Commun Transp Res

Research Article

# STPredictor: Ship Trajectory prediction with instruction-aligned large language models

Siyu Teng[1,2], Ying Yang[1,2], Liang Zhao[3], Baoding Zhou[1,2], Long Chen[4], Ran Yan[5,✉], Jiasong Zhu[1,2,✉]

Cite this article: https://doi.org/10.26599/COMMTR.2026.9640017

**ABSTRACT:** Accurate ship trajectory prediction is crucial to ensure maritime safety. Most existing ship trajectory predictors face two issues: reliance on post-clustered trajectories and limited interpretability in decision processes. In this research, we address these challenges by proposing an explainable Ship Trajectory Predictor (STPredictor), which is facilitated by strong reasoning capabilities of large language models (LLMs). We reformulate the ship trajectory prediction as a language modeling problem, encoding heterogeneous maritime scenarios as natural-language prompts, and employing supervised fine-tuning to design LLMs specifically for the prediction task. Furthermore, we integrate the Chain-of-Thought (CoT) process into the inference pipeline to enhance the transparency and reliability of predictions, and include explanatory requirements in the inference stage to make the decision process align with human instructions. To comprehensively benchmark STPredictor against strong baselines, we construct two large-scale datasets from global Automatic Identification System (AIS) records, including a geospatial-domain dataset and a draught-domain dataset. Extensive experiments based on these datasets demonstrate the superior performance and interpretability of STPredictor in the trajectory prediction task. These findings indicate that LLMs can effectively encode rich interaction information for understanding complex maritime scenarios, thereby laying a solid foundation for reliable and interpretable decision-making in maritime safety.

**KEYWORDS:** large language models (LLMs); Ship trajectory prediction; AIS data; Maritime safety; Fine-tuning

## 1 Introduction

The vast maritime network not only serves as the solid backbone of international supply chains but also supports national energy security and industrial productivity, making it a critical infrastructure for global commerce (Liu et al., 2025a; Yang et al., 2026b). However, as the global fleet expands and routes become denser, maritime traffic complexity rises markedly. Combined with dynamic ocean conditions and rising operational demands, this trend elevates the risk and difficulty of maritime safety management (Yang et al., 2026a).

Against this background, Intelligent Maritime Systems (IMS) are increasingly recognized as a key pathway toward safer and more efficient maritime operations, with ship trajectory prediction serving as the core perceptual and cognitive foundation for decision-making (Chen et al., 2021a; Chen et al., 2021b; Guo et al., 2024). However, achieving such advanced prediction requires overcoming the limitations of current data-driven "black-box" methods, which struggle with interpretability and robustness in unconventional scenarios. The core motivation of STPredictor is to address these challenges by incorporating explicit guidance that promotes human-like behavior.

Research on ship trajectory prediction has evolved through several paradigms, each addressing certain aspects of the problem while exhibiting distinct limitations (Fan et al., 2025). First, kinematic model-based methods rely on physical ship motion equations but struggle to model multi-ship interactions and nonlinear behaviors (Liu et al., 2019; Li et al., 2024b). Second, classical machine learning approaches leverage historical AIS data to identify motion patterns, yet struggle to capture long-range spatiotemporal dependencies (Gao et al., 2023a; Jordan and Mitchell, 2015; Zhang et al., 2020). Third, deep learning methods have achieved strong performance by learning temporal dynamics directly from data; however, they often suffer from error accumulation in multi-step trajectory prediction (Sekhon and Fleming, 2020). Recent studies further enhance interaction modeling through spatiotemporal attention mechanisms and Graph Neural Networks (GNNs), improving the representation of ship group behaviors (Gao et al., 2023b; Zhang et al., 2025a). Despite these advances, existing methods still cannot systematically integrate explicit maritime knowledge, such as the International Regulations for Preventing Collisions at Sea (COLREGs), into the data-driven pipeline.

To address these challenges, we propose Ship Trajectory Predictor (STPredictor), an instruction-aligned method that integrates maritime constraints and observed scenarios to enable interpretable and human-guided trajectory prediction. To validate the robustness of the proposed method and other selected baselines, we construct two AIS datasets for comprehensive

[1] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China. [2] State Key Laboratory of Green and Long-Life Road Engineering in Extreme Environment (Shenzhen), Shenzhen 518060, China. [3] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China. [4] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. [5] School of Civil and Environmental Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore.

✉ Corresponding author. E-mail: R. Yan, ran.yan@ntu.edu.sg; J. Zhu, zjsong@szu.edu.cn

Received: December 2, 2025; Revised: January 20, 2026; Accepted: March 6, 2026

evaluation, the detailed process is listed in *Appendix A*. The first is a geospatial-based dataset designed to evaluate performance across diverse traffic conditions, encompassing three typical complex waterways: the Singapore Strait (SG) as a representative of congested narrow channels, the Persian Gulf (PG) water area in the Indian Ocean with its complex open-water scenarios, and the Zhoushan Archipelago (Zhoushan) featuring complex inshore waterways (Zhao et al., 2025; Song et al., 2023). The second is a draught-based dataset that characterizes the heterogeneity of ship dynamic properties through measured draught. The filtered frames are stratified into three draught intervals (0-5 m, 5-10 m, and 10-25 m), enabling systematic evaluation across heterogeneous ship dynamics and maneuverability characteristics. The contributions of this research are summarized as follows:

1) We propose STPredictor, an LLM-based method for ship trajectory prediction. By reformulating trajectory sequences into structured natural language, STPredictor facilitates holistic interpretation of complex interaction contexts and offers a new paradigm for high-precision trajectory prediction.

2) We design an interpretable prediction framework that jointly outputs trajectories and textual explanations. By integrating the CoT process, the framework mitigates the "black-box" limitation of conventional learning-based methods and enables STPredictor to explicitly follow the execution of instructions.

3) We construct two AIS datasets augmented with geospatial and draught information. These datasets address the absence of critical scenario attributes in existing resources and provide a benchmark foundation for future research on IMS.

The remainder of this research is organized as follows. Section 2 provides a systematic review of recent advances in ship trajectory prediction, covering traditional prediction methods, learning-based methods, and emerging maritime applications of LLMs. Section 3 formulates the task definition of the trajectory prediction problem and presents the proposed instruction-aligned behavioral prediction framework. This section also explains how the Chain-of-Thought (CoT) process is utilized to align high-level human instructions, thus improving the reasoning and interpretive capabilities of the method. Section 4 provides an in-depth analysis and discussion of the experimental results, exploring their implications and potential impact. Section 5 concludes the research by summarizing the main findings and outlining future research directions. Finally, the *Appendix A* presents the construction and utilization details of the datasets used for validation and the implementation details.

## 2 Related Work

### *2.1. Traditional Prediction Method*

Traditional ship trajectory prediction methods can be broadly grouped into physics-based dynamic models, probabilistic filters, and geometric search approaches (Das et al., 2024; Teng et al., 2023; Wu et al., 2022). While these methods lay the foundation for early research in trajectory prediction, they face significant limitations when addressing the complexity and uncertainty of maritime scenarios (Wang et al., 2025b). Dynamic models, based on physical principles, simulate ship dynamic parameters, such as speed over ground (SoG), course over ground (CoG), heading, and rudder angle, using differential equations. For example, Chen et al. (2023a) apply a reduced-order dynamic mode decomposition algorithm to model ship maneuvering dynamics and enable short-term trajectory prediction. Representative studies integrate error-compensation mechanisms or simplified maneuvering models into physics-based predictors to improve short-term trajectory accuracy (Skulstad et al., 2021; Sutulo and Soares, 2024). Such dynamic methods excel in short-term predictions for a single ship in limited environments, but they require precise hydrodynamic parameters and struggle to scale to multi-ship scenarios or integrate environmental factors, resulting in significant long-term trajectory prediction errors in IMS (Li et al., 2025a).

To handle uncertainty, other researchers adopt probabilistic models, including the Kalman Filtering (KF) and the Particle Filtering (PF). Davari et al. (2017) apply KF to ship motion state estimation and demonstrate improved robustness under noisy AIS measurements. Zhang et al. (2021) address instability in the error of the Extended Kalman filter (EKF) for complex scenarios with rapidly changing ship motion states, proposing a ship trajectory tracking algorithm based on Square-Root Covariance Kalman Filter to improve the accuracy and stability of AIS data. The PF method, which uses Monte Carlo sampling to model multimodal distributions, can capture the maneuvering intentions of ships. Particle filtering-based methods further capture maneuvering uncertainty and multimodal motion intentions in complex maritime environments (Jia et al., 2025). Although these methods are effective at filtering noise, they still struggle to model complex interactions between ships and lack robustness in congested waters.

Additionally, geometric search methods are also a mainstream area of research in trajectory prediction. Geometric search methods combine rule-based collision avoidance with global path planning to ensure feasible and interpretable trajectories (Yu and Roh, 2024; Tang et al., 2024). This framework compensates for the lack of accurate lateral motion control in the underactuated system through model prediction, aligning the actual trajectory of the underactuated planning craft with the predicted trajectory.

Although these methods are interpretable, they suffer from high computational costs in large-scale scenarios and poor adaptability to dynamic environments; for instance, dynamic methods (Skulstad et al., 2021; Sutulo and Soares, 2024) and probabilistic models (Davari et al., 2017; Zhang et al., 2021; Jia et al., 2025) scale poorly in multi-ship scenarios, while geometric search methods (Yu and Roh, 2024; Tang et al., 2024) incur prohibitive costs when prediction in expansive or changing waters. Their design focuses more on trajectory feasibility than on probabilistic accuracy, which may lead to overly conservative or unrealistic predictions. These limitations highlight the challenge that traditional maritime trajectory prediction methods face in achieving both joint prediction and instruction alignment to meet collision avoidance requirements in complex real-world maritime environments.

### *2.2. Learning Based Methods*

Ship trajectory prediction, as a core technology of IMS, has evolved significantly from traditional machine learning methods

to deep learning approaches. Traditional learning methods (Farahnakian et al., 2025), such as LSTM (Xin et al., 2018), GRU (Dey and Salem, 2017), Transformer (Gao et al., 2023b), and RNN (Schuster and Paliwal, 1997), rely on manual feature engineering to extract key indicators such as SoG, CoG, heading, and spatial grid states, driving linear regression or clustering models for short-term predictions. While these methods offer advantages in computational efficiency and interpretability, they are highly sensitive to environmental noise, such as ocean currents and wind/wave disturbances, and suffer from limited generalization ability. Liu et al. (2022a) propose an LSTM-based trajectory prediction framework that embeds ship conflict modeling and a mixed loss function to achieve accurate and robust spatiotemporal ship trajectory prediction for maritime Internet of Things, alleviating the challenges posed by explosive AIS traffic and stringent maritime safety requirements. C-LSTM (Li et al., 2019) integrates spatial feature extraction with bidirectional LSTM (BiLSTM) temporal modeling, combined with the adaptive momentum search optimization algorithm. This method reduces the prediction error by 18.7% in multi-ship encounter scenarios compared to LSTM (Xin et al., 2018). LR-LSTM (Kim et al., 2020) combines linear regression coefficients with LSTM hidden states, reducing endpoint errors by 22.3% in long-range trajectory data.

The wave of deep learning has driven a transport revolution in the end-to-end learning paradigm (Ai et al., 2024; Tang et al., 2024). RNN (Schuster and Paliwal, 1997), GRU (Dey and Salem, 2017), and their gated variants mitigate long-term dependency degradation through neural-state propagation. Building upon these foundations, improved LSTM encodes clustered historical AIS sequences and reduces autocorrelation error by 75% in ship trajectory prediction compared to EKF (Tang et al., 2022). However, such methods remain highly sensitive to sensor noise, exhibit limited spatial reasoning ability, and may generate trajectories that violate physical or navigational constraints.

These limitations motivate the introduction of attention mechanisms, which dynamically allocate computational focus to behaviorally critical trajectory segments. Learning-based methods introduce attention mechanisms and recurrent architectures to model long-term temporal dependencies and multi-ship interactions, achieving improved prediction accuracy on AIS data (Wang et al., 2024; Yin et al., 2025; Chen et al., 2023b; Li et al., 2024a; Liu et al., 2025b). However, such methods still struggle to model heterogeneous ship behaviors and face significant challenges in explaining complex maneuvering intentions.

Transformer (Gao et al., 2023b) excels at long-range modeling by self-attention layers. Luo et al. (2025) introduce a multimodal trajectory prediction method, in which electronic navigational charts are segmented into water/land areas and integrated with real AIS data to train the method. The introduction of segment recurrence captures long-term dependencies in AIS data, facilitating the prediction of future ship trajectories. Results show that the proposed method achieves excellent prediction performance in various operating conditions of the ship and outperforms the relevant unimodal methods in prediction tasks.

Recent studies increasingly emphasize explicit spatiotemporal interaction modeling and hybrid architectures that incorporate physical priors. GNNs emerge as a core model (Zhang et al., 2025a), viewing ships as interactive agents embedded in a dynamic maritime environment. By encoding physical constraints, such methods help bridge the gap between purely data-driven learning and mechanism-informed modeling. Recent GNN-based approaches incorporate physical priors and spatiotemporal interaction modeling to bridge data-driven learning and mechanism-informed prediction (Zhang et al., 2025a, 2025b). These methods collectively integrate data flexibility with physical interpretability, forming an emerging foundation for autonomous navigation in IMS.

Despite the progress of learning-based trajectory prediction, a substantial gap persists in instruction-aligned prediction and cross-modal interpretability. While recurrent variants such as C-LSTM and LR-LSTM (Xin et al., 2018; Kim et al., 2020) enhance prediction accuracy, they remain limited in explaining complex maneuvering intentions. Likewise, attention-based (Liu et al., 2022b; Yin et al., 2025) and graph-based methods (Zhang et al., 2025a) offer improved spatiotemporal reasoning but lack grounding in natural language, restricting their use in interactive decision-making. This disconnect between trajectory generation and decision-level contextualization motivates our exploration of instruction-aligned trajectory generation using large language models (LLMs). Our goal is to enable context-aware, natural-language-driven navigation, providing a new pathway toward interpretable and interactive maritime decision-support systems.

### *2.3. LLM for Maritime Research*

Recent studies have demonstrated that LLMs integrated with domain knowledge graphs enable structured reasoning and decision support in complex engineering systems (Liu et al., 2025c). Researchers begin to position fine-tuned LLMs as the cognitive brain for robotics, building Vision-Language-Action (VLA) models and Vision-Language-Navigation (VLN) models to achieve the seamless integration of high-level semantic reasoning and low-level control (Li et al., 2025b; Nie et al., 2025). Representative embodied LLM frameworks decouple high-level semantic reasoning from low-level control, enabling interpretable decision-making and robust action generation in complex environments (Hong et al., 2024; Huang et al., 2025; Zhou et al., 2025). These studies demonstrate how LLMs can embed semantic reasoning into the control loop and couple symbolic inference with physical decision-making, thereby strengthening the integration between high-level cognition and low-level action. Improvements in prompt design, model architecture, and training methodology further advance the generalization capability, scalability, and interpretability of embodied cognitiv – physical systems.

With the exceptional potential of LLMs in spatiotemporal reasoning and causal modeling, researchers begin to explore the redefinition of motion prediction problems through language modeling approaches, enabling the fusion of semantic understanding and decision-making (Wang et al., 2025a). Language-based trajectory prediction methods reformulate motion sequences as discrete tokens, enabling autoregressive generation with implicit modeling of interactions and causal constraints (Seff et al., 2023; Lan et al., 2024). These advances

highlight the universal potential of LLMs in spatiotemporal reasoning. By leveraging their pre-trained knowledge, LLMs can implicitly encode complex traffic semantics, thereby offering a novel solution for generalizable prediction in open-world scenarios. Overall, these studies mark a paradigm shift in trajectory prediction: models are no longer confined to geometric fitting and local feature learning, but instead leverage the spatiotemporal semantic transfer ability of language models to achieve holistic modeling of traffic context, interaction intentions, and causal structures. This language-based approach to motion prediction provides a more general and interpretable cognitive foundation for autonomous decision-making in open environments.

Despite significant progress made by LLMs in robotics and unmanned systems, their application in maritime scenarios remains in its infancy and faces notable limitations. MarineGPT (Zheng et al., 2023) represents an early attempt to develop vertical LLMs capable of marine scenario description and scientific inquiry. However, it does not incorporate systematic modeling of ship motion characteristics or domain-specific navigation rules. BM-RAGAM (Chen et al., 2024) combines BM25 with semantic retrieval to enhance accuracy on marine question-answering tasks, but its framework remains primarily text-centric and lacks integration with embodied functions such as ship control or trajectory prediction. Although the system emphasizes hallucination reduction, it remains confined to cognitive-level reasoning rather than operational, instruction-driven maritime actions. As a result, current LLM-based approaches have yet to address the core challenges of trajectory prediction in IMS. Motivated by the modeling principles outlined in (Liu et al., 2023), we aim to bridge this gap by integrating language comprehension, navigation semantics, and motion modeling into a unified framework, further developing an interpretable, generalizable, and rule-compliant trajectory predictor to advance LLM-based IMS capabilities in complex ocean environments.

## 3 Task Definition

This section reformulates the prediction problem to support experimental design and consistent evaluation. The results are presented to provide a benchmark for predicting ship trajectories in the geospatial and draught datasets.

### *3.1. Definitions and Problem Statements*

From a classical perspective, a ship trajectory can be denoted as an ordered sequence of waypoints:

$$traj = \langle p_1, p_2, \ldots, p_i, \ldots, p_N \rangle \tag{1}$$

where each waypoint is defined as $p_i = [x_i, y_i]^{\mathrm{T}}$, where $(x_i, y_i)$ are expressed in the local coordinate frame. Here, $i$ denotes the index of the waypoint, $N$ is the total number of waypoints for the trajectory, and $(x_i, y_i)$ specifies the spatial position of the ego-ship.

Figure 1 illustrates the trajectory prediction process in maritime environments. Based on this representation, the trajectory prediction task over a horizon of future steps K can be formulated as follows:

$$\hat{p}_{t+1:t+K} = \mathcal{F}(S_y; O; S_t) \tag{2}$$

where $\hat{p}_{t+1:t+K}$ denotes the predicted K-step waypoint sequence. $\mathcal{F}(\cdot)$ denotes a parametric decision function that maps recent information to a distribution over future ship states. $S_y$ denotes the body-fixed coordinate frame used to express relative motion, $O$ summarizes the surrounding scenario context, including port-related information, speed limits, and the historical trajectory $\{p_{t-H+1:t}\}$, $S_t$ denotes the ego-ship state at time $t$, including Latitude, Longitude, SoG, CoG, and other kinematic attributes.

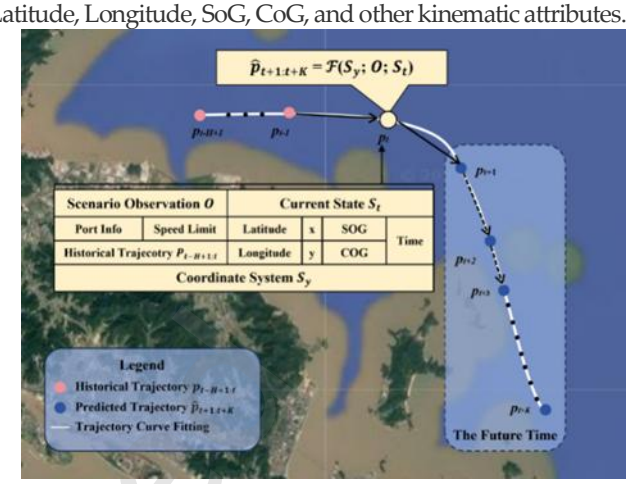



**Figure 1.** The introduction of ship trajectory prediction. Pink and blue points show the waypoints in historical and future trajectories. The data in the yellow table constitutes the input required by STPredictor, $\mathcal{F}(\cdot)$ stands for the predicted policy, predicting the future trajectory from the historical trajectory.

LLMs excel at integrating heterogeneous information, while explicit instructions are critical for strategy selection in maritime navigation. Accordingly, we incorporate a high-level human instruction $I$ into STPredictor to encode task-level constraints and operator-specified operational preferences. Overall, the input information $X_t$ at time $t$ can be formally expressed as follows:

$$X_t = \{S_y; I; O; S_t\} \tag{3}$$

In our prediction framework, the output $Y_t$ should capture not only the predicted trajectory $\Gamma$ but also the intrinsic reasoning process $I_c$ that informs and justifies each navigation decision. To enhance the transparency and cognitive grounding of such decisions, we incorporate the CoT process (Wei et al., 2022), a paradigm that elicits structured intermediate reasoning steps from LLMs and has demonstrated effectiveness in supporting compositional inference and multi-stage decision-making. In STPredictor, $I_c$ is exhibited in textual form, aiming to record the logical inferences and decision rationale formulated during trajectory prediction. Consequently, it provides a transparent and traceable cognitive explanation alongside the strategic output. The overall decision output can be represented as:

$$Y_t = \{\Gamma, I_c\} \tag{4}$$

where $\Gamma$ denotes the predicted navigation trajectory, and $I_c$ represents the CoT process formed during trajectory reasoning. These two components constitute an interpretable decision framework built upon LLMs.

Given the integrated input representation $X_t$,which contains the recent observations together with CoT sequence, the decision function $\mathcal{F}_{LLM}(\cdot)$ outputs a logarithmic probability distribution over all tokens in the unified latent space. This distribution reflects the model's confidence in each element of the generated content, thereby offering a quantitative assessment of the cognitive reliability associated with the predicted results:

$$Y_{log} = \mathcal{F}_{LLM}(X_t) \tag{5}$$

To obtain the final output $Y_t$ from the logarithmic probability distribution $Y_{log}$, the method undergoes a series of probabilistic decoding pipeline. First, a temperature scaling operator $\Phi(\cdot,\gamma)$ is applied directly to the log-probabilities to control the sharpness of the token distribution, thereby controlling the reliability and diversity of the output sequence. Next, the Softmax function $S(\cdot)$ normalizes the output sequence into a valid probability distribution. Finally, a nucleus sampling strategy $P\{\cdot,p_{nuc}\}$ samples tokens from the smallest subset whose cumulative probability mass reaches $p_{nuc}$, achieving a balance between precision and diversity. This process can be formally expressed as below:

$$Y_t = P\left\{S\left(\Phi\left(Y_{log},\gamma\right)\right),p_{nuc}\right\} = P\left\{S\left(\Phi(\mathcal{F}_{LLM}(X_t),\gamma)\right),p_{nuc}\right\} \quad (6)$$

where parameter $\gamma$ controls the diversity of the decoding process by reshaping the token distribution, thereby controlling the level of exploratory behavior exhibited by the method. Smaller $\gamma$ suppresses randomness and biases generation toward the most confident predictions, yielding outputs that are stable, while larger $\gamma$ encourages more exploratory and adventurous content in generation, allowing the method to consider a broader set of candidate tokens and produce more diverse, imaginative continuations. In parallel, the parameter $p_{nuc}$ denotes the sampling probability threshold, which defines the nucleus of the sampling space by retaining only the smallest subset of tokens. This constraint prunes the long tail of unlikely candidates, preventing noisy outputs while maintaining sufficient diversity for naturalistic generation.

Inspired by the InstructChain (Zhang et al., 2024a), this research proposes its extended inference process in the maritime domain. STPredictor optimizes the log-probabilities $Y_{log}$ associated with both human instruction chains $I_c$ and expert trajectories $\Gamma$, thereby establishing a navigation decision mechanism that unites human-like reasoning.

The advantage of STPredictor lies in two capabilities. First, it can generate navigation trajectories that comply with maritime safety standards and regulations while simultaneously producing reference processes that align with the cognitive logic of human captains. Second, our method represents an evolution from mere task execution to the autonomous thinking endowed with genuine understanding and explanatory competence, achieved through the interpretable integration of cognitive reasoning and probabilistic generation.

### *3.2. Instruction Alignment*

To enable imitation learning of human operators, it is necessary to construct the input-output mapping pairs $(X_t,Y_t)$ for fine-tuning LLMs within the modeling process. As formulated in Equation 3, the input data $X_t$ are jointly represented through natural language descriptions and scenario elements, which consist of an instruction part $I_t$ and an observation part $I_o$.

$$X_t = \left\{\underbrace{S_y,I}_{I_t},\underbrace{O,S_t}_{I_o}\right\} \quad (7)$$

where the input can be decomposed by the instruction $I_t$ and observation $I_o$. $I_t$ is designed to constrain and guide the decision-making process and overall maneuvering behavior, ensuring that the generated navigation prediction aligns with human intent while adhering to maritime regulations and safety requirements. $I_t$ is further decomposed into two categories of elements: the system description $S_y$ and the prediction instruction $I_p$. The detailed formulation is shown below:

$$I_t = \{S_y,I_p\} \quad (8)$$

where $S_y$ plays a pivotal role in establishing the mapping between physical feasibility and trajectory executability, which is designed to accurately provide a unified coordinate and constraint system to support the physical realization of high-level prediction commands.

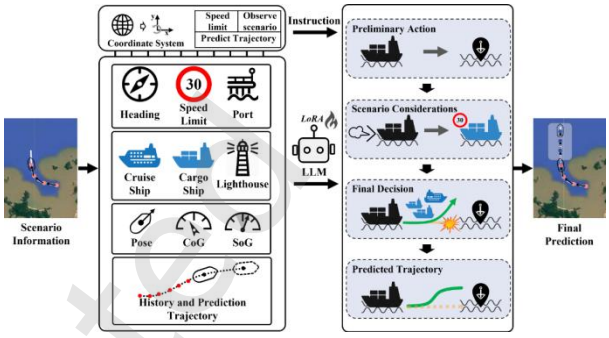



**Figure 2.** Framework of STPredictor for trajectory prediction in maritime scenarios.

To further enhance the STPredictor's predictive accuracy and environmental adaptability, all earth-fixed coordinate information is transformed into the body-fixed coordinate, which serves as the primary reference frame for representing the ship's relative motion variables. In the body-fixed local coordinate frame, the ship's forward direction aligns with the positive X-axis, and the positive Y-axis is perpendicular to it, pointing to the left side. The CoG is defined as the counterclockwise angle measured from the positive X-axis. This definition ensures the consistency and interpretability of trajectory prediction and control methods across different reference frames, enabling the system to maintain a coherent attitude and traceable physical parameters under dynamic maritime transport.

The prediction instruction $I_p$ describes executable suggestions, which range from coarse-grained decisions to fine-grained action controls. By introducing a human-like behavioral paradigm, $I_p$ incorporates safety and compliance considerations, such as rule adherence and speed regulation, into the learning process. Consequently, STPredictor is compelled to proactively avoid potential risks during policy generation, ensuring both the feasibility and safety of the resulting trajectories.

In addition, the observation $I_o$ is formulated to review all essential information for ship trajectory prediction. It serves as a critical component supporting decision-making and interpretable trajectory generation. $I_o$ comprises two major elements: the environmental observation $O$ and the current ship state $S_t$. These two components complement each other by representing the maritime scenario from both the external dynamic environment and the internal kinematic conditions. The detail is expressed in the following formulation:

$$I_o = \{O,S_t\} \quad (9)$$

where $O$ shows the observation data, which primarily includes the ship's current scenario, historical trajectory and final port locations, all of which are critical to accurate trajectory prediction.

The ego-state $S_t$ captures the ship's intrinsic motion state, including position, CoG, heading, SoG, estimated time of arrival, and the recent motion history of the ego-ship.

The above elements describe both the instantaneous state and motion history of the ship, supplying the method with continuous temporal constraints and dynamic boundary conditions. By integrating the above information, the method achieves a coherent balance between decision consistency and control constraints, which enables the predicted trajectory to exhibit temporal continuity and kinematic smoothness, thereby ensuring its physical feasibility and navigational stability in real-world maritime environments.

Through the above process, the input $X_t$ is constructed. Subsequently, a standardized design is applied to STPredictor's output to generate $Y_t$, thereby enhancing the interpretability and robustness of the trajectory prediction method. As shown in Equation 4, the output $Y_t$ is coupled with the predicted chain $I_c$, which is further utilized to compute the final predicted trajectory $\Gamma$. $I_c$ functions as a semantic bridge linking the method's internal cognitive reasoning with the external trajectory generation process. It records the logical inferences and decision-making rationale formed during trajectory synthesis. This mechanism not only provides an interpretable representation of the predicted results but also substantially improves the transparency and comprehensibility of the overall decision-making process.

To enhance the interpretability and robustness of the predicted chain $I_c$ in STPredictor, a four-stage reasoning and constraint mechanism is designed as follows:

1) Preliminary action: Establishes a global maneuver scheme by determining the next waypoint, predicted heading, and target speed, while accounting for

hydrodynamic effects and maneuvering constraints.

2) Scenario consideration: Constructs an interpretable risk map centered on the ego-ship, predicts short-term relative motions, and formulates COLREGs-consistent navigational instructions.

3) Final decision: Synthesizes multi-layer environmental and regulatory constraints to ensure the resulting strategy remains lawful, dynamically consistent, and operationally secure.

4) Predicted trajectory: Embeds the derived intentions and constraints into a multi-objective optimization process to produce smooth, efficient, and physically feasible trajectories.

Based on this multi-stage reasoning pipeline, the predicted trajectory can be formally expressed as a discrete trajectory sequence, as shown below:

$$T = \langle p_1, p_2, \cdots, p_t, \cdots, p_n \rangle, p_t = [x_t, y_t]^\top \ (t \in [0, n]) \quad (10)$$

The whole pipeline of STPredictor, as shown in Figure 2, aligns the learned strategies more closely with human navigation habits while maintaining stable performance in previously unseen scenarios.

For model implementation of STPredictor, LLaMA (Touvron et al., 2023) is adopted as the open-sourced foundation model and fine-tuned following the optimization pipeline proposed in LLaMA2-Accessory (Zhang et al., 2024b) to meet the semantic representation requirements of maritime trajectory prediction. The standardized input $X_t$ undergoes preprocessing by the function $P(\cdot)$, which performs spatiotemporal alignment, noise filtering, and numerical normalization, yielding the latent processed parameters. These are subsequently transformed through the embedding function $E(\cdot)$ (Vaswani et al., 2017), which denotes the token embedding module of the LLaMA that maps the tokenized textual representation of $P(X_t)$ input into continuous encoded vectors $X$. The overall process can be described as follows:

$$X = E\big(P(X_t)\big) \quad (11)$$

After the input text $X_t$ is formatted as encoded vectors $X$, it is subsequently processed through the mapping policy $\mathcal{F}_m(\cdot)$, which leverages the Low-Rank Adaptation (LoRA) technology to enhance the efficiency of model fine-tuning by leveraging low-rank matrix approximations. The output of this process is the logarithmic probability distribution $Y_{log}$ for the vocabulary. The report of this formulated process is as follows:

$$Y_{log} = \mathcal{F}_m\big(X, Att(X, A, B, b), wR(X)\big) \quad (12)$$

where $Att(\cdot)$ denotes the attention module, responsible for adaptively capturing interdependencies among key trajectory points in spatiotemporal sequences. The matrices $A$ and $B$ represent the low-rank approximation components introduced by the LoRA method. $b$ denotes a bias vector that mitigates extreme attention distributions, while $R(\cdot)$ refers to RMSNorm (Zhang and Sennrich, 2019), which stabilizes training dynamics and accelerates convergence. The weight parameter $w$ is the weight coefficient associated with the RMSNorm. Through fine-tuning of the parameters $A$, $B$, $b$, and $w$, the method progressively evolves into a trajectory prediction powered with both semantic understanding and physical consistency, aligning with human-like navigational behavior.

## 4 Experiment

Maritime transport is inherently non-stationary due to tidal constraints, port-operational cycles, and episodic traffic controls, which impairs the transferability of learned interaction priors at the system level. Accordingly, we establish an evaluation benchmark comprising two constructed datasets and a diverse of metrics designed to dense-traffic scenarios and environmental disturbances. Detailed information on our datasets construction and usage is provided in the *Appendix A*. The benchmark quantifies the method's performance under strong disturbances and high-density traffic conditions, with emphasis on trajectory consistency, attitude and maneuverability robustness, and environmental adaptability. This standardized framework provides a reproducible, comparable, and transferable academic and engineering benchmark for assessing IMS under complex maritime conditions.

### *4.1. Definitions and Problem Statements*

We conduct a quantitative experiment of nine methods using five evaluation metrics, considering both global and local perspectives. Previous research (Liu et al., 2025b; Wang et al., 2024) has highlighted that relying on just one or two metrics to assess prediction performance may lead to biases and distorted results. Therefore, to achieve a comprehensive evaluation, this study incorporates a multi-dimensional metric system to quantify method performance.

4.1.1. Root mean squared error (RMSE)

Due to significant dynamic variations in the maritime environment, ships experience continuous attitude fluctuations and speed disturbances under different sea conditions. RMSE offers a distinct advantage in assessing the performance of trajectory prediction methods. By calculating the per-dimension distance between the predicted and label trajectories, RMSE not only maintains overall prediction accuracy in each axis but also reveals the error diffusion characteristics of the method. The expression for RMSE is given below:

$$RMSE = \sqrt{\frac{1}{2K}\sum_{d\in\{x,y\}}\sum_{k=1}^{K}\left(p_k^d - \hat{p}_k^d\right)^2} \tag{13}$$

where $p_k^d$ and $\hat{p}_k^d$ are the $k$th points of the label and predicted trajectory points in dimension $d \in \{x, y\}$ at step $k$.

4.1.2. Average distance error (ADE)

In trajectory prediction validation, ADE is used to assess the overall spatial accuracy of predicted trajectories. This metric calculates the average spatial distance between the predicted and label trajectories at each time step, providing a comprehensive view of the error distribution throughout the entire trajectory. The formulation for ADE is given below:

$$ADE = \frac{1}{K}\sum_{k=1}^{K} \| p_k - \hat{p}_k \|_2 \tag{14}$$

where $p_k$ is the $k$th points of labeled trajectory point, and $\hat{p}_k$ is the $k$th points of the predicted trajectory point.

4.1.3. Average heading error (AHE)

The CoG is a key parameter influencing ship maneuvering stability and trajectory consistency. It directly reflects the deviation between the current navigation direction and the reference trajectory. For commercial ships navigating in complex maritime conditions, maintaining precise heading control is crucial not only for energy efficiency and trajectory smoothness but also for preventing yaw, drift, and collision risks. To quantify the method's performance in heading prediction, we introduce the AHE as an evaluation metric. The formulation for AHE is given below:

$$AHE = \frac{1}{K}\sum_{k=1}^{K} | \theta_k - \hat{\theta}_k | \tag{15}$$

where $\theta_k$ and $\hat{\theta}_k$ are the labeled and predicted CoG of the trajectory point, k is the number of instances.

4.1.4. Final distance error (FDE)

FDE focuses on the accuracy at the endpoint of the predicted trajectory, serving as an important indicator of whether the ship can accurately reach the intended destination. FDE calculates the spatial distance difference between the predicted trajectory's endpoint and the labeled target point, reflecting the method's final positioning ability throughout the entire planning and control process. The formulation for FDE is given below:

$$FDE = \| p_K - \hat{p}_K \|_2 \tag{16}$$

where $K$ represents the index of the final point in the trajectory. In our experiments, the prediction horizon is fixed to $K = 6$. Accordingly, $p_K$ is the final point of the labeled trajectory, $\hat{p}_K$ is the final point of the predicted trajectory.

4.1.5. Final heading error (FHE)

Final Heading is a crucial parameter influencing multi-ship coordination. For autonomous ships that need to collaborate with tugboats, pilot boats, or port handling systems, the accuracy of the final heading angle directly determines their ability to safely and efficiently complete berthing, docking, or heading adjustment tasks. To assess this parameter, FHE is introduced as a key evaluation metric to measure the heading deviation at the endpoint of the predicted trajectory. The formulation for FHE is given below:

$$FHE = | \theta_K - \partial_K | \tag{17}$$

where $\theta_K$ and $\partial_K$ are the labeled and predicted CoG of the trajectory point.

### *4.2. Implementation Details*

We systematically trained and validated STPredictor, alongside 8 existing trajectory prediction methods in the geospatial and the draught-based dataset. The construction and implementation details of the two datasets are provided in the *Appendix A*. We leverage LLaMA (Touvron et al., 2023) as the cornerstone of our proposed method. Four of these are traditional learning-based methods, including LSTM (Xin et al., 2018), GRU (Dey and Salem, 2017), RNN (Schuster and Paliwal, 1997), and Transformer (Gao et al., 2023b), which serve as end-to-end prediction baselines for conventional deep learning methods. The other four methods (Wang et al., 2024; Liu et al., 2025b; Chen et al., 2023b; Li et al., 2024a) are SOTA trajectory prediction methods from the past three years, designed to evaluate the performance under certain conditions and compared with our proposed method.

### *4.3. Main Results of Two Datasets*

This section presents a comparative analysis of the predictive performance of our method and eight representative methods on the geospatial and draught-based datasets, to examine the practical effectiveness and generalization capability of trajectory prediction under heterogeneous maritime scenarios.

To explicitly evaluate the generalization capability of trajectory predictors across heterogeneous water areas and diverse ship draughts, we intentionally forbade any AIS-based trajectory clustering in our datasets, thereby avoiding performance inflation induced by route-specific priors. To ensure the integrity of the experiment, we randomly selected 5,000 frames from the constructed dataset to form the test dataset. To further mitigate random variation, each predictor is employed to predict the test dataset five times, with the mean of these predictions taken as the final result.

**Table 1.** Comparison of the predicted points of the proposed and baseline methods on the different water areas.

| Method | Region | RMSE | ADE | FDE | AHE (°) | FHE (°) |
|---|---|---|---|---|---|---|
| GRU (Dey and Salem, 2017) | PG | 576.2442 | 302.4635 | 605.6841 | 89.22 | 96.61 |
| | SG | 580.4352 | 331.4464 | 590.1041 | 118.38 | 128.02 |
| | Zhoushan | 654.7592 | 329.3382 | 583.8387 | 82.48 | 67.62 |
| LSTM (Xin et al., 2018) | PG | 1673.546 | 995.3737 | 1680.04 | 136.99 | 130.75 |
| | SG | 1229.715 | 590.2351 | 968.0493 | 106.33 | 109.84 |
| | Zhoushan | 2339.512 | 1601.943 | 2700.832 | 98.81 | 96.46 |
| RNN (Schuster and Paliwal, 1997) | PG | 2808.556 | 2298.598 | 3895.944 | 132.09 | 132.05 |
| | SG | 1213.415 | 572.9135 | 956.5177 | 107.77 | 108.03 |
| | Zhoushan | 3342.504 | 3164.28 | 5408.023 | 78.57 | 76.73 |
| Transformer (Gao et al., 2023b) | PG | 659.6749 | 371.1243 | 622.9227 | 93.53 | 94.55 |
| | SG | 618.1335 | 249.6278 | 465.6737 | 124.18 | 138.55 |
| | Zhoushan | 892.2796 | 866.0907 | 1565.837 | 99.62 | 92.97 |
| DAA-SGCN (Wang et al., 2024) | PG | 585.7157 | 315.1178 | 543.9891 | 83.28 | 89.91 |
| | SG | 578.9946 | 324.3664 | 569.4081 | 122.83 | 88.20 |
| | Zhoushan | 720.4289 | 470.8258 | 903.5876 | 74.65 | 73.67 |
| KAN-LSTM (Liu et al., 2025b) | PG | 537.5864 | 242.4522 | 472.3925 | 93.20 | 85.95 |
| | SG | 488.0789 | 210.1525 | 306.8171 | 111.51 | 161.12 |
| | Zhoushan | 648.0208 | 396.6598 | 702.3364 | 83.72 | 69.18 |
| MMTP (Chen et al., 2023b) | PG | 522.5373 | 274.2757 | 507.8498 | 86.68 | 94.86 |
| | SG | 502.9317 | 243.0091 | 368.6798 | 74.81 | 83.54 |
| | Zhoushan | **603.3753** | 314.8713 | 572.7237 | 58.88 | 66.17 |
| BiIFNet (Li et al., 2024a) | PG | 539.1776 | 244.4186 | 437.386 | 99.46 | 114.01 |
| | SG | 535.8449 | 344.694 | 562.0756 | 96.85 | 105.11 |
| | Zhoushan | 656.4987 | 439.2614 | 755.6652 | 68.90 | 70.48 |
| STPredictor | PG | **299.6217** | **50.1605** | **108.5616** | **2.53** | **3.95** |
| | SG | **472.1775** | **104.4753** | **210.8482** | **4.80** | **6.38** |
| | Zhoushan | 644.2640 | **181.2342** | **385.6677** | **7.42** | **8.62** |

Table 1 presents the performance of nine methods evaluated across three representative maritime environments. The four SOTA baselines (Wang et al., 2024; Liu et al., 2025b; Chen et al., 2023b; Li et al., 2024a) all exhibit a noticeable reduction in accuracy when confronted with the unclustered, highly heterogeneous ship trajectories of these water areas. In contrast, STPredictor demonstrates consistently superior performance in PG and SG, indicating that the explicit incorporation of navigation-strategy cues and target-port pose priors provides effective directional intent constraints for maritime trajectory prediction. These priors effectively regularize CoG estimation, allowing STPredictor to maintain AHE and FHE within 10°, which is critical for practical maritime tasks such as port approaching, docking alignment, and collision-aware navigation, and to outperform the best-performing baseline method, MMTP (Chen et al., 2023b), by at least a sixfold margin in angular accuracy.

In the Zhoushan Archipelago, characterized by narrow passages and dense island formations, MMTP (Chen et al., 2023b) shows a localized advantage. Its trajectory-direction vector attention mechanism mitigates the effect of dis- continuous or large-amplitude ship motions by suppressing noise during turns and accelerations. As a result, MMTP achieves a lower RMSE than STPredictor in this hydrodynamically complex area, despite STPredictor still leading in overall angular precision. This region-specific reversal highlights the sensitivity of RMSE to turbulent flow, sharp heading transitions, and short-term track fragmentation, conditions that MMTP's directional attention is particularly suited to handle.

Following this water area analysis, we further evaluated STPredictor, together with eight additional baselines, on the draught-based dataset, again without applying maritime-route clustering, in order to assess the robustness of trajectory prediction across ships with varying inertia and maneuvering characteristics. The objective is to examine how varying ship inertia and navigation stability across draught intervals affect trajectory-prediction accuracy.

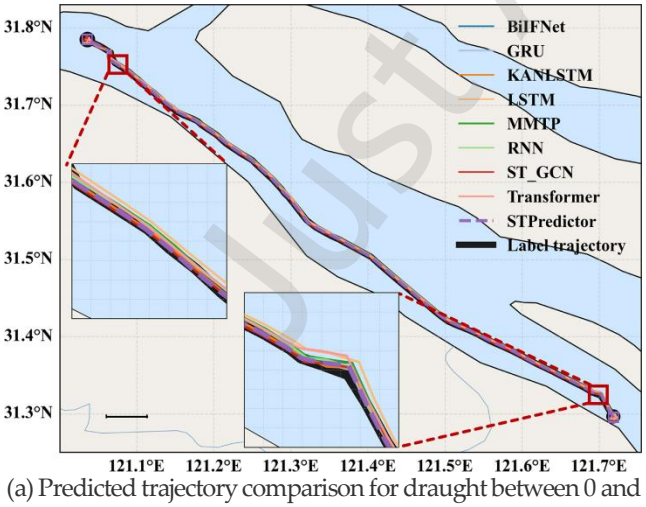
As shown the bold type in Table 2, our proposed method again achieved the best results. Since larger ships have deeper draughts and greater inertia, which render aircraft positions relatively stable, most methods demonstrated superior performance in 10–25 draught to the other two intervals. Concurrently, owing to this dataset's diverse nature derived from global AIS data, all methods exhibited a slight reduction in accuracy compared to Table 1.

For details, STPredictor consistently outperforms all baselines across every draught interval, achieving the lowest prediction error in all metrics. Quantitatively, STPredictor reduces RMSE by roughly 20–30% relative to the strongest baseline MMTP (Chen et al., 2023b) in all draught intervals. Specifically, STPredictor keeps AHE within 5.59–6.60° and FHE within 7.70–9.08°, whereas the best baseline still yields AHE/FHE above 47°/52°. Similarly, ADE and FDE are markedly lower, in 0–5 m draught, STPredictor

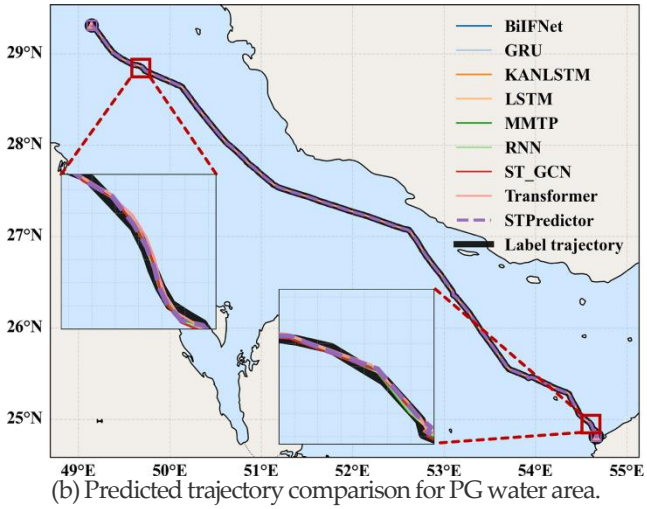
reaches an ADE of 168 m, nearly half of MMTP's 309m. These results demonstrate that integrating the observed scenario with human instructions enables STPredictor to more accurately infer maritime intentions, thereby delivering more precise predictions of waypoints and CoG.

**Table 2.** Comparison of the predicted points of the proposed and baseline methods in different draughts.

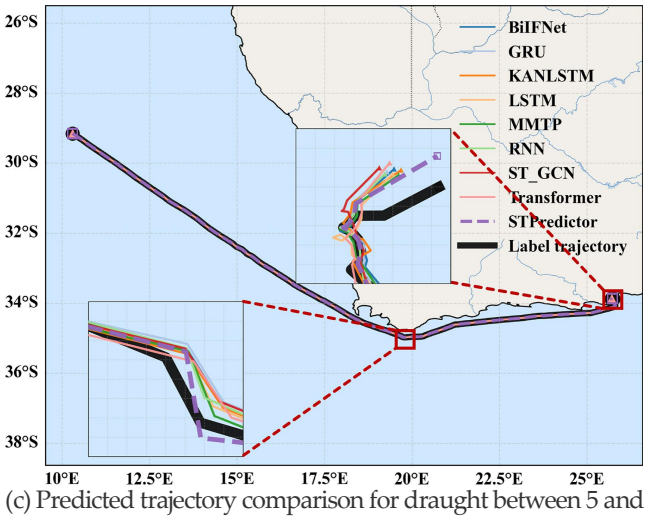
| Method | Draught (m) | RMSE | ADE | FDE | AHE (°) | FHE (°) |
|---|---|---|---|---|---|---|
| GRU (Dey and Salem, 2017) | 0-5 | 992.0621 | 824.1394 | 1110.7000 | 102.12 | 116.21 |
| | 5-10 | 907.9890 | 720.7502 | 1016.5729 | 73.90 | 79.33 |
| | 10-25 | 840.7305 | 713.4836 | 941.2712 | 66.60 | 67.31 |
| LSTM (Xin et al., 2018) | 0-5 | 1088.2235 | 612.1069 | 1023.1205 | 82.45 | 84.98 |
| | 5-10 | 1168.4928 | 489.1608 | 923.3544 | 75.54 | 80.40 |
| | 10-25 | 1013.4395 | 409.2130 | 782.7334 | 61.56 | 65.53 |
| RNN (Schuster and Paliwal, 1997) | 0-5 | 1233.1430 | 518.5061 | 985.9833 | 95.39 | 115.34 |
| | 5-10 | 1112.5019 | 436.6652 | 834.6043 | 84.34 | 89.93 |
| | 10-25 | 982.4847 | 405.2914 | 776.5489 | 70.22 | 87.31 |
| Transformer (Gao et al., 2023b) | 0-5 | 1041.7024 | 424.0508 | 824.7490 | 83.25 | 105.34 |
| | 5-10 | 974.8021 | 382.1189 | 742.2947 | 60.32 | 63.40 |
| | 10-25 | 955.4421 | 500.3441 | 790.1373 | 67.06 | 82.66 |
| DAA-SGCN (Wang et al., 2024) | 0-5 | 827.0852 | 355.2761 | 651.1896 | 77.07 | 92.22 |
| | 5-10 | 936.3820 | 325.3996 | 576.3868 | 69.07 | 85.89 |
| | 10-25 | 846.2980 | 286.6206 | 519.0388 | 57.00 | 77.82 |
| KAN-LSTM (Liu et al., 2025b) | 0-5 | 912.3227 | 636.3599 | 856.4736 | 59.39 | 109.50 |
| | 5-10 | 835.0072 | 558.9144 | 783.8911 | 53.43 | 63.45 |
| | 10-25 | 773.1547 | 547.9980 | 725.8251 | 49.98 | 58.31 |
| MMTP (Chen et al., 2023b) | 0-5 | 870.4016 | 308.6013 | 593.8337 | 83.01 | 86.94 |
| | 5-10 | 825.8353 | 334.7798 | 576.0638 | 50.59 | 55.17 |
| | 10-25 | 785.9218 | 248.0471 | 476.5866 | 47.38 | 52.93 |
| BiIFNet (Li et al., 2024a) | 0-5 | 920.0853 | 393.9556 | 741.1998 | 65.85 | 98.95 |
| | 5-10 | 830.6423 | 343.2512 | 628.7451 | 81.78 | 83.94 |
| | 10-25 | 789.4985 | 303.2982 | 528.7020 | 53.91 | 73.81 |
| STPredictor | 0-5 | **713.9187** | **168.1718** | **336.8572** | **6.60** | **9.08** |
| | 5-10 | **653.4171** | **153.9200** | **308.3100** | **6.04** | **8.31** |
| | 10-25 | **605.0158** | **142.5185** | **285.4722** | **5.59** | **7.70** |



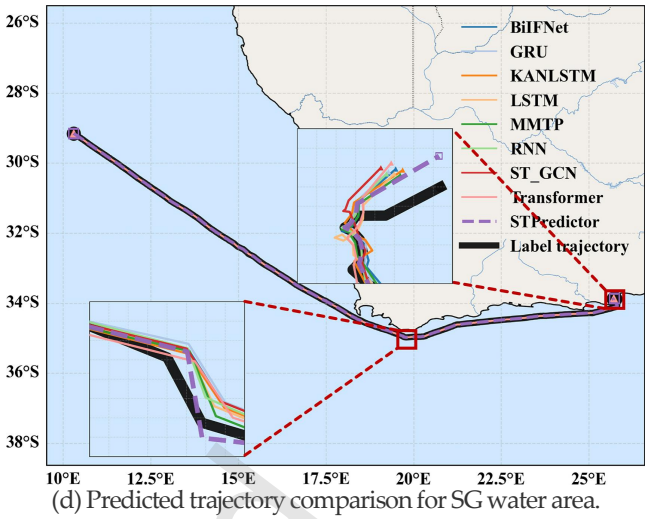

(a) Predicted trajectory comparison for draught between 0 and 5m.



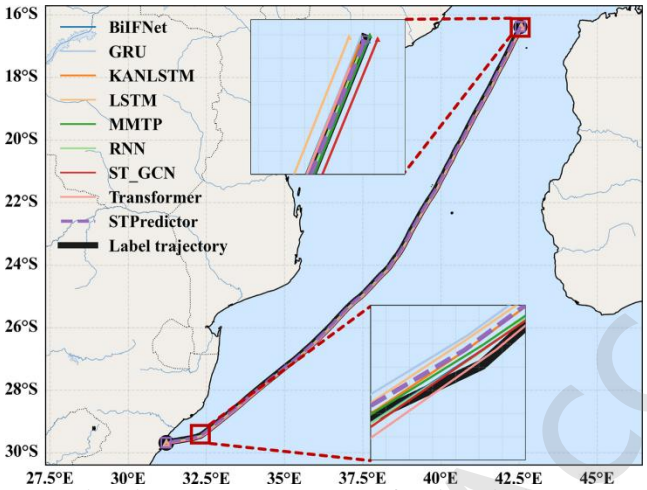

(b) Predicted trajectory comparison for PG water area.



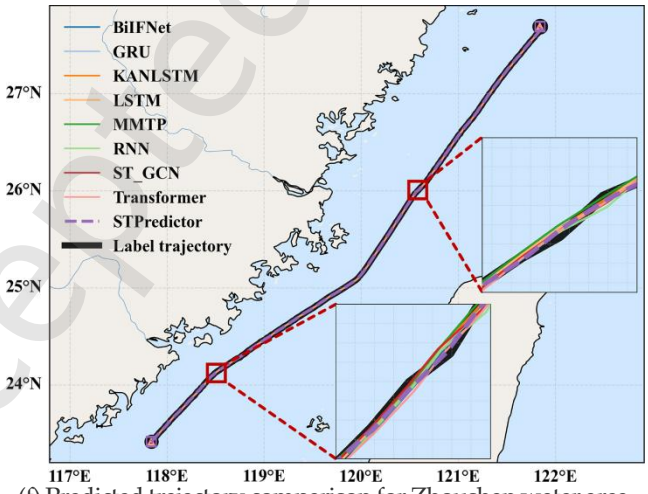

(c) Predicted trajectory comparison for draught between 5 and 10m.




(d) Predicted trajectory comparison for SG water area.




(e) Predicted trajectory comparison for draught between 10 and 25m.




(f) Predicted trajectory comparison for Zhoushan water area.

**Figure 3.** The comparison of predicted and actual trajectories with STPredictor and selected baseline methods in our two datasets, the draught results are listed in the left column, the geospatial results are presented in the right column.

These gains can be attributed to two key design features of STPredictor. First, the method explicitly incorporates navigation-strategy cues and target-port pose information, providing strong directional intent priors that substantially reduce heading ambiguity, especially in cases of noisy or discontinuous AIS ship trajectories. Second, its spatiotemporal consistency modeling enables the network to capture navigation evolution patterns that generalize across diverse global traffic scenarios, enhancing robustness to variations in environment and ship characteristics together, these mechanisms demonstrate that explicitly integrating semantic navigation intent with spatiotemporal modeling provides a principled and practically effective solution for safety-critical maritime trajectory prediction, enabling superior generalization, noise resilience, and directional stability across diverse water areas and ship draught conditions.

### *4.4. Visual analysis of individual trajectory cases*

To enable a detailed evaluation of model behavior and to interpret the origins of performance differences observed in the quantitative results, we further perform visualization and quantitative evaluation on representative ship trajectories selected from six subsets across the geospatial and draught-based datasets. In this subsection, we retain only the first predicted waypoint at each frame and concatenate these one-step-ahead predictions over time to form an integrated rolling trajectory for visual comparison. Accordingly, FDE and FHE are computed using the terminal point of this rolling trajectory. Figure 3 illustrates the visualized prediction results of STPredictor and the other eight well- trained baselines on these trajectories, where local trajectory segments are magnified to better highlight both spatial deviation and CoG inconsistency among different methods. Table 3 and Table 4 provide the corresponding quantitative metrics for six selected trajectories from two datasets.

Table 3 presents the predicted result of three selected trajectories from the geospatial-based dataset, focusing on how different methods handle directional stability and long-term spatial consistency under heterogeneous maritime conditions. STPredictor consistently achieves the lowest RMSE, ADE, and FDE across the selected trajectories. In the PG trajectory with 415 waypoints, STPredictor reaches an RMSE of 149.13 m, outperforming KAN-LSTM at 155.10m and substantially surpassing MMTP at 174.94 m. Its advantage is even more pronounced in angular prediction; STPredictor reduces FHE to

0.14°, which is over four times lower than the next-best baseline, MMTP at 0.59°, demonstrating the effectiveness of semantic navigation intent and port-pose priors in CoG estimation. Similar gains appear in the SG trajectory (115 waypoints), where STPredictor achieves an RMSE of 166.35 m, improving upon MMTP's 228.66 m by approximately 27%, and reducing AHE from MMTP's 9.04° to 0.32°, reflecting its strong directional stability in shipping lanes.

As shown in Figure 3(d), the SG trajectory with 115 waypoints represents a markedly more complex maritime trajectory under which all error metrics increase. Even in this challenging setting, our STPredictor still delivers the best overall performance, achieving an RMSE of 166.35 m, an improvement of approximately 27% over MMTP's 228.66 m, and reducing the AHE from 9.04° to just 0.32°, evidencing its strong directional stability in congested shipping environments where precise CoG control is critical for safe navigation.

For the Zhoushan trajectory with 286 waypoints, STPredictor continues to outperform baseline methods with an RMSE of 142.78 m, yielding a similar improvement range of 14–30% over baseline methods. Notably, it reduces ADE to 115.84 m, compared with BiIFNet's 178.12 m and MMTP's 209.23 m. Although Transformer and KAN-LSTM momentarily achieve the smallest FHE, STPredictor retains the most balanced performance across all metrics, maintaining both low spatial error and consistently low angular drift. These results illustrate that STPredictor not only corrects short-term directional deviations more effectively but also maintains superior long-term stability, especially in regions with strong currents, multi-island occlusions, or high-frequency maneuvering.

Table 4 lists the analysis result of three trajectories from the draught-based dataset. We observe that STPredictor maintains

**Table 3.** Comparison of the predicted points of the proposed and baseline methods on the three selected trajectories from different water areas in the geospatial-based dataset.

| Region | Method | RMSE | ADE | FDE | AHE(°) | FHE(°) |
|---|---|---|---|---|---|---|
| **PG** (415 waypoints) | GRU | 175.4490 | 191.0114 | 96.8701 | 96.36 | 0.79 |
| | LSTM | 172.6968 | 182.7632 | 37.9650 | 100.33 | 0.69 |
| | RNN | 167.2242 | 172.7227 | 21.6971 | 99.53 | 1.11 |
| | Transformer | 373.4964 | 493.6781 | 452.5558 | 101.64 | 10.16 |
| | DAA-SGCN | 164.2353 | 176.1480 | 91.0009 | 102.75 | 65.83 |
| | KAN-LSTM | 155.0961 | 162.7957 | 25.3921 | 95.92 | 1.00 |
| | MMTP | 174.9409 | 198.4376 | 69.7788 | 95.46 | 0.59 |
| | BiIFNet | 161.1960 | 180.4296 | 56.0567 | 100.39 | 209.50 |
| | STPredictor | **149.1280** | **143.4310** | **0.0007** | **91.65** | **0.14** |
| **SG** (115 waypoints) | GRU | 219.9644 | 232.5434 | 71.0862 | 78.78 | 10.91 |
| | LSTM | 224.2621 | 228.1290 | 65.9538 | 73.38 | 3.06 |
| | RNN | 210.5281 | 218.6979 | 61.1368 | 76.02 | 0.84 |
| | Transformer | 246.3166 | 268.5327 | 88.3878 | 85.77 | 42.87 |
| | DAA-SGCN | 277.1260 | 335.4730 | 208.5458 | 93.85 | 76.69 |
| | KAN-LSTM | 226.5217 | 235.4612 | 26.9385 | 96.58 | 6.10 |
| | MMTP | 228.6633 | 243.2324 | 20.3035 | 70.74 | 9.04 |
| | BiIFNet | 218.5120 | 220.5165 | 31.2918 | 78.92 | 38.97 |
| | STPredictor | **166.3541** | **150.2202** | **0.0009** | **58.11** | **0.32** |
| **Zhoushan** (286 waypoints) | GRU | 184.8102 | 183.6903 | 32.1733 | 95.68 | 26.39 |
| | LSTM | 183.0929 | 185.2806 | 27.5104 | 96.08 | 30.73 |
| | RNN | 201.1377 | 202.4795 | 41.0098 | 97.98 | 18.34 |
| | Transformer | 210.2387 | 235.9177 | 269.9336 | 96.91 | **0.03** |
| | DAA-SGCN | 183.1591 | 196.2259 | 118.6098 | 99.54 | 18.33 |
| | KAN-LSTM | 180.1729 | 185.7332 | 55.6186 | 95.88 | **0.03** |
| | MMTP | 193.4806 | 209.2313 | 41.9504 | 92.83 | 8.11 |
| | BiIFNet | 166.0987 | 178.1203 | 129.2767 | 88.61 | 32.72 |
| | STPredictor | **142.7753** | **115.8373** | **0.0025** | **69.38** | 0.43 |

**Table 4.** Comparison of the predicted points of the proposed and baseline methods on the three selected trajectories from different draught intervals in the draught-based dataset.

| Draughts (m) | Method | RMSE | ADE | FDE | AHE(°) | FHE(°) |
|---|---|---|---|---|---|---|
| **0–5** (186 waypoints) | GRU | 81.5156 | 93.4224 | 47.8861 | 39.69 | 144.86 |
| | LSTM | 158.4396 | 218.6459 | 193.2190 | 37.05 | 124.11 |
| | RNN | 71.6345 | 77.2997 | 32.2848 | 40.00 | 113.40 |
| | Transformer | 197.4756 | 274.5410 | 246.1386 | 36.90 | 63.74 |
| | DAA-SGCN | 120.0059 | 152.1310 | 180.5976 | 41.26 | 90.94 |
| | KAN-LSTM | 65.1779 | 73.3121 | 54.3917 | 44.82 | 109.15 |
| | MMTP | 86.4783 | 109.5839 | 69.5970 | 36.42 | 113.21 |
| | BiIFNet | 60.9076 | 65.8013 | 62.1123 | 46.61 | 120.14 |

| Draughts (m) | Method | RMSE | ADE | FDE | AHE(°) | FHE(°) |
|---|---|---|---|---|---|---|
| | STPredictor | **47.0915** | **22.2148** | **0.0009** | **20.76** | **8.51** |
| **5–10** (705 waypoints) | GRU | 244.8765 | 287.3349 | 1308.0421 | 101.60 | 17.98 |
| | LSTM | 228.1279 | 253.4044 | 1165.5239 | 103.38 | 9.25 |
| | RNN | 232.0008 | 258.8898 | 1339.3294 | 105.99 | **6.28** |
| | Transformer | 261.5346 | 336.0785 | 1359.7170 | 92.16 | 31.19 |
| | DAA-SGCN | 231.7453 | 258.0085 | 1521.5877 | 92.49 | 27.23 |
| | KAN-LSTM | 205.8697 | 245.3528 | 1017.9552 | 96.28 | 18.03 |
| | MMTP | 186.7191 | 216.8813 | 1015.7734 | 100.55 | 7.16 |
| | BiIFNet | 185.1100 | 208.1295 | 1181.0858 | 92.36 | 17.62 |
| | STPredictor | **120.4852** | **115.1290** | **892.2144** | **82.11** | 7.58 |
| **10–25** (1324 waypoints) | GRU | 52.7226 | 58.1346 | 68.9080 | 90.21 | 26.60 |
| | LSTM | 64.7531 | 75.8436 | 35.7285 | 87.73 | 9.58 |
| | RNN | 61.8002 | 69.1725 | 28.6234 | 90.25 | 21.83 |
| | Transformer | 127.2641 | 159.4459 | 349.6189 | 70.61 | 2.17 |
| | DAA-SGCN | 76.3539 | 92.5305 | 62.6603 | 90.80 | 50.22 |
| | KAN-LSTM | 54.6434 | 59.5349 | 25.1049 | 86.33 | 33.72 |
| | MMTP | 83.9271 | 110.1757 | 41.5900 | 87.00 | 33.72 |
| | BiIFNet | 58.4085 | 67.4702 | 48.2224 | 92.12 | 16.72 |
| | STPredictor | **42.1618** | **32.5461** | **0.0017** | **61.18** | **1.24** |

consistent advantages across ships with varying draught intervals, indicating strong generalization across ships with different inertia and maneuvering characteristics. For the shallow-draught trajectory with 186, STPredictor achieves an RMSE of 47.09 m, representing a 23% improvement over the strongest baseline BiIFNet with 60.91 m and nearly a 30% gain over KAN-LSTM with 65.18 m. In addition, STPredictor reduces AHE to 20.76°, which is almost half of the best competing method, MMTP with 36.42°. The FHE drops from baseline values exceeding 100° in most methods to 8.51°, demonstrating the method's ability to stabilize CoG predictions in scenarios characterized by rapid speed changes and highly erratic shallow-water navigation.

In the medium-draught interval with 705 waypoints, the performance gap between STPredictor and baseline methods becomes even more pronounced. STPredictor achieves an RMSE of 120.49 m, improving on BiIFNet with 185.11m and MMTP with 186.72 m by approximately 35%. It's ADE of 115.13 m that demonstrates a major reduction relative to all baselines, many of which exceed 200–300 m. Moreover, STPredictor achieves the lowest AHE, contrasting sharply with the 92–106° range of the baselines, while maintaining a competitive FHE of 7.58°. Although the RNN achieves a slightly lower FHE of 6.28°, its spatial prediction errors are substantially higher. This indicates that STPredictor not only preserves high-frequency direction cues but also provides a more balanced and globally consistent prediction across long-range trajectories typically associated with medium-draught merchant ships.

For deep-draught trajectory with 1324 waypoints, typically constrained to well-defined navigation corridors with minimal maneuvering freedom, STPredictor again achieves the best overall performance. It reaches an RMSE of 42.16 m, outperforming GRU, KAN-LSTM, and BiIFNet by 20–30%, and reducing ADE to 32.55 m, yielding a more than 40% improvement compared to all competing methods. Meanwhile, the FHE of STPredictor also reaches 1.24°, outperforming Transformer and significantly reducing long-distance directional drift. These results demonstrate that STPredictor remains robust even when predicting trajectories that are long, structured, and dominated by inertia-driven ship dynamics.

### *4.5. Ablation Experiments*

To further examine the generalization boundary of STPredictor under domain shift, we conduct an ablation experiment across two datasets. In line with the evaluation protocol that avoids route-specific clustering priors, each setting trains STPredictor on a source domain and directly tests it on a target domain, thereby exposing the model to distribution shifts in traffic scenarios and motion dynamics. The detailed results are listed in Table 5.

In the geospatial-based dataset, STPredictor performs best under in-domain evaluation, whereas cross-area transfer consistently degrades in all metrics. The most severe drops occur when transferring into SG, indicating that SG exhibits region-specific motion patterns that are difficult to generalize from other areas.

The same result in the draught-based dataset. Similarly, while in-domain performance is stable within each draught group, cross-draught transfer results in clear error growth. This suggests that differences in maneuverability and inertial dynamics across draught datasets significantly affect the CoG estimation learned by STPredictor.

Taken together, the ablation experiments provide direct evidence that STPredictor exhibits strong scenario understanding and robust motion-dynamics modeling capability. This explains why STPredictor exhibits strong robustness under matched-domain evaluation, whereas cross-domain transfer remains challenging without explicit domain adaptation. Meanwhile, Compared with Table 1 and Table 2, although STPredictor degrades under domain shift, it still outperforms several baseline methods.

Overall, the experimental results validate that the integration of navigation intention instructions, port-pose priors, and a fine-tuned LLM-based architecture provides a principled and practically effective solution for maritime trajectory prediction, enabling stable, interpretable, and high-precision predictions across heterogeneous maritime environments.

## 5 Conclusion

This research presents STPredictor, an interpretable maritime trajectory prediction method that formulates intention estimation and multi-step prediction as a language-modeling task. Experiments show robust generalization and directional stability across heterogeneous environments. By leveraging supervised fine-tuning and the inherent commonsense reasoning and self-explanatory capabilities of LLMs, STPredictor generates not only accurate predictions but also coherent CoT rationales aligned with human instructions and final intentions. Extensive experiments on our uncluttered real-world datasets demonstrate that STPredictor consistently outperforms established baselines, particularly under heterogeneous water areas and varying ship draught conditions. However, STPredictor has its inherent challenges. First, the experimental evaluation relied primarily on a 2021 shipping dataset dominated by cargo ships, which, while sufficient to validate generalization across water areas and draught intervals, limits exposure to extreme maneuvering behaviors exhibited by other ship types. Second, although STPredictor demonstrates superior prediction accuracy and stability in offline evaluation, its inference latency remains notably higher than that of other baselines. An individual prediction requires 8.4 seconds, whereas LSTM, RNN, GNN, and Transformer baselines complete inference in 0.84, 0.74, 1.52, and 2.03 milliseconds, respectively.

**Table 5.** Ablation experiment in cross-domain transfer of STPredictor across the geospatial-based and draught-based dataset. Each row reports the performance of the fine-tuned STPredictor on the target domain. Source and Target denote the training and test datasets.

| Source Domain | Source→ Target | RMSE | ADE | FDE | AHE(°) | FHE(°) |
|---|---|---|---|---|---|---|
| **Cross-water-area** | | | | | | |
| PG | PG → PG | **299.6217** | **50.1605** | **108.5616** | **2.53** | **3.95** |
| | PG → SG | 999.2053 | 890.1385 | 1746.3051 | 27.54 | 35.28 |
| | PG → Zhoushan | 758.7231 | 564.3729 | 1076.2956 | 19.42 | 27.11 |
| SG | SG → SG | **472.1775** | **210.8482** | **104.4753** | **4.80** | **6.38** |
| | SG → PG | 1146.7837 | 869.5942 | 1781.2747 | 24.04 | 32.62 |
| | SG → Zhoushan | 1099.8125 | 809.9954 | 1702.1754 | 25.47 | 32.69 |
| Zhoushan | Zhoushan → Zhoushan | **644.2640** | **181.2342** | **385.6677** | **7.42** | **8.62** |
| | Zhoushan → SG | 999.2053 | 890.1385 | 1746.3051 | 27.54 | 35.28 |
| | Zhoushan → PG | 744.8863 | 456.1481 | 848.4964 | 13.24 | 18.07 |
| **Cross-draught** | | | | | | |
| 0–5 m | 0–5 → 0–5 | **713.9187** | **168.1718** | **336.8572** | **6.60** | **9.08** |
| | 0–5 → 5–10 | 980.4320 | 412.6800 | 872.5400 | 16.80 | 23.90 |
| | 0–5 → 10–25 | 1042.2150 | 445.9200 | 938.1100 | 19.60 | 27.40 |
| 5–10 m | 5–10 → 5–10 | **653.4171** | **153.9200** | **308.3100** | **6.04** | **8.31** |
| | 5–10 → 0–5 | 952.7800 | 401.3500 | 846.7600 | 15.90 | 22.60 |
| | 5–10 → 10–25 | 905.6400 | 372.4800 | 804.9200 | 14.90 | 21.10 |
| 10–25 m | 10–25 → 10–25 | **605.0158** | **142.5185** | **285.4722** | **5.59** | **7.70** |
| | 10–25 → 0–5 | 1018.3300 | 438.2700 | 923.5400 | 18.70 | 26.30 |
| | 10–25 → 5–10 | 938.5600 | 389.4100 | 841.8800 | 16.40 | 23.20 |

This substantial gap arises from the large parameter footprint of LLMs, autoregressive decoding, and the quadratic scaling of global attention with context length. These factors introduce sequence-level dependencies that inflate end-to-end latency, particularly for long historical trajectories or prompt descriptions, thereby constraining real-time deployment. Nevertheless, several emerging techniques offer promising avenues for reducing computational overhead. Knowledge distillation, structured pruning, and post-training quantization have shown strong potential for accelerating LLM inference with minimal degradation in accuracy. Integrating these compression techniques will be critical for enhancing runtime efficiency and enabling real-time application aboard maritime platforms.

In summary, this research demonstrates through extensive experimental validation that STPredictor can substantially advance maritime trajectory prediction by unifying accuracy, directional stability, and interpretable instruction-aligned reasoning, establishing a coherent bridge between semantic decision-making and safety-critical motion prediction. Future efforts will focus on expanding environmental coverage, broadening platform applicability, and adopting compression techniques for LLMs to achieve real-time performance, ultimately advancing the deployment of trustworthy, explainable autonomy in maritime navigation.

## Author contributions

Siyu Teng: Visualization, Methodology, Conceptualization, Writing – original draft, Validation, Data curation. Ying Yang: Conceptualization, Writing – original draft. Liang Zhao: Data curation, Writing – review & editing. Baoding Zhou: Validation. Long Chen: Funding acquisition, Validation, Software. Ran Yan: Conceptualization, Data curation, Writing – review & editing. Jiasong Zhu: Funding acquisition, Conceptualization, Writing – review & editing.

## Replication and data sharing

The source codes and replication package are available on ETS

data at 10.26599/ETSD.2026.9190006. .

## 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, with the following exception: Prof. Ran Yan serves as a Young Editorial Board Member of Communications in Transportation Research. This editorial role has not influenced the design, conduct, or reporting of this research.

## Acknowledgements

This work was supported in part by the Joint Funds of the National Natural Science Foundation of China under U24B20162, in part by the National Natural Science Foundation of China under Grant 62373356, and in part by the National Key Research and Development Program of China under Grant 2022YFB2602101.

## References

Ai, Y., Teng, S., Yang, Q., Wu, Y., Li, Y., Gao, Y., Meng, F., Yang, S., Tian, B., Chen, L., Wang, F.-Y., 2024. iMAPeM: A New Paradigm for Implementing Intelligent Mining With Humans in the Loop. *IEEE Trans Syst Man Cybern Syst* 1–12. https://doi.org/10.1109/TSMC.2024.3396139

Chen, C.-Z., Liu, S.-Y., Zou, Z.-J., Zou, L., Liu, J.-Z., 2023a. Time series prediction of ship maneuvering motion based on dynamic mode decomposition. *Ocean Eng* 286, 115446. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115446

Chen, G., Wang, W., Xue, Y., 2021a. Identification of Ship Dynamics Model Based on Sparse Gaussian Process Regression with Similarity. *Symmetry* 13, 1956. https://doi.org/10.3390/sym13101956

Chen, J., Zhang, J., Chen, H., Zhao, Y., Wang, H., 2023b. A TDV attention-based BiGRU network for AIS-based vessel trajectory prediction. *iScience* 26, 106383. https://doi.org/10.1016/j.isci.2023.106383

Chen, Q., Zhou, W., Cheng, J., Yang, J., 2024. An Enhanced Retrieval Scheme for a Large Language Model with a Joint Strategy of Probabilistic Relevance and Semantic Association in the Vertical Domain. *Appl Sci* 14, 11529.

Chen, Y., Yang, S., Suo, Y., Zheng, M., 2021b. Ship Track Prediction Based on DLGWO-SVR. *Sci Program* 9085617. https://doi.org/10.1155/2021/9085617

Das, T., Goerlandt, F., Pelot, R., 2024. A mixed integer programming approach to improve oil spill response resource allocation in the Canadian arctic. *Multimodal Transp.* 3, 100110. https://doi.org/https://doi.org/10.1016/j.multra.2023.100110

Davari, N., Gholami, A., Shabani, M., 2017. Multirate Adaptive Kalman Filter for Marine Integrated Navigation System. *J Navig* 70, 628–647.

Dey, R., Salem, F.M., 2017. Gate-variants of Gated Recurrent Unit (GRU) neural networks, in: *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. pp. 1597–1600. https://doi.org/10.1109/MWSCAS.2017.8053243

Fan, T., Chen, J., Chung, E., 2025. Integrating micro and macro traffic control for mixed autonomy traffic. *Commun. Transp. Res.* 5, 100188. https://doi.org/10.1016/j.commtr.2025.100188

Farahnakian, Farshad, Nevalainen, P., Farahnakian, Fahimeh, Vähämäki, T., Heikkonen, J., 2025. Maritime vessel movement prediction: A temporal convolutional network model with optimal look-back window size determination. *Multimodal Transp.* 4, 100191. https://doi.org/https://doi.org/10.1016/j.multra.2025.100191

Gao, D.W., Wang, Q., Zhu, Y.S., Xie, L., Zhang, J.F., Yan, K., Zhang, P., 2023a. A novel long sequence multi-step ship trajectory prediction method considering historical data. *Proc Inst Mech Eng Part M J Eng Marit Env.* 237, 166–181. https://doi.org/10.1177/14750902221109718

Gao, K., Li, X., Chen, B., Hu, L., Liu, J., Du, R., Li, Y., 2023b. Dual Transformer Based Prediction for Lane Change Intentions and Trajectories in Mixed Traffic Environment. *IEEE Trans Intell Transp Syst* 24, 6203–6216. https://doi.org/10.1109/TITS.2023.3248842

Guo, Y., Yan, R., Qi, J., Liu, Y., Wang, S., Zhen, L., 2024. LNG bunkering infrastructure planning at port. *Multimodal Transp.* 3, 100134. https://doi.org/https://doi.org/10.1016/j.multra.2024.100134

Hong, Y., Zheng, Z., Chen, P., Wang, Y., Li, J., Gan, C., 2024. Multiply: A multisensory object-centric embodied large language model in 3D world, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 26406–26416.

Huang, C.P., Wu, Y.H., Chen, M.H., Wang, Y.C.F., Yang, F.E., 2025. ThinkAct: Vision-Language-Action Reasoning via Reinforced Visual Latent Planning.

Jia, C., Ma, J., Kouw, W.M., 2025. Multiple Variational Kalman-GRU for Ship Trajectory Prediction With Uncertainty. *IEEE Trans Aerosp Electron Syst* 61, 3654–3667. https://doi.org/10.1109/TAES.2024.3491053

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 255–260. https://doi.org/10.1126/science.aaa8415

Kim, T., Sharda, S., Zhou, X., Pendyala, R.M., 2020. A stepwise interpretable machine learning framework using linear regression (LR) and long short-term memory (LSTM): City-wide demand-side prediction of yellow taxi and for-hire vehicle (FHV) service. *Transp Res Part C Emerg Technol* 120, 102786. https://doi.org/https://doi.org/10.1016/j.trc.2020.102786

Lan, Z.X., Li, H.B., Liu, L.S., Fan, B., Lv, Y.S., Ren, Y.L., Cui, Z.Y., 2024. Traj-LLM: A New Exploration for Empowering Trajectory Prediction with Pre-trained Large Language Models.

Li, H., Xing, W., Jiao, H., Yuen, K.F., Gao, R., Li, Y., Matthews, C., Yang, Z., 2024a. Bi-directional information fusion-driven deep network for ship trajectory prediction in intelligent transportation systems. *Transp Res Part E Logist Transp Rev* 192, 103770. https://doi.org/10.1016/j.tre.2024.103770

Li, S.J., Liu, T.X., Liu, J.L., Xu, C.Q., He, J.W., 2024b. Research on trajectory tracking control of cascade MPC based on ship kinematics and dynamics. *J Ship Res.* https://doi.org/10.19693/j.issn.1673-3185.04141

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).* pp. 1389–1394. https://doi.org/10.1109/ICTIS.2019.8883596

Li, Y., Bai, F., Lyu, C., Qu, X., Liu, Y., 2025a. A systematic review of generative adversarial networks for traffic state prediction: Overview, taxonomy, and future prospects. *Inf Fusion* 117, 102915. https://doi.org/10.1016/j.inffus.2024.102915

Li, Y., Teng, S., Wang, J., Ai, Y., Tian, B., Xuanyuan, Z., Bing, Z., Knoll, A.C., Wang, F.-Y., Chen, L., 2025b. AutoMine: A Multimodal Dataset for Robot Navigation in Open-Pit Mines. *J Field Robot* 42, 1523–1536. https://doi.org/10.1002/rob.22469

Liu, F., Shafique, M., Luo, X., 2025a. Next leap in the sustainable transport revolution: Identifying gaps and proposing solutions for hydrogen mobility. *Commun Transp Res* 5, 100180. https://doi.org/10.1016/j.commtr.2025.100180

Liu, J., Shi, G., Zhu, K., 2019. Vessel Trajectory Prediction Model Based on AIS Sensor Data and Adaptive Chaos Differential

Evolution Support Vector Regression (ACDE-SVR). *Appl Sci* 9, 2983. https://doi.org/10.3390/app9152983

Liu, R.W., Liang, M., Nie, J., Lim, W.Y.B., Zhang, Y., Guizani, M., 2022a. Deep Learning-Powered Vessel Trajectory Prediction for Improving Smart Traffic Services in Maritime Internet of Things. *IEEE Trans Netw Sci Eng* 9, 3080–3094. https://doi.org/10.1109/TNSE.2022.3140529

Liu, W., Shen, H., Hu, Y., Hsieh, T.-H., Wang, S., Han, B., 2025b. Polar ship trajectory prediction based on Kolmogorov-Arnold Networks and LSTM. *Ocean Eng* 336, 121702. https://doi.org/10.1016/j.oceaneng.2025.121702

Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F., Qu, X., 2023. Can language models be used for real-world urban-delivery route optimization? *Innovation* 4.

Liu, Y., Wu, F., Lyu, C., Li, S., Ye, J., Qu, X., 2022b. Deep dispatching: A deep reinforcement learning approach for vehicle dispatching on online ride-hailing platform. *Transp Res Part E Logist Transp Rev* 161, 102694. https://doi.org/10.1016/j.tre.2022.102694

Liu, Y., Zhou, Y., Liu, Y., Xu, Z., He, Y., 2025c. Intelligent fault diagnosis for CNC through the integration of large language models and domain knowledge graphs. *Engineering* 53, 311–322. https://doi.org/10.1016/j.eng.2025.04.003

Luo, J., Xiao, Yi, Li, Y., Xiao, Ye, Yao, W., 2025. Multimodal deep learning framework for vessel trajectory prediction. *Ocean Eng* 336, 121766. https://doi.org/https://doi.org/10.1016/j.oceaneng.2025.121766

Nie, D., Guo, X., Duan, Y., Zhang, R., Chen, L., 2025. WMNav: Integrating Vision-Language Models into World Models for Object Goal Navigation.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 45, 2673–2681. https://doi.org/10.1109/78.650093

Seff, A., Cera, B., Chen, D., Ng, M., Zhou, A., Nayakanti, N., Refaat, K.S., Al-Rfou, R., Sapp, B., 2023. MotionLM: Multi-agent motion forecasting as language modeling, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* pp. 8579–8590.

Sekhon, J., Fleming, C., 2020. A spatially and temporally attentive joint trajectory prediction framework for modeling vessel intent, in: *Proceedings of Machine Learning Research.* pp. 318–327.

Skulstad, R., Li, G., Fossen, T.I., Vik, B., Zhang, H., 2021. A Hybrid Approach to Motion Prediction for Ship Docking—Integration of a Neural Network Model Into the Ship Dynamic Model. *IEEE Trans Instrum Meas* 70, 1–11. https://doi.org/10.1109/TIM.2020.3018568

Song, Z., Zhao, K., Ji, S., Jin, J.G., 2023. Simulation and optimization of transfer system for ore terminal with complex waterways. *Multimodal Transp.* 2, 100107. https://doi.org/https://doi.org/10.1016/j.multra.2023.100107

Sutulo, S., Soares, C.G., 2024. Nomoto-type manoeuvring mathematical models and their applicability to simulation tasks. *Ocean Eng* 304, 117639. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117639

Tang, C., Xia, J., Wang, W., Shang, H., Ji, W., Bao, Q., Yang, D., 2024. A dynamic window model prediction of artificial potential field method for improving the coincidence of actual and predicted trajectory of underactuated planing craft. *Ocean Eng* 313, 119351. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.119351

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, 136–145.

Teng, S., Hu, X., Deng, P., Li, B., Li, Y., Ai, Y., Yang, D., Li, L., Xuanyuan, Z., Zhu, F., others, 2023. Motion planning for autonomous driving: The state of the art and future perspectives. *IEEE Trans Intell Veh* 8, 3692–3711.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., others, 2023. Llama: Open and efficient foundation language models. ArXiv Prepr. ArXiv230213971.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.

Wang, R., Shang, T., Yang, D., Yan, R., 2025a. Empowering econometric methods with machine learning for policy making: A comparative study in maritime transportation. *Transp Res Part Policy Pr.* 200, 104635. https://doi.org/10.1016/j.tra.2025.104635

Wang, S., Li, Y., Xing, H., Zhang, Z., 2024. Vessel trajectory prediction based on spatio-temporal graph convolutional network for complex and crowded sea areas. *Ocean Eng* 298,

117232. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117232

Wang, S., Wang, R., Ren, F., Chen, M., Yan, R., 2025b. A privacy-preserving federated transfer learning with ship mapping (FTL-SM) framework for accurate ship energy efficiency prediction. *Adv Eng Inf.* 68, 103569. https://doi.org/10.1016/j.aei.2025.103569

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., others, 2022. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst* 35, 24824–24837.

Wu, F., Lyu, C., Liu, Y., 2022. A personalized recommendation system for multi-modal transportation

systems. *Multimodal Transp.* 1, 100016. https://doi.org/https://doi.org/10.1016/j.multra.2022.100016

Xin, L., Wang, P., Chan, C.-Y., Chen, J., Li, S.E., Cheng, B., 2018. Intention-aware Long Horizon Trajectory Prediction of Surrounding Vehicles using Dual LSTM Networks, in: *2018 21st International Conference on Intelligent Transportation Systems.* pp. 1441–1446. https://doi.org/10.1109/ITSC.2018.8569595

Yang, L., Yuan, M., Liu, Y., Qu, X., Hu, Z., Zhang, Z., Zhao, X., Fang, S., 2026a. Optimization of task scheduling and resource allocation for autonomous vehicle testing in vehicle-road-cloud collaborative systems. *Expert Syst. Appl.* 299, 129943. https://doi.org/https://doi.org/10.1016/j.eswa.2025.129943

Yang, Y., Zhan, J., Xu, M., Liu, Y., Qu, X., 2026b. Toward climate-neutral urban mobility: understanding shared e-scooter carbon emission patterns through multi-city evidence in Europe. *Transp. Res. Part Policy Pract.* 203, 104736. https://doi.org/https://doi.org/10.1016/j.tra.2025.104736

Yin, J., Yu, Z.W., Wu, H.F., 2025. Ship trajectory prediction based on LSTM model with multi-scale convolution and attention mechanism. *Ocean Eng* 338, 122055. https://doi.org/10.1016/j.oceaneng.2025.122055

Yu, D.H., Roh, M.I., 2024. Method for anti-collision path planning using velocity obstacle and A* algorithms for maritime autonomous surface ship. *Int J Nav Arch. Ocean Eng* 16, 100586. https://doi.org/10.1016/j.ijnaoe.2024.100586

Zhang, B., Sennrich, R., 2019. Root Mean Square Layer Normalization.

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 Emerg Technol* 118, 102729. https://doi.org/10.1016/j.trc.2020.102729

Zhang, J.N., Liu, Z.X., Gan, Y.H., Liu, Y.S., Dong, J.Y., 2025a. Track-to-track association from diverse source ship trajectory based on an improved graph neural network. *Appl Ocean Res* 159, 104598. https://doi.org/10.1016/j.apor.2025.104598

Zhang, R., Guo, X., Zheng, W., Zhang, C., Keutzer, K., Chen, L., 2024a. Instruct Large Language Models to Drive like Humans.

Zhang, R., Han, J., Liu, C., Gao, P., Zhou, A., Hu, X., Yan, S., Lu, P., Li, H., Qiao, Y., 2024b. LLaMA-Adapter: Efficient Fine-tuning of Language Models with Zero-init Attention, in: *International Conference on Learning Representations (ICLR).*

Zhang, X.L., Liu, J., Chen, C.C., Gong, P.Z., Wu, Z.D., Guo, L., 2025b. Modeling temporal continuity of spatial interactions for vessel trajectories prediction in maritime transportation systems. *Eng Appl Artif Intell* 158, 111378. https://doi.org/10.1016/j.engappai.2025.111378

Zhang, Y., Xue, W., Sun, L., Shen, J., 2021. Extended state Kalman filter-based path following control of underactuated autonomous vessels. *Trans Inst Meas Control* 43, 3311–3321. https://doi.org/10.1177/0142331221994410

Zhao, L., Xu, M., Liu, L., Bai, Y., Zhang, M., Yan, R., 2025. Intelligent shipping: integrating autonomous maneuvering and maritime knowledge in the Singapore-Rotterdam Corridor. *Commun Eng* 4, 11.

Zheng, Z., Zhang, J., Vu, T.A., Diao, S., Tim, Y.H.W., Yeung, S.K., 2023. MarineGPT: Unlocking Real-World Ocean Secrets to the Public.

Zhou, Z., Zhu, Y., Wen, J., Shen, C., Xu, Y., 2025. Vision-Language-Action Model with Open-World Embodied Reasoning from Pretrained Knowledge.



## Author biography

**Siyu Teng** received the M.S. degree in Software Engineering from Jilin University, Changchun, China, and the Ph.D. degree in Computer Science from Hong Kong Baptist University, Hongkong, China, in 2020. He is currently an Assistant Professor with the College of Civil and Transportation Engineering, Shenzhen University.

**Ying Yang** received the B.Eng. degree in transportation engineering from Guangdong University of Technology, Guangzhou, China, in 2024. She is currently pursuing the M.S. degree with Shenzhen University, Shenzhen, China. Her research interests include autonomous driving perception, traffic behavior analysis, and deep learning.

**Liang Zhao** received the PhD degree in civil engineering from Zhejiang University, Hangzhou, China, and the Bachelor degree in civil engineering from Central South University, Changsha, China. He is currently a research fellow in Zhejiang University. His research interests include artificial intelligence, autonomous ships, and AI for maritime.

**Baoding Zhou** received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015. He is currently an Associate Professor with the College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China. His research interests include indoor localization and mapping, mobile computing, and intelligent transportation.

**Long Chen** received the B.Eng and Ph.D. degree in Electrical and Electronic Engineering from Wuhan University, Wuhan, China, in 2007 and 2013. He is currently a Professor with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

**Ran (Angel) Yan** received the M.Phil. degree and the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, China, in 2020 and 2022, respectively. She has been an Assistant Professor with the School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, since March 2023.

**Jiasong Zhu** received the B.Eng. degree from Huazhong University of Science and Technology, Wuhan, China, and the M.Eng. degree from Wuhan University, Wuhan, China, in 2003. He received the Ph.D. degree from The University of Hong Kong, Hong Kong, China, in 2008. He is currently a Professor with the Department of Transportation Engineering, College of Civil and Transportation Engineering, Shenzhen University.

## Appendix A

In this Appendix, we provide detailed descriptions of the dataset construction and data formulation pipeline used in STPredictor. Specifically, we first introduce the construction of the geospatial-based and draught-based datasets, including their scenario coverage and dataset composition in Section 1.1. We then present statistical analyses and distribution characteristics of the datasets across representative water areas and draught categories in Seciton 1.2. Finally, we describe the data processing and structured input–output design for LLM-based trajectory prediction, including instruction design in Seciton 1.3.1, scenario description prompts in section 1.3.2, and the interpretable prediction format in Section 1.3.3.

### A1 Dataset Construction

Learning-based spatiotemporal models have demonstrated strong capability in capturing nonlinear temporal dynamics and spatial interactions within complex transportation systems (Zhang et al., 2024c). Maritime activities inherently exhibit rich dynamic characteristics, including variations in ship speed, turning behavior, and environmental interferences. These dynamics can only be effectively captured and represented through real-world data. To address the data scarcity challenge in trajectory prediction, this research constructs two high-fidelity trajectory datasets aligned with human navigation habits, including a geospatial-based dataset with diverse maritime conditions and a draught-based dataset with different ship dynamics.

The geospatial-based dataset characterizes the diversity of maritime operating conditions, distinguished primarily by the geodetic coordinates of the underlying water areas. In detail, it comprises three representative areas: the SG water area in the South China Sea, the PG water area in the Indian Ocean, and the Zhoushan water area in the East China Sea. This configuration captures ship responses to multiple perturbations such as wind forces, ocean currents, and rudder-angle variations, faithfully mirroring human navigational decision-making and control patterns under complex sea conditions.

#### A 1.1. Dataset Description

The first dataset is geospatial-based and comprises three representative water areas. These areas correspond to global trunk shipping lanes, high-dynamic open-sea regions, and typical harbor waterways, forming a hierarchical and representative experimental benchmark.

Among them, the SG water area, which is one of the world's busiest international maritime passages, connecting the Malacca Strait with the South China Sea, and serves as a key hub on the Asia-Europe main route. The strait features narrow channels, pronounced depth variations, and strong tidal and lateral wave effects, with

dense crossings and anchorage zones forming a high-risk, high-density environment. Ships frequently engage in meeting, overtaking, and crossing maneuvers, resulting in trajectories with pronounced nonlinearity and uncertainty. This dataset is thus employed to evaluate each method's generalization capability and stability under ultra-dense traffic and high-disturbance conditions, representing the canonical characteristics of global main shipping routes.

The second is the PG water area, which is located in the Middle East and is a shallow, semi-enclosed basin bordered by Iran to the north and the Arabian Peninsula to the south. It exhibits highly variable meteorological and oceanic conditions, with strong convective winds and circulation patterns in summer, and cyclone-driven disturbances in winter. As a core node in global energy routes, the Persian Gulf sustains heavy oil and gas traffic and significant interregional trade. This dataset is used to evaluate method performance in medium-density transport corridors, where long-duration passages along fixed shipping lanes give rise to trajectory patterns representative of offshore transportation routes.

The third is the Zhoushan water area, which is located at the western edge of the East China Sea near the mouth of Hangzhou Bay. The archipelago is one of China's most critical maritime transport hubs. The waterways are narrow, with dense islands, strong tidal currents, and frequent movements of large ships, forming a typical high-density, high-complexity port environment. This dataset is used to evaluate the adaptability and computational robustness of different trajectory prediction methods under terrain constraints and multi-ship cooperative navigation conditions, offering critical insights into collision-avoidance decision-making for IMS.

Through systematic experiments across these three representative maritime regions, we try to reveal performance differences, adaptation boundaries, and generalization potentials among nine trajectory prediction methods across international main routes, open-sea zones, and port environments.

To further investigate the dynamic discrepancies and trajectory variations of ships under distinct physical constraints, this research constructs the second dataset, which is composed of three distinct draught and is used to train and validate the prediction method across ships of different sizes, including shallow-draught (0–5 m), medium-draught (5–10 m), and deep-draught (10–25 m). These divisions reflect significant differences in ship composition, maneuvering characteristics, and hydrodynamic responses, thereby providing a multidimensional foundation for learning multi-scale dynamic features and developing cross-type adaptive prediction mechanisms.

1) Shallow-draught: Small cargo ships operating with high maneuverability and strong tidal influence. Used to assess short-horizon, high-frequency maneuvers under disturbed flows and to learn localized behaviors (collision avoidance and course changes) for real-time planning.

2) Medium-draught: Medium container ships on

medium to long-range routes with balanced maneuverability and inertia, showing moderate speed adjustments and smooth heading transitions. Evaluates robustness and generalization in commercial traffic, focusing on mid-scale inertial continuation, route keeping, and cooperative avoidance.

3) Deep-draught: Large container ships on deep-sea routes and in deepwater ports with strong inertia and limited maneuverability, exhibiting slow, coupled trajectory changes. Evaluates accuracy and stability under long- delay, low-maneuverability constraints and supports studies on autonomous berthing and large-ship cooperative avoidance.

### A 1.2. Data analysis

This research constructs two large-scale trajectory datasets based on distinct maritime regions and draught categories, providing a solid foundation for comparing the performance and applicability of ten trajectory prediction methods. The datasets help uncover the method's generalization capabilities across various environmental and dynamic conditions, offering valuable insights for the design and application of IMS. Detailed information about the datasets is provided in Table 1 and Table 2.

AIS data often suffers from irregular time intervals and varying signal reception due to factors like system adaptive transmission frequency and satellite coverage. Before model training and fine-tuning, the raw AIS datasets undergo a systematic preprocessing pipeline, including noise removal, missing data completion, and outlier filtering. Additionally, speed-constrained Cubic Hermite Spline (Lekkas and Fossen, 2014) is applied to fit the filtered AIS data, resampling all trajectories with 5-minute intervals. This interpolation ensures kinematic continuity and physical plausibility by using SoG and CoG as first-derivative constraints. The result is time-aligned and kinematically consistent trajectories, providing stable input for sequential modeling.

**Table 1.** Metrics comparison of three water area datasets.

| Water areas | Time | Trajectories | Cover areas | Original points | Revised points* |
|---|---|---|---|---|---|
| SG | June–December 2021 | 224 | longitude: 100.0–110.0<br>latitude: -4.0–6.0 | 29,434 | 895,660 |
| PG | February–December 2021 | 402 | longitude: 45.0–55.0<br>latitude: 23.0–32.0 | 96,616 | 1,298,500 |
| Zhoushan | January–December 2021 | 1,179 | longitude: 118.0–125.0<br>latitude: 27.0–32.0 | 212,501 | 2,080,224 |

Note*: Revised points are interpolated and optimized versions of the original points, with a 5-minute interval.

**Table 2.** Metrics comparison of ships with different draughts.

| Draught depth | Period | Trajectory No. | Original points | Revised points* |
|---|---|---|---|---|
| 0–5 m | June–December 2021 | 365 | 39,948 | 229,770 |
| 5–10 m | January–December 2021 | 3,294 | 1,956,789 | 4,303,282 |
| 10–25 m | January–December 2021 | 2,324 | 2,157,539 | 4,571,036 |

Note*: Revised points are interpolated and optimized versions of the original points, with a 5-minute interval.

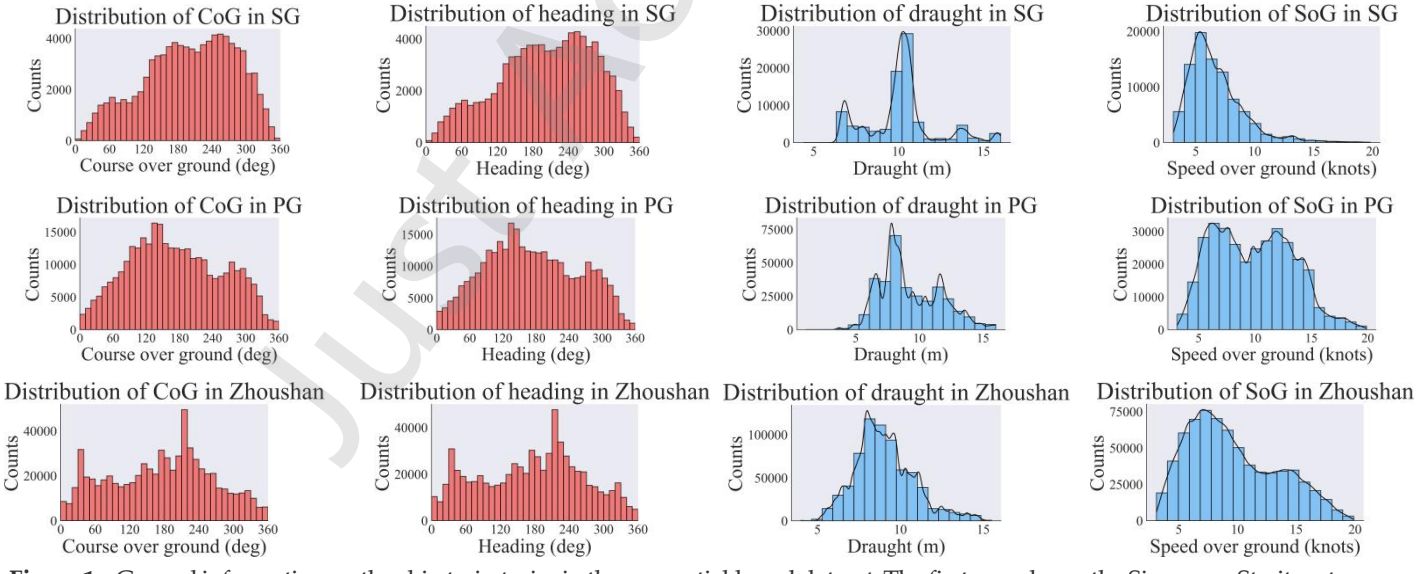
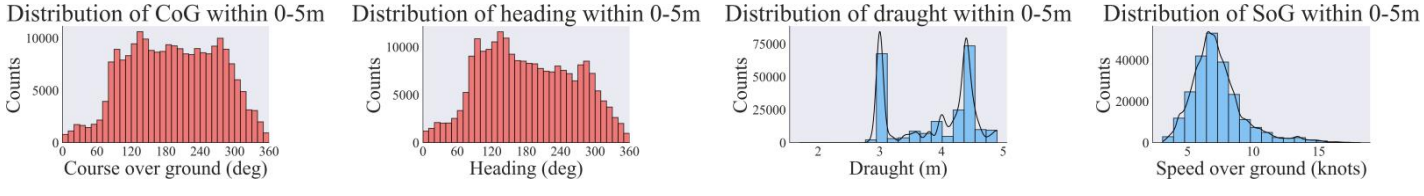



**Figure 1.** General information on the ship trajectories in the geospatial-based dataset. The first row shows the Singapore Strait water area, the second is the review of waypoints in the PG water area, and the last row shows the Zhoushan Archipelagic water area.



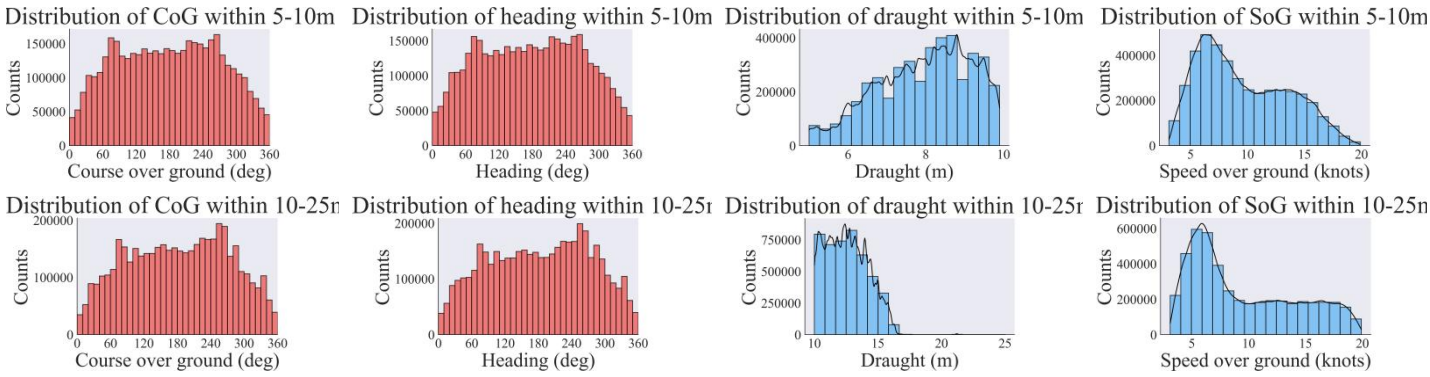



**Figure 2.** Comprehensive overview of waypoint distributions in the draught-based dataset. The first row illustrates small ships with shallow draughts ranging from 0 to 5 meters. The second row summarizes waypoints for medium ships with moderate draughts between 5 and 10 meters. The final row depicts large ships with deep draughts of 10 to 25 meters

The shipping scenario in the PG and Zhoushan water areas is more complex, with dense ports and frequent anchoring and berthing activities, requiring finer filtering and resampling during preprocessing. In contrast, the Singapore Strait, with fewer geographical and channel restrictions, exhibits more continuous trajectories and provides a larger volume of valid data. After preprocessing, the three datasets retain 1,805 complete trajectories, increasing from 338,551 raw AIS points to 4,274,384 resampled trajectory points after preprocessing. The detailed information is listed in Table 1.

The draught-based dataset is illustrated in Table 2. Ships with varying draughts exhibit significant differences in dynamic characteristics and maneuvering behavior, requiring designed physical constraints and motion modeling. For shallow-draught ships (0-5 m), high maneuverability and frequent trajectory changes are observed, often displaying strong randomness and noise. Medium-draught ships (5–10 m) tend to follow fixed routes with higher trajectory regularity and directional consistency. Deep-draught ships (10–25 m) are typically large, with significant inertia, large turning radii, and smoother trajectories. Differential sampling and smoothing strategies were applied to preserve the authenticity and comparability of their dynamic features. Ultimately, the three datasets comprise 5,983 trajectories, covering a wide range of ship types from small workboats to large oil tankers, providing a rich and balanced sample base for multi-scale trajectory prediction model training.

In the geospatial-based dataset, all trajectory points across the three water areas are systematically organized. As shown in Figure 1, the distribution of heading and CoG reveals consistent trends, indicating that ships generally follow stable navigation patterns in main shipping routes. However, significant differences are observed in the distribution of draught and SoG: Singapore Strait exhibits a concentrated speed distribution in the low-to-mid range, reflecting its narrower fairways, high traffic density; PG shows a more dispersed speed distribution with more variable course changes, indicating its diversified operational nature; and Zhoushan Archipelago displays the highest concentration of draught values, signifying a more homogenous ship type structure with a focus on medium to large cargo ships. These distribution differences influence trajectory prediction accuracy during model training: Singapore Strait's high-density and stable trajectories lead to faster convergence and higher prediction accuracy; the complex course patterns in the PG improve the model's generalization ability in non-regular waters; and Zhoushan Archipelago's consistent speed and draught characteristics enhance stability and recognition of specific ship dynamics. These complementary dataset characteristics lay a robust foundation for robust evaluation and transfer learning under multi-scenario, multi-scale conditions.

Additionally, from a ship dynamics perspective, an auxiliary dataset based on draught classification was constructed, as shown in Figure 2. The draught heavily influences a ship's physical properties and maneuvering inertia: ships with draughts under 5 meters are typically small, responsive, and exhibit frequent trajectory changes; medium-draught ships operate along fixed routes, with stable and regular trajectories; and deep-draught ships display significant inertia, large turning radii, and smooth, stable trajectories.

In summary, the geospatial-based dataset reveals the macro-level effects of environmental complexity and navigation conditions on trajectory distribution, while the draught partition dataset captures the micro-level differences in ship dynamics. Together, they form a high-fidelity shipping data system that balances environmental diversity and dynamic authenticity, providing a solid data foundation for performance evaluation and generalization studies in trajectory prediction methods.

### A 1.3. Data Processing

Following the filtering protocol in (Teng et al., 2025), the revised AIS data removes invalid or anomalous records, such as erroneous latitude/longitude and abnormal SoG/CoG. The revised data contains the full trajectory information from departure and pilotage to nearshore transit and berthing, ensuring the physical authenticity and statistical consistency of the trajectory samples. To enable a comprehensive understanding of ship state, we propose a structured trajectory data framework, which is centered around a local coordinate system, integrates ship navigation environment information with motion semantic features, offering high fidelity and strong semantic consistency to support subsequent trajectory prediction and autonomous decision-making modeling. Each frame corresponds to a dynamic state description of the ship at a specific moment, simulating a closed-loop process of a navigation system in maritime environments.

Regarding coordinate system construction, we adopt a local

coordinate system with the ship itself as the origin. The ship's forward direction is defined as the positive X-axis, and the positive Y-axis is perpendicular to the X-axis, pointing to the port side. CoG is defined as the counterclockwise angle from the positive X-axis. This definition aligns with kinematic conventions in robot path planning, ensuring that the method can understand motion direction and spatial relationships geometrically during training. This coordinate system not only improves the comparability and transferability of data across different scenarios but also provides a unified reference frame for consistency learning in multimodal methods for trajectory understanding and behavior generation.

**A 1.3.1. Instruction Design**

To leverage the broad knowledge and strong logical reasoning capabilities of LLMs, we introduce a key high- level semantic instruction into the dataset, enhancing the robustness and cognitive consistency of ship trajectory prediction. This instruction describes the goals, constraints, and environmental context of navigational tasks, allowing the method to not only predict sequential trajectories but also understand navigational intent and operational logic, thereby enabling semantic-level perception of navigation tasks.

The instruction provides interpretable cognitive guidance for the decision process of the prediction method. The instruction "Achieve an energy-efficient acceleration of the ship" expresses an acceleration intent focused on energy conservation, while "Maintain a steady cruise at the current speed and a safe distance of the coastline" describes an operational context centered on safety and stability. These natural language serves not only as input prompts but also as semantic abstractions of navigational strategies and scenario constraints, enabling the method to consider both goal-directedness and context-awareness during prediction.

By embedding semantic instructions into the input structure, the method's decision process evolves from a data- driven model to a task-cognitive one. This structured-semantic fusion design allows LLMs to not only learn the mapping patterns of state space but also understand the logical structure and operational constraints of navigational tasks, offering behavior consistency and interpretability closer to human cognition.

To further enhance the accuracy of the predicted trajectory, four representative task instructions were designed based on different navigational strategies: Cruising, Speeding, Decelerating, and Maneuvering (Yang et al., 2021). Each strategy is paired with multiple semantic instruction templates to accommodate task expressions under different operational conditions, covering both offshore and nearshore scenarios. These instruction templates not only define key parameters such as time span, sampling intervals, and speed range but also describe driving intent, operational constraints, and safety boundaries in natural language. This formalized semantic embedding allows the method to perceive and understand high-level semantic information beyond numerical feature learning, establishing a deep coupling between navigational behavior and semantic cognition. This design enables the prediction method to exhibit stronger adaptability, interpretability, and safety in varying sea conditions and complex traffic environments, laying the theoretical and methodological foundation for cognitive decision-making, as shown in Table 3. To ensure the adaptability of LLMs to these instructions and provide a structured reasoning output template for LLM-based prediction methods, two design strategies were employed:

1) Cruising: Maintain the current speed and heading while ensuring safety constraints, focusing on trajectory smoothness and safe distance from shorelines or shallows. This involves minimizing risks of deviation and grounding, considering tidal and nearshore flow conditions.

2) Speeding: Increase speed within speed limits and acceptable traffic density to meet timeliness goals or leverage favorable current conditions. This requires ensuring the physical feasibility of acceleration and the rate of turn.

3) Decelerating: Reduce speed due to entering speed-restricted zones, congested waters, or approaching a way-point, emphasizing a balance between energy consumption and safety, while reserving dynamic margin for potential risk.

4) Maneuvering: Actions focused on altering heading or trajectory, including route correction, collision avoidance, or compensation for crosscurrents and sidewinds. This typically involves controlled rates of turn and lateral displacement.

Expression-diversity instructions are introduced to more comprehensively capture the richness and variability of human natural language. For this purpose, we generate 12 distinct linguistic variants for each high-level instruction using ChatGPT-5.2. These variants preserve semantic equivalence while differing in wording and syntactic structure. Such a design broadens the linguistic coverage and interpretive flexibility of the instruction set, enabling STPredictor to more robustly process complex inputs that exhibit diverse expression styles, thereby better approximating language navigation commands.

Information-enriched instructions are designed to increase the precision and explicitness of natural language expressions. Each instruction is augmented with shipping-related situational details. This semantically reinforced design allows STPredictor to form a deeper understanding of the surrounding operational context, thereby enhancing both the safety and interpretability of the method in complex maritime environments.

The high-level semantic instruction achieves both semantic diversity and information density, providing a robust linguistic foundation for generalized and interpretable LLM-driven trajectory prediction. The detailed information is listed in Table 3. An example of general-purpose instruction prompts used during the fine-tuning stage of STPredictor is shown in Figure 3.

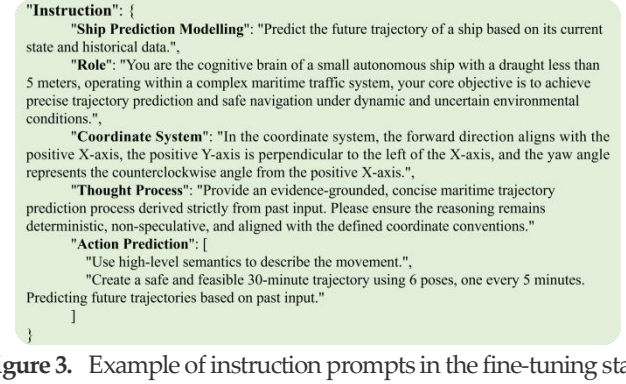



**Figure 3.** Example of instruction prompts in the fine-tuning stage of STPredictor.

### A 1.3.2. Scenario Description

In the trajectory prediction dataset, the input part plays a central role in perception and state modeling. Its significance lies not only in the structured representation of raw AIS data but also in the cross-domain translation from geographic semantic space to ship dynamics space. This section integrates two major information dimensions: Map Insights, reflecting static semantic features of water areas and operational conditions, and Ego-ship, which reflects the dynamic response characteristics under different draught levels. Their combination provides the method with a comprehensive spatiotemporal perception foundation, ensuring cognitive consistency across multiple maritime regions and ship types.

1) Map Insights Module: This module defines the static constraint of the maritime environment. It incorporates region-specific speed limits, dataset boundaries, and geographic elements essential for trajectory prediction, such as port locations, channel orientations, shallow water boundaries, and anchorage zones. By integrating these elements, the method can sense traffic density and environmental constraints during prediction. This module not only provides physical constraints but also acts as a "safety prior" input during model fine-tuning, ensuring that predicted trajectories remain physically realizable and compliant with regional maritime regulations and geographic conditions.

2) Ego-ship Module: This module characterizes the ship's static attributes and dynamic features, serving as the core driving factor for trajectory prediction. It aggregates key dynamic parameters such as the ship's identification, type, pose, heading, CoG, and SoG. All state variables are modeled in a local coordinate system centered at the ego ship, where the X-axis is defined as the forward direction, and the Y-axis points to the port side. This design aligns with the kinematic conventions in trajectory prediction and effectively eliminates scale dependency from geographic coordinates, enabling the method to concentrate on learning relative motion patterns rather than absolute positions.

3) Historical Trajectory Module: This module records the ship's dynamic evolution over time in a time-series format. Each input instance includes 12 historical sampling points, uniformly sampled at 5-minute intervals, covering a total temporal span of 60 minutes, which capture the ship's attitude changes and speed fluctuations. Unlike the discrete positional data in traditional AIS

systems, this research uses continuous point sequences in the local coordinate system to represent motion trajectories, enabling time modeling and dynamic trend inference within a unified geometric space.

Overall, the Input module constructs a multi-layered semantic input system by deeply integrating water region constraints, ship dynamic states, and historical trajectories. It meets the traditional needs of trajectory prediction methods for spatiotemporal feature modeling while providing a structured bridge for LLMs to establish intrinsic connections between semantic understanding, dynamic reasoning, and safe decision-making. Based on this design, the method achieves consistent motion cognition and behavior prediction under multi-area scenarios and multi-draught levels, providing strong data support for trajectory reasoning and decision-making in trajectory prediction. A representative example of the scenario description prompts used in STPredictor is presented in Figure 4.


```
"Input": {
        "Map Insights": {
          "Speed Limit": 25.0,
          "Speed Unit": "knot"
        },
        "Ego Ship": {
          "Pose": [0.0, 0.0],
          "LatLon": [32.7895, 120.8410],
          "Course over Ground": 0.0,
          "Speed over Ground": [6.33, 0.0],
          "Port Name": "NAN TONG",
          "Port LatLon": [32.0305, 120.8572],
          "Port Coordinates": [-84172.03, -1528.25],
          "Arrival Time": "2021-07-01",
          "Historical Trajectory": [...]
        }
}
```


**Figure 4.** Example of scenario description prompts in the fine-tuning stage of STPredictor.


```
"Output": {
        "Thought Process": {
          "1. Preliminary Action": "Predicted trajectory based on current state. Offshore-Cruising: Maintain a steady cruise at 6.33 kn and a safe distance from the coastline, mitigating nearshore currents and grounding risks while maintaining track to the designated waypoint.",
          "2. Speed Considerations": {
            "Speed Limit": "Safe",
            "Current Speed": 6.33,
            "Limit": 25.0,
            "Speed Unit": "knot",
            "Recommendation": "Continue monitoring speed and proceed with caution."
          },
          "3. Final Action": "Uniform to Follow the predicted."
        },
        "Trajectory": [[-935.37, 270.28], [-1849.93, 570.45], [-2764.46, 870.7], [ -3669.5, 909.61], [-4572.15, 883.26], [-5474.78, 857.0]]
    }
}
```


**Figure 5.** Example of CoT prompts in the fine-tuning stage of STPredictor.

### A 1.3.3. Prediction Format

The output part aims to unify the trajectory prediction with an interpretable expression, forming a composite output structure consisting of Thoughts and Trajectory. The goal is not only to provide a quantitative prediction of future trajectories but also to construct a Semantic Decision Space, where the method can understand and reference through navigational scenarios while predicting trajectories. The core idea is to couple the CoT process from Appendix A 1.3.1, using structured semantics to depict decision-making logic under predefined rules and environmental constraints, while representing executable predicted trajectories as discrete future position sequences. This extends the trajectory prediction task from a traditional numerical fitting

problem to a cognitive-intent reasoning task with semantic interpretation. The output part is mainly composed of the thought process and the predicted trajectory.

1) Thoughts Module: This module records the intermediate reasoning process and decision points of the method in a structured language format. It consists of a three-part structure that outlines the method's decision-generation logic: Preliminary Action, Speed Considerations, and Final Decision.

- Preliminary Action: It uses discrete strategy categories to abstract the primary navigational intent in the current situation. These categories are defined as shown in Table 3.
- Speed Considerations: It provides structured fields for speed compliance evaluation and recommendations. It acts as a safety filter layer between the preliminary decision and the predicted trajectory.

• Final Decision: It summarizes the execution advice in the form of an instruction, which helps condense the strategy and speed evaluation into executable action instructions.

This structured design offers dual advantages: On one hand, it provides a natural interface for method interpretability, allowing direct observation of the method's reasoning and decision logic from the output, thus improving the credibility of predictions. On the other hand, this module acts as a supervision signal during model training, constraining the method's output consistency at the semantic level. This facilitates consistent learning in scenario understanding, risk assessment, and strategy selection.

2) Trajectory module: This module handles the core quantitative prediction task, providing the future predicted trajectory in the form of discrete future sampling points. This module uses the local coordinate system as a reference. Each predicted point is represented as a 2D pose vector (x, y), reflecting spatial continuity and directional stability. This time-series form aligns with the ship's dynamic constraints and turning inertia, making it

easier for the method to capture dynamic patterns of trajectory evolution during training, using attention mechanisms.

The output section constructs a system that balances interpretability and learnability through the thought process, which enables the model to learn underlying kinematic laws while simultaneously capturing high-level navigational intent and safety logic, thereby achieving a coherent fusion of semantic reasoning and motion prediction. An example of prediction format is shown in Figure 5 with Offshore-Cruising guidance scenario.

**Table 3.** Examples of Navigation Instructions in our two datasets. The position [x, y] and the range R are measured by meters in body-fixed coordinate, T is the current time marked by seconds, ship heading ψ° is marked by angle, and the units of velocity V, start velocity Vstart , end velocity Vend, current velocity Vcur, limited velocity Vlimit are knots.

| Type | Strategy | The demonstration of prompt instruction |
|---|---|---|
| Offshore | Cruising | Maintain a steady cruise at the [$V_{cur}$] and a safe distance off the coastline, mitigating nearshore currents and grounding risks while maintaining track to the designated waypoint. |
| | Speeding | Achieve an energy-efficient acceleration of the ship from [$V_{start}$ ] to [$V_{end}$ ] knots within [$T$] s. |
| | Decelerating | Stage throttle back to arrest speed from [$V_{start}$] to [$V_{end}$] knots over [$T$] s in [$R$] m, avoiding potential overshoot in following seas. |
| | Maneuvering | Execute heading change [ψ°] while sustaining [$V$] knots; keep coastline distance >= [$R$] m during the turn. |
| Nearshore | Cruising | Proceed at controlled cruise [$V_{cur}$ ] knots (<= [$V_{limit}$ ] knots); align with fairway centerline and keep Cross-Track Error <= [$R$] m. |
| | Decelerating | Initiate early reduction to [$V_{start}$] knots at distance to breakwater [R] m; keep shoreline separation >= [$R$] m. |
| | Maneuvering | Complete the turn with minimal leeway; preserve coastline separation >= [$R$] m and monitor a safety under-keel clearance. |
| | Speeding | Increase speed smoothly toward [$V_{end}$] knots from the current speed in anticipation of a maneuver at waypoint [$x$, $y$]. |



**References**

Lekkas, A.M., Fossen, T.I., 2014. Integral LOS Path Following for Curved Paths Based on a Monotone Cubic Hermite Spline Parametrization. *IEEE Trans Control Syst Technol* 22, 2287–2301. https://doi.org/10.1109/TCST.2014.2306774

Teng, S., Wang, J., Yan, R., Peng, M., Zhu, M., Chen, L., 2025. Ampilot: An Open Benchmark for Autonomous Transportation with Large Language Models in Open-Pit Mines. *SSRN Electron J*. https://doi.org/10.2139/ssrn.5114374

Yang, D., Wu, L., Wang, S., 2021. Can we trust the AIS destination port information for bulk ships–Implications for shipping policy and practice. *Transp Res Part E Logist Transp Rev* 149, 102308. https://doi.org/10.1016/j.tre.2021.102308

Zhang, S., Zhang, J., Yang, L., Chen, F., Li, S., Gao, Z., 2024c. Physics guided deep learning-based model for short-term origin–destination demand prediction in urban rail transit systems under pandemic. *Engineering* 41, 276–296. https://doi.org/10.1016/j.eng.2024.04.020