

Article

A Sea State Estimation and Uncertainty Awareness Model Integrating Multi-Scale Convolution and DropKey-Transformer

Ting Cui ¹, Runze Mao ^{1,*} , Xinyu Guo ¹, Peihua Han ²  and Houxiang Zhang ² 

¹ Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212003, China; 232241803901@stu.just.edu.cn (T.C.); xinyu.guo@stu.just.edu.cn (X.G.)

² Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6025 Alesund, Norway; peihua.han@ntnu.no (P.H.); hozh@ntnu.no (H.Z.)

* Correspondence: mao.rz1993@just.edu.cn

Abstract

Accurate sea state estimation is of great significance to marine engineering safety and disaster prevention and mitigation. This paper proposes a new sea state estimation architecture integrating multi-scale 1D-CNN, transformer encoder, and the MC-DropKey mechanism. First, the front-end multi-scale 1D-CNN is used in sequence modeling to extract high-quality local features. Then, the back-end transformer encoder is used to capture global long-range temporal dependencies. In addition, the MC-DropKey mechanism is introduced. It achieves dynamic uncertainty quantification while maintaining high-precision estimation. Comprehensive comparative and ablation experiments show that the proposed model significantly outperforms the baseline models in wave height and wave direction estimation tasks, achieving the lowest MAE of 0.117 m and CAE of 5.212°. Moreover, the uncertainty intervals output by the model effectively quantify estimation confidence, reaching a 96.67% prediction interval coverage probability. The obtained results provide strong technical support for scientific scheduling and decision-making in complex marine environments.

Keywords: sea state prediction; multi-scale 1D-CNN; transformer encoder; MC-DropKey; uncertainty quantification

1. Introduction

The ocean covers 71% of the Earth's surface. It serves as the lifeline of global shipping trade. Under severe sea conditions, huge waves and strong winds pose serious threats to ship structural safety, navigation stability, and port operation efficiency [1,2]. Therefore, high-precision real-time monitoring and estimation of sea state parameters such as wave height and wave direction is a prerequisite for protecting life and property at sea. It is also the core foundation for modern intelligent ships to achieve route optimization, weather routing, and energy efficiency management [3,4].

Currently, the main methods for obtaining sea state information rely on traditional sensor networks. These include ocean buoys [5], satellite remote sensing [6], and wave radars [7]. However, they all face physical and economic bottlenecks that are hard to overcome in practical applications. It is difficult for them to achieve a balance among high precision, wide coverage, and low cost. Against this background, the ship-as-a-sensor concept, also known as the wave buoy analogy (WBA) [8,9], has quickly become a research hotspot in marine engineering and intelligent shipping. When a ship sails in wind and



Academic Editor: Angelo Rubino

Received: 5 June 2026

Revised: 27 July 2026

Accepted: 28 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

waves, its unavoidable six-degree-of-freedom motion response is essentially a dynamic filtering of the surrounding wave field. Modern ships are commonly equipped with high-precision inertial measurement units (IMUs) [10]. Using vast amounts of ship motion data to infer sea state information can not only enable data collection at a very low cost but also completely break the spatial and temporal limitations of traditional sensors.

The inversion of sea states from ship motions is a complex nonlinear problem [11]. Existing research mainly focuses on two approaches: physics-based models and data-driven methods [12]. Traditional physics-based models [13,14] are primarily based on linear hydrodynamic theory. They perform estimation by establishing a mapping relationship between the ship motion spectrum and the wave spectrum [15]. For example, Bourkaib et al. [16] proposed an adaptive Kalman filtering method. It introduces time-varying ship speed and adaptively estimates the measurement noise covariance. This method estimates wave spectra in real time from ship motion responses. Nielsen et al. [17] proposed a brute-force spectral estimation method based on ship motion measurements. It directly solves point wave spectra from ship motion data through iterative residual calculation. Such methods have clear physical mechanisms. However, under high sea conditions, the ship's violent motions trigger nonlinear hydrodynamic effects [18]. This causes the inversion accuracy to decrease significantly.

In recent years, data-driven methods [19] have gradually become mainstream due to their strong nonlinear fitting capability. For example, Cheng et al. [20] proposed a densely connected convolutional neural network called SSENET. It combines channel attention and feature attention modules to extract hierarchical features from ship motion data. However, this method is sensitive to changes in data distribution. The importance of the attention modules is also unclear. Bisinotto et al. [10] used real ship motion data and ERA5 reanalysis data. They selected an Inception neural network architecture through hyperparameter search. However, this model relies heavily on large amounts of high-quality labeled real ship data. The ERA5 labels themselves also contain uncertainty. This leads to larger estimation errors under low-frequency waves and high sea conditions. Recently, several transformer-based and hybrid attention models have been applied to sea state estimation. Selimović et al. [19] adopted an attention-based neural network to capture salient features from ship motion responses. Kwon et al. [21] employed transformer architectures for estimating directional wave parameters from floating production, storage, and offloading (FPSO) motion signals. Although the above studies have verified the application potential of attention mechanisms, existing solutions typically focus on modeling global dependency relationships. They have not yet addressed the multi-frequency characteristics of ship motions and uncertainty estimation.

Therefore, when facing complex and variable real ocean environments, existing models still have the following major problems. First, real sea states are formed by the superposition of waves of different frequencies. Wave component coupling causes ship motions to exhibit multi-scale characteristics [22]. Traditional models find it difficult to simultaneously capture high-frequency impacts and low-frequency slow rolling. Second, ship attitude is influenced by the ship's own inertia and historical motion states [23]. When processing long sequences, conventional networks are prone to information attenuation or gradient vanishing. This makes it hard to identify the most critical historical moments that contribute most to the estimation. Third, sensor data during actual voyages is easily affected by vibration, interference, or even temporary failure. Most existing models are deterministic models. They output only a single estimated value. When faced with missing information or strong noise, they tend to produce overconfident and incorrect predictions. They lack probabilistic output capability [24,25]. This makes them insufficiently reliable for navigation decision-making where safety is the top priority. Meanwhile, existing methods have not yet

addressed the above problems simultaneously within a unified framework. Most studies focus on only one or two of them. They cannot meet the requirements for reliable estimation under complex and extreme sea conditions.

Based on the above physical logic and practical engineering pain points, this paper proposes an uncertainty-aware hybrid model. The proposed model integrates multi-scale one-dimensional convolutional neural network (multi-scale 1D-CNN), transformer encoder, and Monte Carlo DropKey (MC-DropKey). Unlike existing transformer-based approaches that either focus solely on global attention or apply uncertainty quantification as an afterthought, the proposed model addresses the above problems simultaneously through a tightly coupled architecture of various components. The specific implementation of each component is as follows. To address local motion differences caused by multi-frequency waves, the model constructs a multi-scale 1D-CNN. To decouple transient impact and long-period swell characteristics, it abandons a single convolution kernel. Instead, it designs three parallel 1D convolution branches. Through channel concatenation of multi-scale feature maps, the model constructs local feature representations covering the full frequency domain.

Next, to compensate for the shortcomings of CNNs in long-sequence modeling, this paper feeds the feature maps extracted by the multi-scale convolution into a transformer encoder. With its multi-head self-attention mechanism, the model automatically performs calculations and focuses on the most critical historical moments that contribute most to the current sea state estimation. At the same time, residual connections and layer normalization are combined. This fully preserves the temporal dependencies across the global heading direction.

Finally, to break the limitation that traditional models can only output deterministic scalar values, this paper introduces the MC-DropKey mechanism into the attention computation of the transformer. This mathematically simulates the physical scenarios where sensor signals are damaged or suffer from strong noise pollution during actual voyages. Therefore, the model outputs not only the mean sea state estimation but also the epistemic uncertainty through variance. This enhances the algorithm's robustness and decision-making safety under severe sea conditions. The contributions of this study are as follows:

1. A new sea state estimation architecture integrating multi-scale 1D-CNN and transformer is proposed. It effectively solves the problems of ship multi-frequency motion coupling and long-term dependency modeling.
2. The MC-DropKey uncertainty quantification mechanism is introduced. This equips the model with probabilistic assessment capability for local sensor signal loss. It achieves a leap from deterministic scalar estimation to interval distribution prediction.
3. Each component works collaboratively to simultaneously solve three problems that were previously addressed only in isolation: multi-frequency coupling, long-term dependencies in ship motion sequences, and sensor uncertainty.

The structure of the paper is as follows: Section 2 introduces the proposed model that integrates multi-scale 1D-CNN, transformer encoder, and MC-DropKey. Section 3 describes the datasets, experimental setup, and performance metrics used to evaluate the model. Section 4 presents the results and discussion. Section 5 concludes the paper.

2. Methodology

2.1. Overview

Figure 1 presents the overall framework of the proposed model in this study for sea state estimation. This model proposes an architecture that integrates multi-scale 1D-CNN,

transformer encoder, and MC-DropKey. It is used for sea state estimation including wave height and wave direction. To address the differences in ship motion scales caused by waves of different frequencies, the model first uses multi-scale 1D-CNN to process one-dimensional ship motion time series signals such as roll, pitch, and heave. After detrending and standardization, the data enters three parallel branches with small, medium, and large convolution kernels. These branches capture high-frequency, medium-frequency, and low-frequency signals, respectively. The outputs are then concatenated in the channel dimension to form a local feature matrix.

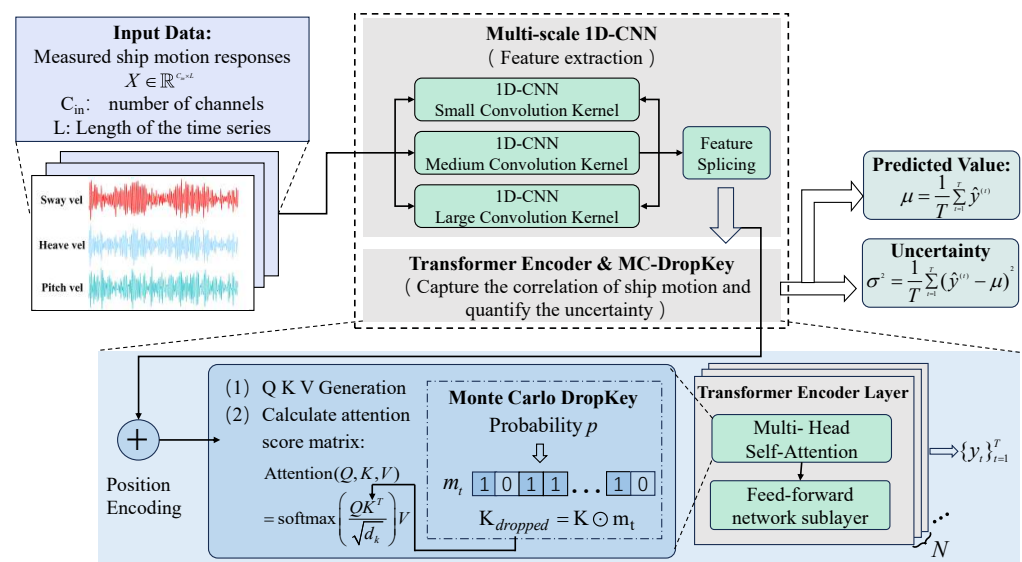


Figure 1. The framework of the proposed model for sea state estimation.

Then, to establish global temporal dependencies, the convolution output undergoes dimension transposition and then enters the transformer encoder. The model injects positional encoding to preserve temporal information. Multi-head self-attention calculates dependency weights across long sequences. This allows local features to corroborate each other. Then, global average pooling fuses features from all time steps into a global description vector. A fully connected layer maps this vector to sea state values.

Finally, to quantify estimation confidence, the model introduces the MC-DropKey mechanism for the transformer [26]. During training, some key features are randomly masked before computing attention scores. This forces the network to complete information from context under missing data conditions. It serves as a regularization effect. During estimation, this mechanism remains active. Multiple stochastic forward passes are performed on the same sample. Each pass generates a different mask and produces a deterministic estimated value. The mean of the estimation ensemble is taken as the sea state estimate. The variance is taken as the uncertainty indicator.

2.2. Multi-Scale 1D-CNN

In the entire sea state estimation model, the multi-scale 1D-CNN undertakes the core task of spatial and local feature extraction. When a ship sails at sea, its motion state is highly complex. It is often influenced by multiple different wave groups simultaneously. High-frequency waves bring sudden impact loads. Long-period large waves cause the ship to roll slowly and continuously. A single-scale convolution kernel finds it difficult to capture these two distinctly different motion characteristics at the same time. Therefore, the model designs a parallel multi-scale convolution architecture [27]. From data input,

detrending, and standardization to multi-scale feature fusion, it constructs a complete local feature extraction pipeline.

2.2.1. Data Preprocessing

The model input comes directly from ship motion measurement data. Assume the number of input channels is C_{in} . Each channel represents an independent one-dimensional time series signal. These signals are truncated within a fixed time window length L . This forms a basic training sample. In mathematical expression, a training sample is constructed as a two-dimensional vector of shape (C_{in}, L) , denoted as

$$X \in \mathbb{R}^{C_{in} \times L} \quad (1)$$

The rows represent different motion degrees of freedom (channels). The columns represent time steps. To eliminate interference from low-frequency DC components such as sensor drift and engine room vibration, and to ensure stable gradient updates during deep learning model training, the data must undergo two strict preprocessing steps before entering the convolution layer. The first step is detrending. For all time points in each channel, the mean of that channel is calculated. Then, each data point subtracts this mean, the formula is as follows:

$$x_i^{(c)} = x_i^{(c)} - \frac{1}{L} \sum_{j=1}^L x_j^{(c)} \quad c = 1, 2, 3 \quad (2)$$

where $x_i^{(c)}$ represents the time series signal of ship motion measurement. This step effectively removes the constant bias in the signal and makes the data symmetric around zero. The second step is Z-score normalization. Based on the detrended data, the model normalizes each channel. This eliminates scale differences caused by different physical quantities with different dimensions, such as angle and displacement. The formula is as follows:

$$x_i^{(c)} = \frac{x_i^{(c)} - \mu^{(c)}}{\sigma^{(c)}} \quad (3)$$

where $\mu^{(c)}$ and $\sigma^{(c)}$ are the mean and standard deviation calculated from all samples in that channel. This ensures that test or estimation data can be scaled using the same set of benchmarks.

2.2.2. Parallel Multi-Scale 1D-CNN Layer

Three parallel 1D convolution branches are designed, the three branches share the same input. But each is equipped with convolution kernels of different sizes. The kernel sizes k are set to a , b , and c , respectively, and the stride is set to s for all branches. This ensures consistency in the time dimension. The standardized sample $X \in \mathbb{R}^{C_{in} \times L}$ is fed into each branch. Each branch uses C_{out} convolution kernels to generate specific feature maps. The formula is as follows:

$$F_k = \text{ReLU}(\text{Conv1d}(X, \theta_k, s_k)) \quad (4)$$

where F_k represents the output feature map, θ_k represents the convolution kernel, and s_k represents the stride.

Branch 1 ($k = a$): It has C_{out} filters of size a with input channels C_{in} . The filters slide along the time axis. They focus on capturing high-frequency details, such as instantaneous slamming and short-period oscillations when the ship is hit by waves. The output feature map matrix of this branch is $F_a \in \mathbb{R}^{C_{out} \times L}$.

Branch 2 ($k = b$): It has C_{out} filters of size b with input channels C_{in} . The convolution kernel size matches medium-frequency fluctuations. It is mainly used to capture roll or pitch characteristics close to the ship's natural period. The output feature map matrix of this branch is $F_b \in \mathbb{R}^{C_{out} \times L}$.

Branch 3 ($k = c$): It has C_{out} filters of size c with input channels C_{in} . These filters extract macro trends over a longer time span. They are specifically used to capture slow undulations and long-period rolling caused by long swells. The output feature map matrix of this branch is $F_c \in \mathbb{R}^{C_{out} \times L}$.

After the three branches complete feature extraction independently, the model concatenates the feature maps along the channel dimension. The formula is as follows:

$$F_{concat} = \text{Concat}([F_a, F_b, F_c], \text{dim} = 0) \quad (5)$$

Since the feature map shape output by each convolution path is (C_{out}, L) , the shape of the concatenated new feature matrix F_{concat} becomes $(3 \times C_{out}, L)$. In this fused matrix, the first C_{out} channels represent high-frequency impact features from the small convolution kernel. The middle C_{out} channels represent medium-frequency natural oscillation features. The last C_{out} channels represent low-frequency swell trends.

2.3. Transformer Encoder

The multi-scale 1D-CNN extracts spatial local features containing full-frequency domain information. The model then sends the fused feature matrix into the transformer encoder. The multi-scale 1D-CNN captures local features at specific moments. The core task of the transformer encoder is to establish long-span global temporal connections on this basis. The internal data flow and feature processing mechanism of this module can be divided into the following key steps, as shown in Figure 2.

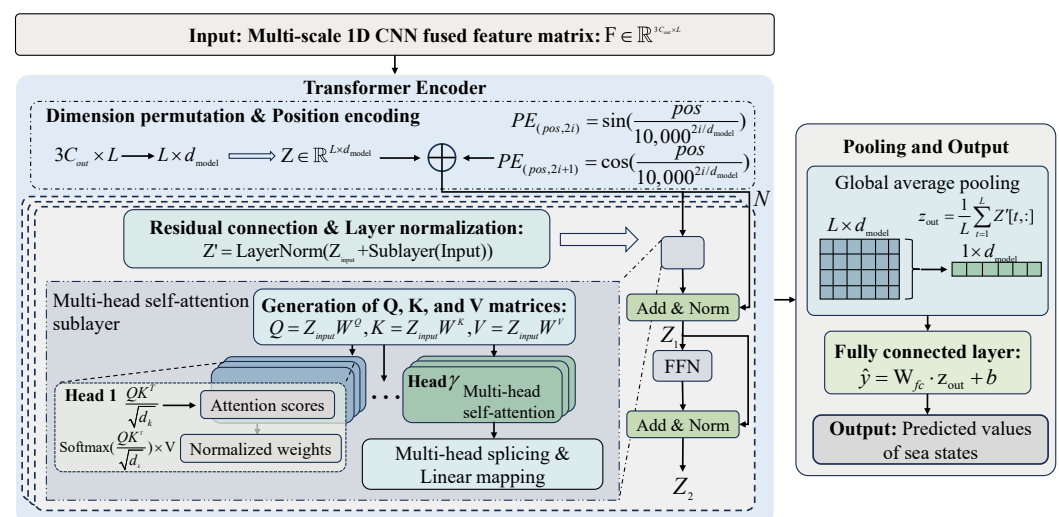


Figure 2. The framework of the transformer encoder.

2.3.1. Position Encoding

The fused feature map output by the multi-scale 1D-CNN is a two-dimensional matrix. To fit the self-attention mechanism inside the transformer, the data first undergoes dimension transposition. The model converts the time steps L into sequence length, then converts the channel number C_{out} into the word vector dimension at each time position. The output of the convolution layer is $F \in \mathbb{R}^{3C_{out} \times L}$. Let $3 \times C_{out} = d_{model}$. After dimension transposition, it becomes $Z \in \mathbb{R}^{L \times d_{model}}$. The matrix shape becomes (L, d_{model}) .

The self-attention mechanism has permutation invariance. It cannot perceive the temporal order in a sequence. Therefore, positional information must be injected artificially. Corresponding temporal position encoding is generated and added to the original sequence features. This ensures that the subsequent attention computation does not lose the temporal evolution pattern of ship motion. Normally, sin/cos positional encoding is used. For position index pos and dimension index i , the following formulas are used:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (6)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (7)$$

A matrix $P \in \mathbb{R}^{L \times d_{\text{model}}}$ with exactly the same shape as Z is generated and added to the original sequence to obtain the final input data:

$$Z_{\text{input}} = Z + P \quad (8)$$

2.3.2. Multi-Head Self-Attention Sublayer

The data fused with positional information enters the encoder layer. The transformer encoder is formed by stacking N identical layers. Each layer consists of two sublayers: a multi-head self-attention sublayer and a feed-forward network (FFN) sublayer. Each sublayer is surrounded by a residual connection followed by layer normalization.

First is the multi-head self-attention sublayer. The input matrix $Z_{\text{input}} \in \mathbb{R}^{L \times d_{\text{model}}}$ is copied into three copies. Unlike the standard transformer where $d_{\text{model}} = \gamma \times d_k$, our implementation projects the input to an inner dimension $d_{\text{inner}} = \gamma \times d_k$ before splitting into heads. This design provides each of the γ attention heads with d_k dimensions for independent feature learning, which can ensure sufficient capacity while maintaining the compact model feature dimension d_{model} . The projection process is as follows:

$$Q = Z_{\text{input}} W^Q, \quad W^Q \in \mathbb{R}^{d_{\text{model}} \times d_{\text{inner}}} \quad (9)$$

$$K = Z_{\text{input}} W^K, \quad W^K \in \mathbb{R}^{d_{\text{model}} \times d_{\text{inner}}} \quad (10)$$

$$V = Z_{\text{input}} W^V, \quad W^V \in \mathbb{R}^{d_{\text{model}} \times d_{\text{inner}}} \quad (11)$$

The model first uses Q and K to calculate the attention score matrix. This matrix represents the correlation strength between all different time steps within the time window. Then, the Softmax function is used to normalize the score matrix, converting the scores in each row into a probability distribution that sums to 1. These probability weights are used to perform a weighted sum on the value matrix V . This extracts the key features of the current time step. The formulas are as follows:

$$\text{Attention Scores} = \frac{QK^T}{\sqrt{d_k}} \quad (12)$$

$$\text{Output} = \text{Softmax}(\text{Attention Scores}) \times V \quad (13)$$

To enable the model to learn features more comprehensively from different subspaces, the model adopts a multi-head attention mechanism. It sets γ such attention heads for parallel computation. Each head independently outputs a feature matrix of shape (L, d_k) . The outputs of these γ heads are concatenated along the feature dimension to form a matrix of shape (L, d_{inner}) . The concatenated matrix passes through a linear layer $W^O \in$

$\mathbb{R}^{d_{\text{inner}} \times d_{\text{model}}}$, which compresses the feature dimension back from d_{inner} to d_{model} to produce an output matrix consistent with the input dimension.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_\gamma) W^O \quad (14)$$

After completing multi-head attention feature extraction, the model adds residual connection and layer normalization at the output end. This is to prevent gradient vanishing in deep networks and to preserve the local features extracted by the front-end CNN. The formula is as follows:

$$Z_1 = \text{LayerNorm}(Z_{\text{input}} + \text{MultiHead}(Q, K, V)) \quad (15)$$

where Z_1 is the output of the multi-head self-attention sublayer. It serves as the input to the FFN sublayer.

2.3.3. Feed-Forward Network (FFN) Sublayer

The FFN is a two-layer fully connected network applied independently and identically to each position. ReLU activation is used in the middle. Its formula is as follows:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (16)$$

where the first layer is an expansion layer: $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{ff}}}$, $b_1 \in \mathbb{R}^{d_{\text{ff}}}$. It increases the dimension from d_{model} to d_{ff} . ReLU activation is used in the middle. The activation function is

$$\text{ReLU} = \max(0, z) \quad (17)$$

The second layer is a compression layer: $W_2 \in \mathbb{R}^{d_{\text{ff}} \times d_{\text{model}}}$, $b_2 \in \mathbb{R}^{d_{\text{model}}}$. It reduces the dimension back to d_{model} . For the output Z_1 of the multi-head self-attention sublayer, the FFN sublayer performs residual connection and layer normalization to obtain

$$Z_2 = \text{LayerNorm}(Z_1 + \text{FFN}(Z_1)) \quad (18)$$

where Z_2 is the output of this encoder layer. It serves as the input to the next layer. After the process is repeated N times, the output Z of the last layer is the final encoded representation. After processing is carried out through the encoding layer, the model averages the output matrix along the time axis. It performs global average pooling. The formula is as follows:

$$z_{\text{out}} = \frac{1}{L} \sum_{t=1}^L Z[t, :] \quad (19)$$

where z_{out} represents the global vector of the overall heading state. Finally, this vector is connected to a fully connected layer. It maps the global features to the required prediction targets. The output dimension D_{out} is determined by the corresponding regression task. This outputs the final sea state prediction value:

$$\hat{y} = W_{\text{fc}} \cdot z_{\text{out}} + b, \quad \hat{y} \in \mathbb{R}^{D_{\text{out}}} \quad (20)$$

where W_{fc} is the weight matrix and b is the bias term of the fully connected layer.

2.4. MC-DropKey

To quantify estimation uncertainty, the MC-DropKey [28] mechanism is introduced. The operation of this mechanism is divided into two key stages: training and estimation.

During the training phase, the main objective of DropKey is to perform regularization to prevent the model from overfitting. This operation takes place within the transformer encoder, specifically before calculating the attention scores. The model randomly generates a mask vector $m \in \{0, 1\}^L$ with a probability of p , and then performs the DropKey operation, as shown in the formula:

$$K_{\text{dropped}} = K \odot m \quad (21)$$

where K_{dropped} represents the processed key matrix, and it is used to calculate the attention score. Through the random occlusion mechanism, the network is forced to learn global correlations in the absence of local feature information, thereby enhancing the generalization ability of the model.

In the estimation stage, the MC-DropKey mechanism remains active. For a given input sample X , the model performs T random forward passes. In each inference process, the model randomly generates a new key mask $m^{(t)}$. It then calculates a deterministic estimation result $\hat{y}^{(t)}$ based on the currently masked network structure. After repeating the above forward pass steps T times, a set of estimation values Y is obtained:

$$Y = \{\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(T)}\} \quad (22)$$

Based on this set of estimation values obtained from T random inferences, the model's final output contains two key statistics. The first is the mean μ . It is obtained by taking the arithmetic average of the T estimation results. This represents the model's core estimate of the sea state estimation:

$$\mu = \frac{1}{T} \sum_{t=1}^T \hat{y}^{(t)} \quad (23)$$

The second is the variance σ^2 . It is used to quantify the dispersion of the T estimation results around the mean. This reflects the model's uncertainty about the current estimation result:

$$\sigma^2 = \frac{1}{T} \sum_{t=1}^T (\hat{y}^{(t)} - \mu)^2 \quad (24)$$

3. Experiment

3.1. Dataset

The dataset used in this study is generated by the high-fidelity ship maneuvering simulator at the Ålesund laboratory of the Norwegian University of Science and Technology (NTNU). The experimental system is equipped with a full-scale ship bridge. It integrates synchronously linked steering gear, thrusters, and environmental control consoles. The platform deeply couples an engineering-verified hydrodynamic core module with industry-standard actuator models. The simulated ship is equipped with a redundant propulsion system consisting of four tunnel thrusters (Thr1–Thr4) and two main azimuth thrusters (Thr5–Thr6). This system can accurately reflect dynamic positioning operations under complex sea conditions. Figure 3 shows a simplified structural diagram of this vessel.

During the data generation process, the simulation platform uses the unimodal JON-SWAP wave spectrum combined with wind and current components to generate highly realistic environmental loads. To reflect the complex variability of open sea weather, the simulated environment covers multiple sea states from level 0 to level 5. It also includes random transitions between different sea states. The platform has built-in model-based feedback control and feedforward disturbance compensation algorithms. A real-time visual system synchronously renders the ship's closed-loop motion response. All ship state variables are recorded at a high sampling frequency of 20 Hz. Ultimately, a large dataset containing 1,517,307 continuous time series samples is collected. This dataset fully covers

the entire closed-loop interaction process from environmental excitation input and control command execution to ship dynamic response output.

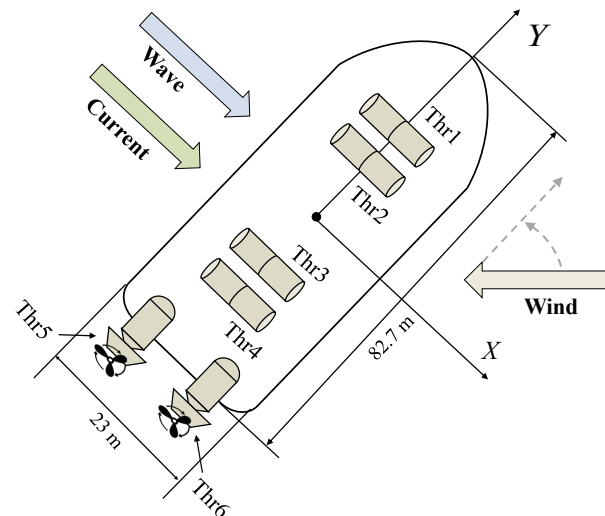


Figure 3. A simplified structural diagram of the vessel.

The initial dataset brings together multi-source heterogeneous information including ship motion, maneuvering commands, and external environmental disturbances. After feature selection and physical consistency optimization, 25 features are finally determined. They are structurally divided into four levels. The first level is motion state. The second level is maneuvering and propulsion input. The third level is environmental disturbance. The fourth level is physically enhanced features. The dynamic response range of the important features of the dataset is shown in Table 1. For the specific research task of sea state estimation in this study, the model does not need to use all 25 features. At the input end, three core ship motion parameters from the first-level motion state features are selected as model inputs. These are sway velocity, heave velocity, and pitch velocity. At the output end, two core wave environment parameters from the third-level environmental disturbance features are selected as estimation targets. These are wave direction (Dir) and significant wave height (H_s).

Table 1. The dynamic response range of the physical degrees of freedom of the dataset channels.

Category	Channel	Unit	Range
Position and Translation	north_pos	m	[−28.67, 29.85]
	east_pos	m	[−15.91, 10.29]
	down_pos	m	[−2.81, 2.70]
Posture and Rotation	yaw_angle	deg	[−16.85, 12.84]
	roll_angle	deg	[−3.23, 4.58]
	pitch_angle	deg	[−5.00, 4.93]
Velocity	surge_vel	m/s	[−1.43, 1.55]
	sway_vel	m/s	[−1.21, 1.13]
	heave_vel	m/s	[−1.90, 1.99]
	yaw_vel	deg/s	[−1.71, 1.61]
	roll_vel	deg/s	[−1.71, 2.21]
	pitch_vel	deg/s	[−3.72, 3.37]
Environment	wave_height	m	[0.03, 3.95]
	wave_dir	deg	[3.57, 337.61]
	wind_vel	m/s	[0.00, 13.22]

To meet the time series input requirements of the model, this experiment performs resampling and independent sample construction on the continuous time series of the original dataset. The original 20 Hz high-frequency data are downsampled to 1 Hz, as ship motions and energetic wave components exhibit dominant periods between 4 and 20 s, corresponding to frequency content below 0.25 Hz. This rate thus satisfies the Nyquist criterion while ensuring computational efficiency. Then, a non-overlapping time window sliding strategy is used to segment the continuous data. Each time window length is set to 960 s. This is because sea state parameters typically maintain stationary characteristics within 10 to 30 min. The 16 min window length falls exactly within this physically coherent interval, ensuring that each sample captures a self-contained sea state realization. Based on this sampling and segmentation strategy, 200 mutually independent and temporally non-overlapping time series samples are extracted from the original dataset, serving as the dataset for this experiment.

For dataset splitting, the hold-out method with random shuffling is used for data partitioning. The order of all independent samples is globally shuffled at random. This eliminates potential local temporal correlation interference that may exist in the segmentation of continuous time series. The shuffled dataset is divided into training set, validation set, and test set. A total of 60 samples are drawn as the test set. This set is used for the objective evaluation of the model's final estimation accuracy and uncertainty quantification. A total of 100 samples are selected as the training set. The remaining 40 samples serve as the validation set. The specific distribution of sea state features estimated in the subsequent experiment is shown in Figure 4.

Although the present study uses the same simulator dataset as our previous ship-motion prediction study [29], the research objective, input-output mapping, model architecture, and evaluation metrics are fundamentally different.

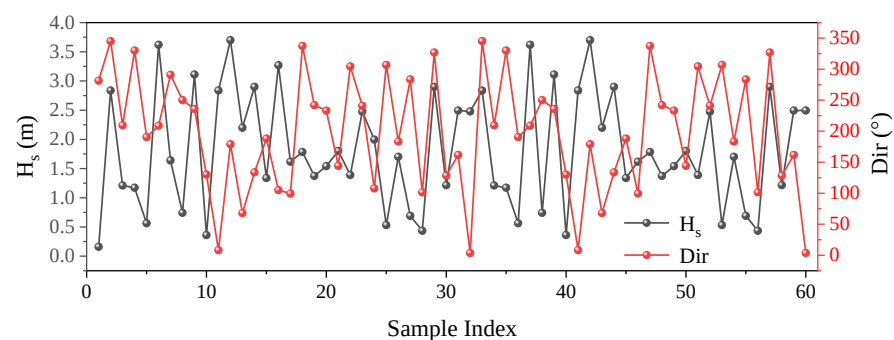


Figure 4. Distribution of estimated sea conditions characteristics.

3.2. Experimental Setup

To verify the performance of the proposed model, the parameter configuration and the selection of estimation labels are specifically designed based on the physical and mathematical characteristics of sea state time series data. The specific parameter configuration is shown in Table 2. Among them, the convolution kernel sizes of the multi-scale 1D-CNN are 3, 7, and 11, which can, respectively, correspond to the features of high-frequency, medium-frequency, and low-frequency waves. For the transformer encoder, the model embedding dimension d_{model} is set to 192, matching the feature map dimensions of the concatenated output of the multi-scale CNN. The encoder consists of four layers with eight attention heads, allowing the model to capture global temporal dependencies. The Drop-Key probability is set to 0.15 to provide suitable regularization while retaining important temporal attention information.

For the selection of estimation labels, the sea state data to be estimated are wave height and wave direction. Therefore, differentiated label processing and estimation strategies are adopted for estimation variables with different physical properties.

Table 2. Parameter settings of the proposed model.

Module	Parameter	Value
Multi-scale 1D-CNN	Number of input channels C_{in}	3
	Number of output channels C_{out}	64
	Small convolution kernel a	3
	Medium convolution kernel b	7
	Large convolution kernel c	11
Transformer encoder	Input feature dimension d_{model}	192
	Inner dimension of multi-head projection d_{inner}	512
	Number of attention heads γ	8
	Dimensions of each head d_k	64
	Number of encoder blocks N	4
	DropKey probability p	0.15
	Number of forward propagation iterations T	150

For the wave height estimation task, wave height is a standard continuous numerical variable. Its true observation values can be used as supervised labels. This guides the model to perform conventional continuous numerical regression learning. For the wave direction estimation task, it is a periodic circular angular quantity with a value range between 0° and 360° . It has an inherent defect: the angular boundaries of 360° and 0° are physically identical but numerically discontinuous. If the original angle values are directly used as estimation output, the model is highly prone to producing drastic estimation mutations at the $0^\circ/360^\circ$ boundary when fitting this circular periodic characteristic.

To solve the modeling defect caused by angular periodicity, this paper performs trigonometric function decoupling preprocessing on the wave direction data in the experiment. The experiment converts the original wave direction angle θ into sine value $\sin \theta$ and cosine value $\cos \theta$. These are used as the dual-output regression targets for joint training. After conversion, $\sin \theta$ and $\cos \theta$ become continuous and bounded small values. This eliminates the jump discontinuity in the original angular space. After the model completes training and outputs the estimated $\sin \theta$ and $\cos \theta$ sequences, the experiment uniformly reconstructs the final true wave direction angle through the two-parameter arctangent formula:

$$\theta = \text{atan2}(\sin \theta, \cos \theta) \quad (25)$$

To validate the performance of the proposed model, four representative models are selected as baselines in this work for comparison: single-scale CNN, LSTM, temporal fusion transformer (TFT), and TimesNet. CNN and LSTM, respectively, represent typical methods for traditional local spatial feature extraction and recurrent temporal modeling. TFT and TimesNet serve as advanced multivariate time series prediction models. To ensure fairness in the comparative experiments, the parameter configurations of the baseline models are aligned and equivalently designed with the corresponding components of the proposed model in terms of structural capacity, feature dimensions, and regularization strength. The specific parameter configurations are shown in Table 3.

The proposed method and all baseline models were implemented using the PyTorch 2.4.1 framework. The global random seed was uniformly set to 2026. The networks were trained using the AdamW optimizer with an initial learning rate of 3×10^{-4} , a weight decay of 1×10^{-4} , and a batch size of 4. The maximum number of training epochs was set to 220, and an early stopping criterion was applied with a patience of 45 epochs. The

proposed model contains approximately 1.2 million trainable parameters. Inference on a single sample (960 time steps $\times$ 3 input channels) takes 8–12 ms on average using an Intel Core i5-13500H CPU, which can meet the real-time requirements for practical deployment. The proposed model outputs three prediction components: normalized significant wave height $\hat{h}_{\text{norm}}$, wave direction sine component $\hat{d}_{\text{sin}}$, and wave direction cosine component $\hat{d}_{\text{cos}}$. The loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{\text{height}} + \mathcal{L}_{\text{dir}} + \lambda \mathcal{L}_{\text{circle}} \quad (26)$$

$$\mathcal{L}_{\text{height}} = \text{SmoothL1}(\hat{h}_{\text{norm}}, h_{\text{norm}}) \quad (27)$$

$$\mathcal{L}_{\text{dir}} = \text{MSE}\left(\begin{bmatrix} \hat{d}_{\text{sin}}, \hat{d}_{\text{cos}} \end{bmatrix}, \begin{bmatrix} d_{\text{sin}}, d_{\text{cos}} \end{bmatrix}\right) \quad (28)$$

$$\mathcal{L}_{\text{circle}} = \text{MSE}(\hat{d}_{\text{sin}}^2 + \hat{d}_{\text{cos}}^2, 1) \quad (29)$$

where $\mathcal{L}_{\text{height}}$ supervises the normalized wave height prediction, and $\mathcal{L}_{\text{dir}}$ controls the directional components. The circle constraint $\mathcal{L}_{\text{circle}}$ enforces the unit-norm constraint, and $\lambda = 0.1$ is the weighting coefficient for the circle constraint. The coefficients of $\mathcal{L}_{\text{height}}$ and $\mathcal{L}_{\text{dir}}$ are both set to 1.0 because wave height and wave direction are the two main prediction targets of the proposed model.

Table 3. Hyperparameter settings of each baseline model.

Baseline Model	Parameter	Value
CNN	Kernel size	5
	Hidden layer dimension	96
	Dropout	0.1
LSTM	Hidden size	128
	Number of stacking layers	2
	Dropout	0.15
TFT	Hidden size	96
	Local encoder layers	1
	Dropout	0.1
TimesNet	Channels	96
	Block kernel sizes	3, 7, 15
	Dropout	0.1

3.3. Performance Metrics

To comprehensively and objectively evaluate the model's estimation performance for wave height and wave direction, this paper selects three performance metrics based on the different physical properties of the estimated variables. These are mean absolute error (MAE), circular absolute error (CAE), and coefficient of determination R^2 .

MAE is used to measure the average absolute difference between estimated wave height and true wave height. The smaller the value of this metric, the higher the estimation accuracy of the model. Its calculation formula is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (30)$$

where N is the total number of samples in the test set. y_i is the true wave height observation value of the i -th sample. $\hat{y}_i$ is the wave height value estimated by the model.

CAE is used to evaluate the estimation performance of wave direction. This metric calculates the minimum angle between the estimated wave direction and the true wave

direction on a circle. This avoids evaluation penalties when crossing the 360° boundary. Its calculation formula is as follows:

$$CAE = \frac{1}{N} \sum_{i=1}^N \min(|d_i - \hat{d}_i|, 360^\circ - |d_i - \hat{d}_i|) \quad (31)$$

where d_i is the true wave direction angle of the i -th sample. $\hat{d}_i$ is the wave direction angle estimated by the model. The smaller the CAE value, the more accurate the model's wave direction estimation.

R^2 is used to reflect the goodness of fit between the model's estimated values and the true observations of wave height or wave direction. Its value typically ranges from 0 to 1. The closer the value is to 1, the better the model's fitting performance. Its calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (32)$$

where $\bar{y}$ is the mean of all true observation values of wave height or wave direction.

In addition to the performance metrics used to measure point estimation accuracy, this paper also introduces an estimation uncertainty evaluation metric based on the MC-DropKey mechanism. During the estimation stage, the MC-DropKey mechanism remains active. By performing T stochastic forward passes on the same input sample, it can output a statistically meaningful distribution of estimation results. This paper uses the variance of these T estimation results to quantify their dispersion around the mean. This directly reflects the model's uncertainty about the current estimation result.

To evaluate the quality of the estimated uncertainty intervals, this study introduces two complementary quantitative metrics: prediction interval coverage probability (PICP) and mean prediction interval width (MPIW). PICP measures the actual proportion of true observation values that fall within the constructed uncertainty intervals. It reflects the reliability of uncertainty quantification. Its definition is as follows:

$$PICP \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i \in [\mu_i - \lambda\sigma_i, \mu_i + \lambda\sigma_i]\} \quad (33)$$

where μ_i and σ_i are the mean and standard deviation of the T random estimations for the i -th sample, respectively. λ is the scaling factor. For example, $\lambda = 1.96$ corresponds to a 95% nominal confidence level. $\mathbf{1}\{\cdot\}$ is the indicator function. The closer the PICP value is to the nominal confidence level, the better the calibration of the uncertainty estimation.

MPIW quantifies the average width of the prediction intervals. It is used to measure the sharpness or informativeness of the uncertainty estimation. Its calculation method is as follows:

$$MPIW \equiv \frac{1}{N} \sum_{i=1}^N 2\lambda\sigma_i \quad (34)$$

4. Results and Discussion

4.1. Predicted Mean

Figures 5 and 6, and Table 4, respectively, show the estimation curves and performance metrics for wave height and wave direction using the above model, and they provide a comparison of the performance of the proposed model with the baseline models on the wave height and wave direction estimation tasks. The proposed model shows advantages in estimation for different physical quantities. For wave height estimation, the proposed model achieves the lowest MAE of 0.117. For wave direction, its CAE is 5.212, which is only about half of that of CNN, TFT, and TimesNet, and is much lower than that

of the LSTM model. In terms of R^2 , the proposed model reaches the highest value of 0.986 for wave direction estimation, while also reaching 0.970 for wave height estimation. Although the overall performance of advanced multivariate time series baseline models (TFT, TimesNet) is superior to the traditional LSTM, the combined accuracy of the two is still not as good as that of the model proposed in this paper. The obtained results indicate that the proposed model demonstrates good estimation and fitting capabilities for data with both sea state characteristics.

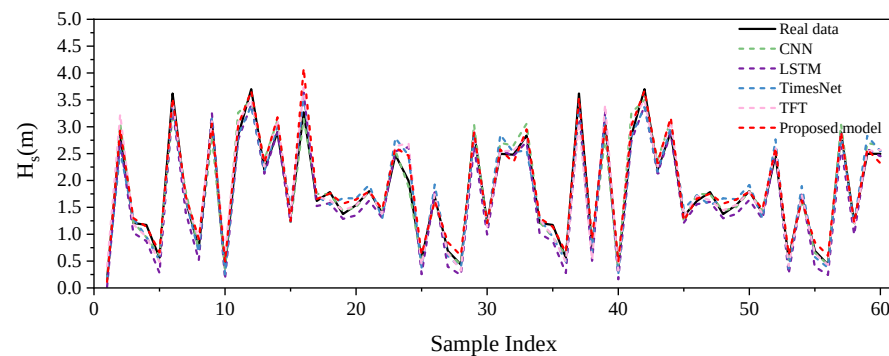


Figure 5. The estimation results of wave height by different methods.

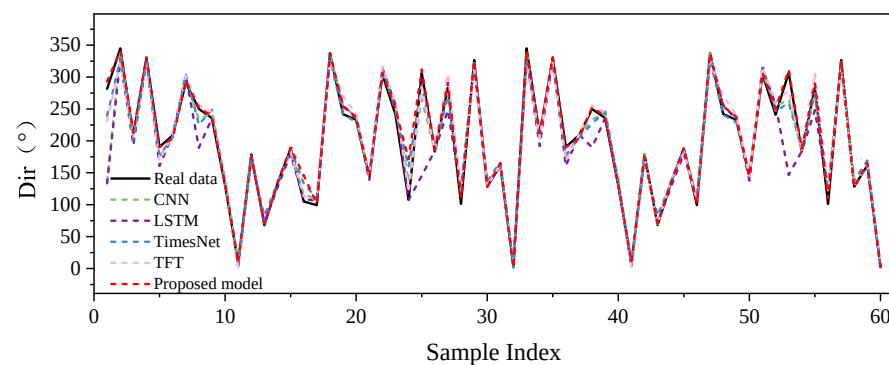


Figure 6. The estimation results of wave direction by different methods.

Table 4. Performance comparison of different models for significant wave height and wave direction estimation.

Sea State	Metrics	Proposed Model	CNN	LSTM	TimesNet	TFT
H_s (m)	MAE	0.117	0.122	0.177	0.144	0.124
	R^2	0.970	0.972	0.955	0.965	0.964
Dir (°)	CAE	5.212	10.145	17.521	10.847	10.735
	R^2	0.986	0.943	0.829	0.973	0.976

It is worth noting that the CNN baseline model is slightly ahead in the wave height R^2 metric (0.972). In contrast, MAE measures point-by-point absolute deviation, which directly shows the estimation accuracy for each specific moment. The proposed model introduces the MC-DropKey mechanism with moderate random regularization perturbation. Although its R^2 is marginally lower by 0.002, it successfully gains better MAE control and dynamic uncertainty quantification capability.

In practical marine engineering, accurately estimating the absolute sea state of individual samples at specific moments is far more critical than merely fitting a global trend. Therefore, the proposed model balances low MAE with dynamic uncertainty quantification capability and can provide higher comprehensive engineering decision-making value.

To verify the effectiveness of each module in the proposed model and their contributions to the overall estimation performance, an ablation experiment is conducted. Four different model configurations are constructed for comparative analysis:

- (1) The full model proposed in this paper.
- (2) The transformer encoder module is removed (multi-scale CNN): Only the multi-scale 1D-CNN is used, and it is connected to the back-end regression layer.
- (3) The front-end multi-scale 1D-CNN module is removed (transformer): The raw time series is directly fed into the transformer encoder for processing.
- (4) The MC-DropKey mechanism is removed (full model without DropKey): Both the multi-scale 1D-CNN and the transformer encoder are retained, using standard deterministic attention without random key masking.

The results are shown in Figure 7a,b. In the wave height and wave direction estimation tasks, removing any feature extraction module leads to performance degradation. When the multi-scale CNN module is removed, the estimation error expands significantly (H_s MAE rises to 0.170, Dir CAE to 43.214°). This indicates that a pure transformer architecture struggles to process raw angular sequences containing high-frequency fluctuations. This result also explains why the proposed model achieves much better performance in wave direction estimation, with a CAE that is only half of that of the baseline models. The multi-scale CNN extracts features at different scales, allowing the subsequent transformer encoder to better learn the temporal patterns in the input signal. The multi-scale CNN alone performs better but still falls short of the full model, as it lacks the transformer encoder's ability to model long-range temporal dependencies across the entire sequence. When the MC-DropKey mechanism is removed, the deterministic point estimation errors slightly decrease. However, this marginal improvement in deterministic point precision comes at the cost of losing uncertainty quantification capabilities. The full model optimally sacrifices a negligible amount of point estimation precision to incorporate stochastic inference, thereby gaining the essential ability to output the distribution of prediction intervals.

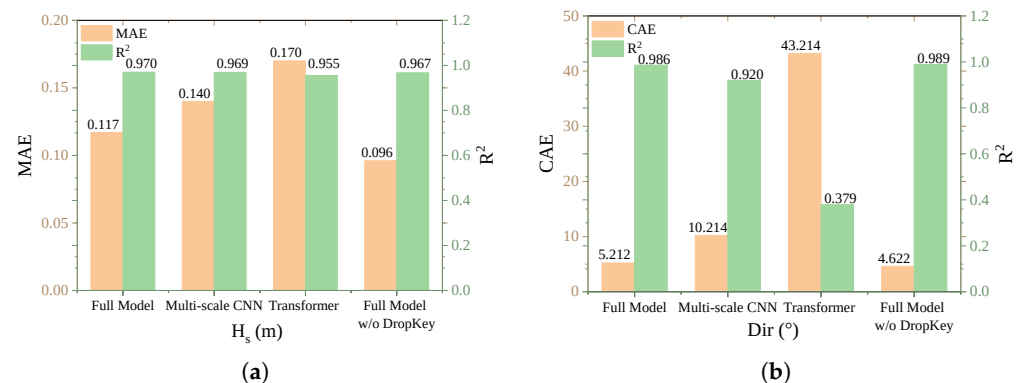


Figure 7. The ablation experiment results of the proposed sea state estimation model are as follows: (a) wave height estimation result and (b) wave direction estimation result.

After deep integration of each module, the complementary advantages of local feature extraction, global temporal correlation modeling, and uncertainty quantification can be simultaneously utilized. This is also the key to the model's ability to improve estimation performance under complex sea states.

To examine whether the model performance is affected by random initialization and dataset splitting, repeated validation experiments were conducted. The model proposed in this paper was independently trained and evaluated under five distinct random seeds, with each seed corresponding to a new dataset splitting scheme. As summarized in Table 5, the proposed model achieves stable and robust estimation performance across multiple

independent repeated experiments. The MAE for wave height is 0.111 ± 0.013 m, and for wave direction, the CAE is $8.770^\circ \pm 1.316^\circ$. The small standard deviations indicate that the model is not highly sensitive to changes in random initialization or dataset splitting. These results support the stability and reproducibility of the proposed model.

Table 5. Performance robustness of the proposed model across 5 independent random seeds and dataset splits.

Metrics	Mean	Standard Deviation	Min	Max
H_s MAE (m)	0.111	0.013	0.094	0.122
Dir CAE ($^\circ$)	8.770	1.316	6.630	9.998

4.2. Uncertainty

The MC-DropKey mechanism is introduced to perform multiple stochastic forward passes, generating a predictive standard deviation σ . Based on this standard deviation, 95% prediction intervals (PIs) are constructed for the sea state estimation results. The results are shown in Figures 8 and 9, and Table 6.

As shown in Table 6, for wave height estimation, the PICP reaches 96.67%, and the MPIW is 1.01 m. For wave direction estimation, the PICP is also 96.67%, and the MPIW is 45.47° . The PICP values of both sea state parameters exceed the 95% nominal confidence threshold, which proves that the uncertainty interval of the model output is highly reliable and can fully encompass the true observed values. In addition to the quantitative metrics, Figures 8 and 9 visually present the 95% prediction intervals and the prediction uncertainty σ curves for wave height and wave direction, respectively. For wave height estimation, the uncertainty values of the overall sequence remain at a relatively low level. At nodes such as 41, 59, 37, and 16, the uncertainty value peaks. For wave direction estimation, its local peak occurs near sample indices 48, 25, 40 and 55.

To further explore the relationship between model uncertainty and sea state complexity, this study conducts a quantitative statistical analysis based on ship motion signals. The normalized spectral entropy (NSE) of the ship motion time series signals input into the model is calculated as the complexity evaluation metric. The closer the entropy value is to 1, the higher the degree of chaos in the ship motion, and the more complex the sea conditions are. The closer the entropy value is to 0, the calmer the sea conditions.

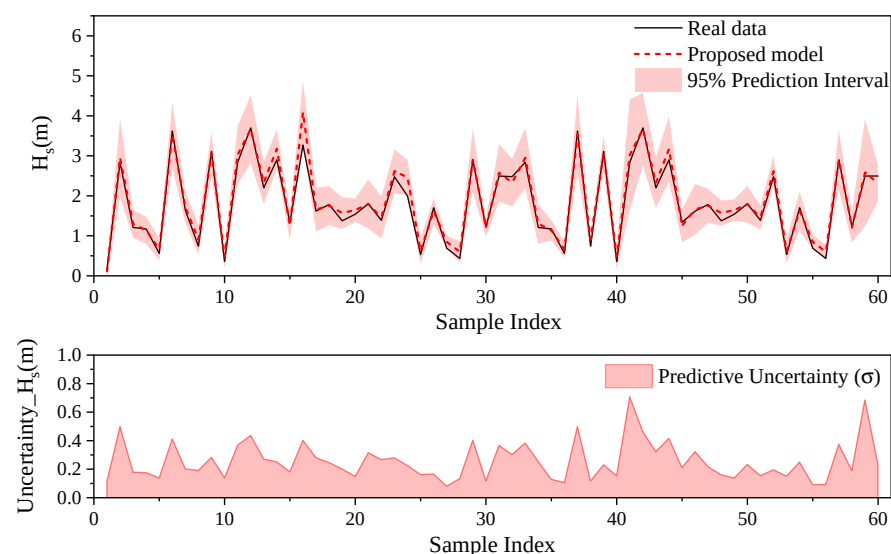


Figure 8. The uncertainty of the predicted wave height by the proposed model.

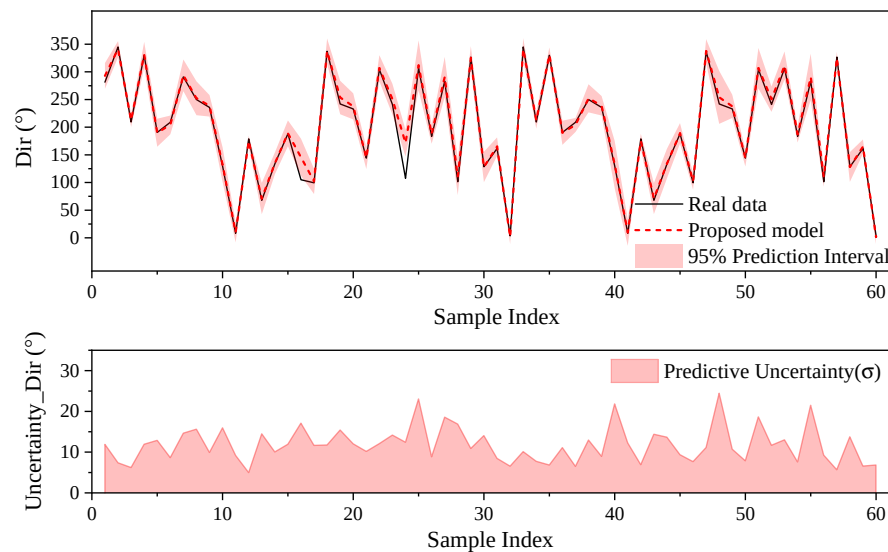


Figure 9. The uncertainty of the predicted wave direction by the proposed model.

Table 6. Quantitative uncertainty evaluation metrics of the proposed model at a 95% confidence level.

Sea State	PICP (%)	MPIW
H_s (m)	96.67%	1.01 m
Dir (°)	96.67%	45.47°

The results are shown in Figure 10. Based on this result, an analysis of event overlap degree was conducted. The 10 samples with the highest spectral entropy values in the sample sequence were defined as complex sea state events, and the distribution of these samples in the high uncertainty of model prediction was examined. Among these complex sea state events, 80% corresponded to the uncertainty peaks of wave height, and 70% corresponded to the uncertainty peaks of wave direction. Similarly, the 10 samples with the lowest entropy values were mostly located in the low-value regions of wave height and wave direction uncertainty. The obtained results indicate that part of the model output uncertainty also reflects the environmental complexity in space.

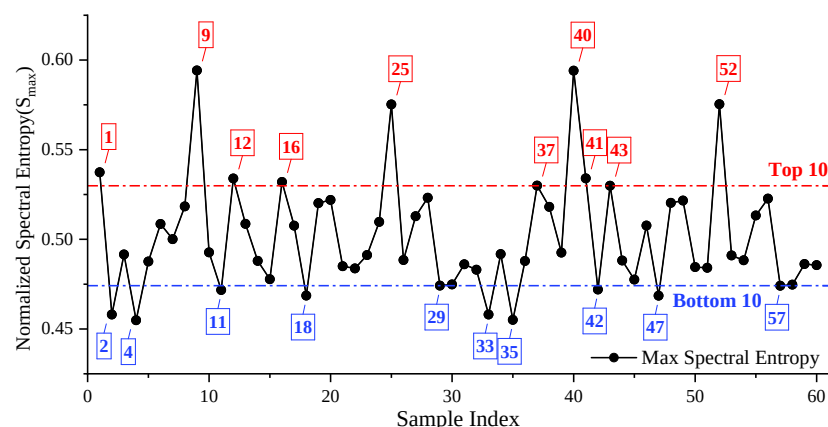


Figure 10. Normalization of the maximum spectral entropy for all samples.

5. Conclusions

To address the highly nonlinear and non-stationary characteristics of time series under complex sea states, this paper proposes a new estimation architecture integrating a

multi-scale 1D-CNN, transformer encoder, and MC-DropKey mechanism. Comprehensive experiments and comparative analyses show that the proposed architecture achieves good performance in both wave height and wave direction estimation tasks. Specifically, it achieves the lowest MAE of 0.117 m for wave height and a CAE of 5.212° for wave direction, alongside high R^2 values of 0.970 and 0.986, respectively. Meanwhile, the ablation study fully confirms the rationality and necessity of this architecture. In addition, the proposed model possesses practical engineering value in terms of uncertainty assessment. The prediction confidence intervals generated by MC-DropKey can fully encompass the true values. The PICP for both wave height and wave direction tasks reaches 96.67%.

It should be pointed out that the dataset used in this study comes from a high-fidelity simulator rather than real operational measurement data. Although the simulator integrates a verified hydrodynamic core module and industry-standard actuator models, there is inevitably a certain deviation between simulation conditions and actual sea conditions. Examples include environmental disturbances such as unmodeled wave nonlinearities and sensor noise characteristics. Therefore, the quantitative results reported in this study demonstrate the model's performance under the controlled conditions. The general applicability of the proposed method on actual ships has not yet been verified.

Nevertheless, the proposed model is essentially data-driven and does not rely on specific assumptions of a simulator. Its generalization ability is relatively strong. If sufficient labeled real ship data at sea are available for training or fine-tuning, the model can be transferred to real ship data. Future work will focus on collecting real ship motion data and synchronous reference sea state ground truth, evaluating domain adaptation performance under real environmental noise, and exploring transfer learning techniques to further bridge the gap between simulation and reality. Ultimately, the proposed method can be reliably deployed on actual ships.

Author Contributions: Writing—original draft preparation, investigation, methodology, and software, T.C.; conceptualization, writing—review and editing, supervision, and methodology, R.M.; investigation and resources, X.G.; writing—review and editing, supervision, and methodology, P.H.; writing—review and editing, and supervision, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author after executing the appropriate data-sharing agreement.

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

References

1. Liu, C.; Hu, Q.; Wang, X.; Yin, J. Event-triggered-based nonlinear model predictive control for trajectory tracking of underactuated ship with multi-obstacle avoidance. *Ocean Eng.* **2022**, *253*, 111278. [\[CrossRef\]](#)
2. Meng, Y.; Rigall, E.; Chen, X.; Gao, F.; Dong, J.; Chen, S. Physics-Guided Generative Adversarial Networks for Sea Subsurface Temperature Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 3357–3370. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Xiao, F.; Zhang, X.; Cheng, X.; Xie, X.; Zhang, J. TSCFNet: Temporal Spectral Feature Cross Fusion Network for Imbalanced Sea State Estimation in Autonomous Ships. *IEEE Trans. Intell. Transp. Syst.* **2026**, *27*, 3257–3270. [\[CrossRef\]](#)
4. Kim, G.Y.; Lim, C.; Oh, S.J.; Nam, I.H.; Lee, Y.M.; Shin, S.C. Deep Learning-Based Prediction of Ship Roll Motion with Monte Carlo Dropout. *J. Mar. Sci. Eng.* **2025**, *13*, 2378. [\[CrossRef\]](#)
5. Song, M.; Liu, S.; Li, W.; Chen, S.; Li, W.; Zhang, K.; Yu, D.; Liu, L.; Wang, X. A continuous space location model and a particle swarm optimization-based heuristic algorithm for maximizing the allocation of ocean-moored buoys. *IEEE Access* **2021**, *9*, 32249–32262. [\[CrossRef\]](#)
6. Koner, P.K.; Harris, A.; Maturi, E. A physical deterministic inverse method for operational satellite remote sensing: An application for sea surface temperature retrievals. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5872–5888. [\[CrossRef\]](#)

7. Rikka, S.; Uiboupin, R.; Kõuts, T.; Vahter, K.; Pärt, S. A Method for Significant Wave Height Estimation From Circularly Polarized X-Band Coastal Marine Radar Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 844–848. [\[CrossRef\]](#)
8. Mounet, R.E.; Nielsen, U.D.; Brodtkorb, A.H.; Tannuri, E.A.; de Mello, P.C. Simultaneous sea state estimation and transfer function tuning using a network of dynamically positioned ships. *Appl. Ocean Res.* **2022**, *129*, 103367. [\[CrossRef\]](#)
9. Han, P.; Li, G.; Cheng, X.; Skjong, S.; Zhang, H. An Uncertainty-Aware Hybrid Approach for Sea State Estimation Using Ship Motion Responses. *IEEE Trans. Ind. Inform.* **2022**, *18*, 891–900. [\[CrossRef\]](#)
10. Bisinotto, G.A.; Sparano, J.V.; Simos, A.N.; Cozman, F.G.; Ferreira, M.D.; Tannuri, E.A. Sea state estimation from full-scale FPSO motions using neural networks: Assessment based on in-service records and metocean reanalysis. *Ocean Eng.* **2026**, *359*, 125911. [\[CrossRef\]](#)
11. Dahl, J.M.; Grilli, A.R.; Steele, S.C.; Grilli, S.T. Estimation of Sea State Parameters from Measured Ship Motions with a Neural Network Trained on Experimentally Validated Model Simulations. *J. Mar. Sci. Eng.* **2026**, *14*, 179. [\[CrossRef\]](#)
12. Li, X.; Mao, R.; Han, P.; Zhang, H. Sea state estimation for ship operations: A systematic review of methods, datasets, and challenges. *Ocean Eng.* **2026**, *364*, 126747. [\[CrossRef\]](#)
13. Nielsen, U.D.; Dietz, J. Estimation of sea state parameters by the wave buoy analogy with comparisons to third generation spectral wave models. *Ocean Eng.* **2020**, *216*, 107781. [\[CrossRef\]](#)
14. Nielsen, U.D.; Brodtkorb, A.H.; Sørensen, A.J. Sea state estimation using multiple ships simultaneously as sailing wave buoys. *Appl. Ocean Res.* **2019**, *83*, 65–76. [\[CrossRef\]](#)
15. Majidian, H.; Wang, L.; Enshaee, H. Part. A: A review of the real-time sea-state estimation, using wave buoy analogy. *Ocean Eng.* **2022**, *266*, 111684. [\[CrossRef\]](#)
16. Bourkaib, R.; Kok, M.; Seyffert, H.C. Unidirectional and multi-directional wave estimation from ship motions using an Adaptive Kalman Filter with the inclusion of varying forward speed. *Probabilistic Eng. Mech.* **2025**, *80*, 103773. [\[CrossRef\]](#)
17. Nielsen, U.D.; Brodtkorb, A.H.; Sorensen, A.J. A brute-force spectral approach for wave estimation using measured vessel motions. *Mar. Struct.* **2018**, *60*, 101–121. [\[CrossRef\]](#)
18. Lee, J.H.; Ko, D.; Choi, J.H. Comparative Studies of Physics-and Machine Learning-Based Wave Buoy Analogy Models Under Various Ship Operating Conditions. *J. Mar. Sci. Eng.* **2025**, *13*, 1823. [\[CrossRef\]](#)
19. Selimović, D.; Hrzić, F.; Prpić-Oršić, J.; Lerga, J. Estimation of sea state parameters from ship motion responses using attention-based neural networks. *Ocean Eng.* **2023**, *281*, 114915. [\[CrossRef\]](#)
20. Cheng, X.; Li, G.; Ellefsen, A.L.; Chen, S.; Hildre, H.P.; Zhang, H. A Novel Densely Connected Convolutional Neural Network for Sea-State Estimation Using Ship Motion Data. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5984–5993. [\[CrossRef\]](#)
21. Kwon, D.S.; Jin, C.; Kim, M.; Kim, S.J. Transformers and neural networks for estimation of parameters of multi-directional waves from rich statistics of FPSO motion signals. *Appl. Ocean Res.* **2026**, *166*, 104927. [\[CrossRef\]](#)
22. Wang, R.; Yuan, M.; Sun, M.; Cai, L.; Ye, Y.; Song, X.; Fu, S.; Li, X.; Tian, X. Multi-Sea-State Joint Frequency-Domain Identification of hydrodynamic parameters for float-over vessels. *Ocean Eng.* **2026**, *359*, 126022. [\[CrossRef\]](#)
23. Li, Y.; Geng, X.; Sun, Q.; Zhou, Z.; Zhang, S. A Coupling Factor Analysis of Predictive Models for Ship Motion Emphasizing Feature Selection and Attention Mechanism-Enhanced Gated Recurrent Unit Networks. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–13. [\[CrossRef\]](#)
24. Montazeri, N.; Jensen, J.J.; Nielsen, U.D. Uncertainties in ship-based estimation of waves and responses. In Proceedings of the OCEANS 2015—MTS/IEEE Washington, Washington, DC, USA, 19–22 October 2015; pp. 1–5.
25. Nielsen, U.D.; Iwase, K.; Mounet, R.E.; Storhaug, G. Uncertainty-associated directional wave spectrum estimation from wave-induced ship responses using Machine Learning methods. *Ocean Eng.* **2024**, *313*, 119543. [\[CrossRef\]](#)
26. Yilma, A.A.; Yang, C.L.; Meles, T.Y.; Nguyen, T.P.Q. Deep evidential transformer with Monte Carlo dropout for uncertainty-aware remaining useful life prediction. *Comput. Ind. Eng.* **2026**, *216*, 111979. [\[CrossRef\]](#)
27. Wang, X.; Xia, M.; Deng, W. MSRN-Informer: Time Series Prediction Model Based on Multi-Scale Residual Network. *IEEE Access* **2023**, *11*, 65059–65065. [\[CrossRef\]](#)
28. Luengo, D.; Martino, L.; Bugallo, M.; Elvira, V.; Särkkä, S. A survey of Monte Carlo methods for parameter estimation. *EURASIP J. Adv. Signal Process.* **2020**, *2020*, 25. [\[CrossRef\]](#)
29. Guo, X.; Mao, R.; Han, P.; Li, Z.; Zhang, H. A Physics-Informed Non-Markovian Deep Learning Model for Robust Ship Motion Prediction Under Non-Ideal Observations. *J. Mar. Sci. Eng.* **2026**, *14*, 1065. [\[CrossRef\]](#)

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