JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000864

Machine Learning Approximations for Fast and Accurate Prediction of Nonlinear Four-Wave Interactions in Spectral Wave Models

Jialun Chen[1], Thomas A. A. Adcock[1], Qingxiang Liu[2], Ronald Clark[3], and Tianning Tang[1,4]

[1]Department of Engineering Science, University of Oxford, Oxford, UK, [2]State Key Laboratory of Physical Oceanography, Ocean University of China, Qingdao, China, [3]Department of Computer Science, University of Oxford, Oxford, UK, [4]Department of Mechanical and Aerospace Engineering, University of Manchester, Manchester, UK

Key Points:

- An AI-driven framework is developed to enable efficient and accurate prediction of nonlinear four-wave interactions
- The proposed model was integrated into the WW3 framework and resulted in a stable model integration
- The proposed framework will help to unlock the vast potential of new generation of operational wave models

Correspondence to:
J. Chen,
jialun.chen@eng.ox.ac.uk

Citation:
Chen, J., Adcock, T. A. A., Liu, Q., Clark, R., & Tang, T. (2026). Machine learning approximations for fast and accurate prediction of nonlinear four-wave interactions in spectral wave models. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH000864. https://doi.org/10.1029/2025JH000864

Received 8 JUL 2025
Accepted 18 DEC 2025

Author Contributions:
Conceptualization: Jialun Chen, Thomas A. A. Adcock, Qingxiang Liu, Ronald Clark, Tianning Tang
Data curation: Jialun Chen
Formal analysis: Jialun Chen, Thomas A. A. Adcock, Qingxiang Liu, Tianning Tang
Funding acquisition: Jialun Chen
Investigation: Jialun Chen, Thomas A. A. Adcock, Tianning Tang
Methodology: Jialun Chen, Thomas A. A. Adcock, Qingxiang Liu, Ronald Clark, Tianning Tang
Project administration: Jialun Chen
Resources: Jialun Chen, Thomas A. A. Adcock, Tianning Tang
Software: Jialun Chen, Tianning Tang

Abstract Operational wave forecasting requires a delicate balance between the accuracy and computational speed of the spectral wave model used, in which the nonlinear wave–wave interaction "source" term plays an important role. The exact formulation of these nonlinear four-wave interactions requires solving the six-dimensional Boltzmann integral, an extremely time-consuming process that has challenged researchers for over half a century. To match the computational speed required by practical applications, almost all state-of-the-art operational wave models rely on simplified approximations such as Discrete Interaction Approximation (DIA) with known deficiencies. In this study, we consider four machine-learning models to estimate the nonlinear four-wave interaction term: a multilayer perceptron, a U-Net, and two variants of these architectures. The model is trained on ERA5 reanalysis wave spectra covering a wide range of geographical locations and thus of various sea-states. To evaluate its robustness, the model is first tested on ERA5 data at unseen locations. The results show that the proposed model is capable of predicting full nonlinear interaction at a comparable computational speed to DIA. We further assessed the model's performance on time-evolving wave growth simulations, which show that the proposed model generalizes well beyond its training conditions and shows strong potential for application in more complex wave scenarios. The present work focuses on deep-water conditions, but could be extended to finite depth. This new machine learning based approach not only paves the way for fast and accurate predictions of nonlinear source terms, but also heralds a new machine learning based pipeline for operational wave forecasting.

Plain Language Summary Accurate modeling of nonlinear four-wave interactions requires solving a six-dimensional Boltzmann integral—an extremely time-consuming process that has challenged researchers for over half a century. To enable practical application, state-of-the-art wave models have relied on simplified approximations of nonlinear interaction, failing to capture certain physics and thus leading to inaccurate wave forecasting. This work developed a machine learning framework to enable efficient and accurate prediction of nonlinear interactions and result in a stable time-evolving wave growth simulation. The proposed approach offers a potentially viable solution to this long-standing challenge.

1. Introduction

Numerous wave models have been developed since the 1960s, evolving over three generations to reach their current state for wind-wave forecasting and hindcasting, including models such as WAM (WAMDI Group, 1988), WAVEWATCH III (WW3) (Tolman, 1991), and traditionally for coastal regions, SWAN (Booij et al., 1999). Ideally, wave modeling requires solving the exact nonlinear four-wave interaction term, or source term, which is a six-dimensional Boltzmann integral—an extremely time-consuming process that has challenged researchers for over half a century (Hasselmann et al., 1985; Tracy & Resio, 1982; Van Vledder, 2006). To reduce the computation time in applications, almost all state-of-the-art operational wave models rely on the Discrete Interaction Approximation (DIA)—a simplified approximation developed by Hasselmann in the 1980s (Hasselmann et al., 1985). However, DIA is known to introduce inaccuracies, as it significantly reduces the number of interacting quadruplet configurations by considering only a specific combination of wave components and its mirror image. Specifically, DIA leads to unrealistically high energy at high frequencies with broad directional distributions (Hasselmann et al., 1985; Liu et al., 2019; Tolman, 2013).

Over the years, considerable efforts have been made to develop more accurate and computationally efficient parameterizations of the nonlinear wave–wave interaction term. Several studies have attempted to extend DIA by

Supervision: Jialun Chen, Thomas A. A. Adcock, Qingxiang Liu, Ronald Clark, Tianning Tang
Validation: Jialun Chen
Visualization: Jialun Chen, Tianning Tang
Writing – original draft: Jialun Chen
Writing – review & editing: Jialun Chen, Thomas A. A. Adcock, Qingxiang Liu, Ronald Clark, Tianning Tang

increasing the number of discrete interactions considered (e.g., Hashimoto & Kawaguchi, 2001; Tolman, 2013; Van Vledder, 2001), offering computational efficiency and the ability to capture some key features of the exact nonlinear interaction. In particular, the Generalized Multiple DIA (GMD; Tolman, 2013) can result in a good agreement for JONSWAP spectral distributions with a standard set of parameters at 1.5 to 3.5 times the cost of DIA, depending on the choice of GMD configurations. Diffusion-type operator methods (Zakharov & Pushkarev, 1999) are very computationally efficient, but have been shown to be less accurate than DIA (Benoit, 2005). Subsequently, a two-scale approximation (TSA) was introduced by Resio and Perrie (2008) and Perrie et al. (2013), providing accurate estimations of the nonlinear interaction for parametric spectra. However, further investigations are required for its application to more complex sea-states. More recently, Alday and Ardhuin (2023) applied a "coarse" GQM integration setting proposed in Gagnaire-Renou et al. (2010), achieving much higher accuracy than DIA at 7.5 times its computational cost.

Several studies have been developed from exact formulations by simplifying or reducing the integration space of the Boltzmann integral (e.g., EXACT-NL (WAMDI Group, 1988), the Webb–Resio–Tracy (WRT) method (Resio & Perrie, 1991; Tracy & Resio, 1982; Van Vledder, 2006; Webb, 1978), and the RIAM method (Komatsu & Masuda, 1996)). Despite these efforts, such methods remain too computationally expensive for direct use in operational wave forecasting models.

With recent advances in AI and the growing availability of real-world data, machine learning appears to be a potential alternative to overcome the limitations mentioned above. The use of neural networks to estimate nonlinear wave–wave interactions was first proposed by Krasnopolsky et al. (2001, 2002). In their approach, both the wave spectra and the corresponding nonlinear interaction terms were decomposed using Legendre and Fourier decompositions. A multilayer perceptron (MLP) neural network was then trained to learn the relationship between these decomposed representations. Once trained, the network's output is reverted back to the corresponding nonlinear interaction term. In a later study, Tolman et al. (2005) replaced the decomposition step by employing a two-dimensional Empirical Orthogonal Functions for the expansion of the spectra, resulting in a more accurate estimation of the interactions. To mitigate error accumulation over wave growth, an internal quality control mechanism was proposed by Krasnopolsky et al. (2008) and Tolman (2009), allowing the model to revert to full solutions when the error in MLP predictions becomes significant. However, the high computational cost associated with frequent full-solution updates limits the practicality of this approach. Wahle et al. (2009) and Puscasu (2014) further applied a direct mapping of the wave spectra onto the nonlinear interaction source term, demonstrating the feasibility of using neural networks to estimate nonlinear interactions for both single- and multi-modal spectra. The authors are also aware that there is a recent independent study (Ikuyajolu et al., 2025) that investigates the use of a very deep MLP model within a global simulation setting. The present work differs in terms of the underlying model architecture and integration strategy with the WW3 framework.

Previous machine learning studies have primarily employed simple feedforward neural network architectures, and further investigation is required before such approaches can be considered viable alternatives within the WW3 framework. A key limitation is that the predictive performance of these models is highly sensitive to the representativeness of the training data, which is challenging given that complex sea-states will occur and these can frequently be of particular practical interest (e.g., crossing seas (Adcock et al., 2011; Cavaleri et al., 2012)). Extrapolation beyond the domain covered during training may lead to substantial errors (Krasnopolsky et al., 2008). Additionally, most prior studies have focused on accurately mapping discrete wave spectra to their corresponding nonlinear interaction terms, yet mapping to test spectra with visual agreement between the predicted and exact nonlinear interaction terms does not guarantee better model behavior or stable integration due to the strong nonlinearity of the interactions (Tolman, 2010; Tolman & Grumbine, 2013). Although Tolman et al. (2005) explored the use of MLP-based approaches for a time integration scheme within a wave model, the MLP model struggled to reproduce accurate wave growth, as the model spectra progressively diverged from the training data set. Therefore, any parameterization of the nonlinear interaction term should be validated within a standard wave model to ensure that the simulated wave growth is consistent with results from exact formulations. Such validation is essential to assess whether machine learning-based models can truly generalize the physical process and be applied reliably in operational applications.

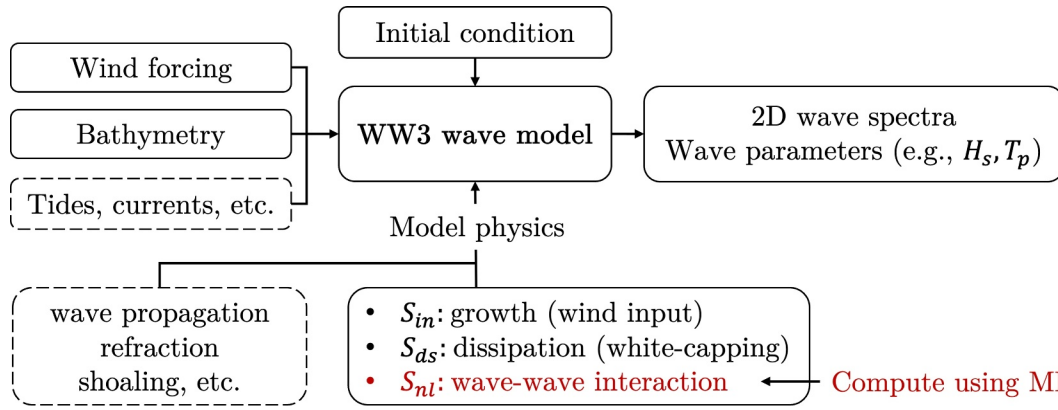

Figure 1. Schematic workflow of the WW3 wave model framework. In this study, a simplified single-grid-point setup in deep water is considered, excluding tidal effects, currents, and wave propagation, for instance. The original formulation for nonlinear wave–wave interactions is replaced by a machine learning model.

The shortcomings of basic neural network models have motivated the exploration of alternative machine learning architectures. We make the assumption that the two-dimensional wave spectrum and its corresponding nonlinear interaction can be considered as a classical 2-D mapping problem in image processing. This perspective opens the door to leveraging state-of-the-art techniques from computer vision as alternative solutions. Accordingly, we propose four machine learning based-models: a multilayer perceptron, a U-Net, a variant trained with the DIA as an additional input, and a post-processed filtered variant to address three key challenges in predicting nonlinear wave–wave interactions: (a) the high computational cost of traditional methods, (b) the limited generalization capability of existing models, (c) the instability during time evolution when integrated with current operational wave model framework and (d) uncertainty qualification of the predictions.

This paper is organized as follows. Section 2 outlines the theoretical background of the wave model and its governing equations. Section 3 presents four machine-learning models, outlines their architectures and their integration into the WW3 wave model. Section 4 introduces the data sets used for training and evaluation. Section 5 discusses the prediction performance of the models, with a focus on their ability to generalize in time-evolving wave growth test cases and their limitations in operational conditions. Finally, Section 6 provides a summary of key findings and directions for future research.

2. Wave Model Theoretical Background

The prediction of wind waves depends on accurately representing key source terms, including wind input, wave dissipation, nonlinear interactions, bathymetry, wave propagation, bottom friction and other physical processes (Komen et al., 1994). Figure 1 illustrates the workflow of the WW3 wave model. In the present work, we consider an idealized duration-limited wave growth case under deep water conditions, excluding effects such as tides, currents, and wave propagation. The simplest form of the spectral balance equation can be expressed as:

$$\frac{\partial F}{\partial t} = S_{tot} = S_{in} + S_{ds} + S_{nl}, \tag{1}$$

where S_{in} is the wind input source term, S_{ds} is the white-capping dissipation term and S_{nl} is the nonlinear four-wave interactions term. S_{tot} is the cumulative sum of these three key source terms, and F is the local wave energy spectrum.

The theory of nonlinear wave–wave interaction term is well-established, and this term plays an important role in wave modeling by re-distributing the energy, providing a stabilization mechanism for the shape of the spectrum and its evolution from short to long waves (K. Hasselmann et al., 1973). The basic equation describing the nonlinear wave–wave interactions was first introduced by K. Hasselmann (1962), known as the Boltzmann integral. This integral describes the rate of change of wave action density $n_i = n(\mathbf{k}_i) = F(\mathbf{k}_i)/\omega_i$ at a given wave number $\mathbf{k}_1$ due to resonant interactions with three other wave components $\mathbf{k}_2$, $\mathbf{k}_3$, $\mathbf{k}_4$, which is given by:

$$\frac{\partial n_1}{\partial t} = \iiint G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\,(n_1 n_3 (n_4+n_2) - n_2 n_4 (n_1+n_3)) \\ \delta(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4)\delta(\omega_1+\omega_2-\omega_3-\omega_4)\,d\mathbf{k}_2 d\mathbf{k}_3 d\mathbf{k}_4. \quad (2)$$

where the radian frequencies ω_i are related with the wavenumber $\mathbf{k}_i$ through linear dispersion relation. The term G is the complicated coupling coefficient given by Webb (1978), Herterich and Hasselmann (1980) and Zakharov (1998), and the δ functions reflect the resonance conditions, ensuring the conservation of wave energy, action and momentum.

3. Machine Learning Framework

In this study, we investigate four types of machine learning-based models: MLP and U-Net, together with their variants ML$^+$ and ML-F, where ML denotes the base model. The superscript $^+$ (e.g., U-Net$^+$) indicates that the model is trained with DIA as an additional input. For the machine learning–embedded WW3 integration scheme, models are denoted WW3–ML, and when a post-processed filtering step is applied, WW3–ML-F. For example, a WW3–U-Net-F$^+$ indicates this is a U-Net embedded WW3 integration scheme with DIA as an extra input and also with a post-processed filtering step applied during integration. This filtering step effectively mitigates spurious low energy transfer at large directional angles, improving stability over longer integration time steps. Details of the filtering process are provided in Section 5.2. We will first look at the multilayer perceptron—a method that has been widely used in previous studies on predicting nonlinear interactions. Then, we will look at the problem from an image processing perspective, where the U-Net architecture will be deployed. In the present work, the full nonlinear wave–wave interactions are calculated according to the WRT method as the target output for the machine learning model.

The multilayer perceptron (MLP) neural network model was inspired by biological nervous systems and functions as a universal approximator capable of mapping sets of input values to output values (Hornik et al., 1989). This model was applied in most of the previous machine learning-based studies for predicting nonlinear interactions (e.g., Krasnopolsky et al., 2001; Puscasu, 2014; Tolman et al., 2005; Wahle et al., 2009). However, MLPs are inherently limited in their ability to capture spatial dependencies and local structures in high-dimensional input data such as wave spectra. Because MLPs treat input features as flattened 1-dimensional vectors, they often struggle to generalize well across arbitrary spectral shapes, especially when extrapolating beyond the training distribution as highlighted by Tolman et al. (2005).

To address these limitations, we employ a U-Net model, a fully convolutional encoder–decoder architecture, tailored to learn the nonlinear mapping from the two-dimensional wave spectrum to the corresponding nonlinear interaction term. The U-Net architecture was originally proposed by Ronneberger et al. (2015) for medical image segmentation and is well suited for processing grid-structured data. It builds upon the foundational convolutional neural network framework (LeCun et al., 1989) and introduces a symmetric encoder–decoder structure with skip connections that link corresponding layers in the encoder and decoder. These connections preserve spatial information, facilitate more efficient optimization and mitigate the vanishing gradient problem during training (Li et al., 2018; Veit et al., 2016).

3.1. Multilayer Perceptron Neural Network Architecture

The MLP model used in this study consists of an input layer, three hidden layers, and an output layer. The configuration follows the earlier study on nonlinear interactions by Wahle et al. (2009) with first and third hidden layers containing 80 neurons each, while the middle bottleneck layer has 39 neurons. Preliminary experiments with alternative configurations indicate that these variations do not lead to significant improvements in model performance unless a very deep MLP architecture is used. Both the input wave spectrum and the target nonlinear interactions are represented as one-dimensional vectors of length 960 (24 directions × 40 frequencies based on the ERA5 data set; see Section 4 for a detailed description of the data set). The input layer passes this vector to the hidden layers, where each neuron is fully connected to every neuron in the subsequent layer. Neuron outputs are computed by applying a rectified linear unit (ReLU) activation function (Jarrett et al., 2009) to the weighted sum of inputs plus a bias term. This process is repeated through all hidden layers. The final output layer applies a linear transformation to produce the predicted nonlinear interaction term. The goal of the training process is to learn an

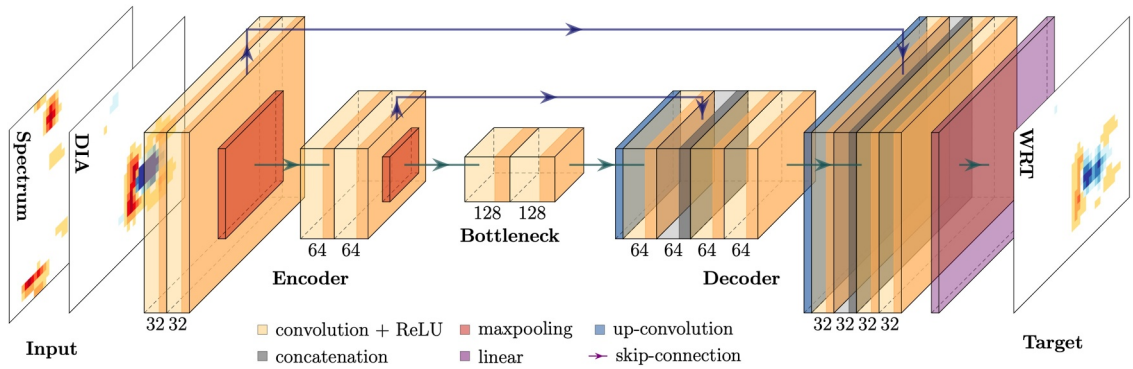

Figure 2. Fully convolutional encoder-decoder architecture with skip connections, designed to map input spectra and DIA terms to the corresponding WRT target terms. The input is processed through two successive 2D convolutional layers with ReLU activation (orange). Downsampling is performed using max-pooling layers (red), while upsampling is achieved through transposed convolutions (blue). Skip connections (purple arrow) help retain spatial information, and concatenation layers (gray) merge encoder and decoder features. The number of feature channels is indicated below each block.

approximation function by optimizing the model parameters—weights and bias terms. For more detailed on the architecture of MLP, one can refer to Goodfellow et al. (2016).

3.2. U-Net-Based Architecture

As depicted in Figure 2, the proposed U-Net architecture comprises two convolutional encoder blocks, each consisting of two consecutive 2D convolutional layers with 3×3 kernels and same padding to preserve spatial dimensions. Each convolutional layer uses the ReLU activation function and He initialization (He et al., 2015), followed by batch normalization for stable and efficient training. These layers are responsible for extracting local spatial features from the input features. Spatial downsampling is performed after each encoder block using a 2×2 max-pooling layer, which reduces the spatial resolution by a factor of two. The number of feature channels doubles at each level, starting from 32, then 64, and reaching 128 in the bottleneck, allowing the network to capture increasingly abstract features. The bottleneck serves as the transition between the contracting and expanding paths and operates at the lowest spatial resolution. Following this, two decoder blocks progressively upsample the feature maps using transposed convolutions with a stride of 2. At each stage, the upsampled output is concatenated with the corresponding encoder feature maps via skip-connections, followed by two convolutional layers. These connections enable the decoder to recover fine spatial details that are lost during pooling operations in the encoder. The number of feature channels is halved at each decoder level (from 128 to 64, and then to 32), and a final convolution layer with a single kernel and linear activation produces the output. It is noted that the DIA input in Figure 2 can be optionally included in the U-Net model to enhance prediction performance. This new model is denoted as U-Net^{+}, as discussed in the results section.

3.3. Uncertainty Quantification

To estimate the uncertainty of model predictions, a quantile-based loss function is developed, applicable to all ML models in this study. A 95% prediction interval (PI) indicates that roughly 95% of the true values are expected to lie within the predicted range. In conventional quantile regression (Or et al., 2020), separate models are trained for upper and lower bounds. For any quantile level $q \in (0, 1)$, the quantile loss is defined as

$$\rho_q(y, \hat{y}) = \begin{cases} q \cdot (y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (q-1) \cdot (y - \hat{y}) & \text{if } y < \hat{y}, \end{cases} \tag{3}$$

where y is the nonlinear interactions calculated using the WRT method and $\hat{y}$ is either the upper or lower quantile prediction. The loss penalizes over-and underestimation according to the specified quantile level q. For instance, when $q = 0.025$, underestimates of the output $\hat{y}$ are penalized less (with a coefficient of $q = 0.025$) than overestimates (with an equivalent coefficient of $1 - q = 0.975$). This allows the model to learn the 2.5th percentile, which corresponds to the lower bound of 95% PI. Inversely, $q = 0.975$, on the opposite side, the model learns the 97.5th percentile, which corresponds to the upper bound of 95% PI.

Table 1
Relative Computational Cost of Different Methods for Computing Nonlinear Interactions

Method	DIA	MLP	U-Net	U-Net$^+$	U-Net-F$^+$	WRT
Relative time	1.0	2.0	4.3	5.5	5.6	102.3

The formulation adopted in this study follows the same principle but predicts a mean $\tilde{y}$ and an associated deviation v, which define symmetric lower and upper bounds $\tilde{y} - v$ and $\tilde{y} + v$, respectively. The joint loss is defined as:

$$\mathcal{L}(y, \tilde{y}, v) = \rho_{0.025}(y, \tilde{y} - v) + \rho_{0.975}(y, \tilde{y} + v). \tag{4}$$

where y and $\tilde{y}$ are the WRT and predicted nonlinear interactions. The loss simultaneously penalizes over-and underestimation of the lower (2.5th percentile) and upper (97.5th percentile) bounds. Minimizing this loss trains the network to adjust both $\tilde{y}$ and v, so that the interval $[\tilde{y} - v,\ \tilde{y} + v]$ achieves approximately 95% coverage of the target values.

3.4. Training and Evaluation

The model architecture presented in this study was implemented in Python using the Keras (Chollet et al., 2017) and TensorFlow (Abadi et al., 2016) libraries. The training process applied Adam optimizer (Kingma & Ba, 2014), an adaptive variant of the stochastic gradient descent algorithm, to minimize the mean squared error between the predicted and target WRT terms and quantile loss (Equation 4) for prediction intervals. Training is performed using a batch size of 64 and a learning rate of 10^{-3}, with early stopping to prevent overfitting. The hyperparameters involved in the U-Net (such as the kernel size, the learning rate, the batch size, etc.) are tuned empirically and further optimization could potentially improve the model's accuracy.

It is noted that proper normalization is essential during training. Previous studies have addressed this by applying non-dimensional or normalized spectra and nonlinear interaction terms to ensure proper scaling behavior of the neural networks (Krasnopolsky et al., 2002; Tolman et al., 2005). In this study, the ERA5 data set contains multi-modal spectra, which makes it difficult to directly normalize nonlinear interactions based on spectral parameters (e.g., peak frequency, direction), since the maximum energy transfers often occur around lower spectral peaks. Hence, this work uses the absolute value of DIA to normalize both the DIA input and the WRT target, followed by division by their respective standard deviations. Spectral inputs are similarly normalized by their maximum values and then scaled by their standard deviations.

The training times for the MLP and U-Net models are approximately 1 and 6 hr, respectively, using a computer equipped with an 8-core CPU, a 14-core GPU, and 16 GB of RAM. As shown in Table 1, once trained, the computational cost of the MLP model is approximately double that of DIA, and the U-Net model is 4.3 times slower. When DIA is used as an additional input (i.e., U-Net$^+$), the model runs 5.5 times slower, yet still achieves roughly 20 times the cost of the WRT method. The additional filtering step adds only a minor computational overhead. It should be noted that the reported computation times are based on offline evaluations of the integration scheme and can vary depending on the operational configuration and data input/output overhead. The current implementation focuses on a single-grid offline model integration scheme, the performance and scalability in an online, global setup will be explored in future work.

To evaluate the prediction performance of different models, the normalized root mean squared error, $\mathcal{E}$, is considered in this study and defined as:

$$\mathcal{E} = \sqrt{\frac{1}{nm\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[y(f_i,\theta_j) - \tilde{y}(f_i,\theta_j)\right]^2}, \tag{5}$$

where $y(f_i,\theta_j)$ and $\tilde{y}(f_i,\theta_j)$ denote the WRT and predicted nonlinear interactions at the i-th frequency and j-th direction; $f_i (i = 1,\ldots,n)$ and $\theta_j (j = 1,\ldots,m)$ correspond to the frequency and directional components; $n = 40$ and $m = 24$ represent the total number of frequency and direction intervals in each ERA5 training and testing data set. The normalization factor σ^2 is the sample variance of the corresponding WRT term, computed over all frequency and direction spaces. A normalized error $\mathcal{E} = 1$ indicates that the prediction error is equivalent to predicting uniformly zero values. Further, the two-dimensional root mean squared error $\epsilon(f,\theta)$ averaged over N test cases can be expressed as:

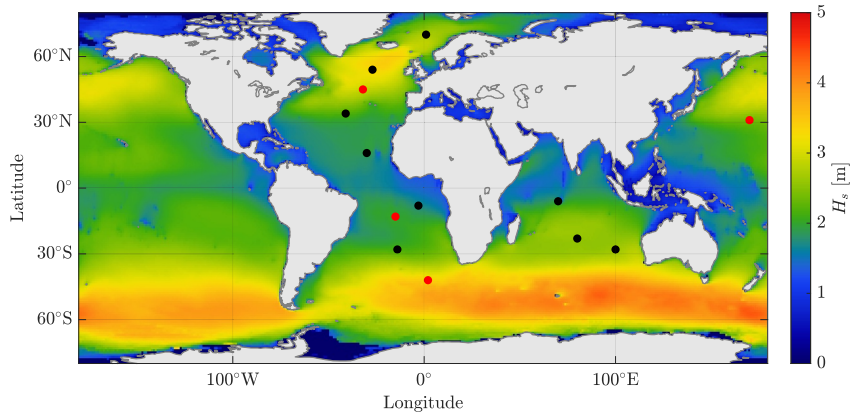

Figure 3. Training (black dots) and testing (red dots) locations selected from a wide range of geographical locations. The colormaps show the global hindcast of significant wave height H_s averaged over the year 2024.

$$\epsilon(f,\theta) = \sqrt{\frac{1}{N}\sum_{s=1}^{N}\left[y_s(f,\theta) - \tilde{y}_s(f,\theta)\right]^2}. \tag{6}$$

where the index $s = 1,\ldots,N$ denotes the test cases used for evaluation, with N denoting the total number of test samples (e.g., 2,928 samples at a single test location during 2024 with 3 hourly intervals).

3.5. Machine Learning Model Integration With WW3

As shown in Figure 1, the original WRT term in the WW3 scheme is replaced by the ML model trained on the ERA5 data set to reproduce wave growth and spectral evolution over time. Specifically, the modified WW3 model scheme retains the same source terms (i.e., wind input and dissipation). To avoid treating the dominant diagonal derivative terms of the source function required in implicit schemes, the integration of the source terms in the wave energy balance equation is performed using an explicit scheme (Liu et al., 2021; WAVEWATCH III Development Group, 2019). To ensure compatibility with WW3, the ML model is implemented as a callable function within the source code at every internal WW3 integration step, which is set to 60 s in this study. Over a 48-hr simulation, this results in 2,880 consecutive calls to the machine learning model to compute the nonlinear interaction. The input to the ML model is the wave spectrum at a grid point at each time step, preprocessed to match the normalization used during training. The model outputs the predicted nonlinear source term, which is then de-normalized and passed back to the WW3 integration scheme to update the wave spectrum. This setup enables a direct comparison between conventional methods and the ML-based models, offering insight into their compatibility and robustness in operational conditions.

As mentioned previously, one of the critical challenges in integrating the machine learning model into the WW3 framework is to ensure stable wave growth over time (Tolman & Grumbine, 2013), particularly the emergence of spurious low-energy components at wider directional angles (see Figure 8a). This issue has also been observed in earlier studies by Tolman (2009); Tolman et al. (2005), where MLP-based models result in the accumulation of small nonlinear transfer values, necessitating implicit treatment to maintain model robustness. Since the machine learning model is applied at every timestep, even small prediction errors can accumulate and gradually distort the spectral shape—especially under long-duration simulations. This makes it essential to ensure physical consistency and numerical stability throughout the integration.

To address this, we apply a post-processing step at each timestep using a filter derived from the DIA method. This DIA-based filter identifies regions where the nonlinear energy transfer is effectively zero and sets the corresponding ML-predicted values to zero. Notably, the DIA method is known to produce a broader range of nonlinear interactions than the exact WRT method, ensuring that regions with WRT-based energy transfer are unaffected. This filtering step helps suppress unrealistic energy growth in large directional angles, leading to more stable and physically consistent spectral evolution during the simulation. Although this filtering step is specifically tailored for the U-Net-based models and differs from the DIA-based filter proposed by Tolman (2011), both

approaches leverage the DIA configuration to enhance stability and accuracy by suppressing spurious energy transfer.

4. Data Collection and Presprocessing

Two-dimensional wave spectra used to train the machine learning model are obtained from the European Center for Medium-Range Weather Forecasting ERA5 reanalysis data set—a data set that combines in situ observations, satellite measurements, and large-scale simulations. For a more detailed description, see: https://www.ecmwf.int/en/forecasts/documentation-and-support/2d-wave-spectra. Specifically, the data has a resolution of 24 directions (with 15-degree increments) and 30 frequency components, ranging from 0.03 Hz $(f_{\min})$ to 0.55 Hz $(f_{\max})$. The frequencies follow a geometric progression with each successive component increasing by a factor of 1.1: $f(i) = f(i-1) \times 1.1$, where i is the frequency index. It is noted that wave energy decays slowly at the rear face of the spectrum and would still contribute to the nonlinear energy transfer process. Therefore, the data set is further extended to 40 frequency components with a f^{-5} high-frequency tail following Benoit (2007), corresponding to the maximum frequency of 1.42 Hz.

To train our machine learning model, input–output pairs are constructed. Each input $\mathbf{x} \in \mathbb{R}^{24\times40\times1}$ consists of a single wave spectrum, or $\mathbf{x} \in \mathbb{R}^{24\times40\times2}$ when the DIA-based nonlinear four-wave interaction is included as an additional input. Both inputs have a resolution of 24 directions and 40 frequencies. Each output set $\mathbf{y} \in \mathbb{R}^{24\times40\times1}$ is the corresponding nonlinear interaction term computed using the WRT method. In this study, we incorporate DIA as an additional input feature, extending beyond prior approaches that rely solely on the wave spectrum. Although DIA simplifies the six-dimensional interaction space by considering only a limited number of resonant quadruplets, it preserves some important characteristics of the exact solution. Specifically, the DIA reproduces the conservation of total spectral energy, wave action, and momentum by redistributing spectral energy across the four resonant wave components (K. Hasselmann, 1963). These conservation properties hold regardless of the number or arrangement of quadruplets, if each satisfies the resonance conditions. Furthermore, DIA downshifts the spectral peak toward lower frequencies and maintains a stable spectral shape at higher frequencies during wave growth, which is consistent with the WRT method (K. Hasselmann et al., 1973). This allows the machine learning model to leverage DIA's underlying physics, as demonstrated in the results section.

The training data set is obtained from 9 distinct deep-water geographic locations (indicated by black dots in Figure 3), spanning the period from 2016 to 2023, with a temporal resolution of three hours (a total of 210,384 spectra). The training data set is split into 80% for training and 20% for validation. These locations are selected to represent a wide range of wave characteristics and sea states across varying weather conditions, including both single-modal and multi-modal wave spectra, enhancing the model's ability to generalize to unseen cases. In this study, we use a more rigorous test data set to ensure the model's generalization. Instead of random splitting the full data set into training and testing, our test data is fully independent at a new location, which might exhibit different weather systems, and also covers a new time period. This gives us full confidence that the model will work well and align with operational requirements in deep water. This independent test data set contains data from 4 different locations (red dots in Figure 3) in 2024, with a 3-hr resolution (11,712 spectra). These test locations and time periods are not included in the training and validation data set to prevent data leakage and ensure a comprehensive assessment of the model's generalization to previously unseen conditions where the wave climate may be significantly different.

Once the model has been validated against unseen wave conditions from ERA5 reanalysis data, a more challenging and critical task is to deploy the model into WW3 and use the ML-embedded WW3 to reproduce wave growth under arbitrary wave conditions, which is crucial for validating the model's ability to capture wave growth and maintain energy balance over time when coupled with wind input and dissipation source terms. Details of this test case are provided in the next section.

5. Result

This section presents the prediction performance of nonlinear four-wave interactions using the proposed four types of machine learning models in comparison to WRT and DIA methods. We first evaluate the performance of the MLP, U-Net and U-Net$^+$ models using test cases from the ERA5 reanalysis data set. To further assess the model's compatibility within the WW3 framework, we integrate previously trained machine learning models

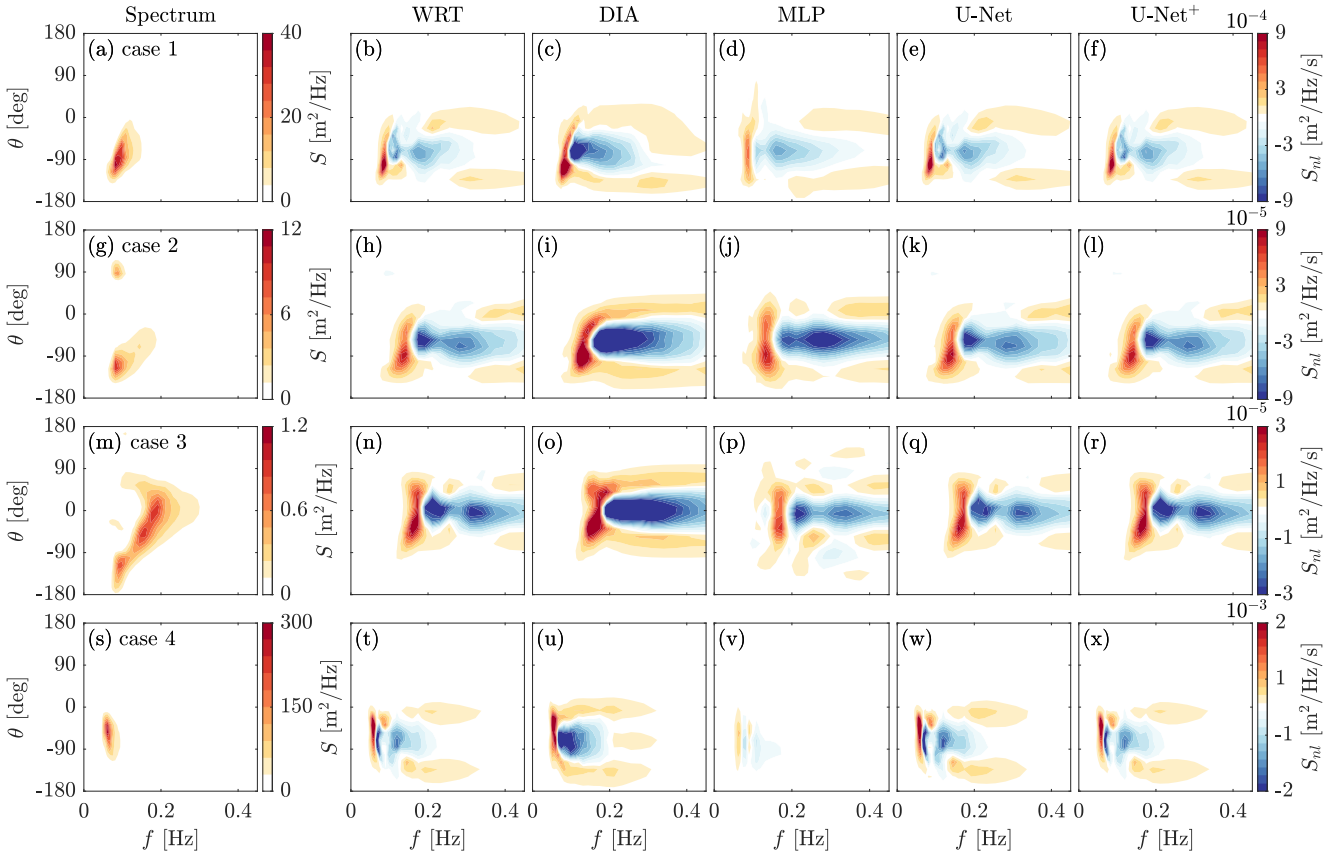

Figure 4. Representative examples of wave spectra and their corresponding nonlinear transfer terms. The left column shows three types of spectra: single-modal (first row), bi-modal (second row), and multi-modal (third row), with the fourth row showing an extreme case. The right columns display the corresponding nonlinear interactions computed using the WRT, DIA, MLP, U-Net and U-Net$^+$ models, respectively.

based on the ERA5 data set into the WW3 and perform a 48-hr simulation close to operational conditions. A post-processing filter, denoted "F", is further applied to enhance prediction accuracy for the integration scheme.

5.1. ERA5 Reanalysis Test Case

The test location at 42°S 2°E is first selected to assess model performance across varying spectral shapes. Figure 4 presents three typical wave spectra—single-modal (first row), bi-modal (second row), and multi-modal (third row), alongside their corresponding nonlinear interaction terms computed using WRT, DIA, MLP, U-Net and U-Net$^+$ models. A more direct comparison of integrated results for each case is presented in Figure 5, where the left panel shows the directionally integrated nonlinear transfer terms, and the right panel shows in frequency-integrated terms. To test models' robustness and generalization capability, two special cases were included: the third row shows a crossing-sea condition, and the fourth row shows an extreme case with H_s = 11.5 m obtained at 31°N 170°E. Notably, the fourth row represents a rare extreme event, with only 159 of the 210,384 training cases (0.08%) and 6 of the 11,712 test cases having $H_s \geq 10$ m.

The DIA method overestimates the magnitude of the positive lobe at low frequencies and the negative lobe across all cases, suggesting excessive energy transfer relative to the WRT term, consistent with previous studies (e.g., Hasselmann et al., 1985; Krasnopolsky et al., 2001). Further, DIA fails to capture the patterns of the negative lobe and the nonlinear energy transfers in the higher-frequency range. Broader directional spreading and larger estimation near the mean directions are evident in the right panel of Figure 5. These discrepancies are attributed to DIA's simplified representation of resonant interactions. The MLP model shows marginal improvements over DIA by partially capturing some high-frequency features. However, it tends to underestimate the positive peaks and introduces spurious energy transfer across broader directional ranges as the spectrum becomes increasingly complex, as shown in Figure 4p. Furthermore, the MLP model exhibits limited generalizability in certain

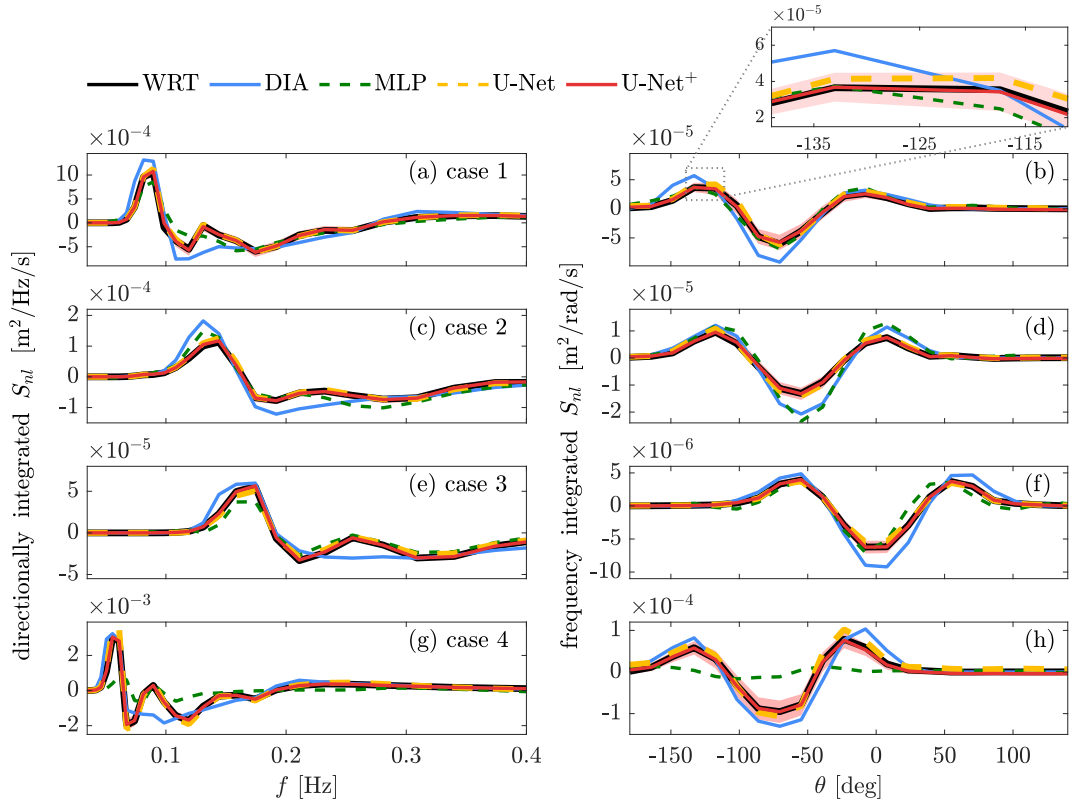

Figure 5. Comparison of nonlinear interaction terms integrated over direction (left column) and frequency (right column) for four cases presented in Figure 4. A zoomed-in view of figure (b) is displayed in the top-right corner. Each subplot compares the WRT (solid black) against DIA (solid blue), MLP (dashed green), U-Net (dashed yellow), and U-Net$^+$ (solid red) predictions. The red shaded regions represent the largest plausible variation based on 95% prediction intervals of U-Net$^+$ model.

scenarios. As shown in Figure 4v, the MLP model performs poorly on the out-of-distribution extreme case, highlighting its limited ability to extrapolate beyond the range of conditions represented in the training data, consistent with findings in Tolman (2009). In contrast, U-Net and U-Net$^+$ models accurately reproduce the WRT terms even under these conditions, indicating that their convolutional encoder–decoder structure captures local and multiscale patterns more effectively. By learning hierarchical representations of both local and global

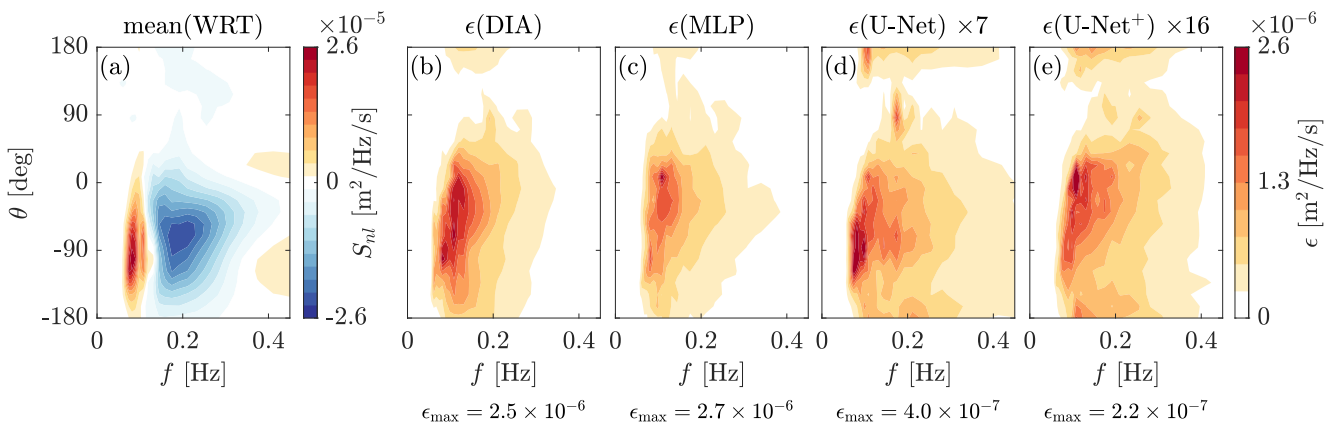

Figure 6. Comparison of root mean square errors (ϵ, Equation 6) at a single test location over a year across different methods: (a) mean of nonlinear interaction terms computed with the WRT method; (b) corresponding error between WRT and DIA; (c–e) errors between WRT and the ML models. Errors from U-Net and U-Net$^+$ are scaled by factors of 7 and 16, respectively, to enhance visibility. The maximum error for each method is shown below its corresponding subplot.

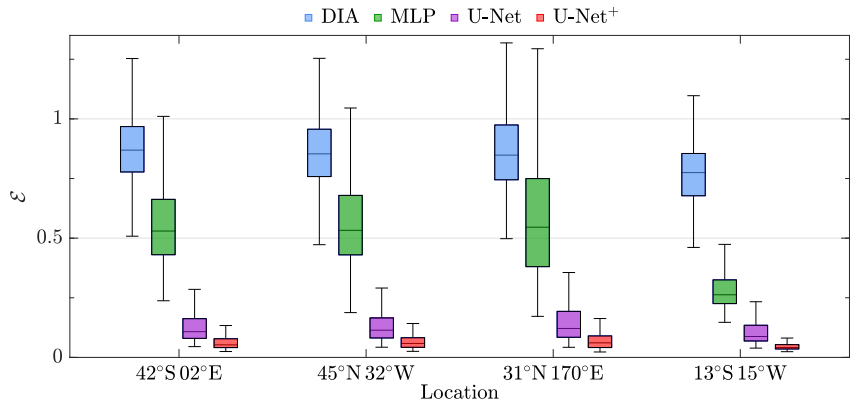

Figure 7. Boxplot of normalized root mean square error ($\mathcal{E}$, Equation 5) for DIA (blue), MLP (green), U-Net (purple), and U-Net$^+$ (red) across four test locations. The central line in each box represents the median error, while the box edges indicate the interquartile range (25th to 75th percentiles) and the whiskers extend to 1.5 times the interquartile range.

features, as discussed in Section 3.2, the U-Net architectures can capture fine-scale spectral or DIA details (e.g., spectral peaks) while retaining broad spectral context, thereby leading to enhanced generalization across all examined cases.

The red shaded regions in each subplot in Figure 5 represent the uncertainty estimated from the 95% prediction intervals (Equation 4) of U-Net$^+$ model. We note that integration is involved to produce the total energy transfer along frequency or direction in Figure 5. The uncertainty after the integration is, however, much trickier to obtain, since the underlying distribution and the dependencies among prediction points are unknown. Therefore, we used a conservative approach by taking the maximum possible variation, assuming that all prediction points could simultaneously reach their respective upper or lower quantile levels. The upper and lower bounds were obtained by integrating the per-frequency or per-direction 95% quantile estimates. The resulting envelopes thus represent the largest plausible variation based on 95% prediction intervals of individual points, and the actual variation is expected to be narrower than those presented here. The consistently narrow width of these intervals demonstrates the model's robustness for these unseen cases. It is also noted that the close agreement between training and test performance further confirms that the model is well-trained and not overfitting.

Notably, the enhanced accuracy of the U-Net$^+$ model compared to U-Net suggests that incorporating DIA as an additional input brings extra useful features that benefit the model accuracy. The inclusion of DIA was motivated by the hypothesis that combining a baseline physics-based estimate (DIA) with a data-driven model (U-Net) could improve accuracy and stability in simulating wave growth. This approach is conceptually analogous to the TSA method, in which the spectrum is decomposed into a broad-scale component, represented by a parametric spectral form and a local-scale residual component. The residual provides the flexibility needed for detailed balance and helps recover essential features of nonlinear interactions, particularly at frequencies above the spectral peak (Resio & Perrie, 2008). In U-Net$^+$, the DIA term is intended to provide a baseline approximation from an oceanography perspective that the network may refine. Additionally, the target WRT term describes the rate of change of the action spectrum due to nonlinear interaction, which exhibits distinct characteristics from the input spectrum itself. Providing an extra example of the spectral change rate term (i.e., DIA) avoids the model relearning the full nonlinear interactions mapping from the spectrum.

Figure 6a shows the mean of nonlinear energy transfer computed using the WRT method at a test location (42°S 2°E) with a total of 2,928 spectral cases. Figure (b) and (c–e) show the root mean square error (ϵ, Equation 6) of all 2,928 cases for the DIA and ML-based models, respectively. To enhance visual contrast, the errors of the two U-Net-based models are scaled by factors of 7 and 16, respectively. This scaling also highlights that the U-Net$^+$ model achieves accuracy close to 16 times better than that of the DIA method. The MLP model yields error levels comparable to the DIA method, reflecting its limited ability to generalize across varying wave characteristics. In contrast, both U-Net-based models achieve significantly lower errors, with the U-Net$^+$ model further reducing errors, particularly near 90°. Notably, the DIA method also exhibits relatively low errors in this

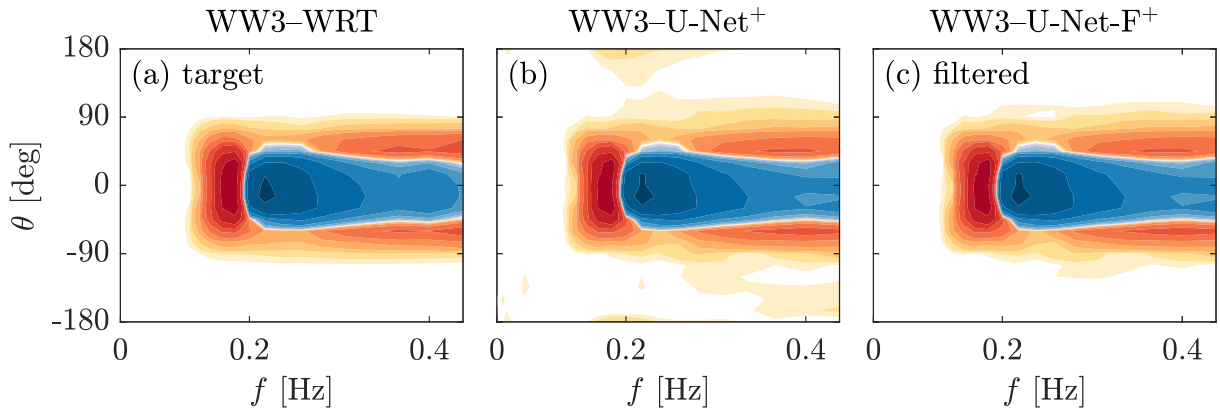

Figure 8. Time-evolving nonlinear interaction terms after 1 hr of model integration computed using (a) the WRT method, (b) the U-Net$^+$ model without filtering, and (c) the U-Net$^+$ model with a DIA-derived filter. The lowest contour levels for the nonlinear interactions are set to $\pm 4 \times 10^{-5}$ with an increment factor of 2 (i.e., $\pm 4 \times 10^{-5} \times 2^{\mu}$, with $\mu = 0, 1, \ldots, 10$). All plots are normalized by the square root of total energy.

region, suggesting that incorporating DIA as an additional input enhances the model's ability to capture spatial variability in the nonlinear energy transfer.

The boxplot in Figure 7 shows the normalized root mean square error ($\mathcal{E}$, Equation 5) for DIA (blue), MLP (green), U-Net (purple), and U-Net$^+$ (red) across four different test locations. Consistently higher errors, along with broader interquartile ranges in DIA results, indicate that DIA exhibits both greater bias and variability relative to all ML models. The wide spread in MLP errors suggests limited generalization across varying conditions. It can be seen that the MLP error is much lower at location 13°S 15°W as the test data set exhibits wave conditions and characteristics similar to those of the training data. The performance of the MLP model is expected to improve if the variability between the training and test sets is reduced, for instance by normalizing the mean wave directions. However, this is outside the scope of the present study and will not be investigated further here. In contrast, the U-Net-based models, particularly U-Net$^+$, demonstrate markedly improved accuracy and reduced variability across all sites. The consistently low errors and variability of the U-Net models underscore their robust generalization capability across diverse spectral conditions in the ERA5 reanalysis data set.

5.2. Time-Evolving Wave Growth Test Case

Previous results demonstrate that direct mapping of nonlinear wave interactions using machine learning achieves satisfactory performance under various wave conditions based on the ERA5 data set. Integrating these machine learning models into the operational forecast requires further investigation, in particular, the model stability for such an ML-embedded WW3 model over long-time simulations. As such, we performed further tests on the practicability of this ML-embedded WW3 model using a standard test case, which was also included in WW3 original validation. This is a duration-limited wave growth test case at a fixed location in deep water.

The initial wave condition selected is a JONSWAP spectrum (Hasselmann et al., 1973) with a peak frequency $f_p = 0.2$ Hz, Phillips coefficient $\alpha = 0.0081$, peak spectral width parameters $\sigma_a = 0.07$, $\sigma_b = 0.09$, peak enhancement factor $\gamma = 2$ and significant wave height of $H_s = 1.1$ m. The source term package ST4 (Ardhuin et al., 2010) is used for estimating the wind input and white-capping dissipation terms as it exhibits minimal bias in significant wave height compared to other source term packages (Stopa et al., 2016). The frequency-direction grid resolution is the same as the ERA5 reanalysis data set. A similar single-grid-point wave growth test is performed following Tolman and Grumbine (2013). The test uses a wind speed of $U_{10} = 20$ m/s, with the wind direction turning from 0° to −60° during the first 8-hr, shifting to 30° between 8 and 18-hr, and remaining constant until 48-hr.

As noted in earlier studies (Tolman, 2009; Tolman et al., 2005), the ML output may produce spurious low energy transfer in the nonlinear interaction space. These initial errors can accumulate and result in larger errors across broader directional ranges over time. To mitigate this issue, a post-processing step was applied at each timestep using a DIA-derived zero filter as shown in the shaded area. This filter sets the ML-predicted values to zero in regions where the DIA output is originally zero. Since DIA has a larger frequency and directional coverage than

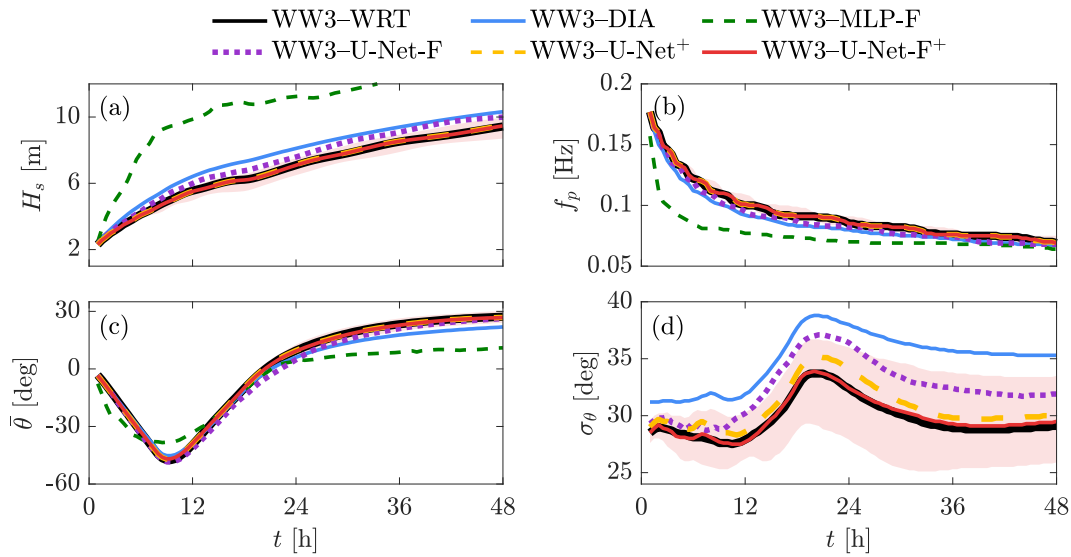

Figure 9. Time evolution of key wave parameters over a 48-hr wind-forced simulation, plotted at 30-min intervals: (a) significant wave height H_s, (b) peak frequency f_p, (c) mean wave direction $\bar{\theta}$, and (d) directional spreading σ_θ (noted that MLP model exhibits a large deviation $\sigma_\theta \in (41°, 59°)$, which is not shown). The upper and lower bounds of red shaded regions indicate maximum and minimum nonlinear transfers of WW3–U-Net-F$^+$ model within the 95% prediction intervals.

the WRT interactions term, this filtering approach effectively suppresses unrealistic low energy transfer across a wider directional range without interfering with valid interactions within the direction and frequency ranges in the WRT term.

Figure 8a shows that spurious energy appears and accumulates after 1-hr model integration. The lowest contour level is set to a small value (i.e., $\pm 4 \times 10^{-5}$) to enhance the visibility of low-energy transfer around U-Net$^+$ output. Although these low energy transfers may not be fully covered by the filter range in each integration time step, applying the filter at each integration step is sufficient to eliminate most of these low energy transfers, resulting in more stable model integration and improved accuracy in both mean wave direction and directional spreading as evident in Figure 8c. A detailed analysis of the model integration performance will be presented later, the results here primarily illustrate the effectiveness of the filtering process.

The time evolution of four key wave parameters over a 48-hr wind-forced simulation is shown in Figure 9, with spectral outputs sampled at 30-min intervals starting from the 1-hr forecast. The significant wave height, H_s, defined as the average height of the highest one-third of waves, is computed as $H_s = 4\sqrt{M_0}$, where M_0 is the zeroth spectral moment; peak frequency, f_p is derived from the 1D frequency spectrum using a parabolic fit at the discrete peak; mean wave direction, $\bar{\theta}$ is the energy-weighted average of wave directions across all frequencies derived from the 2D wave spectrum; wave spreading, σ_θ, is the directional spreading function (Kuik et al., 1988), describes the distribution of wave energy across different directions. The WW3–DIA method overestimates the significant wave height H_s and directional spreading σ_θ, while underestimating the peak frequency f_p and deviating from the reference mean wave direction $\bar{\theta}$. Although the WW3–MLP-F model demonstrates marginal improvements over WW3–DIA on ERA5-based test cases, it clearly fails to reproduce the correct wave growth within the WW3 integration scheme. As shown in Figure 9d, the WW3–MLP-F model significantly overestimates σ_θ, reaching $\sigma_\theta = 48°$ after the first hour, well beyond the displayed y-axis range. This is because of the limited generalization capability of the MLP model. The initial JONSWAP spectrum has different spectral characteristic relative to the ERA5-based training set, leading to large errors in the predictions of nonlinear interaction terms. Given its inability to capture nonlinear evolution and maintain stable model integration, the WW3–MLP-F model is excluded from subsequent analyses. In contrast, the WW3–U-Net-F model closely reproduces the WRT-based wave evolution across all parameters. However, it is observed that there is a tendency to overestimate directional spreading and deviate from other model parameters over longer time steps. This raises the question of whether the WW3–U-Net-F model is inherently stable, which will be investigated further in the following section. By incorporating DIA as an additional input, the WW3–U-Net$^+$ (without DIA-derived filter) and WW3–U-Net-F$^+$

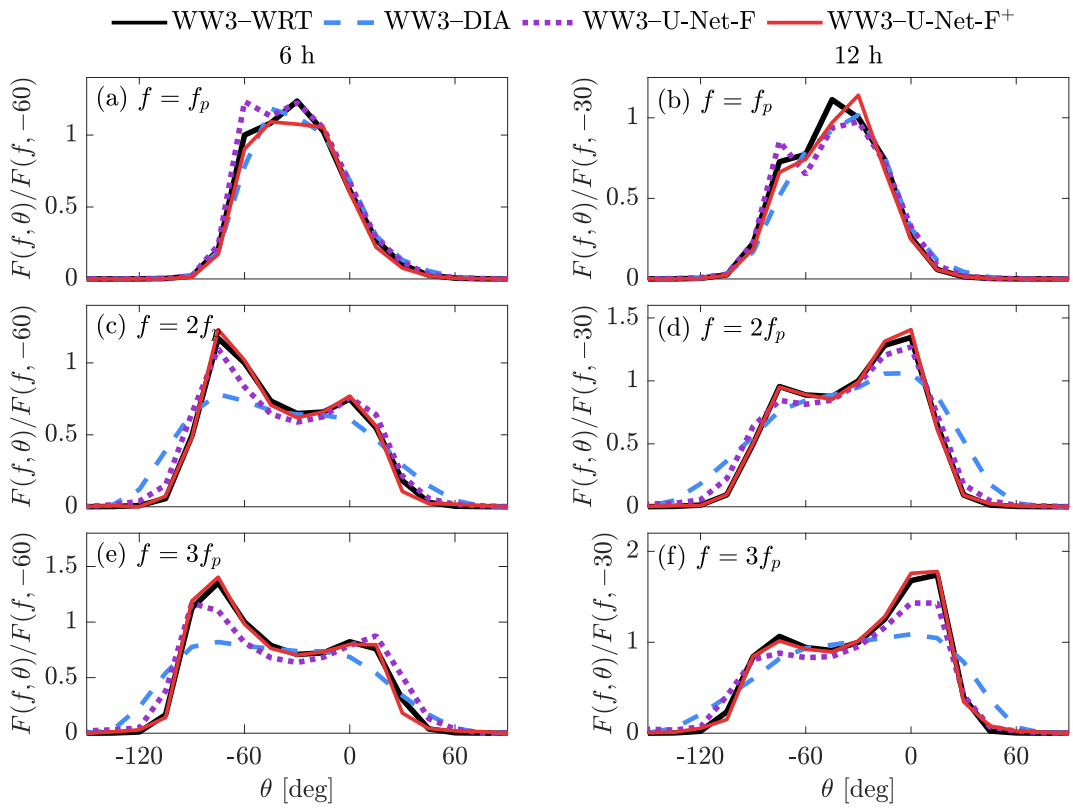

Figure 10. Comparison of the normalized angular distribution of the wave spectrum after 8 and 12-hr model integration. Panel (a, b) corresponds to $f = f_p$, panel (c, d) to $f = 2f_p$ and panel (e, f) to $f = 3f_p$.

models both achieve excellent agreement with WW3–WRT across all metrics, with WW3–U-Net$^+$ slightly overestimating σ_θ due to accumulated errors of spurious small energy.

To visualize the associated uncertainty in wave evolution, red shaded regions produced by the WW3–U-Net-F$^+$ model are shown for each parameter. The upper and lower bounds correspond to the maximum and minimum nonlinear energy transfers at each integration time step. Although the quantile-based approach provides 95% prediction intervals for individual nonlinear energy transfer terms, these intervals represent uncertainty in the rate of energy redistribution within the wave spectrum rather than in the spectral energy itself. Consequently, to approximate the uncertainty bounds of integrated wave parameters over time, we evaluate two bounding scenarios: one corresponding to the case of maximum energy transfer, and the other to minimal energy transfer. These scenarios define the envelope of plausible spectral evolution driven by nonlinear interactions, and are used to quantify the range of variability in the predicted wave parameters across successive time steps.

The upper bound of the simulation predictions is obtained using the 95% upper quantile estimates of positive nonlinear interaction energy transfer (i.e., larger energy gain) and the 95% lower quantile estimates of negative energy transfer (i.e., larger energy loss), leading to a faster downshift of energy toward lower frequencies. Conversely, the lower bound of the simulation predictions corresponds to weaker nonlinear transfers and a slower downshift. It is noted that the selection of maximum and minimum energy transfers does not consider energy conservation but is intended to provide an approximation of the prediction intervals. In Figure 9d, the directional spreading shows the largest uncertainty bound, mainly due to the accumulation of small spurious energy contributions over time, as illustrated in Figure 8. Although a DIA-based filter was applied to mitigate part of this error across successive time steps, slightly broader directional spreading still occurred compared with the WRT method, leading to a wider prediction interval.

The wave spectra at f_p, $2f_p$, and $3f_p$ after 6 and 12-hr of simulation are shown in Figure 10. Spectral energy is normalized by $F(f,\ -60°)$ at 6-hr and $F(f,\ -30°)$ at 12-hr to highlight directional features. While the WW3–DIA method performs reasonably well near the spectral peak at f_p, it fails to capture the bimodal structure at higher

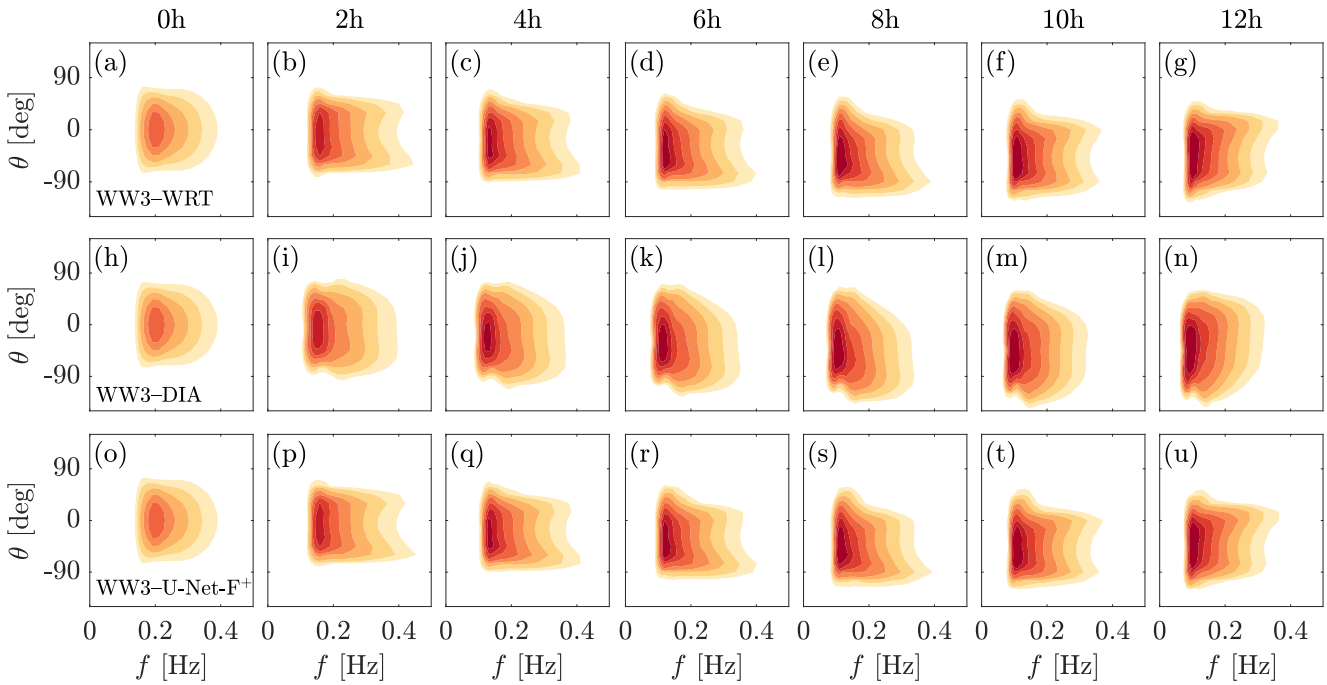

Figure 11. Comparison of 2D logarithm contour plots for spectral evolution under wind forcing over a 12-hr simulation, plotted with 2-hr intervals. The lowest contour level for the spectrum is set to 0.04, with an increment factor of 2. All spectra are normalized by their respective square root of total energy.

harmonics ($2f_p$ and $3f_p$) and tends to produce excessively broad directional spectra. In contrast, the WW3–U-Net-based predictions closely follow the WW3–WRT method at both 6 and 12 hr. Although minor discrepancies are observed near the peak at f_p, the bimodal spectral shapes at $2f_p$ and $3f_p$ are accurately reproduced, highlighting its strong generalization capability in wave growth cases, and overall good stability in wave growth.

The time-evolving wave growth under wind forcing on a logarithmic scale is illustrated in Figure 12 (spectral evolution) and Figure 11 (nonlinear interactions) over the first 12-hr simulation period. In the DIA-based simulation, the wave spectrum remains predominantly unimodal following the shift in wind direction and a slower response to the new wind than the WRT-based simulation. Further, the DIA results exhibit energy transfer across a wide directional range. In contrast, the WRT shows a secondary peak aligned with the new wind direction, resulting in a bimodal spectrum. The ML-based simulation closely reproduces this bimodal behavior, capturing both the directionally evolving energy transfer and the redistribution of spectral energy. As shown in Figure 11, some low energy transfer begins to accumulate at the edges of the nonlinear interaction region after 4 hr of simulation. However, the energy associated with this low energy transfer is minimal and does not have a noticeable impact on the overall spectral growth.

It is noted that the same test case was also performed with a constant wind direction over a 48-hr simulation (results not shown), as well as for a higher peak frequency (e.g., $f_p = 0.3$ Hz), representative of young sea states as the initial condition. The resulting wave direction deviated by approximately $\pm 0.5°$ from the mean wave direction $\bar{\theta} = 0°$, which is expected given that the ML output is not perfectly asymptotic. Nevertheless, all key parameters remained in excellent agreement, further demonstrating the model's robustness in simulating a general wave growth test case.

Conservation of energy is a key property of nonlinear four-wave interactions. Therefore, it is important to evaluate whether the predictions of the ML models preserve the conservation characteristics. Figure 13 shows the percentage ratio of the hourly integrated nonlinear interaction term S_{nl} to the integrated spectral energy S. After the initial growth, both WRT and DIA approach near zero and retain conservation. The early deviations are expected and are associated with the maximum cut-off at high-frequency energy (Tsagareli et al., 2010). In contrast, the MLP model fails to maintain the conservation balance, leading to substantial discrepancies

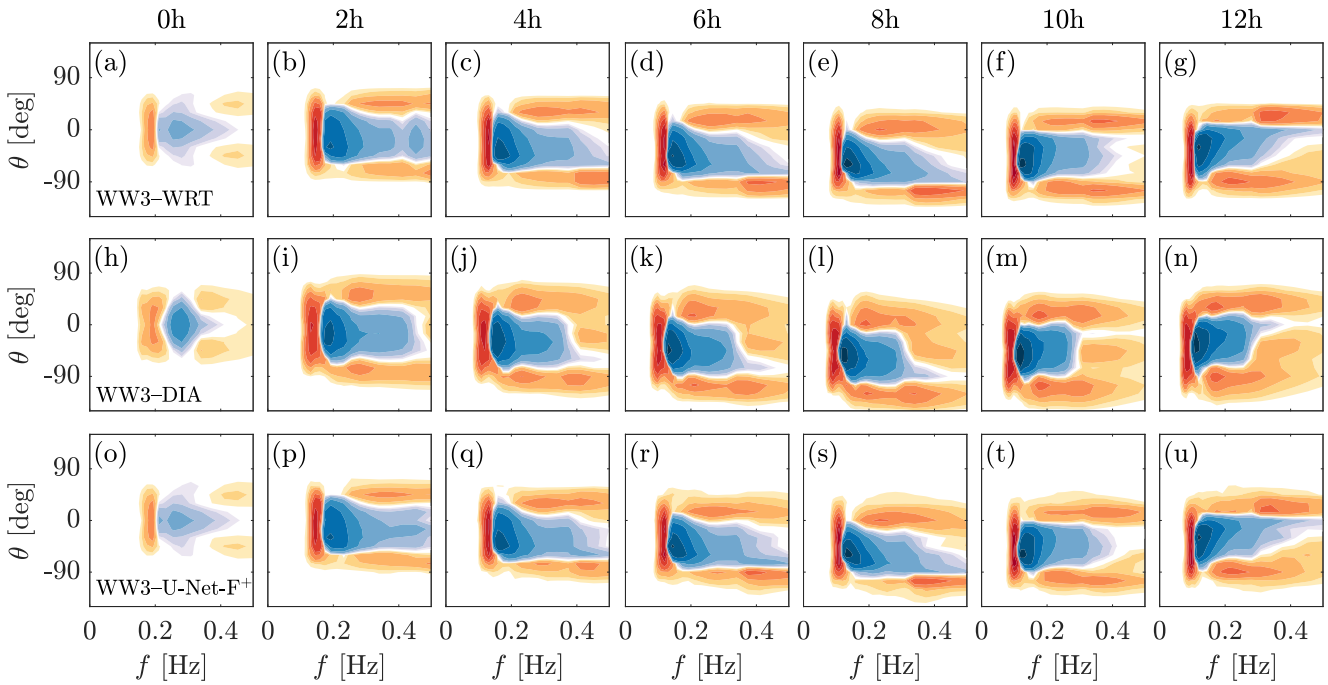

Figure 12. Comparison of nonlinear interactions in logarithm scale for wave spectra shown in same Figure. The lowest contour levels for the nonlinear interaction are set to $\pm 4 \times 10^{-4}$, with an increment factor of 2. All nonlinear interactions are normalized by their square root of total energy.

throughout the simulation (percentage ratio scale shown on the right-hand *y*-axis). Loss of conservation causes small imbalances to accumulate over time, resulting in unrealistic spectral growth or decay.

The WW3–U-Net-F$^+$ preserves the conservation balance more closely than WW3–U-Net-F, indicating that the inclusion of DIA as an additional input helps the network implicitly learn this property from the training data. Minor discrepancies appear during the early stages of integration, but the overall energy loss remains small relative to the total spectral energy. The filtered predictions (WW3–U-Net-F$^+$) exhibit smaller changes in total energy compared with the unfiltered predictions (WW3–U-Net$^+$). This suggests that the improvement from filtering is not linked to energy conservation, since the amount of energy affected by filtering at each integration

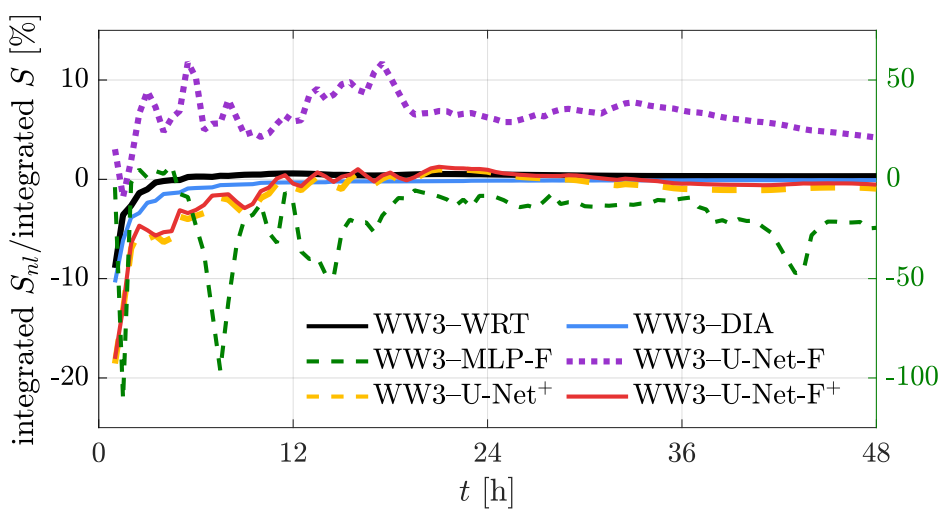

Figure 13. The percentage ratio of the hourly integrated nonlinear interaction term to the integrated spectral energy over a 48-hr simulation for different methods, with the MLP ratio scale shown on the right-hand *y*-axis.

time step is orders of magnitude smaller than the dominant energy transfer. Incorporating such a constraint may further enhance conservation law and will be explored in future work.

5.3. Discussion

From the above studies, the proposed WW3–U-Net-F^+ model shows excellent agreement with the WRT method and is expected to perform reasonably well with more complex scenarios. This is evident by its ability to generalize well in two distinct test settings using a model trained on the ERA5 reanalysis data set: (a) to different time periods and geographic locations with a wide range of sea states within ERA5 data set, and (b) to model integration tests initialized with an arbitrary JONSWAP spectrum, which exhibits characteristics distinct from the ERA5 training data. However, there are several limitations that need to be considered.

All ML models used in this work were trained and evaluated under deep-water conditions. In shallow water, triad interactions may dominate over nonlinear four-wave interactions, and model performance may decrease under such conditions. The framework can, however, be extended to shallow water by applying a depth scaling function similar to that used in DIA. DIA is an inherently deep-water approximation, and finite-depth effects are commonly introduced through depth-scaling terms (Hasselmann et al., 1985), with further extension for extremely shallow water conditions in Tolman (2013). Such approaches may be applied as a post-processing step to extend our work to limited water depths. The current model architecture also allows a more rigorous extension, which would involve expanding the training data set to include shallow and intermediate depth cases, or incorporating depth as an additional input feature so that depth dependence is captured directly by the model.

The present ML architectures are limited to the fixed frequency–direction discretizations used in the training data (24 directions × 40 frequencies in this study based on the ERA5 data set) and require retraining for alternative grids. This limitation could be addressed with operator-learning approaches, such as Fourier Neural Operators, which are capable of generalizing across discretizations and will be explored in future work. Furthermore, training in this study relied on ERA5 reanalysis spectra and the ST4 wind input and dissipation package, which was originally calibrated in combination with the DIA method (Ardhuin et al., 2010). Model performance may differ when coupled with other source-term packages, potentially requiring further adjustment or retraining. The complexity of real-world observational data also presents challenges, such as radar, buoys, or other in situ observations, which are subject to measurement noise and sampling errors.

Another limitation is the interpretability of the ML models. Incorporating DIA as an additional input improves prediction performance, as the network can effectively learn to correct the residual errors between DIA and WRT rather than reproducing the full WRT physics directly. However, the proposed ML models function as a black box, the extent to which they capture underlying physical processes, rather than simply correcting systematic DIA biases remains an open question.

Finally, the proposed U-Net model and its variants are substantially faster than the WRT method, but they remain 4.3 to 5.6 times the cost of DIA. Moreover, while more complex architectures or model configurations could potentially enhance generalization and accuracy, such improvements are often associated with increased computational cost. Consequently, the trade-off between accuracy and efficiency needs to be carefully considered in operational settings.

6. Conclusion

The nonlinear energy transfer arising from resonant four-wave interactions forms a fundamental component of modern spectral wave models. Since the 1980s, the Discrete Interaction Approximation (DIA) has enabled a new generation of operational wave modeling and remains widely used in both forecasting and hindcasting applications. Although alternative methods for more accurate and efficient computation were developed over the years, these are often tailored to specific spectral conditions and lack general applicability.

In this study, a machine learning-based model was developed for estimating the nonlinear four-wave interaction term and was evaluated using both ERA5 reanalysis data and time-evolving wave growth simulations. Once trained, the model achieves results that are comparable with those of the benchmark Webb–Resio–Tracy (WRT) method. For the ERA5 data set, it perfectly reproduces WRT nonlinear source terms with significantly improved accuracy over DIA, while maintaining a similar level of computational efficiency. The proposed model also

demonstrates strong generalization across various wave conditions at different locations and time periods. Notably, it learns to correct the errors of DIA inputs in overestimating the energy shifts in frequency space.

Further to the standard test cases, we also performed a stability test when embedding machine learning models for long-term forecasting with time integration. In such a time-evolving wave growth test case, the model remains stable and provides an accurate estimation of bulk parameters throughout a 48-hr duration with predefined JONSWAP spectrum and wind forcing, even without retraining. It clearly shows that the spectral evolution based on the proposed model accurately aligns with changing wind directions and captures the bi-modal spectral shapes. These results indicate that the proposed model generalizes well beyond its training conditions and shows strong potential for application in more complex wave scenarios. This highlights the possibility of using machine learning models as a replacement for DIA in operational wave models. In this test, the MLP models, however, cannot provide stable predictions, and the results diverge fairly quickly as the simulation runs further. The results also hint at the importance of including similar stability tests for ML-embedded wave forecast models and ensuring proper conservation properties of the interactions. Future efforts will further validate against real-world wave measurement data and more realistic settings to enhance its applicability for operational forecasting.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The ERA5 reanalysis data supporting the analyses in this study are available at https://apps.ecmwf.int/data-catalogues/era5. The code used to build and train the machine learning models, as well as to integrate them into WAVEWATCH III, is available at Chen et al. (2025).

Acknowledgments
This research is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of Schmidt Sciences. Q.L. acknowledges the support from the Shandong Provincial Natural Science Fund for Excellent Young Scientists Fund Program (Overseas) (2023HWYQ-056), the National Natural Science Foundation of China (42106012), the Taishan Scholars Program (tsqnz20221111) and the Fundamental Research Funds for the Central Universities (202441007). The authors are grateful to Jean Bidlot, James N. Steer, Ruslan M. Puscasu and Olawale J. Ikuyajolu for very helpful discussion.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. Retrieved from https://www.tensorflow.org

Adcock, T. A. A., Taylor, P. H., Yan, S., Ma, Q., & Janssen, P. A. E. M. (2011). Did the Draupner wave occur in a crossing sea? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *467*(2134), 3004–3021. https://doi.org/10.1098/rspa.2011.0049

Alday, M., & Ardhuin, F. (2023). On consistent parameterizations for both dominant wind-waves and spectral tail directionality. *Journal of Geophysical Research: Oceans*, *128*(4), e2022JC019581. https://doi.org/10.1029/2022jc019581

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., et al. (2010). Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *Journal of Physical Oceanography*, *40*(9), 1917–1941. https://doi.org/10.1175/2010jpo4324.1

Benoit, M. (2005). Evaluation of methods to compute the non-linear quadruplet interactions for deep-water wave spectra. In *Proceedings of the 5th International symposium on ocean wave measurement and analysis* (pp. 3–7).

Benoit, M. (2007). Implementation and test of improved methods for evaluation of nonlinear quadruplet interactions in a third generation wave model. In *Coastal Engineering 2006* (pp. 526–538). World Scientific. https://doi.org/10.1142/9789812709554_0046

Booij, N., Ris, R. C., & Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research*, *104*(C4), 7649–7666. https://doi.org/10.1029/98jc02622

Cavaleri, L., Bertotti, L., Torrisi, L., Bitner-Gregersen, E., Serio, M., & Onorato, M. (2012). Rogue waves in crossing seas: The louis majesty accident. *Journal of Geophysical Research*, *117*(C11). https://doi.org/10.1029/2012jc007923

Chen, J., Adcock, T., Liu, Q., Clark, R., & Tang, T. (2025). Machine learning approximations for fast and accurate prediction of nonlinear four-wave interactions in spectral wave models. *Zenodo*. https://doi.org/10.5281/zenodo.17154355

Chollet, F. (2017). *Deep learning with python*. Manning Publications.

Gagnaire-Renou, E., Benoit, M., & Forget, P. (2010). Ocean wave spectrum properties as derived from quasi-exact computations of nonlinear wave-wave interactions. *Journal of Geophysical Research*, *115*(C12). https://doi.org/10.1029/2009jc005665

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. MIT Press Cambridge, (Vol. 1(2)).

Hashimoto, N., & Kawaguchi, K. (2001). Extension and modification of Discrete Interaction Approximation (DIA) for computing nonlinear energy transfer of gravity wave spectra. In *Ocean wave measurement and analysis* (pp. 530–539).

Hasselmann, K. (1962). On the non-linear energy transfer in a gravity-wave spectrum Part 1. General theory. *Journal of Fluid Mechanics*, *12*(4), 481–500. https://doi.org/10.1017/s0022112062000373

Hasselmann, K. (1963). On the non-linear energy transfer in a gravity wave spectrum Part 2. Conservation theorems; wave-particle analogy; irrevesibility. *Journal of Fluid Mechanics*, *15*(2), 273–281. https://doi.org/10.1017/s0022112063000239

Hasselmann, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., et al. (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A*. Retrieved from https://www.researchgate.net/publication/256197895_Measurements_of_wind-wave_growth_and_swell_decay_during_the_Joint_North_Sea_Wave_Project_JONSWAP

Hasselmann, S., Hasselmann, K., Allender, J., & Barnett, T. (1985). Computations and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. Part II: Parameterizations of the nonlinear energy transfer for application in wave models. *Journal of Physical Oceanography*, *15*(11), 1378–1391. https://doi.org/10.1175/1520-0485(1985)015<1378:capotn>2.0.co;2

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on computer vision* (pp. 1026–1034).

Herterich, K., & Hasselmann, K. (1980). A similarity relation for the nonlinear energy transfer in a finite-depth gravity-wave spectrum. *Journal of Fluid Mechanics*, *97*(1), 215–224. https://doi.org/10.1017/s0022112080002522

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

Ikuyajolu, O. J., Van Roekel, L. P., Brus, S. R., & Thomas, E. E. (2025). NLML: A deep neural network emulator for the exact nonlinear interactions in a wind wave model. *Authorea Preprints*. https://doi.org/10.22541/essoar.174366388.80605654/v1

Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In *International Conference on computer vision* (pp. 2146–2153).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980

Komatsu, K., & Masuda, A. (1996). A new scheme of nonlinear energy transfer among wind waves: RIAM method-algorithm and performance. *Journal of Oceanography*, *52*(4), 509–537. https://doi.org/10.1007/bf02239052

Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., Janssen, P., et al. (1994). *Dynamics and modelling of ocean waves* (Vol. 532). Cambridge University Press. https://doi.org/10.1017/cbo9780511628955

Krasnopolsky, V. M., Chalikov, D. V., & Tolman, H. L. (2001). Using neural network for parameterization of nonlinear interactions in wind wave models. In *IJCNN'01. International Joint Conference on neural networks. Proceedings (Cat. No. 01CH37222)* (Vol. 2, pp. 1421–1425). https://doi.org/10.1109/ijcnn.2001.939570

Krasnopolsky, V. M., Chalikov, D. V., & Tolman, H. L. (2002). A neural network technique to improve computational efficiency of numerical oceanic models. *Ocean Modelling*, *4*(3–4), 363–383. https://doi.org/10.1016/s1463-5003(02)00010-0

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Tolman, H. L., & Belochitski, A. A. (2008). Neural network approach for robust and fast calculation of physical processes in numerical environmental models: Compound parameterization with a quality control of larger errors. *Neural Networks*, *21*(2–3), 535–543. https://doi.org/10.1016/j.neunet.2007.12.019

Kuik, A. J., Van Vledder, G. P., & Holthuijsen, L. H. (1988). A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography*, *18*(7), 1020–1034. https://doi.org/10.1175/1520-0485(1988)018<1020:amftra>2.0.co;2

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). *Visualizing the loss landscape of neural nets*. Advances in Neural Information Processing Systems, (Vol. 31).

Liu, Q., Gramstad, O., & Babanin, A. (2021). Kinetic equations in a third-generation spectral wave model. *Journal of Fluid Mechanics*, *910*, A50. https://doi.org/10.1017/jfm.2020.1036

Liu, Q., Rogers, W. E., Babanin, A. V., Young, I. R., Romero, L., Zieger, S., et al. (2019). Observation-based source terms in the third-generation wave model WAVEWATCH III: Updates and verification. *Journal of Physical Oceanography*, *49*(2), 489–517. https://doi.org/10.1175/jpo-d-18-0137.1

Or, D. B., Kolomenkin, M., & Shabat, G. (2020). Generalized quantile loss for deep neural Networks. *arXiv preprint arXiv:2012.14348*. https://doi.org/10.48550/arXiv.2012.14348

Perrie, W., Toulany, B., Resio, D. T., Roland, A., & Auclair, J.-P. (2013). A two-scale approximation for wave–wave interactions in an operational wave model. *Ocean Modelling*, *70*, 38–51. https://doi.org/10.1016/j.ocemod.2013.06.008

Puscasu, R. M. (2014). Integration of artificial neural networks into operational ocean wave prediction models for fast and accurate emulation of exact nonlinear interactions. *Procedia Computer Science*, *29*, 1156–1170. https://doi.org/10.1016/j.procs.2014.05.104

Resio, D. T., & Perrie, W. (1991). A numerical study of nonlinear energy fluxes due to wave-wave interactions Part 1. Methodology and basic results. *Journal of Fluid Mechanics*, *223*, 603–629. https://doi.org/10.1017/s002211209100157x

Resio, D. T., & Perrie, W. (2008). A two-scale approximation for efficient representation of nonlinear energy transfers in a wind wave spectrum. Part I: Theoretical development. *Journal of Physical Oceanography*, *38*(12), 2801–2816. https://doi.org/10.1175/2008jpo3713.1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, proceedings, Part III* (Vol. 18, pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Stopa, J. E., Ardhuin, F., Babanin, A., & Zieger, S. (2016). Comparison and validation of physical wave parameterizations in spectral wave models. *Ocean Modelling*, *103*, 2–17. https://doi.org/10.1016/j.ocemod.2015.09.003

Tolman, H. L. (1991). A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. *Journal of Physical Oceanography*, *21*(6), 782–797. https://doi.org/10.1175/1520-0485(1991)021<0782:atgmfw>2.0.co;2

Tolman, H. L. (2009). Practical nonlinear interaction algorithms. In *11th international workshop on wave hindcasting and forecasting & coastal hazards symposium, JCOMM Tech. Rep* (Vol. 52).

Tolman, H. L. (2010). *Optimum discrete interaction approximations for wind waves. Part 4: Parameter optimization*. (Technical Note No. 288). NOAA/NWS/NCEP/MMAB.

Tolman, H. L. (2011). A conservative nonlinear filter for the high-frequency range of wind wave spectra. *Ocean Modelling*, *39*(3–4), 291–300. https://doi.org/10.1016/j.ocemod.2011.05.004

Tolman, H. L. (2013). A generalized multiple discrete interaction approximation for resonant four-wave interactions in wind wave models. *Ocean Modelling*, *70*, 11–24. https://doi.org/10.1016/j.ocemod.2013.02.005

Tolman, H. L., & Grumbine, R. W. (2013). Holistic genetic optimization of a generalized multiple discrete interaction approximation for wind waves. *Ocean Modelling*, *70*, 25–37. https://doi.org/10.1016/j.ocemod.2012.12.008

Tolman, H. L., Krasnopolsky, V. M., & Chalikov, D. V. (2005). Neural network approximations for nonlinear interactions in wind wave spectra: Direct mapping for wind seas in deep water. *Ocean Modelling*, *8*(3), 253–278. https://doi.org/10.1016/j.ocemod.2003.12.008

Tracy, B. A., & Resio, D. T. (1982). *Theory and calculation of the nonlinear energy transfer between sea waves in deep water (WES Report No. 11)*. US Army Corps of Engineers.

Tsagareli, K., Babanin, A. V., Walker, D., & Young, I. (2010). Numerical investigation of spectral evolution of wind waves. Part I: Wind-input source function. *Journal of Physical Oceanography*, *40*(4), 656–666. https://doi.org/10.1175/2009jpo4345.1

Van Vledder, G. P. (2001). Extension of the discrete interaction approximation for computing nonlinear quadruplet wave-wave interactions in operational wave prediction models. In *Ocean wave measurement and analysis* (pp. 540–549).

Van Vledder, G. P. (2006). The WRT method for the computation of non-linear four-wave interactions in discrete spectral wave models. *Coastal Engineering*, *53*(2–3), 223–242. https://doi.org/10.1016/j.coastaleng.2005.10.011

Veit, A., Wilber, M. J., & Belongie, S. (2016). *Residual networks behave like ensembles of relatively shallow networks*. Advances in Neural Information Processing Systems, (Vol. 29).

Wahle, K., Günther, H., & Schiller, H. (2009). Neural network parameterisation of the mapping of wave spectra onto nonlinear four-wave interactions. *Ocean Modelling*, *30*(1), 48–55. https://doi.org/10.1016/j.ocemod.2009.06.001

WAMDI Group. (1988). The WAM model—a third generation ocean wave prediction model. *Journal of Physical Oceanography*, *18*(12), 1775–1810. https://doi.org/10.1175/1520-0485(1988)018<1775:twmtgo>2.0.co;2

WAVEWATCH III Development Group. (2019). *User manual and system documentation of WAVEWATCH III version 6.07 (Tech. Note No. 333)*. NOAA/NWS/NCEP/MMAB.

Webb, D. J. (1978). Non-linear transfers between sea waves. *Deep-Sea Research*, *25*(3), 279–298. https://doi.org/10.1016/0146-6291(78)90593-3

Zakharov, V. E. (1998). *Nonlinear waves and weak turbulence*. American Mathematical Soc, (Vol. 36). https://doi.org/10.1007/978-1-4612-0331-5

Zakharov, V. E., & Pushkarev, A. N. (1999). Diffusion model of interacting gravity waves on the surface of deep fluid. *Nonlinear Processes in Geophysics*, *6*(1), 1–10. https://doi.org/10.5194/npg-6-1-1999