

Extreme Attitude Prediction of Amphibious Vehicles Based on Improved Transformer Model and Extreme Loss Function

Qinghuai Zhang¹, Boru Jia¹, Zhengdao Zhu¹, Jianhua Xiang¹, Yue Liu² and Mengwei Li²

Received: 20 October 2024 / Accepted: 01 April 2025

© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

Amphibious vehicles are more prone to attitude instability compared to ships, making it crucial to develop effective methods for monitoring instability risks. However, large inclination events, which can lead to instability, occur frequently in both experimental and operational data. This infrequency causes events to be overlooked by existing prediction models, which lack the precision to accurately predict inclination attitudes in amphibious vehicles. To address this gap in predicting attitudes near extreme inclination points, this study introduces a novel loss function, termed generalized extreme value loss. Subsequently, a deep learning model for improved waterborne attitude prediction, termed iInformer, was developed using a Transformer-based approach. During the embedding phase, a text prototype is created based on the vehicle's operation log data is constructed to help the model better understand the vehicle's operating environment. Data segmentation techniques are used to highlight local data variation features. Furthermore, to mitigate issues related to poor convergence and slow training speeds caused by the extreme value loss function, a teacher forcing mechanism is integrated into the model, enhancing its convergence capabilities. Experimental results validate the effectiveness of the proposed method, demonstrating its ability to handle data imbalance challenges. Specifically, the model achieves over a 60% improvement in root mean square error under extreme value conditions, with significant improvements observed across additional metrics.

Keywords Amphibious vehicle; Attitude prediction; Extreme value loss function; Enhanced transformer architecture; External information embedding

1 Introduction

Amphibious vehicles, which combine the functionality of cars and boats, possess significant application value owing

Article Highlights

- An analysis of the imbalanced waterborne attitude data in amphibious vehicles was conducted. From the perspective of extreme value theory, the limitations of the traditional mean squared error (MSE) loss function in handling extreme-value data were investigated, revealing its tendency to cause underfitting. To address this, a generalized extreme value loss (GEVL) function was introduced to enhance the model's focus on extreme-value regions.
- An attitude prediction model named iInformer for amphibious vehicles was proposed, incorporating a sparse attention mechanism. During the embedding phase, an information matrix was constructed from textual logs of the vehicle's waterborne operations, thereby improving the model's comprehension of attitude dynamics.
- The effectiveness of the GEVL loss function and the iInformer attitude prediction model was validated through left-turn maneuvering experiments performed on amphibious vehicles in aquatic environments.

✉ Jianhua Xiang
xiangjh@bit.edu.cn

¹ School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

² China North Vehicle Research Institute, Beijing 100072, China

to their ability to traverse obstacles such as rivers, lakes, and seas. However, these vehicles face inherent challenges, including suboptimal streamlined designs and uneven buoyancy distribution in water. This makes their water stability inferior to that of traditional boats, making them more prone to instability risks (Xu et al., 2005). Neural networks offer distinct advantages such as fast response times, strong adaptability to nonstationary and nonlinear data, and quick processing capabilities (Hou and Xia, 2024; Wang et al., 2017). Utilizing a neural network prediction model to forecast an amphibious vehicle's attitude in water in real time, compensatory adjustments can be made in advance. This serves as an effective method to maintain stability and ensure the safety of amphibious vehicles during aquatic operations.

Amphibious vehicle motion data is recorded using sensors installed on the vehicle body, with posture prediction classified as a time series forecasting problem. Recently, the Transformer model has demonstrated exceptional success in fields such as natural language processing and image classification (Vaswani et al., 2017). Their attention mechanism captures relationships between different inputs, making them highly effective in predicting long sequences with multiple inputs (Harrou et al., 2024a; Harrou et al., 2024b; Hittawe et al., 2024). Transformer variants such as PatchTST and Informer have further enhanced the model capabilities in long-sequence time series applications. PatchTST

improves the ability to capture both local and long-term dependencies by dividing time series data into several patches, which serve as input tokens. Informer, however, utilizes a sparse attention mechanism, which reduces model complexity while enhancing its training and prediction efficiency (Zhou et al., 2021).

Despite these advancements, Transformer models face significant challenges when applied to imbalanced datasets (Ding et al., 2019). On one hand, during water-based experiments or normal operations, it is nearly impossible for amphibious vehicles to experience severe swaying or capsizing. Typically, these vehicles run smoothly on the water, with roll and pitch angles remaining within a relatively safe range. For instance, when assessing the roll angle, data within the safe range can be regarded as normal, while data outside this range is considered extreme. According to related literature and experimental datasets, over 95% of the data falls into the normal category, while less than 5% is classified as extreme (Chen et al., 2021; Ma and Chang, 2015; Yin et al., 2018). When using the Transformer architecture to imbalanced datasets, the model often becomes biased toward the majority class, losing its ability to capture extreme values. Conversely, large tilt angles in amphibious vehicles on water can result from various factors such as sea conditions, driving states, and task modes. Among these, driving state information is recorded by sensors and can be quantified as model input for the model. However, factors like sea conditions and task modes are more difficult to quantify and are typically recorded as log files. Traditional transformer architectures for time series forecasting can only process numerical inputs, rendering them unable to model the information within these log files. As a result, the model's ability to identify and predict extreme events is further diminished. This data imbalance severely impacts the model's accuracy in predicting extreme attitudes of amphibious vehicles, causing problems like underfitting (as shown in Figure 1). From a safety standpoint, accurate predictions during extreme conditions are far more critical than those under normal circumstances. Addressing the issue of underfitting and improving the model's prediction accuracy in extreme scenarios is crucial for bolstering its practicality and ensuring the safety of amphibious vehicles during waterborne operations.

Current approaches to addressing data imbalance primarily include data resampling, model adjustments, ensemble learning, and loss function reconstruction (Chawla et al., 2002; Liu et al., 2008; Ross and Dollár, 2017; Sun et al., 2009). Data resampling improves the data structure by oversampling extreme values or undersampling the majority of data. However, in amphibious vehicle aquatic posture datasets, data resampling requires not only oversampling the extreme values but also oversampling the corresponding sea conditions, task modes, and other contextual information. This process is complex and may introduce noise. Model adjust-

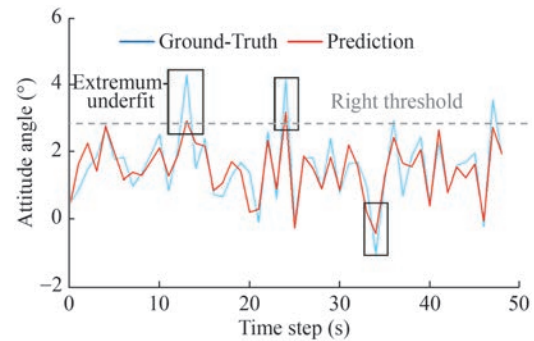


Figure 1 Extreme underfitting model

ments and ensemble learning, while effective, can significantly increase model complexity and prolong training and prediction times, which undermines real-time performance and future incremental learning requirements. Reconstructing the loss function provides an alternative approach. This method preserves the original data characteristics of the dataset without requiring modifications to the data and avoids adding model complexity. This approach is relatively easier to implement and offers higher generalizability.

With these considerations in mind, an analysis of the data imbalance issue is conducted. Inspired by previous research, the reasons for underfitting when dealing with extreme value data using the traditional mean squared error (MSE) loss function are examined from the perspective of extreme value theory. To mitigate this underfitting, the generalized extreme value loss (GEVL) function, based on the Gumbel distribution, is introduced. This loss function enhances the model's attention to extreme value data (Zhang et al., 2023). An improved Transformer-based model, named iInformer, is proposed for predicting the aquatic posture of amphibious vehicles. In the Embedding stage, text prototypes are created using amphibious vehicle waterway driving log data. These logs are further enriched with additional text data containing details about the operational environment and task modes when extreme values occur. This helps the model identify and understand the causes of extreme value occurrences. Numerical data is input as patches to highlight the local patterns of data changes. The Encoder–Decoder stage integrates a sparse attention mechanism to improve training and prediction efficiency. To address the challenges of poor convergence ability and slow convergence speeds introduced by the GEVL function, a teacher forcing mechanism is incorporated. This implementation significantly enhances the model's convergence capability.

2 Extreme value functions

2.1 Extreme events and tailed distributions

To better analyze the prediction deficiencies of time series forecasting models in extreme value scenarios, it is first

necessary to define extreme events. Given a time series of length $T: (\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) = \{(x_1, y_1), (x_2, y_2), \dots, (x_T, y_T)\}$, where x_t and y_t represent the input and output at time t respectively, we introduce the following extreme value label sequence $\mathbf{V}_{1:T} = \{v_1, \dots, v_T\}$:

$$v_t = \begin{cases} 1, & |y_t - \bar{y}| > \varepsilon \\ 0, & |y_t - \bar{y}| \leq \varepsilon \end{cases} \quad (1)$$

Here, a large constant $\varepsilon > 0$ serves as the data threshold, while \bar{y} corresponds to the mean of the sequence $\mathbf{Y}_{1:T}$. At time step t , when $v_t = 0$, output y_t represents a normal event, whereas $v_t = 1$ indicates an extreme event.

Extreme value theory is introduced for further analysis. Extreme value theory is a modeling technique that studies the distribution and characteristics of extreme values in random processes, with a specific focus on heavy-tail phenomena. According to the definition of extreme value theory, in a cumulative distribution function, if the distribution of a random variable X satisfies $\lim_{x \rightarrow \infty} e^{\lambda x} P_r[X > x] = \infty$ for any λ , then X is classified as having a heavy-tail distribution. Otherwise, it is classified as a light-tail distribution (Zhang et al., 2023). Intuitively, as shown in Figure 2, when the probability of tail data occurrence exceeds that of an exponential distribution, the distribution is considered heavy-tail. Conversely, if it falls below the exponential distribution, it is classified as light-tail (Fan et al., 2014). Examples of commonly used light-tail distributions include the Gaussian and Poisson distributions, while heavy-tail distributions include the Gumbel and Weibull distributions.

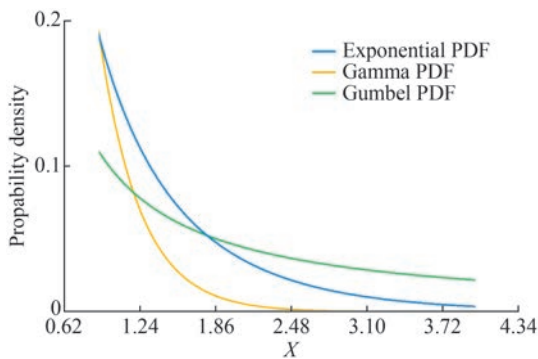


Figure 2 Tail probability density function

2.2 Relationship between gaussian kernel function and square loss

Kernel density estimation (KDE) is a nonparametric statistical method used to estimate an unknown probability density function. The basic idea is to calculate a weighted average of the probability density function within a certain range around each observation, forming the estimated probability density function (Silverman, 1986). The kernel func-

tion serves as the weighting function in this process, with the Gaussian kernel function being one of the most commonly used options (Rosenblatt, 1956). Assuming that y_1, y_2, \dots, y_n are sampled observations from a Gaussian distribution P_y , the KDE with kernel K can be expressed as

$$\hat{p}_{Y, \text{Gaussian}}(y) := \frac{1}{n\tau\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{(y-y_i)^2}{2\tau^2}},$$

where the positive smoothing parameter τ is usually referred to as the bandwidth. The tail properties of KDE are largely determined by its underlying kernel. When the kernel function is heavy-tailed, then for any $t > 0$, its KDE satisfies $\int_{-\infty}^{\infty} e^{ty} K(y) dy = \infty$, and vice versa.

The square loss function is a commonly used loss function in time series prediction. For a general time series $(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T})$, the square loss of time series prediction is defined as follows:

$$L_{\text{square}}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) = \frac{1}{T} \sum_{t=1}^T \|y_t - o_t\|^2 \quad (2)$$

Here, $o_t = f_{\theta}(\mathbf{X}_{1:T})$ is the output of the model f_{θ} at time t . According to probability theory, minimizing the square loss is equivalent to maximizing the Gaussian likelihood function (Hornik et al., 1989):

$$\min_{\theta} \sum_{t=1}^T \|y_t - o_t\|^2 \Leftrightarrow \max_{\theta} \prod_{t=1}^T N(y_t | o_t, \tau^2) \quad (3)$$

Here, τ is the Gaussian variance, while θ represents the parameter of the prediction model. According to the Bayesian theory, the maximum likelihood function can be written as follows:

$$\max_{\theta} \prod_{t=1}^T N(y_t | o_t, \sigma^2) \Leftrightarrow \max_{\theta} p_{Y|X}(\mathbf{Y}_{1:T} | \mathbf{X}_{1:T}, \theta) \quad (4)$$

$$p_{Y|X}(\mathbf{Y}_{1:T} | \mathbf{X}_{1:T}) = \frac{p_{X|Y}(\mathbf{X}_{1:T} | \mathbf{Y}_{1:T}) p_Y(\mathbf{Y}_{1:T})}{p_X(\mathbf{X}_{1:T})} \quad (5)$$

Assuming that the model has sufficient learning capacity, $p_{Y|X}$ approaches the optimal solution (Arjovsky and Bottou, 2017). Consequently, $p_Y, p_X, p_{X|Y}$ also approach their respective optimal solutions. Therefore, it is reasonable to assume that the optimized empirical distribution p_Y has the same form as the Gaussian kernel density estimate based on y_1, y_2, \dots, y_n :

$$\hat{p}_Y(y) = \frac{1}{T\tau} \sum_{i=1}^T K_{\text{Gaussian}}\left(\frac{y-y_i}{\tau}\right) \quad (6)$$

Here, the bandwidth τ is an unknown parameter.

As previously discussed, once a model achieves the minimum square loss, its predicted marginal probability density,

p_γ , closely approximates a Gaussian KDE. The kernel function of the Gaussian KDE, which is based on the Gaussian distribution function, belongs to the light-tail distribution category. According to extreme value theory, the Gaussian distribution effectively predicts the central portion of the data distribution, representing typical and moderate variations. However, it offers limited insight into fluctuations caused by extreme events. The mismatch between the light-tail Gaussian KDE and the heavy-tail nature of real-world distributions can lead to underfitting the predictive model when addressing extreme events.

2.3 Attitude angle data distribution and extreme value loss function

According to extreme value theory, real-world events typically follow a heavy-tail distribution (Haan and Ferreira, 2006). This paper analyzes experimental data on the attitude angles of amphibious vehicles during waterborne operations. Data from 20 experiments under level one and level two wave conditions were selected for statistical analysis, focusing on the vehicle’s roll and pitch angles. The data were sampled at a rate of one reading per second, resulting in a total of 20 000 data points. Figure 3 presents the histogram illustrating the distribution of the vehicle’s waterborne attitude angles.

Extreme values are identified as the larger absolute values of the attitude angles rather than being defined from a strictly physical perspective. It is evident that the distributions of both the roll and pitch angles of the vehicle exhibit heavy-tail characteristics, with a significant proportion of larger absolute values. Among these datasets, the roll angle data exhibits more pronounced fat-tailed characteristics, while the pitch angle data demonstrates more distinct long-tailed behavior. This confirms that the waterborne attitude angles of the amphibious vehicles follow a heavy-tail distribution. The mismatch between the light-tail KDE and the actual heavy-tail distribution data leads to underfitting near the extremes. To address this issue, an extreme value loss function based on the Gumbel KDE is introduced. According to research by Zhang et al. (2023) the Gumbel kernel takes the following form:

$$K_{\text{Gumbel}}(u) = e^{-(1 - e^{-u^\gamma})^\gamma} u^2 \tag{7}$$

Here, $\gamma > 0$ is a hyperparameter used to control the thickness of the Gumbel kernel at the tail. Correspondingly, the Gumbel loss function has the following form:

$$L_{\text{Gumbel}}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) = \frac{1}{T} \sum_{t=1}^T (1 - e^{-\delta_t^\gamma})^\gamma \delta_t^2 \tag{8}$$

Here, $\delta_t = y_t - o_t$, while parameter γ ranges from $(0, +\infty)$.

Utilizing the Gumbel loss function for model learning involves estimating the distribution of p_γ using heavy-

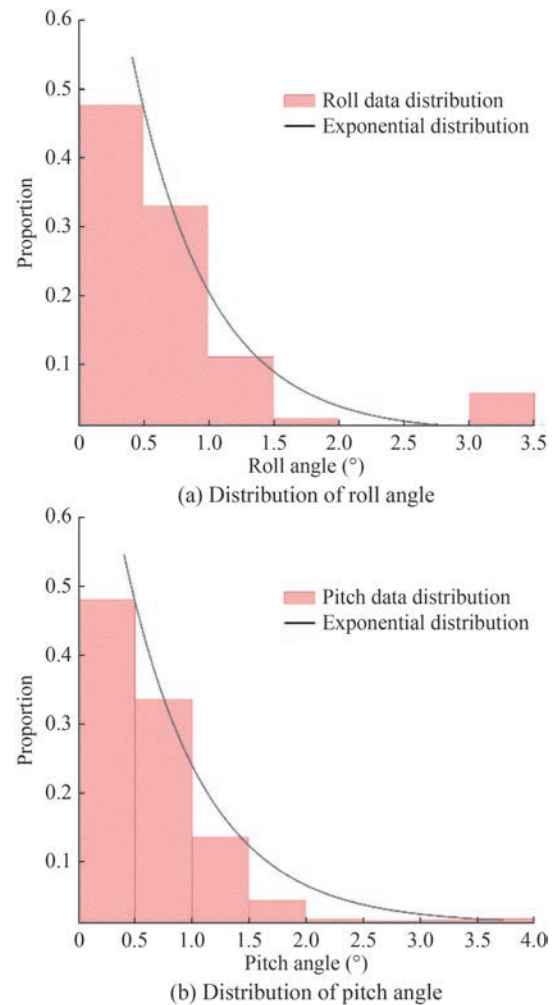


Figure 3 Distribution of attitude angles

tailed KDE, thereby ensuring that the model does not disregard extreme data points.

3 Deep learning network design

The framework of iInformer, an improved prediction model for amphibious vehicle waterborne attitudes based on the Transformer architecture, is depicted in Figure 4. iInformer retains the unique sparse attention mechanism of Informer to enhance computational efficiency and generalization performance. The Encoder architecture closely mirrors that of the Informer.

In the Decoder and Embedding layers, specific improvements are made tailored to the practical requirements of amphibious vehicle waterborne attitude prediction. The two main improvements are as follows:

- 1) In the embedding phase, a text prototype based on the operational logs of amphibious vehicles is constructed. This provides the model with relevant information about waterborne operating conditions.

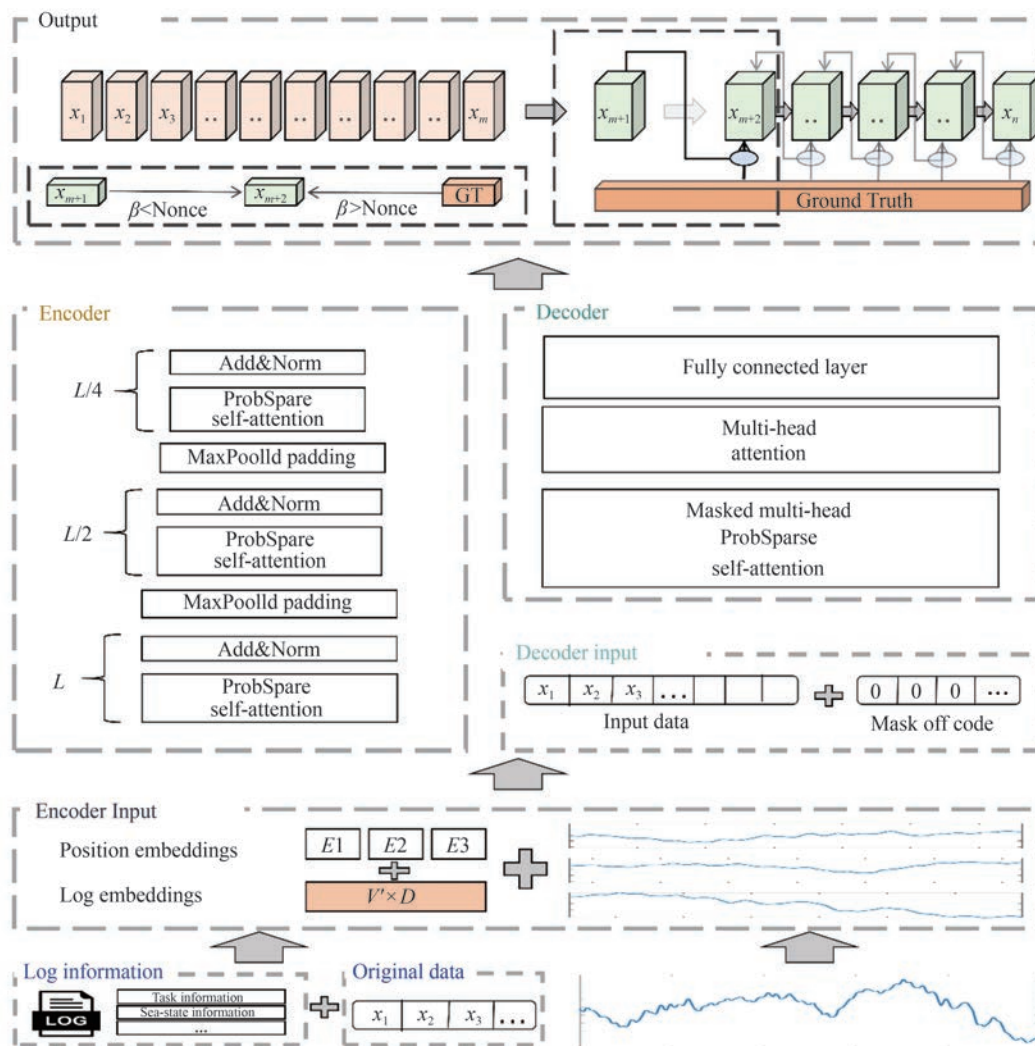


Figure 4 iInformer model overview

2) Segmented data is used as model input to enhance the model’s ability to capture local data characteristics. Furthermore, a teacher forcing mechanism is introduced to address potential convergence issues caused by incorporating GEVL.

These enhancements aim to optimize the iInformer model, enabling it to deliver more accurate and efficient predictions of amphibious vehicle waterborne attitudes.

3.1 Data processing

The amphibious vehicle posture prediction model is fundamentally data-driven, making data processing a key step to ensure that the model operates accurately. The goal of this process is to prepare appropriate input data for the model. Similarly, a time series with a length of T ($X_{1:T}, Y_{1:T}$) is analyzed, where X_t represents the input of the model at time t , X_t contains n -dimensional input data, expressed as $X_t^{1:n}$. Sensor-collected data is first filtered and subjected to noise reduction to ensure its reliability.

Time series data exhibits distinct local features. Segment-

ing the data allows the model to better capture local characteristics (such as periodic patterns and trends), thereby improving its generalization capabilities (Nie et al., 2022; Zhou et al., 2022). Breaking down long sequences into smaller patches also reduces the data processed at any given moment, lowering computational complexity and memory requirements. For a time series with an input length of t X_t , the length of each patch is L_p . The total number of patches within t is calculated as $P = \left\lfloor \frac{t - L_p}{S} \right\rfloor + 1$, where S is the sliding stride, and each patch data is expressed as X_{patch} .

3.2 Embedding architecture

The iInformer model’s embedding process does not include time-feature encoding. While amphibious vehicle data from waterborne operations is inherently temporal, its primary characteristics stem from sequential changes rather than periodic or cyclical time patterns. Vehicle movements exhibit randomness without a repeating temporal cycle,

rendering time-feature encoding unsuitable for this attitude prediction model; in fact, it would introduce extraneous information. Instead, operational factors such as the specific task being executed and the hydrological environment play a more significant role in influencing the vehicle’s waterborne attitude changes. Therefore, a textual prototype of operational conditions, including waterborne missions, sea-state conditions, and driving modes, is incorporated to enhance the model’s predictive accuracy.

This textual prototype is constructed using operational logs from the vehicle’s waterborne activities. These logs provide information on driving modes (e.g., autonomous or manual), wave conditions, and mission tasks. First, the log files are cleaned by removing punctuation, tokenizing the text, and deleting common stop words, leaving only the most meaningful textual data.

The clean textual data is then transformed into vector representations using pretrained word embeddings. These embeddings generally contain a large number of word vectors, while the vocabulary in amphibious vehicle operational logs is smaller and more fixed. To ensure model efficiency and enable real-time predictions, a linear transformation is applied to the pretrained embeddings. For the trained word representation matrix $E \in R^{V \times D}$, where V represents the vocabulary size, and D is the embedding dimension, a linear transformation matrix, $W \in R^{V' \times V}$, is constructed. Here, V' represents the target number, with $V' \ll V$. The pretrained word embedding E is converted into a small number of word embeddings through a linear change matrix $E' = WE$, thereby greatly reducing the number of word vectors. Using word embedding, the word vector of the amphibious vehicle log data, $X_{m,t}$, is generated. In addition to the log text encoding, the model contains X_{patch} , representing the positional encoding of each $X_{p,t}$. The position encoding adopts the classic Transformer position encoding method, and X_p , the corresponding encoding information is finally input $PE = \text{Concat}(X_{p,t}, X_{m,t})$.

3.3 Encoder and decoder

The iInformer architecture employs an encoder–decoder structure, incorporating a multihead sparse attention mechanism instead of the standard attention mechanism used in Transformers. This design optimizes the model’s generalization ability while significantly improving computational efficiency.

1) ProbSparse attention: The traditional Transformer employs a full attention mechanism that calculates attention scores between every **Query** and **Key**, significantly increasing the complexity of model training (Vaswani et al., 2017). By contrast, iInformer replaces this approach with a sparse attention mechanism (Nie et al., 2022; Zhou et al., 2021). Rather than calculating the weighted sum of attention scores for all inputs, iInformer first determines a spar-

sity score. This score filters out the less relevant **Query**, enabling the model to focus on the critical information within the data. This selective process enhances the model’s efficiency by concentrating attention on the most significant inputs.

The formula for sparse attention obtained through the sparse self-attention mechanism is:

$$A(Q, K, V) = \text{softmax}\left(\frac{\overline{Q}K^T}{\sqrt{d}}\right)V \tag{9}$$

Here, $A(Q, K, V)$ is the sparse attention weighted sum matrix, \overline{Q} is the query vector matrix Q after sparse screening, and V is the value vector matrix.

As shown in Figure 4, the encoder of the prediction model consists of three layers of attention mechanisms. Each attention mechanism includes a sparse self-attention layer, a residual connection, and layer normalization. These layers are interconnected through pooling layers. To address issues associated with long-sequence inputs, residual connections and layer normalization play a crucial role. They effectively mitigate gradient vanishing, accelerate model convergence, and enhance overall stability and generalization capability (Shafiq and Gu, 2022).

2) Input of encoder and decoder: Encoder takes the patch and its information encoding as input, which can be expressed as follows:

$$X_{\text{encoder}} = \text{Concat}(PE, X_{\text{patch_embed}}) \in R^{(L_{\text{token}} + L_s) \times d}$$

Here, PE is the encoding information, $X_{\text{patch_embed}}$ denotes the embedding layer representation of each X_{patch} , which is generated by the linear change X_{patch} of the embedded change matrix $W_{\text{embed}} \in R^{L_p \times d_m}$ after block division $X_{\text{patch}} \in R^{L_p \times L_p}$:

$$X_{\text{patch_embed}} = X_{\text{patch}}W_{\text{embed}} \tag{10}$$

The input of the decoder is:

$$X_{\text{decoder}} = \text{Concat}(PE, X_{\text{token}}, X_0) \in R^{(L_{\text{token}} + L_y) \times d} \tag{11}$$

Here, $X_{\text{token}} \in R^{(L_{\text{token}}) \times d}$ represents the real data used to guide the model’s prediction, L_{token} denotes the length of the prior data. Meanwhile, $X_0 \in R^{(L_y) \times d}$ serves as a placeholder for the target sequence, initialized with scalar values of 0. Here, L_y represents the length of the target sequence.

3) Teacher Enforcing: In Section 2.3, the GEVL was introduced for model training. However, as shown in Eq. (7) and Eq. (8), the penalization applied by Gumbel KDE to the model’s prediction results depends on the chosen parameters. To correct prediction errors in the extreme value regions, these parameter values must be increased within a certain range. However, excessively high values can cause the model to overreact to large errors early in the training process, negatively affecting convergence speed and stability.

Poor parameter selection can significantly hinder the model's convergence capability and lead to substantial oscillations during training under the GEVL loss function. Additionally, compared to the MSE loss function, the GEVL loss function slows down the convergence rate. To enhance convergence, a teacher forcing mechanism is implemented into the decoder.

During the generation process, this mechanism introduces a scheduled sampling rate parameter β . When β is less than a randomly generated nonce, the model uses its previous prediction as input for predicting the next time step. Conversely, when β is greater than the nonce, the ground truth value from the previous time step is used as input. By providing correct feedback during training, the teacher forcing mechanism helps establish a stable gradient flow early in training. This approach mitigates error propagation caused by the model's own prediction errors and minimizes the likelihood of cumulative prediction errors within the sequence. As a result, the training process becomes smoother, and the convergence speed improves (Tashkova et al., 2012). It is worth noting that teacher forcing can introduce exposure bias (Bengio et al., 2015). To address this issue, parameter β is linearly decayed throughout the training process. By the time the training reaches 75% of the total epochs, β decays to 0, and the model ceases using ground true values as inputs. This adaptive adjustment allows the model to correct its predictions using teacher forcing, thereby avoiding significant errors and improving both convergence capability and speed.

4 Attitude prediction experiments

This chapter validates the accuracy of the amphibious vehicle's water attitude prediction model using test data obtained from amphibious vehicle trials. The prediction focuses specifically on the vehicle's pitch and roll angles during left turns on the water.

4.1 Overview of evaluation

1) Dataset: The data used in this study is sourced from marine test data of a China-produced amphibious vehicle. The principal design parameters of the vehicle are detailed in Table 1.

During the experiment, the amphibious vehicle performed a left turn in level-one sea conditions, a scenario that induced significant changes in its attitude. Sensor data was collected throughout the experiment, capturing various aspects of the vehicle status, such as roll and pitch angles relevant to the prediction model, as well as additional parameters influencing changes in the vehicle's water attitude. Sensors sampled data at a rate of one reading per second over a 1 000-second period, resulting in 1 000 data points for each sensor parameter.

Table 1 Vehicle primary parameters

Descriptions	Value
Vehicle weight (t)	28.5
Center of mass height (m)	0.85
Draft (m)	1.283
Waterline length (m)	8.293
Head angle (°)	0.738
Half vehicle mass moment of inertia (kg·m ²)	(23 347, 72 105, 72 105)
Vehicle mass moment of inertia (kg·m ²)	(46694, 144210, 144210)

The motion attitude of the vehicle is depicted in Figure 5, illustrating its initial movement in a straight line, progression into the left turn, and eventual stop after completing the amphibious maneuver. Changes in the roll and pitch angles of the amphibious vehicle are further illustrated in Figure 6.

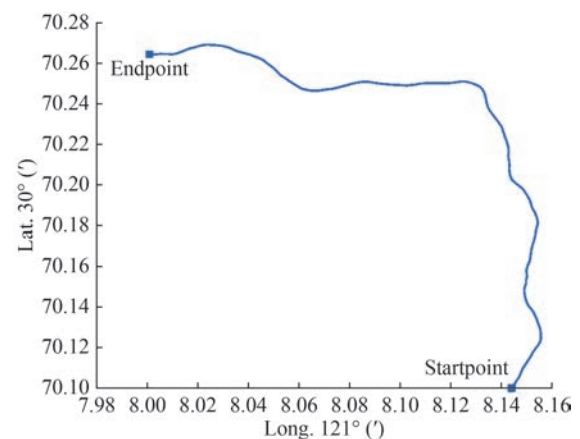


Figure 5 Vehicle trajectory

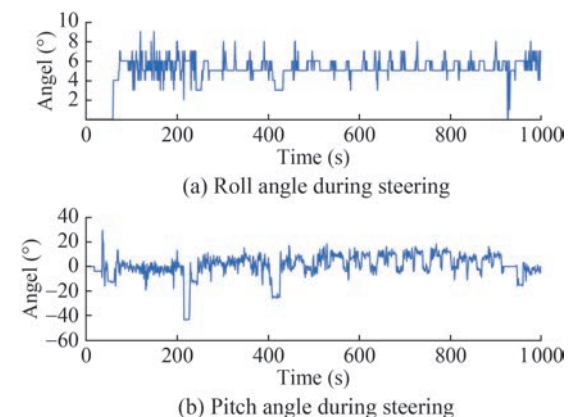


Figure 6 Ground truth data for the left turn process

2) Evaluation indicators: To assess the model performance, metrics such as the root MSE, symmetric mean absolute percentage error (SMAPE), and R^2 coefficient are employed. These metrics offer comprehensive insights into the model's accuracy from various perspectives (Xue et al.,

2020). RMSE, SMAPE, and R^2 are defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \tag{12}$$

$$SMAPE = \frac{100\%}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{14}$$

Here, y_i , \bar{y} , and \hat{y} represent the true value, data mean, and model estimated value, respectively. The value range of SMAPE lies between 0% to 200%. Lower RMSE and SMAPE values indicate higher prediction accuracy. Meanwhile, the R^2 coefficient ranges from 0 to 1, with higher values indicating better prediction performance.

3) **Comparative Analysis:** A comparative analysis is conducted using the combined model iInformer-GEVL, which integrates the improved Informer architecture with the extreme value loss function, to predict the amphibious vehicle attitude on water. This model is contrasted with several typical time series prediction models, including Informer, LSTM, and their variations with the commonly used MSE loss function. Specifically, the models are classified as iInformer-MSE, Informer-GEVL, Informer-MSE, LSTM-GEVL, and LSTM-MSE. These models are evaluated for their accuracy in predicting both the vehicle’s roll and pitch angles during the left-turning maneuver of an amphibious vehicle, effectively validating their predictive performance.

4.2 Effectiveness of the iInformer model

The model’s effectiveness is validated by its performance on the amphibious vehicle’s waterborne roll angle and pitch angle datasets. Figures 7 and 8 present the prediction results of various comparative models for these datasets.

1) **Model Accuracy:** Table 2 compares the prediction accuracy of various models on the amphibious vehicle’s roll and pitch angle datasets. The results indicate that iInformer demonstrates strong performance on both datasets. Compared to the Informer network, iInformer achieves improved results across all metrics by using the same loss function. For the roll angle dataset, the maximum gains in RMSE, SMAPE, and R^2 are 128.96%, 158.61%, and 11.09%, respectively. For the pitch angle dataset, the maximum gains are 47.18%, 64.02%, and 0.57%. Moreover, iInformer exhibits even greater improvements over the traditional LSTM model, with maximum increases of 243.41%, 133.49%, and 1.86% in RMSE, SMAPE, and R^2 across all datasets.

A paired t -test is conducted on the prediction errors of

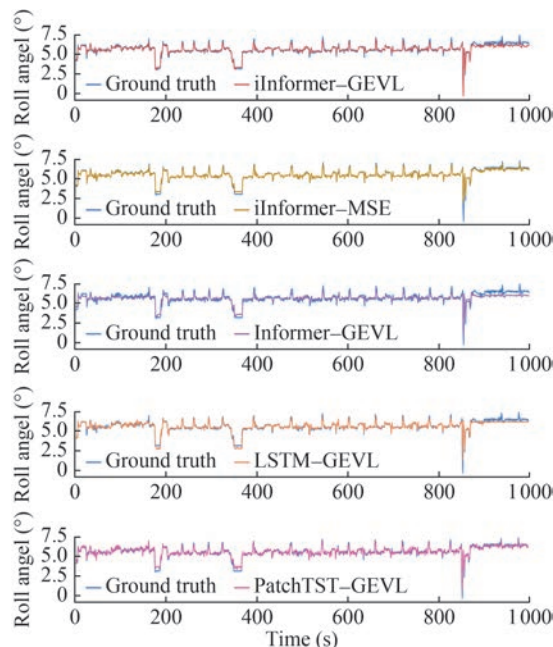


Figure 7 Prediction results of roll angles from the models

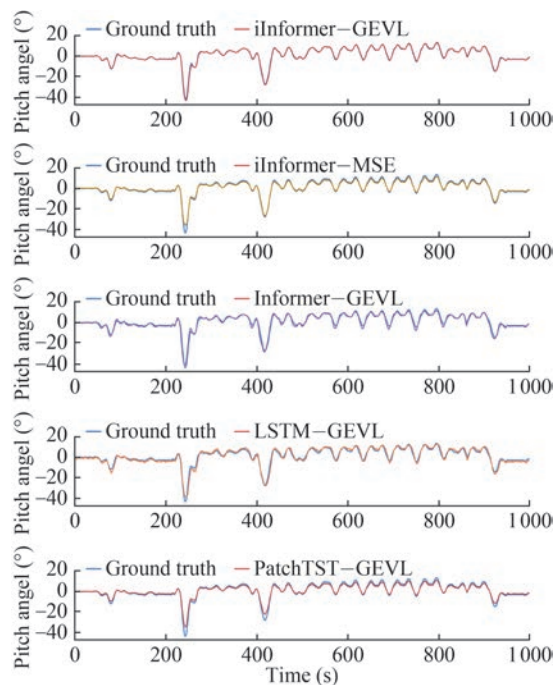


Figure 8 Prediction results of pitch angles from the models

iInformer-GEVL and Informer-GEVL. For the roll angle dataset, the significance level of error changes before and after the improvement is below 0.05, and the same applies to the pitch angle dataset. This confirms, with 95% confidence, that the error reductions are statistically significant, indicating enhanced model accuracy. Furthermore, a box plot analysis of iInformer-GEVL’s prediction errors for both datasets shows that most errors fall within 0.5° , demonstrating a good overall model fit, as shown in Figure 9.

Table 2 The forecast accuracy for the overall dataset is estimated using RMSE, SMAPE, and R^2 metrics. The best score per column under the same conditions is highlighted in bold

Models	Roll angle data			Pitch angle data		
	RMSE	SMAPE	R^2	RMSE	SMAPE	R^2
iInformer-GEVL	0.1941	0.0262	0.9698	0.3720	0.1251	0.9946
iInformer-MSE	0.1336	0.0302	0.9086	0.4945	0.1740	0.9962
Informer-GEVL	0.3251	0.0535	0.8730	0.5475	0.1066	0.9889
Informer-MSE	0.3059	0.0781	0.8562	0.5639	0.2854	0.9893
LSTM-GEVL	0.3821	0.0488	0.6594	0.6724	0.2921	0.9818
LSTM-MSE	0.4588	0.0556	0.3176	0.7934	0.2065	0.9780
PatchTST-GEVL	0.2416	0.0500	0.8825	0.5323	0.2318	0.9901
PatchTST-MSE	0.2379	0.0634	0.8817	0.5478	0.1697	0.9899

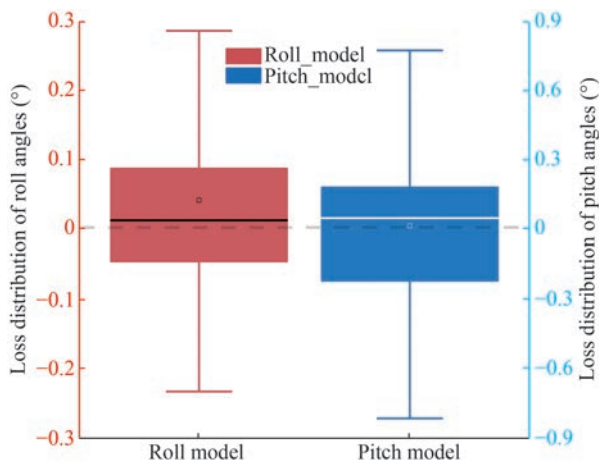
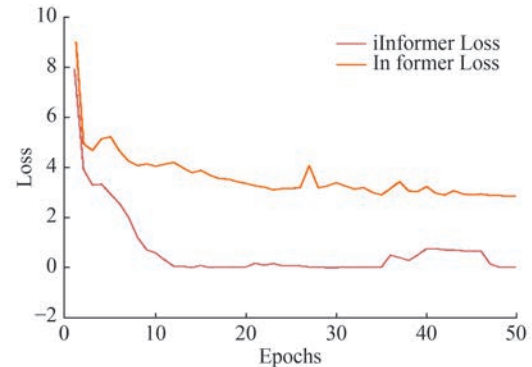


Figure 9 Box plot of model prediction losses

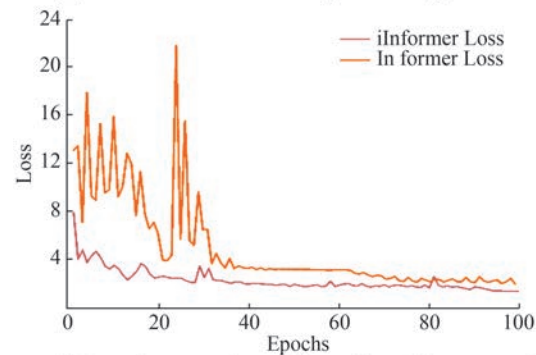
2) Model convergence ability: The introduction of GEVL poses challenges to model convergence. To better fit the extreme portions of the data distribution and increase the penalty for large errors, it is necessary to appropriately raise the parameter values in GEVL. However, higher penalties for large errors can cause the model to overreact during early training, potentially leading to issues such as gradient explosion. Furthermore, larger parameter values can negatively impact the model’s convergence speed and stability. To enhance convergence capability, we propose the iInformer architecture, which employs a teacher forcing mechanism to guide convergence and fitting during the early training stages. This approach effectively prevents gradient explosion and convergence failure.

As shown in Figure 10, the enhanced iInformer model exhibits significantly improved convergence capabilities. In Figure 10(a), the model uses a larger teacher forcing coefficient early training, enabling it to receive accurate prediction feedback, rapidly reduce losses, and facilitate swift convergence and fitting. As training progresses, the teacher forcing coefficient decays to 0, mitigating the risk of exposure bias. Additionally, the stepwise output approach

enhances the model’s fitting ability, helping it locate better local minima. Moreover, Figure 10(b) illustrates that excessively large values can destabilize the model and hinder convergence. The teacher forcing mechanism in the iInformer model significantly enhances stability, particularly in scenarios with uneven error distributions.



(a) Loss decrease chart with epochs=50, gamma=3



(b) Loss decrease chart with epochs=100, gamma=5

Figure 10 iInformer and Informer model convergence statuses

3) Long-sequence prediction capability: The iInformer model effectively retains Informer’s excellent performance in long-sequence predictions, as depicted in Figure 11.

When predicting sequences spanning tens to hundreds of steps, LSTM models experience significant accuracy degradation, whereas iInformer and Informer models main-

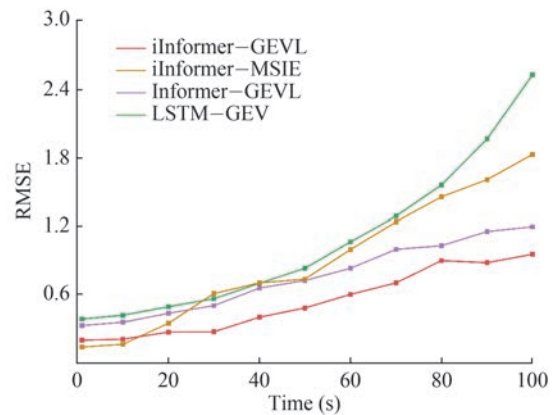


Figure 11 RMSE for different prediction models

tain relatively stable accuracy decay. This capability is critical for forecasting attitude changes over longer periods, ensuring the safety of amphibious vehicles on water.

4.3 Effectiveness of GEVL function

The effectiveness of the extreme value loss function GEVL is analyzed. Table 2 compares the prediction accuracy of different models on the roll angle and pitch angle datasets of an amphibious vehicle. For the pitch angle data, models utilizing the GEVL function generally outperform those using the MSE function under similar conditions. The maximum improvements in RMSE, SMAPE, and R^2 are 32.93%, 94.67%, and 0.38%, respectively. However, for the roll angle data, the benefits of the GEVL function are less apparent, with the MSE outperforming GEVL in some metrics.

The varying performance of the GEVL function is hypothesized to stem from differences in the data distributions of the two datasets. In the roll angle dataset, most data points cluster near the mean, resulting in smaller deviations. According to Eq. (1), this dataset predominantly consists of normal events ($v_i = 0$), with only a few extreme value events ($v_i = 1$). Conversely, the pitch angle dataset exhibits a higher frequency of extreme value events ($v_i = 1$), making the GEVL function more effective for fitting this data.

To validate this hypothesis, the distance of each data point from the mean in both datasets is calculated. Data points with greater distances are extracted to form an extreme value dataset ($v_i = 1$), comprising 15% of the total data. The model's accuracy on this extreme value dataset is then evaluated, with the results presented in Table 3.

Table 3 Estimation of forecast accuracy for the extreme datasets ($v_i = 1$) is estimated using RMSE, SMAPE, and R^2 metrics. The best scores per column under the same conditions are highlighted in bold

Models	Roll angle data			Pitch angle data		
	RMSE	SMAPE	R^2	RMSE	SMAPE	R^2
iInformer-GEVL	0.2439	0.0896	0.9642	0.4798	0.0356	0.9980
iInformer-MSE	0.4016	0.1107	0.8403	0.7424	0.0643	0.9908
Informer-GEVL	0.4040	0.1031	0.8416	0.6663	0.0510	0.9959
Informer-MSE	0.5586	0.1244	0.5713	1.1140	0.0654	0.9900
LSTM-GEVL	0.6622	0.1017	0.8081	0.5589	0.0323	0.9985
LSTM-MSE	0.9481	0.1717	0.1589	1.2587	0.0841	0.9883
PatchTST-GEVL	0.4880	0.1600	0.8825	0.6856	0.0779	0.9898
PatchTST-MSE	0.5962	0.1934	0.8317	0.8869	0.1036	0.9890

The GEVL function significantly improves the prediction accuracy for extreme value datasets in both roll and pitch angles. Specifically, for the roll angle extreme value dataset, the improvements are notable, with RMSE increasing by up to 64.65%, SMAPE by up to 68.82%, and R^2 by up to

349.2%. Similarly, for the pitch angle extreme value dataset, the gains are even more substantial, with RMSE improving by up to 125.2%, SMAPE by up to 160.3%, and R^2 by up to 1.032%.

These experiments demonstrate that integrating the GEVL function into the prediction model for amphibious vehicle water attitude effectively enhances its ability to predict extreme values of attitude angles. This improvement holds significant importance for enhancing the safety of amphibious vehicles during water operations.

5 Conclusions

This paper focuses on establishing an appropriate waterborne attitude prediction model for amphibious vehicles to prevent instability risks. It introduces an extreme value loss function based on heavy-tailed kernel functions designed to overcome the underfitting issues near extreme data points that traditional square loss function models fail to address. Furthermore, the study proposes a novel posture prediction model built on an improved Transformer architecture. During the embedding phase, a text prototype is constructed using the amphibious vehicle's waterway driving log data, enabling the model to better interpret the vehicle's operational environment. To tackle the challenges posed by the new loss function, such as significant training errors and weak convergence, the model incorporates a teacher forcing mechanism for extreme values, ensuring improved stability and convergence during training. Experimental validation using left-turn maneuver data from the amphibious vehicle demonstrates that the extreme value loss function effectively enhances the model's accuracy in predicting extreme values, while also achieving impressive overall prediction performance. The iInformer model significantly improves convergence stability and achieves excellent prediction accuracy, particularly excelling in long-sequence predictions.

Despite the advancements of the proposed model, there is still significant room for improvement. The current dataset is limited to left-turn maneuver data collected under sea state 1 conditions, which restricts the scope of experimental validation. Future work will focus on expanding the dataset to include a wider range of operational scenarios and environmental conditions, enabling more thorough validation and further enhancing the model's accuracy and reliability. This study also introduced an extreme value loss function based solely on the Gumbel kernel. However, other heavy-tailed distributions, such as Weibull, Pareto, and Fréchet, also hold potential as kernel functions for developing alternative extreme value loss functions. These distributions may offer even greater predictive performance, and systematically analyzing them will be a key focus of future research. Beyond amphibious vehicles, the ability to forecast extreme values of navigation posture is critically

important in fields such as aerospace and maritime operations. Exploring the application of this model to these domains will be a direction for future research.

Funding Supported by the National Defense Basic Scientific Research Program of China.

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Arjovsky M, Bottou L (2017) Towards principled methods for training generative adversarial networks. *Machine Learning, ArXiv abs/1701.04862*
- Bengio S, Vinyals O, Jaitly N, Shazeer N (2015) Scheduled sampling for sequence prediction with recurrent neural networks. *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16: 321-357. <https://doi.org/10.1613/jair.953>
- Chen Q, Zheng S, Li M, Zhuo H (2021) Research on prediction error of ship rolling motion. *Ship Engineering* 43(2): 42-47
- Ding DZ, Zhang M, Pan XD, Yang M, He XN, Assoc Comp M (2019) Modeling extreme events in time series prediction. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, USA, 1114-1122*. <https://doi.org/10.1145/3292500.3330896>
- Fan JQ, Fan YY, Barut E (2014) Adaptive robust variable selection. *Annals of Statistics* 42(1): 324-351. DOI: 10.1214/13-AOS1191
- Haan L, Ferreira A (2006) *Extreme value theory: An introduction*. Springer, New York, USA
- Harrou F, Dairi A, Dorbane A, Sun Y (2024a) Enhancing wind power prediction with self-attentive variational autoencoders: A comparative study. *Results in Engineering* 23: 13. <https://doi.org/10.1016/j.rineng.2024.102504>
- Harrou F, Zeroual A, Kadri F, Sun Y (2024b) Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. *Results in Engineering* 23: 14. <https://doi.org/10.1016/j.rineng.2024.102342>
- Hittawe MM, Harrou F, Togou MA, Sun Y, Knio O (2024) Time-series weather prediction in the Red sea using ensemble transformers. *Applied Soft Computing* 164: 16. <https://doi.org/10.1016/j.asoc.2024.111926>
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5): 359-366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Hou X, Xia S (2024) Short-term prediction of ship roll motion in waves based on convolutional neural network. *Journal of Marine Science and Engineering* 12(1): 102. <https://doi.org/10.3390/jmse12010102>
- Liu XY, Wu J, Zhou ZH (2008) Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39(2): 539-550
- Ma X, Chang L (2015) Investigation on nonlinear rolling dynamics of amphibious vehicle under wind and wave load. *Journal of Measurement Science and Instrumentation* 6(3): 275-281
- Nie Y, Nguyen NH, Sinthong P, Kalagnanam J (2022) A time series is worth 64 words: Long-term forecasting with transformers. *Machine Learning, arXiv preprint arXiv:2211.14730*
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* 27(3): 832-837. DOI: 10.1214/aoms/1177728190
- Ross TY, Dollár G (2017) Focal loss for dense object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2980-2988*
- Shafiq M, Gu ZQ (2022) Deep residual learning for image recognition: A survey. *Applied Sciences-Basel* 12(18): 43. <https://doi.org/10.3390/app12188972>
- Silverman BW (1986) Density estimation for statistics and data analysis. *J. Roy. Stat. Soc. Ser. C* 37: 120-121
- Sun Y, Wong AK, Kamel MS (2009) Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence* 23(4): 687-719
- Tashkova K, Silc J, Atanasova N, Dzeroski S (2012) Parameter estimation in a nonlinear dynamic model of an aquatic ecosystem with meta-heuristic optimization. *Ecological Modelling* 226: 36-61. <https://doi.org/10.1016/j.ecolmodel.2011.11.029>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, USA
- Wang G, Han B, Sun W (2017) Short-term prediction of ship motion based on LSTM. *Ship Science and Technology* 39(13): 69-72. (in Chinese)
- Xu G, Zhou J, Yao X, Guo Q (2005) Toss motion of amphibious vehicle in sea-way. *Acta Armamentarii* 26(4): 433-437. (in Chinese)
- Xue YF, Liu YJ, Ji C, Xue G (2020) Hydrodynamic parameter identification for ship manoeuvring mathematical models using a Bayesian approach. *Ocean Engineering* 195: 106612. <https://doi.org/10.1016/j.oceaneng.2019.106612>
- Yin JC, Perakis AN, Wang N (2018) A real-time ship roll motion prediction using wavelet transform and variable RBF network. *Ocean Engineering* 160: 10-19. <https://doi.org/10.1016/j.oceaneng.2018.04.058>
- Zhang M, Ding DZ, Pan XD, Yang M (2023) Enhancing time series predictors with generalized extreme value loss. *IEEE Transactions on Knowledge and Data Engineering* 35(2): 1473-1487. DOI: 10.1109/TKDE.2021.3108831
- Zhou HY, Zhang SH, Peng JQ, Zhang S, Li JX, Xiong H, Zhang WC (2021) Informer: Beyond efficient transformer for long sequence time-series forecasting. *Machine Learning, arXiv: 2012.07436*
- Zhou T, Ma ZQ, Wen QS, Wang X, Sun L, Jin R (2022) FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. *Proceedings of the 39th International Conference on Machine Learning (ICML)*, Baltimore, USA