mechanism, which applies direct L2 regularization constraints to model parameters during gradient updates. This improvement helps prevent model overfitting and enhances the ability to generalize complex patterns during training. By combining this recursive approach with error minimization using the AdamW optimizer, the model ensures high-accuracy predictions for future trajectories.

6. Experiments

In this section, we evaluate the proposed Bi_SSE model against state-of-the-art methods using real-world dataset. We begin by presenting the experimental setup; then, we conduct extensive experiments to demonstrate the superiority of the proposed Bi_SSE model.

6.1. Experimental setup

Dataset. We conduct experimental analysis on the ct_dma (Nguyen and Fablet, 2024; Jiang et al., 2023; Suo et al., 2024a) dataset, which includes AIS observation data of cargo and cruise vessels from January 1, 2019, to March 31, 2019. The study area (Region of Interest, ROI) is rectangular, ranging from latitude 55.5° to 58.0° and longitude 10.3° to 13.0°. In terms of data partitioning, we use the dataset from January 1, 2019, to March 10, 2019, as the training set, the data from March 11, 2019, to March 20, 2019 as the validation set, and the dataset from March 21, 2019 to March 31, 2019 as the test set. After filtering out short trajectory sequences, a total of 11,888 AIS samples were retained. This data partitioning method ensures the model's stability in time series and provides a reasonable experimental basis for evaluating the model's trajectory prediction performance across different periods.

Evaluation metrics. Following the prominent works in VTP (Nguyen and Fablet, 2024; Takahashi et al., 2024; Zhang et al., 2023; Park et al., 2021), three standard metrics are used for the performance evaluation of the proposed Bi_SSE model,including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Haversine distance. MAE reflect the model's absolute errors, and RMSE particularly focuses on the impact of larger errors. Haversine distance, crucial in vessel trajectory prediction, measures spherical distances between geographic coordinates. Lower Haversine distance indicate superior predictive.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{21}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} \tag{22}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and N is the number of samples.

$$d = 2r \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{23}$$

where d denotes the prediction error at time step that is calculated as the haversine distance between the true position and the predicted one. $\Delta\phi$ represents the difference in latitude between the two points, ϕ_1 and ϕ_2 represent the latitudes of two points respectively, $\Delta\lambda$ represents the difference in longitude, and r is the Earth's radius (approximately 6371 km).

Baselines. We compare the proposed Bi_SSE model with deep learning models and recently proposed high-performance VTP models, the baselines are as follows:

1. LSTM (Tang et al., 2022): LSTM is a special version of RNN with a memory cell architecture designed to address the long-term dependency problem of traditional RNNs for VTP.
2. Attention_LSTM (Liu and Ma, 2022): The Transformer model is applied to AIS data to overcome the limitations of traditional methods (such as LSTM) in handling nonlinear trajectory data.
3. Transformer (Takahashi et al., 2024): The Transformer model is applied to AIS data to overcome the limitations of traditional methods (such as LSTM) in handling nonlinear trajectory data.
4. TrAISformer (Nguyen and Fablet, 2024): Extracts long-term temporal patterns of AIS data in high-dimensional space, while addressing the data's heterogeneity and multimodal issues through a newly designed loss function.
5. Mamba (Suo et al., 2024b): Using Mamba for VTP, aiming to address issues such as insufficient prediction accuracy and slow inference speed in models like Transformer.

Parameter settings. The experimental setup in this study includes the following hardware and software configurations: the CPU is an Intel Core i5-13400F, the GPU is an Nvidia RTX 4060Ti with 8GB of memory, the operating system is Ubuntu, and the PyTorch version is 2.3.1 + cu118. For hyperparameters, the batch size is set to 16, the historical data step P is 30 (corresponding to 5 h), and the model is trained for 20 epochs. The learning rate was not provided. The prediction step length Q is set to 6 (1 h), 12 (2 h), and 18 (3 h). The number of encoding layers is 24, and the Bi-Mamba2:SECausal ratio is set to 7:1.

To evaluate the accuracy of the model proposed in this paper for vessel trajectory prediction, we conducted detailed tests for both short-term (3 h) and long-term (10 h) prediction tasks. In addition, we performed experiments and analyses on the model's convergence speed, resource consumption, and performance in real-world scenarios. Furthermore, we compared the performance of Bi-Mamba2 and SECausal models with different proportions.

7. Experiments and results

7.1. Experimental results and analysis

To comprehensively evaluate the performance of the proposed Bi_SSE model, the experimental results are categorized into short-term and long-term trajectory prediction. Specifically, short-term trajectory prediction is designed to forecast vessel trajectories within 1–3 h intervals, while long-term trajectory prediction aims to predict vessel trajectories up to 10 h intervals. The results are shown in Table 1 and Fig. 6.

Short-term trajectory prediction results analysis. Table 1 shows the performance of our proposed Bi_SSE model and baselines on VTP in terms of MAE, RMSE, and Haversine distance across 1, 2, and 3-h prediction intervals. For convenience of comparison,

1. LSTM performs poorly in three-hour predictions (MAE = 3.6211, Haversine = 3.9328). This stems from LSTM's inability to retain information from prolonged sequences and lack of spatial feature extraction, resulting in consistently high Haversine distances across temporal ranges, indicating imprecise modeling of geographic trajectory changes.
2. The Attention_LSTM model shows only a modest improvement over the standard LSTM model, with the three-hour MAE decreasing from 3.6211 to 3.5974, while the Haversine distance even slightly increases. In addition, the one-hour Haversine distance is 1.4785, which is still slightly higher than that of the Transformer (1.0661), indicating that the model fails to effectively capture fine-grained information in the short term.
3. The Transformer model demonstrates enhanced long-term dependency modeling but lacks domain-specific optimization, it achieves notable reductions in three-hour prediction MAE (3.3615) and Haversine distance (3.6479) relative to Attention_LSTM. However, error accumulation persists, and long-term prediction accuracy remains suboptimal. The limited improvement in geographical location prediction accuracy indicates insufficient optimization for spatiotemporal features inherent to vessel trajectory forecasting.
4. TrAISformer performs better for long-term predictions, it outperforms Transformer with an MAE of 3.2611 and Haversine distance of 3.6089 for three-hour predictions. However, the advantage compared to short-term predictions (1-h prediction with an MAE of 0.5884) is reduced. The errors for long-term predictions are still not fully resolved, and the Haversine distance for the 2-h prediction is only slightly lower than Transformer (from 2.2150 to 2.1393),

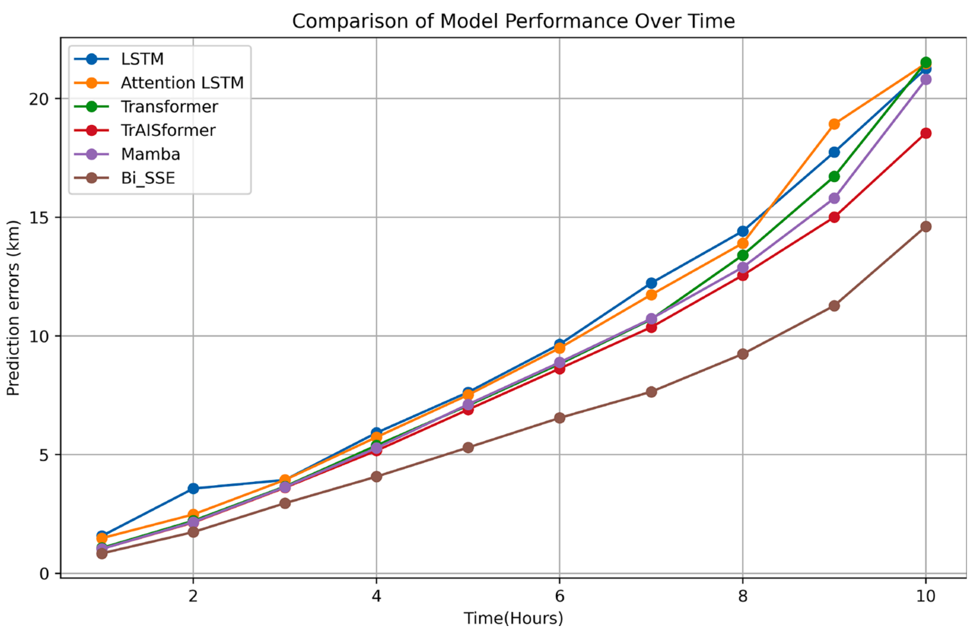


Fig. 6. Comparison of prediction errors for each model at the ten-hour forecast.

Table 1
Performance comparison of the dataset for baselines, MAE and RMSE values are multiplied by 10^2. The unit of Haversine distance is "km". The best results are highlighted in boldface, while the second-best results are underlined.

Model	MAE			RMSE			Haversine		
	1 h	2 h	3 h	1 h	2 h	3 h	1 h	2 h	3 h
LSTM	0.6246	1.8279	3.6211	0.6525	2.9815	6.5990	1.5508	2.5709	3.9328
Attention_LSTM	0.6085	1.7721	3.5974	0.5969	2.6611	6.5150	1.4785	2.4798	3.9340
Transformer	0.5930	1.6315	3.3615	0.5004	2.4772	6.2716	1.0661	2.2150	3.6479
TrAISformer	0.5884	1.6252	3.2611	0.5104	2.4572	6.2554	1.0289	2.1393	3.6089
Mamba	0.5876	1.6320	3.2570	0.5218	2.4574	6.2074	1.0278	2.1589	3.6325
Bi_SSE	**0.5405**	**1.3380**	**2.7334**	**0.4797**	**2.2381**	**5.2743**	**0.8379**	**1.7373**	**2.9527**
Improved (%)	8.03 %	17.67 %	16.07 %	4.13 %	8.91 %	15.03 %	18.47 %	18.79 %	18.18 %

indicating that the model's ability to model geographical locations still needs further optimization.

5. The MAE of Mamba for 1-h predictions (0.5876) is slightly lower than TrAISformer (0.5884), the RMSE (0.5218) is higher which suggests unresolved short-term error fluctuations. For the 3-h prediction, the Haversine distance is 3.6525, significantly higher compared to the 1-h prediction, indicates that while Mamba's long-term dependency modeling surpasses previous models, there remains substantial room for enhancement.

The proposed Bi_SSE model effectively addresses the limitations associated with short-term predictions. Compared to Mamba and TrAISformer, Bi_SSE significantly reduces the MAE for 1-h predictions (0.5405), a decrease of approximately 8.03 %, and RMSE (0.4797), which shows a reduction of about 4.13 %. Moreover, Bi_SSE successfully mitigates the problem of long-term error accumulation. For 3-h predictions, the model achieves significantly lower MAE (2.7334) and Haversine distance (2.9527 km) compared to the baselines, demonstrating its remarkable improvement in long-term dependency modeling. For 1-h predictions, Bi_SSE's Haversine distance is just 0.8379 km clearly outperforming Mamba and TrAISformer, showcasing its ability to model geographic locations. Overall, Bi_SSE exhibits considerable advantages in both short-term and long-term prediction tasks, addressing the shortcomings of previous models in terms of accuracy, spatiotemporal features, and error accumulation.

Long-term trajectory prediction result analysis. To avoid overfitting in short-term predictions (such as the first three hours) and ensure the model's generalization ability in long-term predictions, the extend the prediction horizon to 10 h. By increasing the prediction duration, the model's prediction ability and robustness over a longer time scale can be more comprehensively assessed. The experimental results are shown in Fig. 6. From the 10-h prediction error results, it can be observed that:

1. LSTM exhibits rapid error accumulation over time, reaching 21.2634 km at 10 h, performing the worst. The error remains relatively mild within 3 h, but from 4 h onward (5.9202 km), the error increases significantly, indicating that LSTM has poor long-term dependency modeling capability.
2. Attention_LSTM shows some improvement within initial 3 h. However, after 8 h (13.9059 km), the error starts to rise rapidly, and by 10 h, the error reaches 21.4780 km, which is similar to LSTM. This suggests that Attention_LSTM still struggles with capturing long-term dependencies effectively.
3. Transformer demonstrates enhanced long-term prediction accuracy, outperforming Attention_LSTM and LSTM from 1 h to 9 h. However, the 10-h error reaches 21.5362 km, indicating that while Transformer has better stability in predictions, it has not fully resolved the issue of error accumulation.
4. TrAISformer improves the predictions for the first 6 h, but from the 8th hour onward, the rate of error accumulation increases. Moreover, it displays in sufficient mid-term prediction accuracy, with the 9-h error of 15.0054 significantly higher than Bi_SSE.
5. Mamba performs similarly to TrAISformer in short-term predictions, but from the 6th hour onward (10.7259 km), the rate of error growth accelerates. This indicates that Mamba is suitable for short-term predictions but lacks stability. Compared to TrAISformer, the fluctuations gradually increase, especially between 8 and 10 h, where error accumulation becomes evident, suggesting that Mamba's ability to model long-term trajectories still requires optimization.

The proposed Bi_SSE model exhibits the most stable error growth trend. Its short-term (1-h) error is only 0.8379 km, and the 12-h error is the lowest at 14.6144 km, which is approximately 21.21 % lower than the best-performing TrAISformer. Moreover, the error growth is relatively smooth across all time periods.

Convergence trend analysis. Fig. 7 presents a comparative analysis of loss value fluctuations per epoch during training, revealing the differences in convergence speed. The proposed Bi_SSE and Mamba models exhibit exceptional performance, reaching fitted states in 8 and 5 epochs, respectively. Compared to others, their training efficiency is significantly improved: Bi_SSE converges approximately 8 times faster than LSTM and 10 times faster than Attention_LSTM; 1.5 times faster than Transformer, and 2 times faster than TrAISformer. This efficiency advantage makes the proposed Bi_SSE model more suitable for applications requiring rapid training and low computational resources. In contrast, LSTM and Attention_LSTM have slower convergence speeds, requiring 68 and 87 epochs to reach fitting, respectively. This reflects that traditional LSTM models require more iterations to train on long sequence data, making them less efficient. Transformer and TrAISformer achieve fitting in 12 and 15 epochs, respectively, showing significant improvement over traditional LSTM models, while still slightly lagging behind the Mamba series. The proposed Bi_SSE introduces an SSD operation for bidirectional complementary modeling of input features, slightly decreasing training efficiency compared to the Mamba module.

Resource consumption analysis. Fig. 8 shows the comparison of per-epoch training durations and model parameter sizes across various models. It reveals that traditional LSTM and Attention_LSTM models require extended training periods of 296 s and 334 s, respectively, mainly due to their complex computation process and weaker parallelization capability. Transformer and TrAISformer exhibit training times of 229 and 217 s, respectively, demonstrating enhanced efficiency attributed to their architecture's effective self-attention mechanism. The Mamba model's training duration of 230 s per epoch aligns closely with Transformer, indicating that it strikes a balance between efficiency and

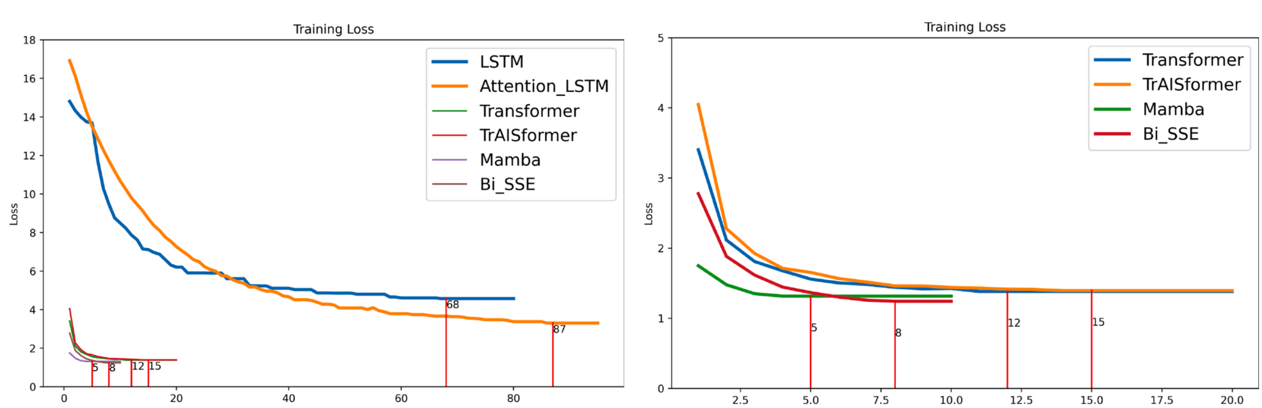


Fig. 7. The training loss values of the six models (left) and the training loss values of the four faster models (right).

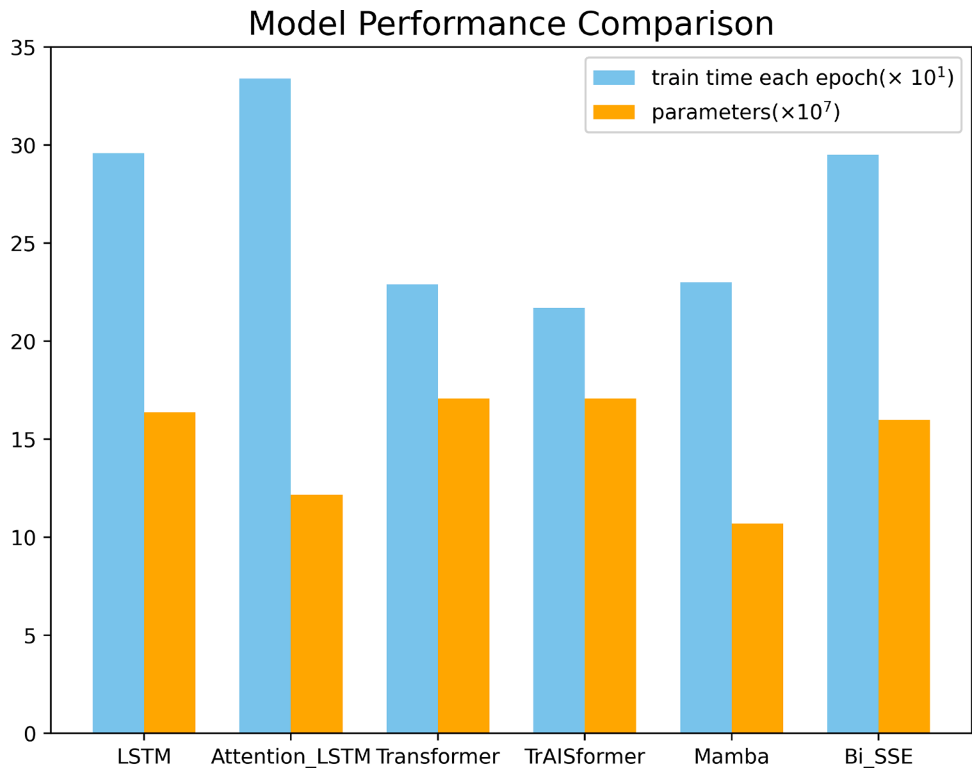


Fig. 8. Training time per epoch and parameter sizes for each model.

performance in sequence modeling. The proposed Bi_SSE model has a training time of 295 s, slightly higher than Mamba and Transformer. This is due to Bi_SSE incorporating the complementary feature modeling process (bidirectional SSD computation). Although the added SSD computation leads to a slight reduction in efficiency, resulting in a slight delay in inference and training speeds compared to Mamba, Bi_SSE significantly improves the model's feature representation ability and prediction accuracy. Considering both the prediction performance and computational efficiency, Bi_SSE shows a good balance between the two, making it particularly suitable for scenarios with high prediction accuracy

Analysis of actual performance. Fig. 9 shows the performance of Bi_SSE and other baselines across 3 real-world scenarios. The objective is to forecast a vessel's 10-h (60 time steps.), 14-h (84 time steps.) and 16-h(96 time steps.) future trajectory based on its preceding 2-h (12 time steps.) movement data. Specifically, Fig. 9(a) represents small-angle turning, Fig. 9(b) indicates sudden large-angle turning, and Fig. 9(c) shows multiple turning directions. We can see that vessels exhibit dynamic behavior, potentially undergoing significant changes in direction, speed, and environmental interactions over extended periods. Transformer, TrAISformer, Mamba, and the proposed Bi_SSE model outperform LSTM and AttentionLSTM in the three scenarios. This can be attributed to the limitations of LSTM in capturing complex long-term dependencies. The proposed Bi_SSE model exhibits superior prediction performance compared to baselines, mainly due to its incorporation of bidirectional state space and squeeze excitation attention with causal self-attention mechanisms, which enable the simultaneous capture of both long-term and fine-grained vessel trajectory features, thereby enhancing prediction accuracy.

The performance of different baselines also varies across different scenarios. Fig. 9(a) shows that the vessel's heading angle changed slightly during the next 10 h. The predictive performance of Transformer, TrAISformer, Mamba, and the proposed Bi_SSE model is superior to the LSTM and AttentionLSTM, mainly due to the LSTM's limitations in capturing long-term dependencies, which are critical for maintaining accuracy in temporal sequences.

From Fig. 9(b), we can see that the predictive performance of TrAISformer and the proposed Bi_SSE model is superior to other baselines when abrupt large-angle turning. The enhancement stems from their innovative architecture, which incorporates discrete high-dimensional AIS data representations and a new loss function. These modifications explicitly address data heterogeneity and multimodality, improving prediction performance.

Fig. 9(c) shows the predicted results for the next 16 h. It can be seen that the vessel turned direction multiple times during operation, and LSTM performance is the worst when changing direction at small angles. The proposed Bi_SSE model achieves remarkable accuracy, with its predicted trajectory closely aligning with the actual path. This is because Bi_SSE model is highly proficient at capturing complex trajectory changes, proving its reliability under long periods and varying trajectory conditions.

Analysis of the impact of different ratios on prediction accuracy. We also explore the performance of Bi_SSE at different ratios by fine-tuning models with various ratios of Bi_Space and SECausal modules. The cumulative average Haversine distances during the initial ten-hour of each model under ratios of 1:1, 2:1, 3:1, 5:1, 7:1 and 11:1 are shown in Fig. 10. The optimal performance is observed at a 7:1 ratio of Bi_Space to SECausal, suggesting that an appropriate module balance enhances the model's ability to capture complex dependencies and fine-grained features, thereby significantly improving prediction performance.

7.2. Ablation experiment

To further analyze the impact of Bi_Space and the SECausal on the overall model performance, we designed five ablation experiments:

1. A model using 24 layers of the Mamba2 module (Mamba2-24) to evaluate its performance in the deep encoding structures.

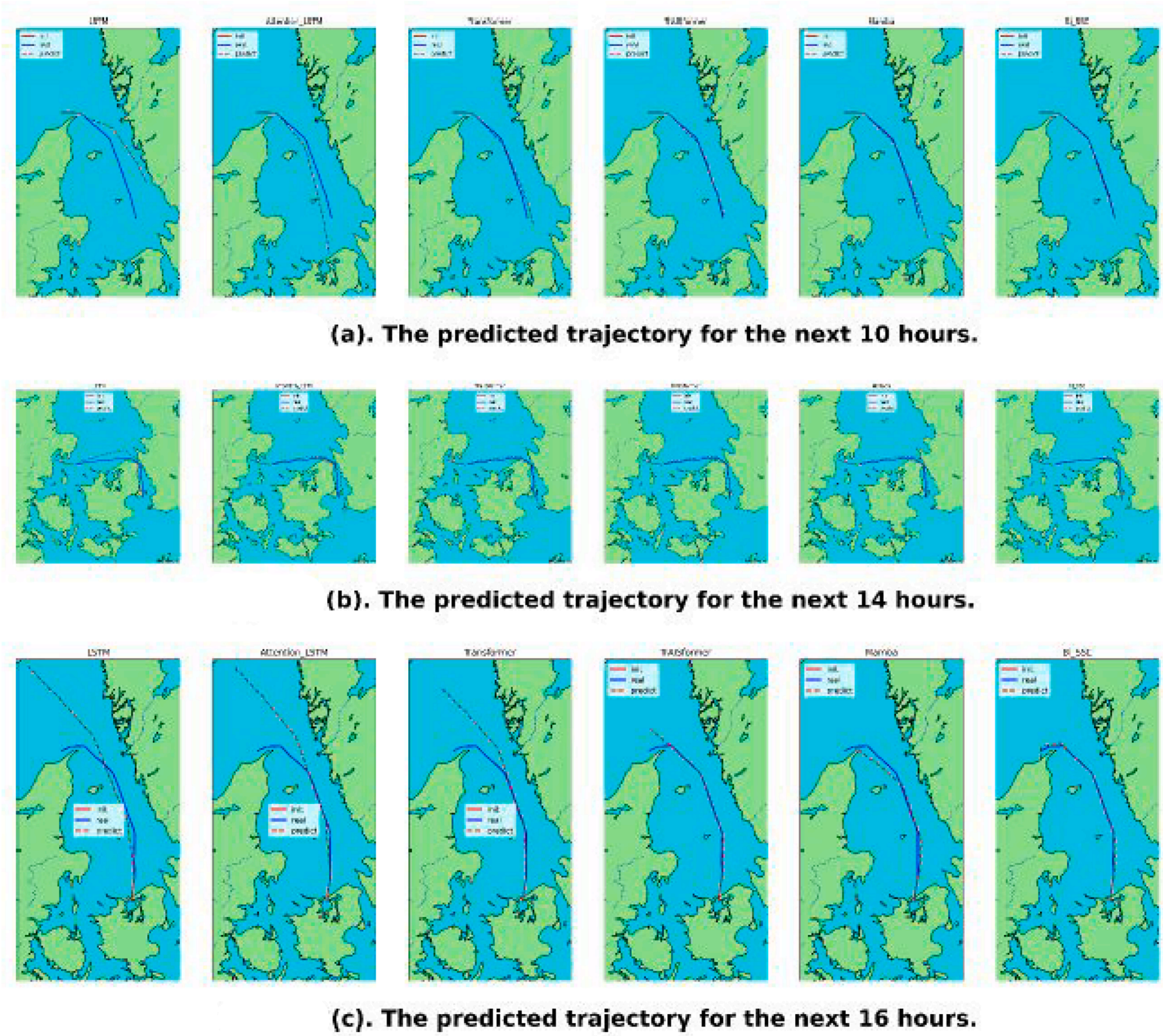


Fig. 9. Visualization of AIS-based vessel trajectory predictions. The red solid line represents the trajectory of the first 2 h, the blue solid line represents the actual trajectory of the vessel, and the red dashed line represents the predicted result.

2. A model with 24 layers of SECausal modules (SECausal-24) to explore its effectiveness without Bi_Space participation.
3. Based on Bi_SSE, remove all SECausal modules and retain only 21 layers of the Bi_Space module (Bi_Space-21) to test the performance.
4. To evaluate the model's performance without enhancements to causal self-attention, all SECausal modules in Bi_SSE were substituted with causal self-attention (Bi_SpaceCausal).
5. To examine the model's performance without improvements to Mamba2, all Bi_Space modules in Bi_SSE were replaced with Mamba2 (Mamba2_SEC).

These experiments enable a systematic evaluation of the role and performance variations of different module proportions in trajectory prediction. Analysis of the experimental results depicted in Fig. 11 yields the following conclusions:

Based on the comparative analysis of the experimental results shown in Fig. 11, the following conclusions can be drawn:

1. SECausal Module: The SECausal module demonstrates stable performance in short-term predictions (e.g., 1-h error of 0.8582 and 2-h error of 1.8084) and exhibits enhanced capability in long-sequence predictions (e.g., 10-h error of 16.4817). This indicates that dynamic feature weighting and causal relationship modeling of the SECausal module effectively improve prediction accuracy, particularly for long-sequence tasks.
2. Combining Mamba2 and SECausal: The integration of Mamba2 with causal attention results in excellent performance in both short-sequence and long-sequence tasks (e.g., 1-h error of 0.8609 and 10-h error of 15.4983). This combination addresses the limitations of individual modules in various modeling aspects, leading to overall performance enhancement. The optimized Bi_SSE model further enhances overall performance (e.g., 1-h error of 0.8507 and 10-h error of 14.5392), particularly excelling in long-sequence prediction tasks. The design of alternately stacking Mamba2 and SECausal modules enables effectively balance local and global dependency modeling while maintaining high computational efficiency.
3. Integration Bi_Space and SECausal: Building upon the combination of Bi_Space and SECausal, the introduction of Bi_Space further reduces prediction errors, demonstrating superior accuracy in shorter time frames (outperforming Bi_SSE in the 1–9 Hour range). This indicates that the complementary feature extraction capability and

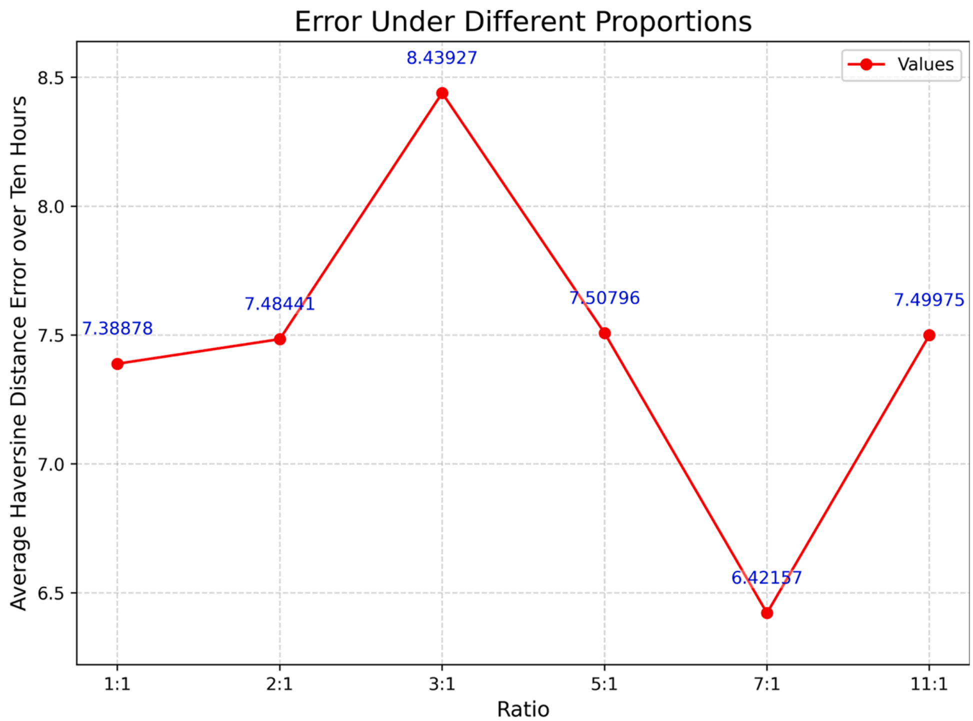


Fig. 10. Comparison of Bi_SSE performance at different ratios during the initial ten-hour.

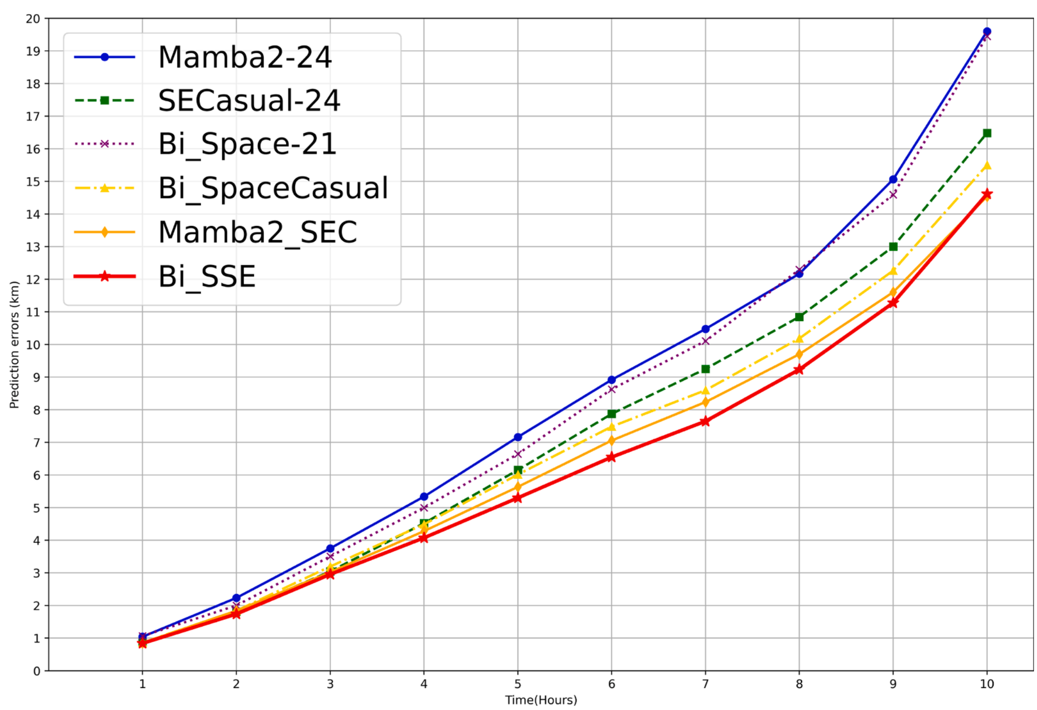


Fig. 11. Comparison of model accuracy in ablation experiments.

finer-grained sequence modeling provided by Bi_Space synergize effectively with the long-range dependency capture of the SECausal, resulting in improved overall prediction performance.

4. Impact of Layer Configuration: The experimental results demonstrate that the layer configuration of Bi_Space and SECausal significantly influences model performance. By employing a reasonable alternation of these modules, the model can better adapt to the features of complex trajectory data, achieving higher accuracy while maintaining lower memory consumption.

8. Conclusion

We propose a novel Bi_SSE model, which integrates Bi_Space and SECausal component for vessel trajectory prediction. Experimental results show that the proposed Bi_SSE model outperforms baselines in both short-term (1–3 h) and long-term (10 h) predictions, particularly excelling in stability and robustness over extended periods, with prediction errors consistently under 15 nautical miles. Ablation experiments further validate the effectiveness of combining Bi_Space and SECausal module, enhancing the model's ability to capture complex dynamics and exhibit great potential in trajectory prediction. Furthermore, the proposed Bi_SSE achieves an optimal balance between training efficiency and parameter scale, optimizing performance and computational resources. For future studies, we will explore integrating ship size, ship type information, and port destination ports data to a multi-source fusion framework, thereby improving the accuracy of prediction and meeting practical application requirements.

CRediT authorship contribution statement

Caiquan Xiong: Methodology; **Jiaming Li:** Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Data curation, Conceptualization; **Xinyun Wu:** Formal analysis; **Donghua Liu:** Methodology; **Qi Wang:** Validation; **Rong Gao:** Resources; **Jinjia Ruan:** Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant number 62206116, National Key Research and Development Program under Grant number 2023YFC3107903.

References

Borkowski, P., 2017. The ship movement trajectory prediction algorithm using navigational data fusion. Sensors 17 (6), 1432.
Chen, X., Hu, R., Luo, K., Wu, H., Biancardo, S.A., Zheng, Y., Xian, J., 2025. Intelligent ship route planning via an a search model enhanced double-deep q-network. Ocean Eng. 327, 120956.
Chen, X., Wu, H., Han, B., Liu, W., Montewka, J., Liu, R.W., 2023. Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach. Eng. Appl. Artif. Intell. 125, 106686.
Chen, X., Zhang, H., Zhao, F., Cai, Y., Wang, H., Ye, Q., 2022. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. IEEE Trans. Instrum. Meas. 71, 1–12.
Dao, T., Gu, A., 2024. Transformers are SSMs: generalized models and efficient algorithms through structured state space duality. In: Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21–27, 2024. OpenReview.net.
Diamant, R., Jin, Y., 2013. A machine learning approach for dead-reckoning navigation at sea using a single accelerometer. IEEE J. Ocean. Eng. 39 (4), 672–684.
Gao, D.-w., Zhu, Y.-s., Zhang, J.-f., He, Y.-k., Yan, K., Yan, B.-r., 2021. A novel MP-LSTM method for ship trajectory prediction based on AIS data. Ocean Eng. 228, 108956.
Gu, A., Dao, T., 2023. Mamba: Linear-time sequence modeling with selective state spaces. arXiv:2312.00752
Gu, Y., Hu, Z., Zhao, Y., Liao, J., Zhang, W., 2024. Mfgtn: a multi-modal fast gated transformer for identifying single trawl marine fishing vessel. Ocean Eng. 303, 117711.
Guo, S., Liu, C., Guo, Z., Feng, Y., Hong, F., Huang, H., 2018. Trajectory prediction for ocean vessels base on k-order multivariate Markov chain. In: Wireless Algorithms, Systems, and Applications: 13th International Conference, WASA 2018, Tianjin, China, June 20–22, 2018, Proceedings 13. Springer, pp. 140–150.
Jiang, D., Shi, G., Li, N., Ma, L., Li, W., Shi, J., 2023. Trfm-ls: transformer-based deep learning method for vessel trajectory prediction. J. Mar. Sci. Eng. 11 (4), 880.
Jin, F., Liu, K., Liu, C., Cheng, T., Zhang, H., Lee, V. C.S., 2024. A cooperative vehicle localization and trajectory prediction framework based on belief propagation and transformer model. IEEE Trans. Consum. Electron. 70 (1), 2746–2758.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems.
Li, H., Xing, W., Jiao, H., Yang, Z., Li, Y., 2024a. Deep bi-directional information-empowered ship trajectory prediction for maritime autonomous surface ships. Transp. Res. Part E 181, 103367.
Li, L., Zhang, H., Gan, Z., 2024b. Factors affecting college students' attitudes towards carpooling. Transp. Saf. Environ. 6 (2), tdad025.
Li, W., Zhang, C., Ma, J., Jia, C., 2019. Long-term vessel motion predication by modeling trajectory patterns with AIS data. In: 2019 5th International Conference on Transportation Information and Safety (ICTIS). IEEE, pp. 1389–1394.
Liu, C., Guo, S., Feng, Y., Hong, F., Huang, H., Guo, Z., 2019a. L-VTP: Long-term vessel trajectory prediction based on multi-source data analysis. Sensors 19 (20), 4365.
Liu, J., Shi, G., Zhu, K., 2019b. Vessel trajectory prediction model based on AIS sensor data and adaptive chaos differential evolution support vector regression (ACDE-SVR). Appl. Sci. 9 (15), 2983.
Liu, T., Ma, J., 2022. Ship navigation behavior prediction based on AIS data. IEEE Access 10, 47997–48008.
Liu, Y., Wu, H., Wang, J., Long, M., 2022. Non-stationary transformers: exploring the stationarity in time series forecasting. Adv. Neural Inf. Process. Syst. 35, 9881–9893.
Loshchilov, I., 2017. Decoupled weight decay regularization. arXiv:1711.05101
Ma, J., Li, F., Wang, B., 2024a. U-mamba: enhancing long-range dependency for biomedical image segmentation. arXiv:2401.04722
Ma, Q., Du, X., Liu, C., Jiang, Y., Liu, Z., Xiao, Z., Zhang, M., 2024b. A hybrid deep learning method for the prediction of ship time headway using automatic identification system data. Eng. Appl. Artif. Intell. 133, 108172.
Nguyen, D., Fablet, R., 2024. A transformer network with sparse augmented data representation and cross entropy loss for AIS-based vessel trajectory prediction. IEEE Access 12, 21596–21609.
Park, J., Jeong, J., Park, Y., 2021. Ship trajectory prediction based on bi-LSTM using spectral-clustered AIS data. J. Mar. Sci. Eng. 9 (9), 1037.
Perera, L.P., Soares, C.G., et al., 2010. Ocean vessel trajectory estimation and prediction based on extended Kalman filter. In: The Second International Conference on Adaptive and Self-Adaptive Systems and Applications. Citeseer, pp. 14–20.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.
Skulstad, R., Li, G., Fossen, T.I., Vik, B., Zhang, H., 2019. Dead reckoning of dynamically positioned ships: using an efficient recurrent neural network. IEEE Robot. Autom. Mag. 26 (3), 39–51.
Sun, G., Hua, Y., Hu, G., Robertson, N., 2021. Mamba: Multi-level aggregation via memory bank for video object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35, pp. 2620–2627.
Suo, Y., Chen, W., Claramunt, C., Yang, S., 2020. A ship trajectory prediction framework based on a recurrent neural network. Sensors 20 (18), 5133.
Suo, Y., Ding, Z., Zhang, T., 2024a. The mamba model: a novel approach for predicting ship trajectories. J. Mar. Sci. Eng. 12 (8).
Suo, Y., Ding, Z., Zhang, T., 2024b. The mamba model: a novel approach for predicting ship trajectories. J. Mar. Sci. Eng. 12 (8), 1321.
Sutulo, S., Moreira, L., Soares, C.G., 2002. Mathematical models for ship path prediction in manoeuvring simulation systems. Ocean Eng. 29 (1), 1–19.
Takahashi, K., Zama, K., Hiroi, N.F., 2024. Ship trajectory prediction using AIS data with transformer-based AI. In: 2024 IEEE Conference on Artificial Intelligence (CAI). IEEE, pp. 1302–1305.
Tang, H., Yin, Y., Shen, H., 2022. A model for vessel trajectory prediction based on long short-term memory neural network. J. Mar. Eng. Technol. 21 (3), 136–145.
Wang, Z., Li, C., Xu, H., Zhu, X., 2024. Mamba YOLO: SSMs-based YOLO for object detection. arXiv:2406.05835
Wu, H., Yang, Y., Xu, H., Wang, W., Zhou, J., Zhu, L., 2024. Rainmamba: enhanced locality learning with state space models for video deraining. In: Proceedings of the 32nd ACM International Conference on Multimedia, pp. 7881–7890.
Xiong, C., Shi, H., Li, J., Wu, X., Gao, R., 2024. Informer-based model for long-term ship trajectory prediction. J. Mar. Sci. Eng. 12 (8), Art. no. 1269.
Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., Liu, Z., 2020. Ais data driven general vessel destination prediction: a random forest based approach. Transp. Res. Part C 118, 102729.
Zhang, J., Wang, H., Cui, F., Liu, Y., Liu, Z., Dong, J., 2023. Research into ship trajectory prediction based on an improved LSTM network. J. Mar. Sci. Eng. 11 (7), 1268.
Zhang, S., Zhang, R., Yang, Z., 2024. Matrrec: uniting mamba and transformer for sequential recommendation. arXiv:2407.19239
Zhang, X., Liu, G., Hu, C., Ma, X., 2019. Wavelet analysis based hidden Markov model for large ship trajectory prediction. In: 2019 Chinese Control Conference (CCC). IEEE, pp. 2913–2918.
Zhu, F., 2021. Ship short-term trajectory prediction based on RNN. In: Journal of Physics: Conference Series. Vol. 2025. IOP Publishing, p. 012023.
Zou, B., Li, W., Hou, X., Tang, L., Yuan, Q., 2022. A framework for trajectory prediction of preceding target vehicles in urban scenario using multi-sensor fusion. Sensors 22 (13), 4808.