Uncertainty Modeling in Significant Wave Height Forecast

by

Harshkumar Mukeshbhai Savaliya
202111052

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

June, 2023

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,

ii) due acknowledgment has been made in the text to all the reference material used.

Harshkumar Mukeshbhai Savaliya

Certificate

This is to certify that the thesis work entitled Uncertainty Modeling in Significant Wave Height Forecast has been carried out by Savaliya Harshkumar Mukeshbhai for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.

Pritam Anand

Dr. Pritam Anand
Thesis Supervisor

Acknowledgments

I am grateful to my supervisor, Dr. Pritam Anand, for all of his help, encouragement, and support during the course of my research study. His knowledge and ideas really inspired how I approached my work and assisted me in overcoming challenges.

I am also thankful to my peers and friends Yash Joshi, Jaymin Vekariya, Gaurav Jha,Vinay Sheth, Vyom Shah, and Rohit Mishra for their constant support and help. They have made my study and life at DAIICT enjoyable and unforgettable.

Last but not least, I want to express my gratitude to my family, especially my parents, for their love, patience, and constant support during this journey. They have been a source of inspiration and motivation for me.

Contents

Abstract

This thesis proposes different variants of LSTM models for point and probabilistic forecasting of significant wave height (SWH), a crucial component of wave energy. SWH forecasting is challenging due to ocean waves' complex and chaotic nature. The thesis applies different decomposition methods, such as wavelet decomposition (WD), empirical mode decomposition (EMD), and variational mode decomposition (VMD), to enhance the performance of LSTM models. The thesis also uses a convolutional neural network (CNN) and a genetic algorithm to improve the feature extraction and hyperparameter tuning of LSTM models. Moreover, the thesis develops a probabilistic forecasting model for SWH using the pinball loss function, which captures the uncertainty and provides confidence intervals for the forecasts. The thesis evaluates the proposed models on seven real-world SWH datasets collected from four different ocean buoys. The results show that the CNN-LSTM model outperforms other LSTM variants in deterministic forecasting, while the probabilistic forecasting model provides reliable and sharp confidence intervals for SWH.

Index Terms: *Probabilistic forecasting, Time-series forecasting, Long short-term Memory, Significant wave height forecasting.*

List of Principal Symbols and Acronyms

τ	Quantile level
τ_1	Upper quantile value
τ_2	Lower quantile value
C	Observed Calibration
L	Predicted lower quantile
t	Actual Calibration
U	Predicted upper quantile
EMD	Empirical Mode Decomposition
GA	Genetic Algorithm
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error
SVR	Support Vector Regression
SWH	Significant Wave Height
VMD	Variational Mode Decomposition
WD	Wavelet Decomposition

List of Tables

List of Figures

CHAPTER 1
Introduction

1.1 Motivation

As global resources are depleted, reducing reliance on fossil fuels and exploring alternative energy sources has become increasingly important. Wave energy is receiving more attention because of its sustainability, environmental friendliness, high energy density, and wide distribution. Compared to wind energy hourly, wave energy is the most concentrated renewable energy source[16]. The water density, wavelength, and actual wave height are the main factors affecting the energy output from an ocean-based wave source. The significant wave height is the main variable parameter for ocean wave energy that plays a crucial role in wave energy generation because the density of water is constant, and the wavelength is generally predictable[5]. In other words, the wave energy may be scheduled, and wave power consumption will be maximized using accurate SWH forecasting[42]. This has a big financial impact on the system's functions and can significantly save costs. A few hours ahead, forecasting this vital wave energy parameter is essential for various ocean engineering applications, such as marine structures, maritime traffic, and the dredging industry. Significant Wave Height is also important for marine operations such as shipping, construction, reducing marine time accidents, etc. [3][22].

Significant Wave Height(SWH) is a crucial component of wave energy. It is obtained by taking the average of one-third highest ocean waves observed in a given time interval. On average, 15% of waves will equal or exceed a significant wave height. The highest ten percent of waves may be 25–30% higher than the significant wave height. Additionally, we can rarely expect a wave that is nearly twice as tall as the significant wave height. That is why we have considered significant wave height(SWH) instead of considering any other wave height measurement. SWH's hourly accurate forecasting can significantly improve decisions made in marine activities and offshore activities. However, because of the complex marine environment and chaotic nature, it is challenging to forecast SWH accurately. In this thesis, we aim to make SWH forecasting more reliable using point forecasting and probabilistic forecasting.

There are different ways to forecast significant wave height. Initially, researchers used numerical models that rely on the action balance equation. These models are hard to implement and require much processing time[2]. Parametric machine learning models were employed for short-term wave height forecasting. These models summarize data with a set of parameters of fixed size. Auto-Regressive (AR), Auto-Regressive Moving Average(ARMA), and Auto-Regressive Integrated Moving Average(ARIMA) are the various parametric models that were used for short-term wave height forecasting. Other than these, Support Vector Machine (SVR)[14] based models have also been effective for short-term prediction of SWH.

The non-parametric models are flexible and expressive. These models efficiently capture the nonlinear pattern of the data. They do not require any assumption about the distribution of the data. Non-parametric methods are more efficient than parametric methods for short-term prediction. Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Gated Recurrent Unit(GRU) are the various non-parametric models used for the forecasting task. RNN and GRU can perform well for short-term prediction, but they suffer from gradient vanishing and exploding problems. To overcome these challenges, researchers developed another variant of RNN called Long Short-Term Memory (LSTM) model.

Apart from point forecasting, a more reliable approach exists for highly random data like SWH. Probabilistic forecasting can provide a range of possibilities and quantify the length of uncertainty in a prediction. This approach is better in terms of decision-making. It can incorporate forecasting risk, avoid overconfidence, and reduce forecast error.

1.2 Contribution

In this study, we provide wave hybrid models for predicting hourly and six-hourly significant wave height(SWH). For the feature extraction and to gain more information about the receiving SWH signal, we have used Convolutional Neural Network(CNN) and different decomposition techniques along with Long Short-

Term Memory(LSTM) architecture. We have used Wavelet Decomposition(WD), Empirical Mode Decomposition(EMD), and Variational Mode Decomposition(VMD) techniques to decompose our SWH signal. To tune the hyperparameter of LSTM, we have used Genetic Algorithm(GA). We have collected data from four different buoys from the national data buoy center. Then we trained different wave hybrid models for one hour and six hours ahead point forecasting. Using commonly used evaluation criteria, we have analyzed the performance of all the wave hybrid models using box plots. Then we concluded that even the best forecasting model does not incorporate uncertainty in SWH.

We have used the proposed probabilistic forecasting model to handle uncertainty. Instead of relying on just a single point, probabilistic forecasting gives us a range in which our next prediction lie, which is more helpful information for decision-making. We have used a quantile regression-based approach to achieve probabilistic forecasting. We used LSTM with a pinball loss function in the probabilistic forecasting model. We have set different evaluation criteria, namely calibration, sharpness, and error, to evaluate our probabilistic forecasting model.

1.3 Thesis Organization

In Chapter 2, we provide a literature review of some of the existing methods developed and applied for point forecasting of SWH.

Chapter 3 discussed a detailed description of the different wave hybrid models. It also describes the different decomposition techniques, genetic algorithms, and CNN architecture used in the wave hybrid models. We also describe the dataset that we use to train and test the wave hybrid models. After that, we have provided the result obtained by these different wave hybrid models. We conclude this chapter with a brief analysis and discussion of the results.

Chapter 4 gives a detailed description of the probabilistic forecasting model. It describes the architecture used for probabilistic forecasting and briefly describes the pinball loss function. After that, it also includes the numerical results obtained by this model on a different dataset. Further, it provides a brief discussion of the obtained result.

In Chapter 5, we summarize this thesis's main contributions and findings.

Chapter 6 We highlight the strengths and limitations of our proposed methods and models for SWH forecasting. We also suggest some directions and recommendations for future research in this field.

CHAPTER 2

Literature Survey

Ocean wave energy is one of the potential source of clean energy which uses the power of waves to generate the energy. One of the challenges associated with the wave energy is to forecast the Significant Wave Height (SWH) efficiently. Most of the models attempt to obtain the point forecast i,e. they forecast a single value for SWH using the historical data. But, very often there are high uncertainties associated with their predictions. In these cases, it is always important to obtain probabilistic forecast which estimates the interval of SWH with a certain confidence.

2.1 Point Forecasting

Point forecasting is a method of predicting a single value for a future outcome. It is a simple and convenient approach that can be applied to forecast SWH. Point forecasting is a method of predicting a single value for a future outcome. To evaluate the performance of point forecasting, we can use metrics such as the root mean squared error (RMSE) and the accuracy, which measure how close the predicted values are to the actual values. The lower the RMSE and the higher the accuracy, the better the point forecasting model is. Below, we present some point forecasting models researchers have employed to predict significant wave height.

- Statistical models:
  Statistical models are straightforward and faster algorithms that rely on empirical relationships between SWH and other variables. The researcher used AR, ARMA, and ARIMA-based machine learning models to improve the short-term prediction of SWH. In [17], researchers have proposed an AR model assuming that current ocean wave height is linearly dependent on the past ocean wave height. Soares, Ferreira, et al. have used the ARMA model to predict SWH at different locations of the Portuguese coast[1]. In [19], they proposed a short-term ocean wave forecasting method using an

auto-regressive moving average (ARMA) model. In [36], they compared AR and ARMA models for short-term wave forecasting using real wave data from different locations. However, these model also has drawbacks and challenges for SWH forecasting. The time series has to be uniform for the ARMA model to produce accurate predictions. The ARIMA model has advantages over the ARMA model for non-stationary time series. However, none of these statistical time-series models has the ability to adequately represent the complex non-linear relationship. Additionally, they presume that the noise in the data comes from the normal distribution, which may not be true, especially for SWH time-series data. This model requires careful selection and tuning of the parameters, which can affect the performance and stability of the model.

- Artificial Neural Network based models:
  Modern non-parametric machine learning models have yielded significantly better results, particularly for short-term SWH forecasting tasks, as they do not make any assumptions about the noise distribution. Researchers have used different neural network architectures to obtain reliable results on SWH. In [21], they used the ANN model for monthly mean SWH prediction. ANN has limited forecasting ability; due to this, it is not a promising approach for SWH forecasting.[18]. In [11], authors have used the Extreme Learning Machine (ELM) along with grouping genetic algorithm for searching the effective features set and obtaining the effective forecast of short-term SWH. The hidden layer weights in the ELM feed-forward neural network are chosen at random. It does not necessitate the costly backpropagation method, unlike the ANN model, for tuning all weights.

- Deep Learning based Models:
  Other non-parametric models like RNN and LSTM can capture the nonlinear and complex relationships between the input and output variables and also handle the long-term and complex relationships between input and output. A sequential learning model called the Recurrent Neural Network (RNN) is able to recognize the temporal pattern in the data.In [6], and [33] researchers have used RNN for SWH. But, the RNN architecture suffers from vanishing gradient problems and fails to learn the long-term dependencies in data. The Gated Recurrent Unit GRU [8], which eliminates the vanishing gradient problem and uses the gate mechanism to control the information flow, enhances the RNN. In [7], authors have used GRU to efficiently fore-

cast SWH. The Long Short Term Memory Network(LSTM) is a more complex neural architecture than GRU and involves more gates for controlling the flow of information. Researchers have used LSTM architecture for ocean wave height forecasting in [35], [43], [18].

- SVR based models:
  Support Vector Regression (SVR) models are popular choice of researchers in hourly forecasting tasks due to its simplicity. Support Vector Machine (SVR) models are derived from the principles of statistical learning theory. SVR-based models are likely to arrive at the global optimal solution, in contrast with various neural network architecture-based machine learning models. In [32], they have shown that the SVR method performs better than the artificial neural network. In [4], researchers have successfully used the SVR model to predict SWH near the west coast of India. Cornejo-Bueno, Nieto Borge, et al. have used an X-radar-based image and SVR model to forecast SWH in [10]. Duan, Han, et al. have used the SVR model and the Empirical Mode Decomposition method in [15]. However, SVR-based methods are not suitable for nonlinear and non-stationary data. This model fails when data is too noisy and has many outliers like SWH. In [27] and [15], they have shown that SVR based model fails to forecast the SWH efficiently. There are two main challenges associated with SVR forecasting. The first one these methods are sensitive to the choice of kernel function and hyperparameters. The second challenge requires an informative set of features for obtaining good forecasts with SVR models. SVR models may obtain poor forecasts with raw SWH time-series signals.

- Hybrid Models:
  In recent work, researchers have started to use wave hybrid models for SWH. In the wave hybrid models, we combine different methods and techniques to improve the overall efficiency of SWH forecasting. Wei Hao, Xiaofang Sun, et al. have proposed a hybrid model using LSTM with the EMD decomposition method for wave prediction in offshore China [23]. In this [38], researchers have used EMD-PSO-LSSVR hybrid model for forecast SWH for the lead times 1,3 and 6 hours. In [37], they have used ANN with wavelet decomposition for 48 hours ahead SWH forecasting. Even with the application of advanced non-linear machine learning methods, predicting SWH accurately remains challenging due to ocean waves' random and chaotic nature.

However, point forecasting also has some limitations, such as ignoring the uncertainty and variability of future outcomes. All previously discussed models fail to consider the future variable's randomness, which can result in poor judgment, risk underestimating, and overconfidence.

2.2 Probabilistic Forecasting

The probabilistic forecasting model provides a range of possible outcomes, which helps for better decision-making and risk management rather than relying on a single point. Instead of providing single-point estimation, probabilistic forecasting gives us a two-point in which our next prediction lies. Evaluation of the probabilistic forecasting model is difficult because the prediction interval cannot be directly compared with actual results. In [24], researchers have used a kernel density estimation-based approach to find the probability density function of SWH and wind speed. Kernel density estimation is a non-parametric approach to estimating the probability density function of a random variable[40]. It involves assigning each data point a weight using a kernel function, which is a symmetric, smooth function, and then adding the weights to produce a smooth curve. The kernel function's bandwidth parameter influences the predicted density's smoothness, determining how wide or narrow the kernel function is. Taylor and Jeon have used the ARMA-GRACH probabilistic forecasting model for wave height prediction [39]. This model focuses on modeling the conditional mean and variance of the series. This model is only efficient when the distribution of the data is known. Another non-parametric approach is based on quantile regression, which does not require any assumption about the input data. Unlike the kernel density function, It does not require smoothing parameters. In [41],[29], and [34], researchers have efficiently used a quantile regression-based approach for wind, solar, and weather prediction.

CHAPTER 3

Proposed Wave Hybrid Point Forecasting Models for SWH

This chapter discussed the various architectures that can be utilized for SWH point forecasting. It also provided an overview of the various decomposition methods used in these experiments. Following that, it established the evaluation criteria that were used to measure the effectiveness of the proposed wave hybrid models. Then, it presented the numerical outcomes for each of the seven datasets used in the proposed method. At last, the results and the outcomes have been examined and discussed.

3.0.1 Simple LSTM architecture

The first architecture we used for SWH one-hour and six-hour ahead point forecasting is Long Short-Term Memory (LSTM)[25]. LSTM is a form of recurrent neural network (RNN) that can manage long-range relationships in sequential input. Natural language, audio, and video are examples of sequential data with a temporal structure that calls for models that can capture the temporal dynamics and interdependence between the parts. LSTM includes feedback loops that let it learn from both recent and previous inputs, in contrast to feedforward neural networks that treat each input independently. In this manner, LSTM avoids the issue of vanishing or exploding gradients that occurs in simple RNNs and may preserve information over a longer duration of time.

A typical LSTM unit is made up of four parts: a memory cell, an input gate, an output gate, and a forget gate. The primary component that stores the internal state of the unit is the memory cell. The amount of current input to add to the memory cell is determined by the input gate. How much of the memory cell will be output to the following layer is decided by the output gate. How much of the preceding memory cell should be kept or erased is decided by the forget gate. These elements work together to control information flow inside the object and sustain a durable state over time. Figure 3.1 depicts the structure of a single LSTM unit.

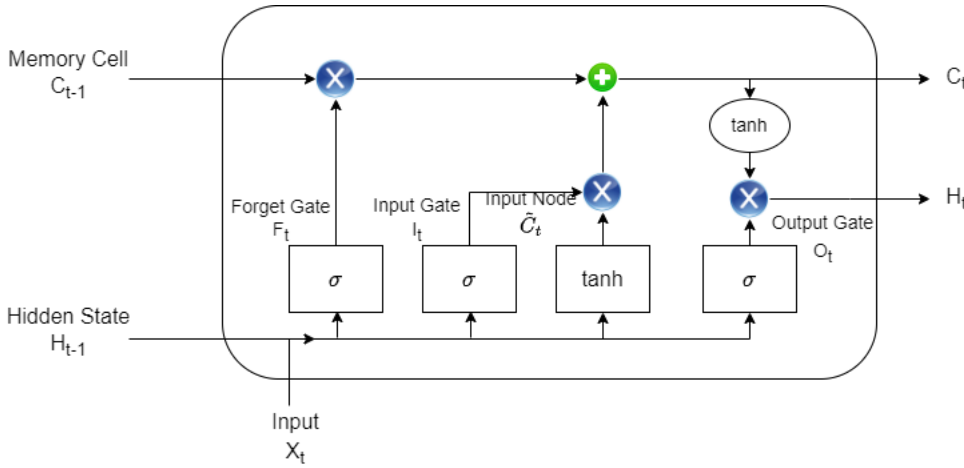


Figure 3.1: LSTM single block architecture

At a particular time instance t, LSTM block has Current Input (X_t), Previous Cell Output (H_{t-1}), Previous Memory Cell (C_{t-1}), Current Cell Output H_t, which is obtained by solving (3.2), Current Memory Cell (C_t), which is obtained by solving (3.1).

$$C_t = F_t * C_{t-1} * \tilde{C}_t \tag{3.1}$$

$$H_t = O_t * tanh(C_t) \tag{3.2}$$

The intermediate gates of LSTM are Input gate (I_t), which is calculated as shown in (3.3), Forget gate (F_t), which is calculated as shown in (3.4), Output gate (O_t), which is calculated as shown in (3.5), candidate gate ($\tilde{C}_t$), which is calculated as shown in (3.6)

$$I_t = \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + b_i) \tag{3.3}$$

$$F_t = \sigma(W_{xf} * X_t + W_{hf} * h_{t-1} + b_f) \tag{3.4}$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + b_o) \tag{3.5}$$

$$\tilde{C}_t = I_t * (tanh(W_{xc} * X_t + W_{hc} * h_{t-1} + b_c)) \tag{3.6}$$

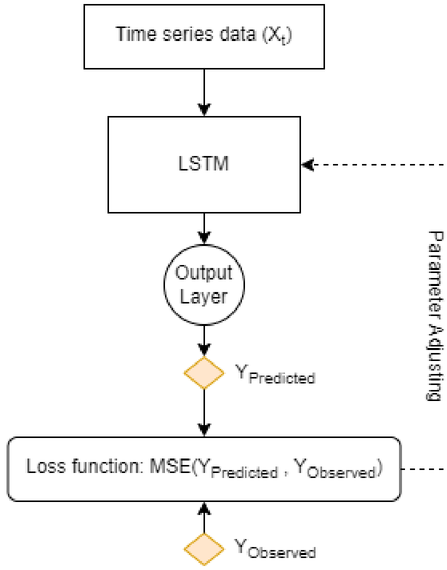


Figure 3.2: Simple LSTM model architecture

For given time-series SWH data $(X_1, X_2, ..., X_t)$, First we have to construct the training set (X, Y) using $X = \begin{bmatrix} X_1, & X_2, & ...., & , X_p \\ X_2, & , X_3, & ... & , X_{p+1} \\ ... & ... & ... & \\ X_{t-p}, & X_{t-p+1}, & ...., & X_{t-1} \end{bmatrix}$ and $Y = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ ... \\ X_t \end{bmatrix}$. Using the training set (X, Y), and the simple LSTM model is trained as shown in Figure 3.2. We have used MSE to calculate the error in this LSTM model. For the prediction of the SWH x_{t+1}. The LSTM model is estimated for the test point $[X_{t-p+1}, X_{t-p+2}, ..., X_t]$.

3.0.2 LSTM with different decomposition methods

In this architecture, we used various decomposition methods, namely Wavelet Decomposition (WD), Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD), along with LSTM. Signal decomposition is a method of breaking down an abstract signal into smaller and more relevant components, which can reveal the data's hidden structure and features. Multiple applications for signal decomposition exist, including noise reduction, feature extraction, etc. Now let's briefly understand all of these decomposition techniques.

1. Wavelet Decomposition[20]
   Wavelets are used in the wavelet decomposition method of signal processing to represent a signal at different sizes and orientations. Wavelet decomposition can be used to extract information from a variety of data, including audio signals, pictures, and climatic data. We can perform wavelet decomposition using a variety of wavelet functions, such as Daubechies, Haar, Morlet, etc. In our experiments, we've used Daubechies wavelet decomposition. Wavelet decomposition has several advantages, unlike other signal processing strategies like the Fourier transform. Fourier transform records the frequency information of a signal, whereas wavelet decomposition records both the frequency and temporal information. While Fourier transforms presuppose a global periodicity, Wavelet decomposition can adapt to the local characteristics of a signal.

2. Empirical Mode Decomposition[26]
   EMD is a signal processing technique that decomposes a signal into a series of intrinsic mode functions (IMF) with well-defined instantaneous frequencies. IMFs are oscillatory functions with a symmetric envelope, the same number of extrema, and zero crossings. EMD does not require prior knowledge of the signal properties or basis functions and can be applied to nonlinear and nonstationary signals. EMD yields a set of IMFs and a trend that can be used to recreate the original signal via summing. While WD might miss the multiscale aspect of the signal, EMD can extract global structure and handle fractal-like data. In contrast to WD, which may induce aliasing or loss of information, EMD can offer a comprehensive and non-redundant representation of a signal.

3. Variational Mode Decomposition[13]
   VMD does not require prior knowledge of the signal properties or basis functions and can be applied to nonlinear and nonstationary signals. VMD produces a series of IMFs and their center frequencies, which can be used to recreate the original signal using summing. The IMFs depict oscillations in the signal at various scales, from high frequency to low frequency. The significant frequency elements of each IMF are captured by the center frequencies. Compared to methods for signal processing like the Fourier transform, wavelet transform, and empirical mode decomposition, VMD has a few advantages. While Fourier transforms presumes a stable frequency spectrum, VMD may handle signals with time-varying frequency and amplitude. Wavelet transform requires a fixed set of basis functions, whereas

VMD can additionally adapt to the local characteristics of a signal. While empirical mode decomposition could result in the addition of modal aliasing or noise sensitivity, VMD can also offer a comprehensive and non-redundant representation of a signal.

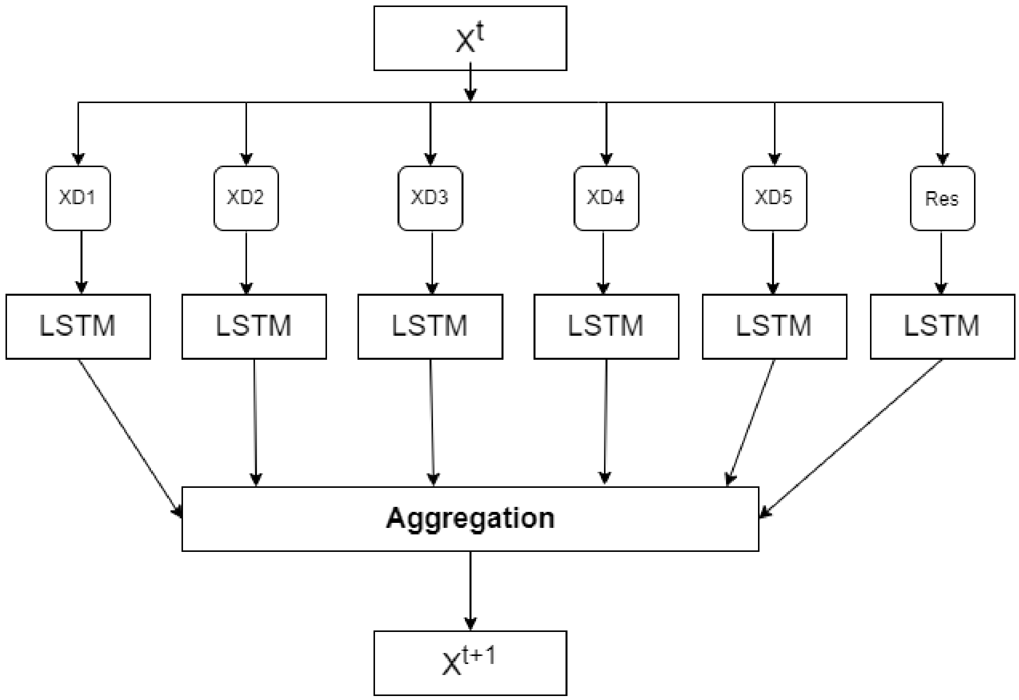


Figure 3.3: LSTM with different decomposition methods

For given time-series SWH data $(X_1, X_2, ..., X_t)$, First, we have to decompose the signal using different decomposition techniques. In the WD approach, we have to apply wavelet transform to the time-series signal to obtain five high-frequency components (D_1, D_2, D_3, D_4, and D_5) that represent the details of the signal at different scales, and one approximate low-frequency signal (A_5) that represents the approximation of the signal at the coarsest scale. In VMD and EMD, we have to decompose the signal into five intrinsic mode functions (IMFs) and one residual signal that contains the remaining low-frequency trend of the data. After that, we must construct the training set (X, Y) for all six decomposed signals, as mentioned in the above section. For all these six training sets (X, Y), we have to train the LSTM model shown in Figure 3.2. We have to combine the outcomes of each signal after training the LSTM model before making a final prediction. The overall architecture is shown in Figure 3.3.

3.0.3 LSTM with genetic algorithm

Deep learning models require parameter adjustment because they can enhance their efficiency and accuracy. An extensive set of hyperparameters controls deep learning models' behavior and learning process. Finding the ideal values for these hyperparameters through parameter tuning will minimize the loss function and increase the model's prediction capacity for the given problem. Parameter tuning is vital because various problems may require different hyperparameter settings to achieve the best outcomes. For instance, a higher learning rate might accelerate a model's convergence on one problem but lead to divergence on another.

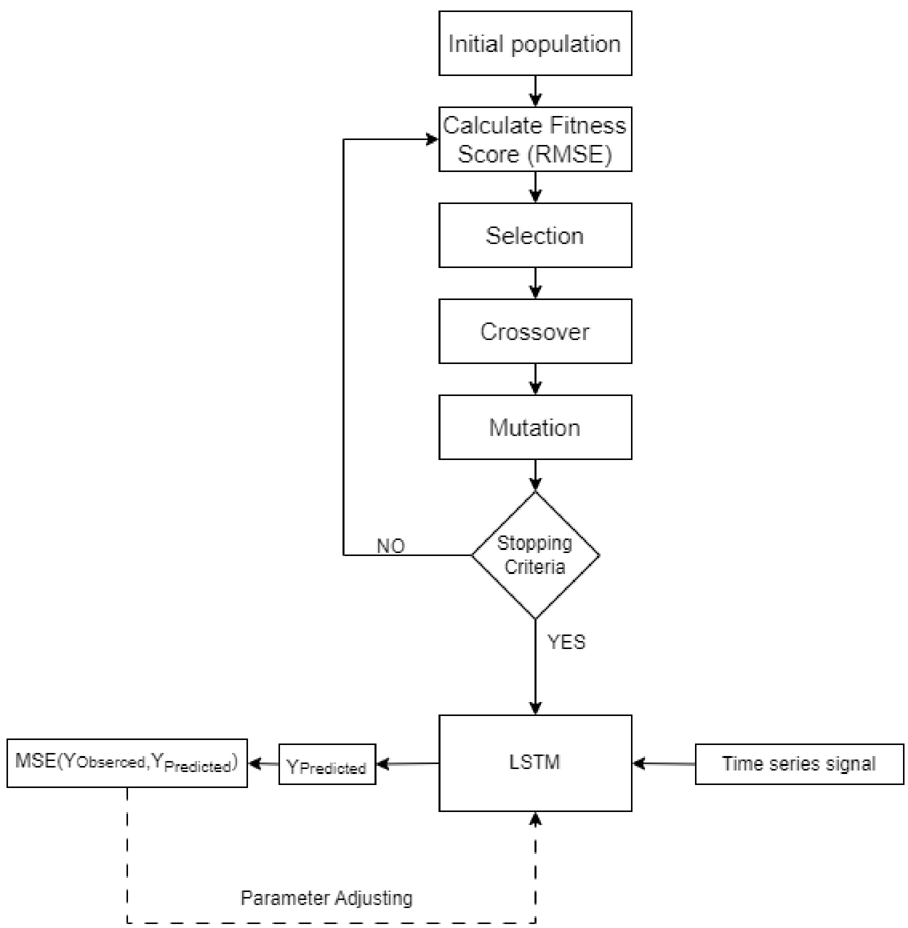


Figure 3.4: Simple LSTM with genetic algorithm based parameter tuning

For SWH point forecasting, we used the LSTM deep learning model. The length of the input signal (window size), the batch size, and the number of hidden units in a single layer are the three primary hyperparameters that determine the performance of the LSTM model. The hidden unit count determines the capacity and complexity of the model, the batch size by the number of samples processed at once, and the window size by the amount of historical data utilized for forecasting. We employed a genetic algorithm[12], an evolutionary optimization method replicating natural selection, to obtain the ideal values for these hyperparameters. The optimal set of hyperparameters that decreases forecasting error is determined via the genetic algorithm. For SWH point forecasting, we have chosen a window size of 100. The batch size and the number of hidden units are determined from the genetic algorithm

Selection, crossover, and mutation are the three main operators used in a genetic algorithm. Selection chooses the best solutions from the present population to create offspring for the following generation. Crossover combines two-parent approaches to produce fresh offspring with traits from both parents. To add diversity and exploration, mutation randomly modifies some aspects of a solution. These operators are continued until a stopping requirement is satisfied, such as reaching the desired fitness level or the maximum number of generations. We have used Root Mean Squared Error(RMSE) as a fitness score. The solution with the lowest RMSE has a higher chance of creating offspring in the next round of the algorithm.

After finding the best parameter from the genetic algorithm, we have to pass these parameters to LSTM for better forecasting of SWH. A working flow of LSTM with genetic algorithm is shown in Figure 3.4. We have used the roulette wheel selection algorithm[30] in our experiment. The crossover probability is 0.4, and the mutation probability is 0.1 in our experiments.

3.0.4 CNN-LSTM Architecture

Convolutional neural networks (CNNs) may learn hierarchical and abstract features from time series data by stacking many convolutional and pooling layers. In contrast to the top layers, which learn high-level features like anomalies and occurrences, the lower layers learn low-level properties like patterns and cycles. In [31], researchers have shown the effectiveness of Convolutional Neural Networks with LSTM on the gold price prediction.

The main objective behind using LSTM with CNN is that CNN is efficient in terms of feature extraction, and LSTM is efficient in capturing long dependencies between the data signal[31]. For the SWH point forecasting, we have used the architecture shown in Figure 3.5. It has two convolution layers with 32 and 64 neurons

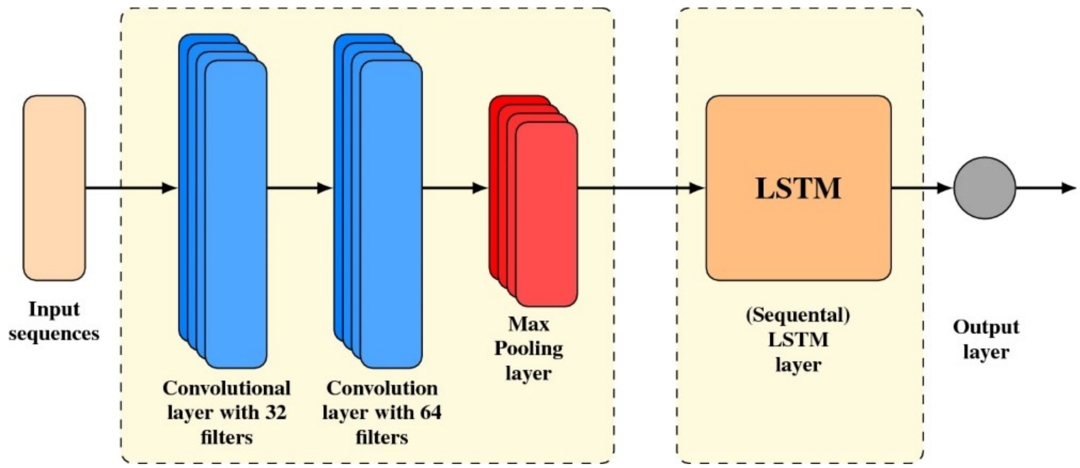


Figure 3.5: LSTM with Convolutional Neural Network

and one max pooling layer. The final output of the CNN layer, which is extracted feature from the raw input signal, is now fed to the LSTM for the forecasting of SWH.

3.1 Dataset Description

We have to collect time-series hourly ocean wave height data from different buoys available, namely 41053, 42001, 42002, and 42035, at National Data Buoy Center(NDBC)(https://www.ndbc.noaa.gov/). Each dataset includes a Timestamp(Date, Month, Year, Hour, and Minutes), Wind direction, Wind speed, Wind Gust, Atmospheric Pressure, Air Temperature, Water Temperature, Dew point, Wave Height, etc.

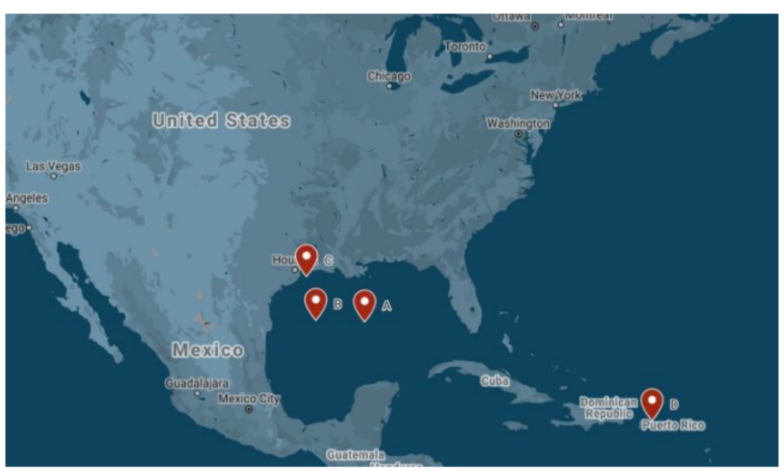


Figure 3.6: Geographics location of different buoy

We have considered only the wave height attribute for prediction. We have listed brief details of chosen ocean buoys dataset in Table 3.1. Datasets A, B, C, and D are used for one-hour ahead time prediction, and E, F, and G are used for six-hour ahead forecasting. The datasets' geographic locations are shown in Figure 3.6. The datasets were checked for missing values and outliers before being used in the study.

Station ID(Dataset)	Coordinates	Time	No. of Samples
41053(A)	66°5′58" W, 18°28′27" N	Feb'21 - Mar'21	670
42001(B)	89°39′25" W, 25°56′31" N	Jan'15 - Apr'15	2869
42002(C)	93°38′46" W, 26°3′18" N	Jan'15 - Apr'15	2872
42035(D)	94°24′45" W, 29°13′54" N	Jan'15 - Apr'15	2852
42001(E)	89°39′25" W, 25°56′31" N	Jan'15 - Apr'15	476
42002(F)	93°38′46" W, 26°3′18" N	Jan'15 - Apr'15	478
42035(G)	94°24′45" W, 29°13′54" N	Jan'15 - Apr'15	479

Table 3.1: Dataset description

3.2 Evaluation Criteria

We have used Root Mean Squared Error (RMSE) and Accuracy as our evaluation parameters to assess the efficiency of our various wave hybrid models. With the notations, y_i as actual SWH for i^{th} test sample, $\hat{y}_i$ as predicted SWH for i^{th} test sample, $\bar{y}_i$ as mean of actual SWH and n as total number of testing samples, we briefly describe our evaluation criteria as follows.

(a) Root Mean Square of Errors (RMSE):
It is obtained by $\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}$.

(b) Mean Absolute Percentage of Errors (MAPE):
It is obtained by $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$.

(c) Accuracy (Acc):
It is the measure for the accuracy of obtained predictions. It is obtained by Acc = 100 - MAPE.

(d) Mean of Absolute Deviations (MAD):
It is obtained by $\frac{1}{n}\sum\limits_{i=1}^{n}|y_i - \hat{y}_i|$.

(e) Normalized Mean Squares of Errors (NMSE):
It is ratio of Sum of Squares of Errors (SSE) and Sum of Squares of Testing samples (SST) and is obtained by $\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$.

3.3 Result Analysis and Discussion

In this section, we present our numerical results. We have listed the performance of LSTM models with no decomposition, WD, EMD, and VMD decomposition methods and also with CNN-LSTM and LSTM with genetic algorithm techniques on our selected SWH datasets with a lead time of 1 hour and 6 hours using Accuracy, RMSE at Table 3.2 and Table 3.4. We have plotted the prediction obtained by the CNN-LSTM model for datasets A, C in Figure 3.7,and E, and G in Figure 3.8.

Dataset	Model	Accuracy	RMSE	NMSE	MAD	MAPE
A	LSTM	92.15	0.1562	0.3613	0.1425	7.84
	WD-LSTM	77.37	0.4398	2.0965	0.3714	22.62
	EMD-LSTM	91.91	0.1826	0.2645	0.137	8.08
	VMD-LSTM	85.69	0.2711	0.7968	0.2492	14.30
	CNN-LSTM	94.08	0.1077	0.1257	0.0886	5.18
	Genetic-LSTM	61.12	0.7204	2.1945	0.3911	37.65
B	LSTM	90.65	0.164	0.1404	0.1151	9.34
	WD-LSTM	93.31	0.1087	0.0501	0.0764	6.68
	EMD-LSTM	81.88	0.2797	0.4087	0.2125	15.11
	VMD-LSTM	94.11	0.098	0.0617	0.074	5.88
	CNN-LSTM	93.2	0.1199	0.0751	0.0849	6.79
	Genetic-LSTM	71.16	0.4685	0.3842	0.2322	19.8
C	LSTM	90.37	0.1503	0.092	0.1092	9.62
	WD-LSTM	93.34	0.1092	0.0485	0.0773	6.65
	EMD-LSTM	81.91	0.2553	0.2655	0.201	18.05
	VMD-LSTM	94.9	0.0858	0.03	0.0595	5.09
	CNN-LSTM	93.57	0.1155	0.0543	0.0781	6.42
	Genetic-LSTM	73.86	0.4964	0.3026	0.2325	16.51
D	LSTM	86.3	0.1433	0.218	0.1021	13.69
	WD-LSTM	89.55	0.0971	0.1001	0.0737	10.44
	EMD-LSTM	71.45	0.2116	0.4754	0.1912	28.54
	VMD-LSTM	93.71	0.0719	0.0549	0.04933	6.28
	CNN-LSTM	91.88	0.0928	0.0914	0.0619	8.11
	Genetic-LSTM	56.49	0.1901	0.2055	0.1053	15.61

Table 3.2: Numerical Results of lead time 1 hrs

Models	Mean accuracy
LSTM	81.33 ± 12.09
WD-LSTM	75.17 ± 18.22
EMD-LSTM	79.96 ± 6.48
VMD-LSTM	86.03 ± 8.28
CNN-LSTM	87.4 ± 8.06
Genetic-LSTM	66.2 ± 8.52

Table 3.3: Average accuracy on all seven data sets.

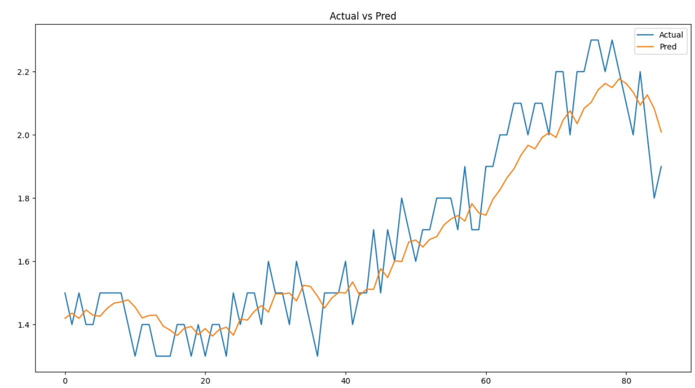
Dataset	Model	Accuracy	RMSE	NMSE	MAD	MAPE
E	LSTM	77.4	0.3973	0.9175	0.3085	22.59
	WD-LSTM	67.66	0.4473	1.1629	0.3875	32.33
	EMD-LSTM	80.27	0.3019	0.5297	0.2418	19.72
	VMD-LSTM	76.66	0.4333	1.0912	0.3319	23.33
	CNN-LSTM	81.91	0.2966	0.5113	0.2297	18.08
	Genetic-LSTM	72.74	0.5584	1.2145	0.4075	27.26
F	LSTM	73.13	0.5014	0.8201	0.3296	26.86
	WD-LSTM	48.22	0.6131	1.2263	0.5414	51.77
	EMD-LSTM	76.84	0.3207	0.3355	0.2431	23.15
	VMD-LSTM	80.73	0.3216	0.3374	0.227	19.26
	CNN-LSTM	84.44	0.262	0.224	0.1883	15.55
	Genetic-LSTM	73.52	0.5917	1.2039	0.3512	26.48
G	LSTM	59.33	0.3258	1.0436	0.2489	40.66
	WD-LSTM	56.77	0.3164	0.9846	0.2628	43.22
	EMD-LSTM	75.5	0.3055	0.9175	0.219	24.49
	VMD-LSTM	76.41	0.1984	0.387	0.1508	23.58
	CNN-LSTM	72.73	0.2587	0.6583	0.1874	27.26
	Genetic-LSTM	54.52	0.2685	0.9233	0.2169	45.48

Table 3.4: Numerical Results of lead time 6 hrs

Now, we shall briefly analyze the numerical results presented in Table 3.2 and Table 3.4. For better visualization of obtained numerical results, we have compared the accuracy obtained by different LSTM models using a box plot in Figure 3.9 for all considered seven datasets.

In the box plot of Figure 3.9, we can compare the median of accuracy obtained by different wave hybrid models. The CNN-LSTM model obtains the highest 91.88 median accuracies, followed by the LSTM with no decomposition model with 86.3 median accuracies. The CNN-LSTM model also excels with other wave hybrid models if we consider the 25^{th} percentile of accuracy values.

(a) Dataset A



(b) Dataset C

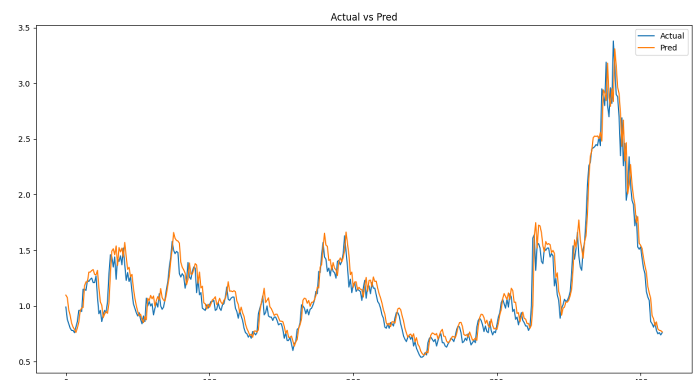
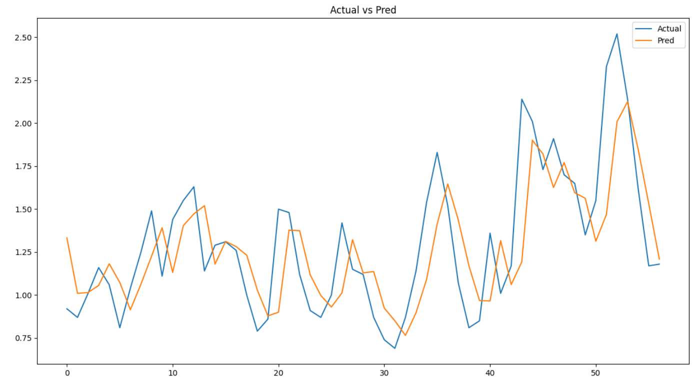


Figure 3.7: Performance of CNN-LSTM model on 1 hour ahead lead time

(a) Dataset E



(b) Dataset G

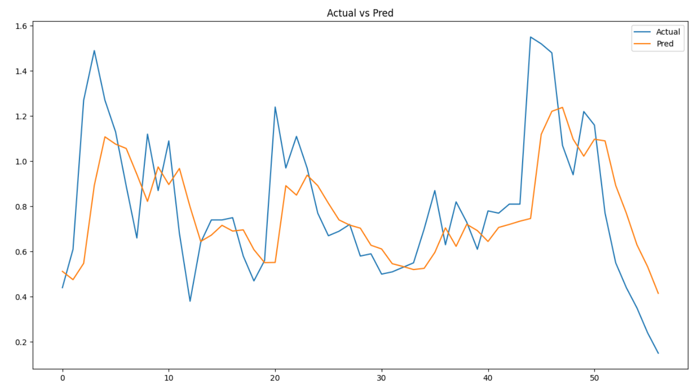


Figure 3.8: Performance of CNN-LSTM model on 6 hour ahead lead time

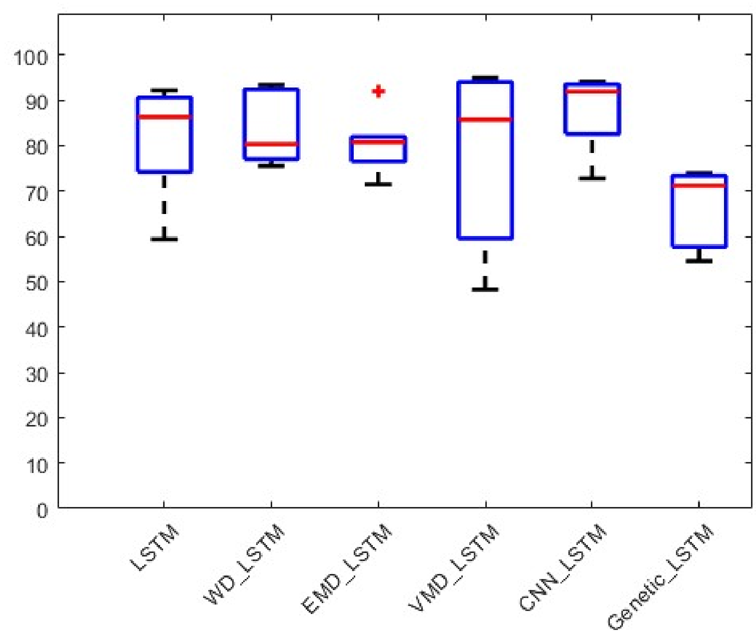


Figure 3.9: Accuracy Box-plot of different hybrid model

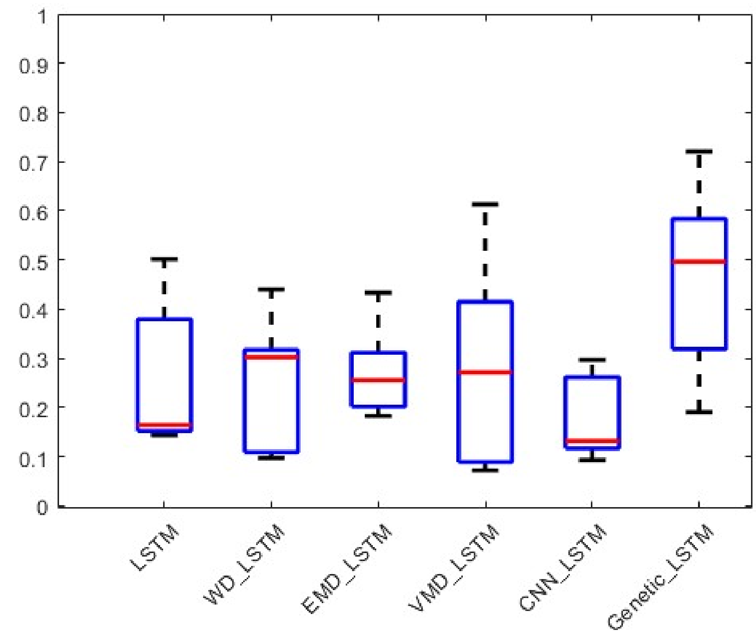


Figure 3.10: RMSE Box-plot of different hybrid model

From Figure 3.9, we can also compare the accuracy obtained by different decomposition methods, namely WD, EMD, and VMD. The VMD-LSTM obtains the highest 85.69 median accuracy, followed by EMD-LSTM with 76.17 and WD-LSTM with 62.15 median accuracy. VMD-LSTM performs better than all the other decomposition techniques if we consider 25^{th} and 75^{th} percentile of accuracy values. If we consider the median RMSE value, also VMD-LSTM has a minimum of 0.1318 median RMSE, followed by WD-LSTM with 0.1089 median RMSE. VMD-LSTM also outperforms the other decomposition methods if we consider 25^{th} and 75^{th} percentile of RMSE.

In box plot of Figure 3.10, we can compare the median of RMSE obtained by different wave hybrid models. The CNN-LSTM model obtains the lowest 0.1318 medians RMSE which is followed by the LSTM with no decomposition model with 0.164 median RMSE. The CNN-LSTM model also excels with other wave hybrid models if we consider the 75^{th} percentile of RMSE values.
We have shown the average accuracy values obtained by different wave hybrid models in Table 3.3. We can observe that CNN-LSTM-based wave hybrid models obtain 87.40 mean accuracy. From this numerical analysis, we can conclude that the CNN-LSTM model performs better than any other LSTM variants.

3.4 Conclusion

In this chapter, we have presented various wave hybrid models for SWH point forecasting. We first applied a simple LSTM model, followed by different decomposition methods, namely WD, EMD, and VMD, in conjunction with LSTM, to extract more information from the raw input signal. We have then employed an evolutionary algorithm, namely a genetic algorithm, to optimize the parameters of LSTM. Furthermore, we have combined CNN and LSTM models to enhance the forecasting performance. We have demonstrated the numerical results of all these wave hybrid models and shown that the CNN-LSTM model outperformed the others.

However, point forecasting has some limitations regarding uncertainty. It only predicts a single value and does not account for the variability and unpredictability of the future SWH. The SWH is highly random and chaotic. Even the best point forecasting wave hybrid model may fail to provide reliable forecasts for future scenarios. Therefore, we need a SWH forecasting model that can also incorporate uncertainty in the input data. The next chapter will introduce the uncertainty model for Significant Wave Hybrid.

CHAPTER 4

Probabilistic Forecasting for Significant Wave Height

Point forecasting models for SWH data produce point estimates that do not account for the uncertainty and randomness inherent in the data. To capture the uncertainty, we developed a probabilistic forecasting model for SWH that assigns a probability to each possible outcome. Probabilistic forecasting is a way of predicting future values of a variable by presenting a probability distribution that emphasizes the outcome's uncertainty and variability.

We proposed a probabilistic forecasting model for SWH in this chapter. The architecture for probabilistic forecasting has first been introduced. After presenting the numerical findings, we briefly examined how the probabilistic forecasting model performed.

4.1 Proposed Methodology for Probabilistic Forecasting

Significant wave height data are highly chaotic and random. Point forecasting does not help us in better decision-making. Instead of point forecasting, we can use a probabilistic forecasting model, which can give a result with a certain confidence. To model the uncertainty in SWH, we have employed the architecture shown in Figure 4.1.To understand architecture better, let's first understand the pinball loss function[28].

- Pinball loss function:
  The pinball loss function, also known as the quantile loss, is a metric used for evaluating the accuracy of a quantile forecast. A quantile forecast predicts the future distribution of a variable at a particular percentile, such as the

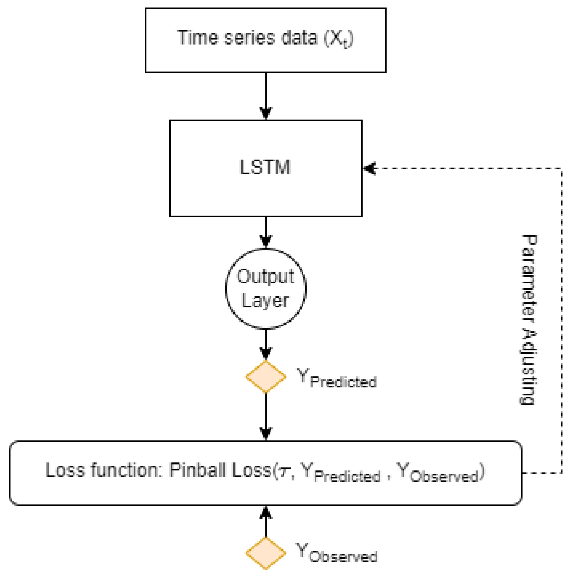


Figure 4.1: Proposed architecture for probabilistic forecasting

95th percentile or the 10th percentile of a given data. By penalizing overestimation and underestimating differently depending on the quantile level, the pinball loss function determines how well the quantile forecast correlates with the true distribution of the outcome. When the quantile forecast is more accurate, the pinball loss function has a smaller value; when it is less precise, it has a greater value.

$$PL_q(\tau, y, \hat{y}) = \begin{cases} (y - \hat{y}) * \tau & y \geq \hat{y} \\ (\hat{y} - y) * (1 - \tau) & \hat{y} > y \end{cases} \tag{4.1}$$

As we can see in equation 4.1, It has three arguments; first, y which is the actual value, $\hat{y}$ which is the observed or predicted value, and τ denotes the quantile forecast's target quantile level. For example, τ = 0.85 indicates that the quantile forecast attempts to forecast the 85th percentile of the variable's future distribution. τ controls the form and slope of the quantile loss function, which penalizes overestimation and underestimating mistakes differently depending on τ.

Our proposed architecture consists of a simple LSTM model and a pinball loss function. To obtain a t percent confidence interval, we train the model twice with τ_1 and τ_2 such that difference between τ_1 and $\tau 2$ is t.

4.2 Evaluation Criteria

We have three criteria to measure the performance of the probabilistic forecasting model. First is calibration, Second is sharpness, and Third is Error. By using notation, C denotes the calibration, U_t is the predicted upper quantile value, L_t is the predicted lower quantile value, τ_1 and τ_2 represent upper quantile and lower quantile respectively, and t is the difference between upper quantile τ_1 and lower quantile τ_2, we have described evaluation criteria as follows.

(a) Calibration[9]:

Calibration is required for reliable and accurate forecasting because it guarantees that predictions are compatible with observed data and represent the uncertainty of future outcomes. It can also tell us about the uncertainty and confidence of future outcomes. The confidence and trust of users and decision-makers who rely on forecasts can also be increased through calibration.

$$C = \frac{1}{n} \sum_{i=1}^{n} u_i \tag{4.2}$$

$u_i = 1$ if the predicted value is between the upper and lower quantile and $u_i = 0$ otherwise.

(b) Sharpness[44]:

Sharpness is a desired attribute of probabilistic forecasts since it provides the level of uncertainty and precision of the predictions. Sharpness is a forecasting property that is independent of actual results. A reliable probabilistic forecast needs to be accurate and precise. However, increasing sharpness may result in a drop in calibration, and vice versa, thus, there is frequently a trade-off between the two.

$$Sharpness = \frac{1}{n} \sum_{i=1}^{n} U - L \tag{4.3}$$

(c) Error:

It is an absolute difference between observed calibration (C) and actual calibration t.

$$Error = |C - t| \tag{4.4}$$

4.3 Results

In this section, we have provided the numerical results obtained using the proposed SWH probabilistic forecasting model. Table 4.1 shows the calibration, sharpness, and error for the different datasets with different lower and upper quantile values. We have also plotted the prediction obtained by different datasets with different t values in Figure 4.2, 4.3, 4.4, and 4.5.

Dataset	Lower quantile	Upper quantile	Calibration	Sharpness	Error
A	0.40	0.80	0.3333	0.1749	0.0666
B	0.05	0.9	0.8392	0.4868	0.0108
B	0.10	0.50	0.3995	0.1281	0.0005
C	0.10	0.90	0.8514	0.5092	0.0514
C	0.40	0.60	0.1627	0.0817	0.0372
D	0.2	0.8	0.61	0.2675	0.0104

Table 4.1: Numerical result of probabilistic forecasting

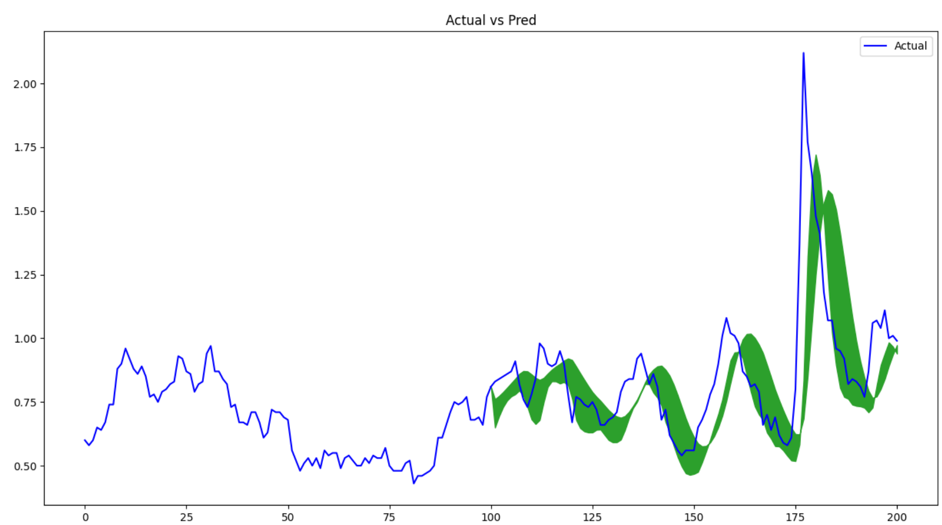


Figure 4.2: Result of probabilistic forecasting on dataset B with t=0.40

4.4 Analysis and Discussion

In this section, we present a brief analysis of the results obtained by the probabilistic forecasting model. The evaluation criteria of probabilistic forecasting are calibration and sharpness, as defined in Section 4.2. Calibration refers to the statistical consistency between the forecasted probabilities and the observed frequencies. A well-calibrated model should produce confidence intervals that contain the actual

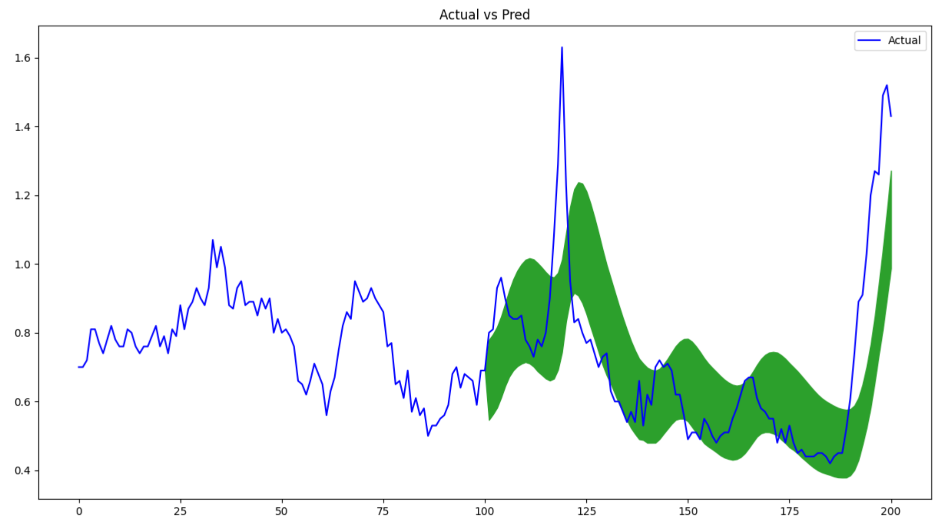


Figure 4.3: Result of probabilistic forecasting on dataset D with t=0.60

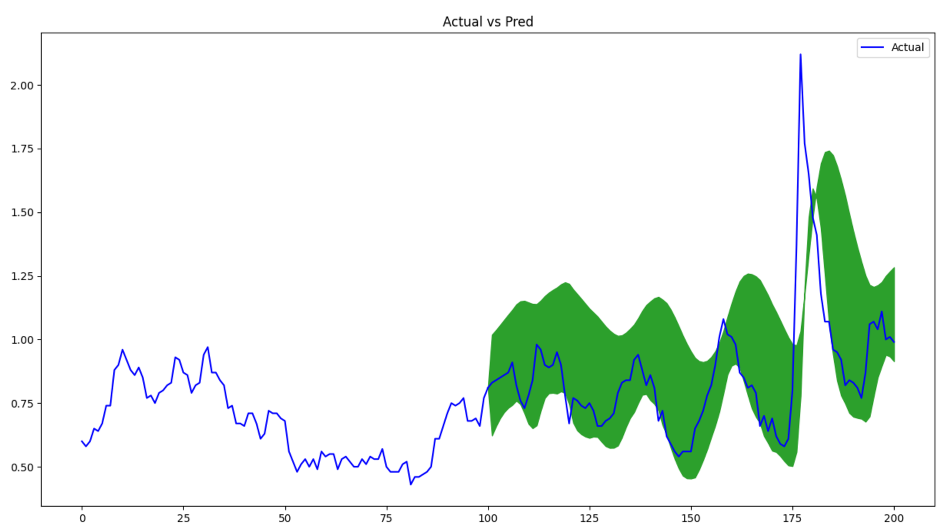


Figure 4.4: Result of probabilistic forecasting on dataset B with t=0.85

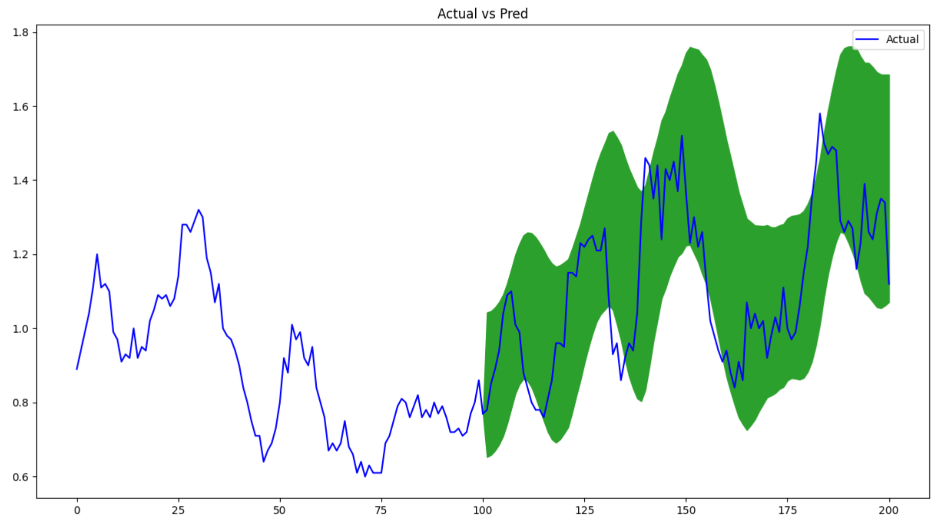


Figure 4.5: Result of probabilistic forecasting on dataset C with t=0.80

values with the expected frequency. Sharpness refers to the concentration of the forecasted probabilities around a central value. A sharp model should produce narrow confidence intervals on average. Therefore, probabilistic forecasting aims to achieve both good calibration and sharpness.

From Table 4.1, we can see that for dataset A, to achieve a 40 percent confidence interval, we have trained our model two times with $\tau_1 = 0.80$ and $\tau_2 = 0.40$, and we got the 0.33 calibration with the 0.1749 sharpness and 0.066 error. For dataset B, to achieve 85 and 40 percent confidence, we have used τ_1 values are 0.90 and 0.50, τ_2 values are 0.05 and 0.10 and got the calibration 0.83 and 0.39, sharpness 0.4868 and 0.1281 with error 0.0108 and 0.0005, respectively. Likewise, Table 4.1 shows the results of datasets C and D.

So, using the probabilistic forecast model for SWH, we can efficiently predict the future outcome with a certain confidence level. We can obtain a range of potential values for the future outcome and the likelihood of each value by applying the probabilistic forecast model for SWH. This gives us more knowledge and flexibility than a single-point prediction, which may be incorrect or misleading. As a result, the probabilistic prediction model for SWH is a more beneficial and accurate method of forecasting future events.

CHAPTER 5
Conclusions and Future Work

SWH is a key parameter for marine activities. It also has an impact on the stability and performance of offshore structures. Moreover, for renewable energy, ocean waves play an important role. Forecasting SWH efficiently helps to produce more energy using ocean waves.

In our study, We have developed different wave hybrid point forecasting models for SWH. At first, we used a simple LSTM model. LSTM is a deep learning model which can capture the long-term dependencies of the given input data signal. Then we used different decomposition methods, namely WD, EMD, and VMD, to get more information from the raw data. These decomposition methods can capture trends and cycle present in the input signal. It can also remove the noise present in data. Deep learning models have many hyperparameters which need to be tuned. The same set of parameters performs differently for the two different problems. We have used an evolutionary genetic algorithm to tune the batch size and the number of hidden units of the LSTM. The received SWH signal is abstract; we have used Convolutional Neural Network to get more information. CNN are the deep learning models which can learn features from the input data. We use CNN for feature extraction and LSTM to capture long-term dependencies. We have trained all these different wave hybrid models on seven real-world ocean wave height datasets. The dataset was collected from different geographic locations. After briefly analyzing the numerical results, we find that the CNN-LSTM model obtains the best performance with a median accuracy of 91.88 and a mean accuracy of 87.40.

Further, we have proposed uncertainty modeling for SWH. Point forecasting for SWH is unsuitable when we have highly random and chaotic data like SWH. Point forecasting may produce inaccurate results. On the other hand, probabilistic forecasting forecast value with a certain confidence. Rather than giving a single point, the probabilistic forecast model provides a range of future outcomes. We have proposed architecture in Chapter 4 for a probabilistic forecasting model. We have used quantile regression based approach for probabilistic forecasting of SWH. We have used the pinball loss function with LSTM deep learning model. The pinball loss function has the parameter τ, which can give targeted quantile estimation. To get t percent confidence, we have to train the LSTM model twice with different τ values.

This study proposes a probabilistic forecasting model for SWH using the pinball loss function. We have shown that our model can capture the uncertainty in SWH and provide reliable confidence intervals. However, some limitations and challenges need to be addressed in future work. First, Our model requires training two different models separately to obtain t percent confidence. This may increase the computational cost and complexity of the model. Moreover, if trained on different data or with different hyperparameters, it may introduce inconsistency between the models.

Second, the pinball loss function tries to optimize sharpness and calibration; the balance between these two quantities is made implicitly, which may result in poor performance. Therefore, we need a loss function that can efficiently optimize sharpness subject to calibration.

We hope our work can inspire further research on probabilistic forecasting of SWH and other renewable energy sources and contribute to developing more efficient and robust wave energy systems.

References

[1] Linear models of the time series of significant wave height on the southwest coast of portugal. *Coastal Engineering*, 29(1):149–167, 1996.

[2] Prediction of significant wave height using regressive support vector machines. *Ocean Engineering*, 36(5):339–347, 2009.

[3] F. A. R. Abdullah, N. S. Ningsih, and T. M. Al-Khan. Significant wave height forecasting using long short-term memory neural network in indonesian waters. *Journal of Ocean Engineering and Marine Energy*, 8(2):183–192, May 2022.

[4] K. Ajeesh and P. C. Deka. Forecasting of significant wave height using support vector regression. In *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, pages 50–53, 2015.

[5] M. Ali, R. Prasad, Y. Xiang, and R. C. Deo. Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms. *Renewable and Sustainable Energy Reviews*, 132:110003, 2020.

[6] C. E. Balas, L. Koç, and L. Balas. Predictions of missing wave data by recurrent neuronets. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 130(5):256–265, 2004.

[7] X. Chen and W. Huang. Spatial–temporal convolutional gated recurrent unit network for significant wave height estimation from shipborne marine radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.

[8] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

[9] Y. Chung, W. Neiswanger, I. Char, and J. Schneider. Beyond pinball loss: Quantile methods for calibrated uncertainty quantification. *CoRR*, abs/2011.09588, 2020.

[10] L. Cornejo-Bueno, J. Nieto Borge, E. Alexandre, K. Hessner, and S. Salcedo-Sanz. Accurate estimation of significant wave height with support vector regression algorithms and marine radar images. *Coastal Engineering*, 114:233–243, 2016.

[11] L. Cornejo-Bueno, J. Nieto-Borge, P. García-Díaz, G. Rodríguez, and S. Salcedo-Sanz. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm – Extreme Learning Machine approach. *Renewable Energy*, 97(C):380–389, 2016.

[12] K. Deb. An introduction to genetic algorithms. *Sadhana*, 24(4):293–315, Aug 1999.

[13] K. Dragomiretskiy and D. Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2014.

[14] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9. MIT Press, 1996.

[15] W. Duan, Y. Han, L. Huang, B. Zhao, and M. Wang. A hybrid emd-svr model for the short-term prediction of significant wave height. *Ocean Engineering*, 124:54–73, 2016.

[16] J. Fernández, S. Salcedo-Sanz, P. Gutiérrez, E. Alexandre, and C. Hervás-Martínez. Significant wave height and energy flux range forecast with machine learning classifiers. *Engineering Applications of Artificial Intelligence*, 43:44–53, 2015.

[17] F. Fusco. Short-term wave forecasting as a univariate time series problem. (ee/2009/jvr/03). December 2009.

[18] S. Gao. A forecasting model for wave heights based on a long short-term memory neural network, 2021.

[19] M. Ge and E. Kerrigan. Short-term ocean wave forecasting using an autoregressive moving average model. pages 1–6, 08 2016.

[20] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2):50–61, 1995.

[21] K. Günaydın. The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. *Ocean Engineering*, 35(14):1406–1415, 2008.

[22] P. Hao, S. Li, and Y. Gao. Significant wave height prediction based on deep learning in the south china sea. *Frontiers in Marine Science*, 9, 2023.

[23] W. Hao, X. Sun, C. Wang, H. Chen, and L. Huang. A hybrid emd-lstm model for non-stationary wave prediction in offshore china. *Ocean Engineering*, 246:110566, 2022.

[24] A. F. Haselsteiner, J.-H. Ohlendorf, and K.-D. Thoben. Environmental contours based on kernel density estimation, 2017.

[25] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[26] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, Mar 1998.

[27] C. Jörges, C. Berkenbrink, and B. Stumpe. Prediction and reconstruction of ocean wave heights based on bathymetric data using lstm neural networks. *Ocean Engineering*, 232:109046, 2021.

[28] R. Koenker and K. F. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156, December 2001.

[29] P. Lauret, M. David, and H. Pedro. Probabilistic solar forecasting using quantile regression models. *Energies*, 10(10):1591, Oct 2017.

[30] A. Lipowski and D. Lipowska. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*, 391(6):2193–2196, 2012.

[31] I. E. Livieris, E. Pintelas, and P. Pintelas. A cnn–lstm model for gold price time-series forecasting. *Neural Computing and Applications*, 32(23):17351–17360, Dec 2020.

[32] J. Mahjoobi and E. Adeli Mosabbeb. Prediction of significant wave height using regressive support vector machines. *Ocean Engineering*, 36(5):339–347, 2009.

[33] S. Mandal and N. Prabaharan. Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, 33(10):1401–1410, 2006.

[34] L. Massidda and M. Marrocu. Quantile regression post-processing of weather forecast for short-term solar power probabilistic forecasting. *Energies*, 11(7):1763, Jul 2018.

[35] M. P and F. Enigo. Forecasting significant wave height using rnn-lstm models. pages 1141–1146, 05 2020.

[36] Y. Peña Sanchez and J. Ringwood. A critical comparison of ar and arma models for short-term wave forecasting. 08 2017.

[37] R. Prahlada and P. C. Deka. Forecasting of time series significant wave height using wavelet decomposed neural network. *Aquatic Procedia*, 4:540–547, 2015. INTERNATIONAL CONFERENCE ON WATER RESOURCES, COASTAL AND OCEAN ENGINEERING (ICWRCOE'15).

[38] G. Tang, H. Du, X. Hu, Y. Wang, C. Claramunt, and S. Men. An emd-pso-lssvm hybrid model for significant wave height prediction. *Ocean Science Discussions*, 2021:1–16, 2021.

[39] J. W. Taylor and J. Jeon. Probabilistic forecasting of wave height for offshore wind turbine maintenance. *European Journal of Operational Research*, 267(3):877–890, 2018.

[40] G. R. Terrell and D. W. Scott. Variable kernel density estimation. *The Annals of Statistics*, 20(3):1236–1265, 2023/05/30/ 1992.

[41] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. *IEEE Transactions on Power Systems*, 32(4):2767–2778, 2017.

[42] S. Yang, Z. Zhang, L. Fan, T. Xia, S. Duan, C. Zheng, X. Li, and H. Li. Long-term prediction of significant wave height based on sarima model in the south china sea and adjacent waters. *IEEE Access*, 7:88082–88092, 2019.

[43] X. Zhang, Y. Li, S. Gao, and P. Ren. Ocean wave height series prediction with numerical long short-term memory. *Journal of Marine Science and Engineering*, 9(5):514, May 2021.

[44] Y. Zhang, J. Wang, and X. Wang. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32(C):255–270, 2014.