Development of Hybrid Models for Short Term Ocean Wave Height Forecasting

by

Shantanu Jain

202011026.

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

MASTER OF TECHNOLOGY

in

INFORMATION AND COMMUNICATION TECHNOLOGY

to

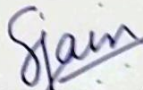**DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY**
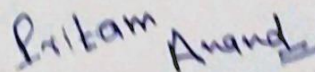

MAY, 2022

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 DA-IICT and has not been submitted elsewhere for a degree,

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

Signature of Student

Certificate

This is to certify that the thesis work entitled Development of Hybrid Models for Short Term Ocean Wave Height Forecasting has been carried out by Shantanu Jain (202011026) for the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology under my supervision.

Prof. Pritam Anand
Thesis Supervisor

Acknowledgement

- I would like to thank the **Dhirubhai Ambani Institute of Information and Communication Technology** in **Gandhinagar** for providing me with access to their supercomputer **PARAM SHAVAK** for all of the computations I needed.

- I would also want to convey my heartfelt gratitude to my faculty mentor, **Dr. Pritam Anand**, for providing invaluable assistance and sharing in-depth information on this topic.

CONTENTS

ii

Abstract

The significant wave height prediction plays a very crucial role in wave power generation. Apart from this, the hourly efficient prediction of SWH can significantly help to improve the decisions in maritime and off-shore activities. But, the highly random and chaotic nature of ocean waves makes the significant wave height prediction task difficult and challenging. In this thesis, we have developed a series of wave hybrid models for hourly significant wave height prediction. Our developed hybrid models use a signal decomposition method along with a regression model. We have used the ε-Support Vector Regression (ε-SVR), Least Squares Support Vector Regression (LS-SVR), Long Short-Term Memory (LSTM) and Large-margin Distribution Machine based Regression (LDMR) model for the regression task. For signal decomposition methods, we have considered the Wavelet Decomposition (WD), Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) method. Apart from this, we have also used the Particle Swarm Optimization (PSO) method to tune the parameters of the used regression model in our wave hybrid models. Till now, the VMD method and LDMR model have not been used in any wave hybrid model. We have evaluated the performance of our developed wave hybrid models on time-series significant wave heights, collected from four different buoys, and located at different geographical locations using the different evaluation criteria. After the brief analysis of the obtained numerical results, we conclude that the LDMR based wave hybrid model outperforms the other regression model based hybrid model. Also, the VMD based wave hybrid models can obtain better performance than other decomposition based hybrid models. Further, we perform two-way ANOVA analysis on obtained numerical results which statistically infer that the use of a particular decomposition method and a particular regression model affect the prediction accuracy significantly but, their effects are independent

LIST OF PRINCIPAL SYMBOLS AND ACRONYMS

SVR	Support Vector Regression
LS-SVR	Least Squares Support Vector Regression
LDMR	Large-margin Distributed Machine based Regression
LSTM	Long Short Term Memory
PSO	Particle Swarm Optimization
WD	Wavelet Decomposition
EMD	Empirical Mode Decomposition
VMD	Variational Mode Decomposition
ANOVA	Analysis of Variance
RMSE	Root Mean Square Error
MAD	Mean Average Deviation
MAPE	Mean Absolute Percentage Error
C	Penalty factor
ϵ	Error band
ξ	Slack variable
K	Kernel function
$\mathbb{R}$	Set of real numbers
α	Support vector
IMF	Intrinsic Mode Function
$\boldsymbol{\phi}$	Mother wavelet function
a	Scaling factor
b	Shifting factor

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Motivation

The renewable energy technologies have exploded in popularity in recent years as a means of addressing the fossil fuel dilemma and mitigating climatic hazards. The wave energy is one of the most promising and clean sources of sustainable energy for meeting future energy demands. When compared to other renewable energy sources such as solar and wind, it offers larger prospects and a greater energy density [8]. In determining the wave power level, the Significant Wave Height (SWH) is a highly important and essential metric. It is obtained by taking the average of one-third highest ocean waves observed in a given time interval. In wave power generation, considerable wave height prediction is extremely important. In order to improve the performance of wave energy converters, it is necessary to have a reliable and efficient prediction of significant wave height a few hours ahead of time [9]. Aside from that, hourly efficient SWH forecast may considerably aid in decision-making in marine and off-shore activities. However, due to the extremely random and chaotic behavior of ocean waves, meaningful wave height prediction is a tough and complex endeavor. Traditionally, researchers have employed energy balance models to anticipate waves across a vast geographical and temporal scale utilizing various ocean factors [10]. However, these models are computationally expensive and sophisticated, limiting their utility to hourly SWH predictions for a single area. For short-term wave height forecasting, researchers use data-driven models that predict SWH for the next few hours using time-series data. The parametric algorithms (time-series models) are the one which makes the assumptions for the form or the curve while training to ease the process of learning.

Their numbers of parameters are of fixed size and independent of size of the training data. Researchers have successfully used parametric time-series models such as Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), and others for SWH forecasting. There is some noteworthy work that makes good use of the AR model for SWH prediction which are [11] and [12]. Ge and Kerrigan [13] effectively employed the ARMA model to compute noise covariance in ocean wave data. Vanem and Walker [14] also utilized the ARIMA model to identify the climate in terms of ocean waves.

The non-parametric algorithms are the one which do not assumes anything and changes their size and magnitude according to the training data during learning process. They are dependent on the size of the training set. Non-parametric current machine learning methods have demonstrated to be more promising than parametric time-series models, especially for short-term forecasting tasks. Researchers employed many neural architecture types to arrive at an accurate short-term prediction of SWH. For SWH prediction, Deo et al. utilized a three-layer feed forward ANN model [15]. Using the ANN model, the author enhanced the SWH prediction in [16]. In [17], the authors utilized the Extreme Learning Machine (ELM) in conjunction with a grouping genetic algorithm to find the most effective features set to anticipate short-term SWH. The daily prediction of SWH was obtained using an ensemble of ELM models in [18]. Researchers have found the Recurrent Neural Network (RNN) to be a particularly successful and popular solution for SWH prediction [19] [20]. In [20] [21], multiple forms of the Gated Recurrent Unit (GRU) were employed to estimate the SWH. For short-term SWH prediction, the Long Short Term Memory Network (LSTM) [3] is a popular RNN design. [22] [23] [24] [25] are some key research studies that apply LSTM models for SWH forecasting.

1.2 Related Work

In predicting tasks, researchers frequently use Support Vector Regression (SVR) models [26]. These models are based on statistical learning theory [26], and in their optimization problem, they can minimize an acceptable trade-off between empirical risk and model complexity. Furthermore, unlike other neural network architecture-based machine learning models, these models may reach the global optimal solution. In a number of studies, several forms of the SVR models have been utilized to provide accurate predictions of the SWH. [27] [28] [29] [30] are a few of them.

Despite the use of sophisticated non-linear machine learning methods, there are some circumstances where effective SWH forecasts are difficult to produce, owing to the very unpredictable and chaotic character of ocean waves. In recent years, researchers have begun to develop hybrid wave models to improve prediction in order to overcome these obstacles. A decomposition approach is used in wave hybrid models to deconstruct time-series data into more useful components. Following that, these attributes were fed into machine learning algorithms in order to produce more accurate predictions. Researchers presented hybrid model versions by evaluating various combinations of decomposition methods and machine learning models and evaluated their behaviour on various SWH datasets. Prahlada and Deka suggested a hybrid model that employs the ANN model in conjunction with the Wavelet Decomposition (WD) approach to forecast SWH with a lead time of up to 48 hours [31]. [21] [32] [33] are some more important hybrid models that employ the wavelet decomposition approach in conjunction with a variety of neural network topologies. The Empirical Mode Decomposition (EMD) [5] approach has also been utilized extensively in conjunction with several machine learning algorithms to create efficient hybrid models. Duan et al. employed the EMD [5] approach in conjunction with

the ϵ-SVR model [1] to produce short-term forecasting of significant wave height [10]. Tang et al. employed the Least Square Support Vector Regression (LS-SVR) [2] models in conjunction with the EMD approach to estimate short-term wave height. Ali and Parasad employed an enhanced ensemble EMD approach in conjunction with an ELM model to anticipate major wave height 30 minutes in advance [34].

1.3 Contribution

In this study, we provide a set of wave hybrid models for predicting hourly significant wave height. A signal decomposition approach and a regression model are used in our proposed hybrid models. We have used ϵ-Support Vector Regression (ϵ-SVR), Least Squares Support Vector Regression (LS-SVR), Large- margin Distributed Machine Regression (LDMR) and Long Short Term Memory (LSTM) models for performing the task of prediction of significant ocean wave height. We have exploited the Wavelet Decomposition (WD), Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) approaches for decomposition of signals. Aside from that, we have employed the Particle Swarm Optimization (PSO) approach to fine-tune the parameters of the regression model used in our wave hybrid models. The VMD technique and the LDMR model have never been employed in a wave hybrid model before. Using multiple assessment criteria, we have tested the efficacy of our created wave hybrid models using time-series significant wave heights gathered from four separate buoys positioned at different geographical locations. Following a quick examination of the numerical findings obtained, we conclude that the LDMR-based wave hybrid model surpasses the other regression model-based hybrid mode. Furthermore, VMD-based wave hybrid models outperform other decomposition-based hybrid models. Furthermore, we run two-way ANOVA analysis on the obtained numerical findings, which statistically indicate that, the adoption of a certain

decomposition method and specific regression model impact prediction accuracy considerably, but their effects are independent.

The organizations of rest of this thesis are as follows: In Chapter 2, we have described about different machine learning and metaheuristic models used in our developed wave hybrid models. In Chapter 3, we have briefly discussed the signal decomposition techniques used in our developed hybrid models. Chapter 4 briefly describes methodologies and implementation details of our developed hybrid models. In Chapter 5, we have presented the numerical results and their brief analysis. Chapter 6 concludes this thesis.

Chapter 2

Machine Learning Models and their Parameter Tuning

In this chapter, we shall briefly describe the used machine learning models in our developed wave hybrid models. Further, we have also described the metaheuristic algorithm for tuning the parameters of the machine learning models.

Given a training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1,2,....,l\}$, SVR models minimize a linear combination of the loss function and regularization term to obtain a linear estimate $f(x) : w^T x + b, w \in \mathbb{R}^n, b \in \mathbb{R}$. For estimating the nonlinear function, it finds $f(x) :$ $w^T \phi(x) + b = K(x^T, A^T)u + b$ in feature space, where K is an appropriate kernel satisfying Mercer condition [36] such that $\phi(x_i)^T \phi(x_j) = K(x_i, x_j)$. The matrix A is $l \times n$ data matrix containing l data points in $\mathbb{R}^n$ and u is $l \times 1$ decision variable required for obtaining kernel generated surface.

2.1 ϵ-Support Vector Regression Model

Support Vector Regression (SVR) is a supervised learning approach for forecasting discrete values. The fundamental tenet of SVR is finding the optimum fit line. The flexibility of SVR allows us to decide how much error in our model is acceptable, and it will locate a suitable hyperplane. The hyperplane with the most points is the best-fitting line in SVR. In ϵ-SVR model, there is ϵ-insensitive loss function which enables the model to ignore the error upto ϵ. The Figure 1 shows the plots of the ϵ-insensitive loss function.

The ϵ-SVR model [26] minimizes ϵ-insensitive loss function along with $\frac{1}{2}w^Tw$ regularization. It finds the solution of the following optimization problem:

$$\min_{w,b} \frac{1}{2}w^Tw + C\sum_{i=1}^{l} L_\epsilon(y_i - (w^T\phi(x_i) + b)) \tag{1}$$

where L_ϵ is the epsilon insensitive loss function and is given by:

$$L_\epsilon(u) = \max(|u| - \epsilon, 0) = \begin{cases} |u| - \epsilon \; ; \; if \; |u| > \epsilon, \\ 0 \qquad ; otherwise \end{cases} \tag{2}$$
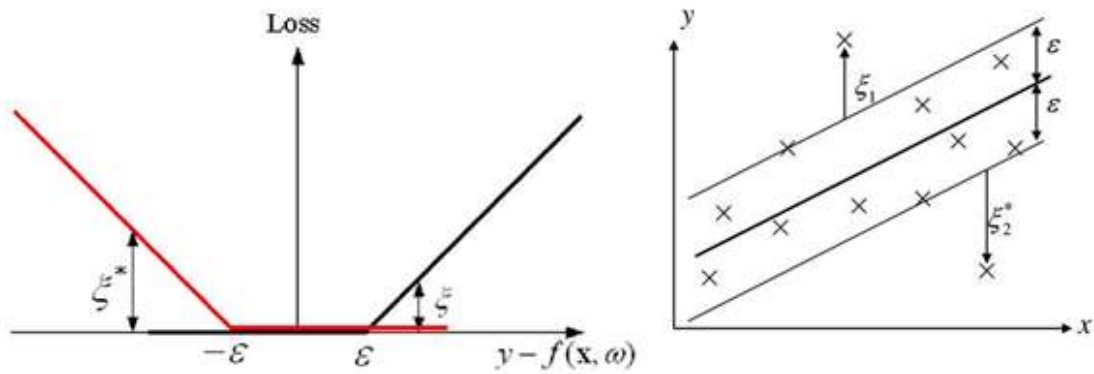
Figure 1: ϵ-insensitive loss function

After introducing the slack variables ξ_i and ξ_i^* for $i = 1,2,\dots,l$ and the user defined penalty factor C, the problem (1) can be reduced as :

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}w^Tw + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$subject \; to \quad y_i - (w^T\phi(x_i) + b) \leq \epsilon + \xi_i,$$

Image Reference: Linear Regression and Support Vector

Regression (adelaide.edu.au)

$$(w^T \phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*,$$

$$\xi_i, \xi_i^* \geq 0 , i = 1,2,3, \dots, l \quad . \tag{3}$$

The solution of the primal problem (3) is computationally expensive. Therefore, we need to derive the equivalent Wolfe dual problem [38] for the primal problem (3). For this, we obtain the Lagrangian function for primal problem (3) as:-

$$L(w,b,\xi,\xi^*,\alpha,\alpha^*,\eta,\eta^*) = \frac{1}{2}w^T w + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}(\eta_i \xi_i + \eta_i^* \xi_i^*)$$

$$- \sum_{i=1}^{l}\alpha_i(\epsilon + \xi_i + (y_i - (w^T\phi(x_i) + b)))$$

$$- \sum_{i=1}^{l}\alpha_i^*(\epsilon + \xi_i^* - (y_i - (w^T\phi(x_i) + b))) \tag{4}$$

where, α, ξ, η are the Lagrangian multiplier vectors. Hence, the dual problem of the primal problem (4) can be written as:

$$\min_{\alpha^{(*)} \in R^{2l}} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(K(x_i,x_j)) + \epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*),$$

$$subject\ to \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0,$$

$$0 \leq \alpha_i, \alpha_i^* \leq C , i = 1,2,3, \dots, l \tag{5}$$

then the solution $(\overline{w}, \overline{b})$ to the above primal problem (4) can be obtained by:

$$\overline{w} = \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)\phi(x_i) \tag{6}$$

To compute the value of b, we will consider the index j where,

$$0 < \alpha_j < C$$

$$\bar{b} = y_j - \bar{w}^T x_j + \epsilon \tag{7}$$

Or we can also consider the index k

$$0 < \alpha_k < C$$

and compute $\qquad \bar{b} = y_k - \bar{w}^T x_k - \epsilon \tag{8}$

2.2 Least Squares Support Vector Regression Model

The typical ϵ-SVR can solve a convex quadratic optimization problem with a unique solution, but its method of solving the problem is quite complex. Suykens et.al. [2] proposed the least squares support vector machine (LS-SVR) by converting inequality requirements to equality constraints and substituting empirical risk deviation with quadratic deviation. LS-SVR preserves the features of SVR while handling the issues including nonlinearity and small sample sizes. Additionally, LS-SVR solves problems more quickly and uses less processing resources.

The LS-SVR model [2] minimizes the quadratic loss function along with $\frac{1}{2}w^T w$ regularization term. It minimizes:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i^2$$

$$subject\ to \quad y_i - (w^T \phi(x_i) + b) = \xi_i, i = 1,2,3, \dots, l \tag{9}$$

The primal problem for the above problem:

$$L(w, \alpha, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i^2 + \sum_{i=1}^{l} \alpha_i (y_i - (w^T \phi(x_i) + b) - \xi_i) \quad (10)$$

where, $C > 0$ is a user defined parameter. The solution of the problem (10) can be obtained by solving the following system of equations.

$$\begin{bmatrix} 0 & e^T \\ e & K(A, A^T) + \frac{2}{C} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (11)$$

After obtaining the (b, α) by solving the above system of equations, we can estimate our regression function for a given x ∈ $\mathbb{R}$ using

$$f(x) = w^T \phi(x) + b = K(x^T, A^T)\alpha + b \quad (12)$$

2.3 Large-margin Distribution Machine Regression Model

In the presence of normal noise in the data, the least square loss function utilized in the LS-SVR model allows it to perform optimally. In the case of uniform noise, the ϵ-insensitive loss function in the ϵ-SVR model allows it to perform optimally. We should also try to reduce the spread of data points within the ϵ -tube. As a result, including both of these loss functions in the suggested formulation allows for a trade-off between sparsity and scatter reduction. It also allows the proposed model to use the entire training set's information while avoiding over-fitting.

For quantifying empirical risk, the LDMR model [4] minimizes the linear combination of the quadratic loss and ϵ-insensitive loss function, as well as the $\frac{1}{2} w^T w$ regularization component. It allows it to outperform both the ϵ-SVR and the LS-SVR models. The LDMR model [4] minimizes:

$$\min_{w,b} \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^{l} (y_i - (w^T \phi(x_i) + b)^2) + C \sum_{i=1}^{l} L_\epsilon (y_i - (w^T \phi(x_i) + b), \quad (13)$$

which can be converted to the following QPP

$$\min_{w,b,\xi_1,\xi_2} \frac{c}{2}\|w\|^2 + \frac{1}{2}\|Y - (Aw + eb)\|^2 + Ce^T(\xi_1 + \xi_2)$$

$$subject \ to \quad Y - (Aw + eb) \leq e\epsilon + \xi_1,$$

$$(Aw + eb) - Y \leq e\epsilon + \xi_2,$$

$$\xi_1, \xi_2 \geq 0 \tag{14}$$

where C, e, c and ϵ are the user specified positive parameters. We prefer to solve Wolfe dual of the primal problem (14) which can be obtained as follows,

$$\min_{\alpha_1,\alpha_2} \frac{1}{2}(\alpha_1 - \alpha_2)^T H\big(c(\alpha_1 - \alpha_2) + Y^T H(cI_0 + H^T H)^{-1}H^T(\alpha_1 - \alpha_2)\big)$$
$$-Y^T(\alpha_1 - \alpha_2) + \epsilon e^T(\alpha_1 + \alpha_2),$$

$$subject \ to \quad 0 \leq \alpha_1, \alpha_2 \leq Ce \tag{15}$$

Here $H = [K, e]$ is a augmented matrix and $I_0 = \begin{bmatrix} I & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 \dots & 0 \end{bmatrix}$, where I is a $n \times n$ identity matrix.

After obtaining the solution of problem (15), the solution vector $u = [w, b]^T$ can be obtained by using $u = [w, b]^T = (cI_0 + H^T H)^{-1}H^T(\alpha_1 - \alpha_2 + Y)$ and regression function using is estimated using (12).

2.4 Long Short Term Memory (LSTM)

Long short-term memory (LSTM) is a type of artificial neural network used in deep learning and artificial intelligence. LSTM features feedback connections, unlike normal feed-forward neural networks. This type of recurrent neural network can handle not just single data points (like

photos), but also complete data sequences (such as speech or video). Because there might be delays of undetermined duration between critical occurrences in a time series, LSTM networks are well-suited to categorizing, processing, and generating predictions based on time series data. LSTMs were created to solve the problem of vanishing gradients that can occur while training standard RNNs. In many cases, LSTM has an advantage over RNNs, hidden Markov models, and other sequence learning approaches due to its relative insensitivity to gap length.
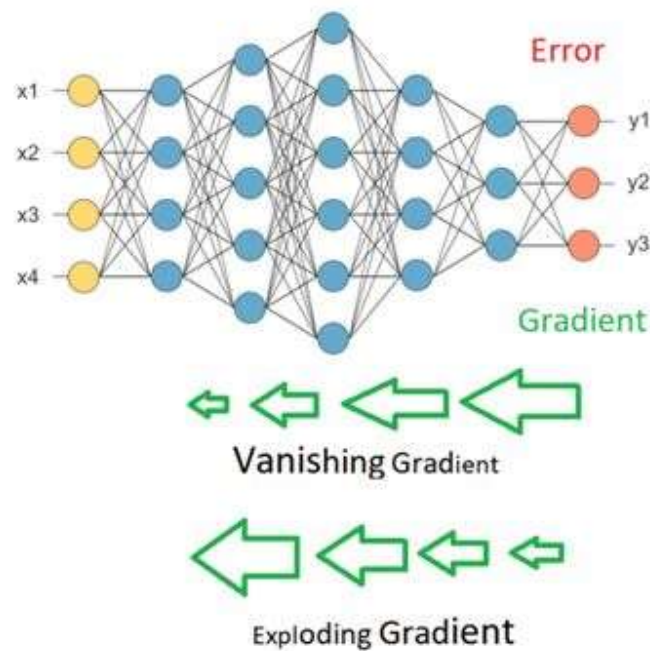
Figure 2: Vanishing and exploding gradient error explanation

To update the network's weights, a neural network employs an algorithm known as Back-Propagation. So, first, BP determines the gradients from the mistake using the chain rule in Calculus, and then it updates the weights (Gradient descent). Because the BP commences from the output layer and goes all the way back to the input layer, updating weights may not be a

problem in a simple neural network, but it may be in a deep neural network. As we return to the gradients, the values may get smaller exponentially, causing a Vanishing Gradient problem, or larger exponentially, causing an Exploding Gradient problem. Figure 2 and Figure 3 gives the pictorial representation of the exploding and vanishing gradients.
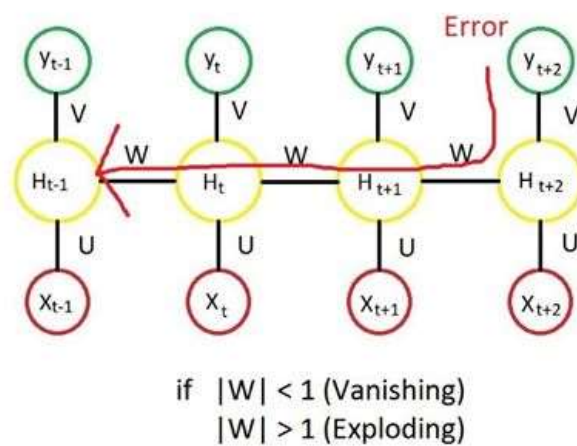
Figure 3: Vanishing and exploding gradient error flow

As a result, we have issues with network training. We have time steps in RNNs, and the current time step value is dependent on the prior time step, thus we must go back in time to perform an update. The Figure 4 below depicts an accurate description of LSTM as well as solution to the aforementioned challenges.

LSTM Network

- ◉ = Element-wise multiplication
- ⊕ = Element-wise addition

$$f_t = \sigma \,(X_t * U_f + H_{t-1} * W_f)$$
$$\bar{C}_t = \tanh \,(X_t * U_c + H_{t-1} * W_c)$$
$$I_t = \sigma \,(X_t * U_i + H_{t-1} * W_i)$$
$$O_t = \sigma \,(X_t * U_o + H_{t-1} * W_o)$$

$$C_t = f_t * C_{t-1} + I_t * \bar{C}_t$$
$$H_t = O_t * \tanh \,(C_t)$$

X_t = Input vector
H_{t-1} = Previous cell Output
C_{t-1} = Previous Cell Memory
H_t = Current cell Output
C_t = Current cell Memory

W, U = weight vectors for forget gate (f), candidate (c), i/p gate (I) and o/p gate (O)

Note : These are different weights for different gates, for simpicity's sake, I mentioned W and U

Figure 4: Working of the LSTM cell network

At any time *t,* the inputs to an lstm cell are X_t (current input), H_{t-1} (previous cell output) and C_{t-1} (previous cell memory), outputs are H_t (current cell output) and C_t (current cell memory) and the intermediate gates are f_t (forget gate), $\bar{C}_t$ (candidate gate), I_t (input gate) and O_t (output gate). Except for the candidate gate, which employs the *tanh* activation function, all of the gates utilize the *sigmoid* activation function. $U = (U_f, U_c, U_i, U_o)$ and $W = (W_f, W_c, W_i, W_o)$ are the weight vectors of the current input and previous output respectively for the forget, candidate, input and output gate operations repectively.

These gates take the current input and previous output, execute element-wise multiplication with their corresponding weight vectors, and then sum the results. After which the activation functions is applied to the corresponding results.

These gate operations generate vectors in between -1 to 1 in case of tanh activation function and 0 to 1 in case of sigmoid activation function and accordingly we obtain our forget, candidate, input and output gate vectors. Using these gate vectors, current memory state C_t and current output state H_t are obtained, which can be studied from Figure 4.

Image Reference: Chapter 10.1: DeepNLP — LSTM (Long Short Term Memory) Networks with Math. | by Madhu Sanjeevi (Mady) | Deep Math Machine learning.ai | Medium

2.5 Parameter Tuning

In 1995, Kennedy and Eberhart proposed particle swarm optimization (PSO) [7]. Initially, they stated in their first research paper that an epistemologist believed that a group of socio-biological animals moving forward together "can profit from the experience of all other members," i.e., when a flying bird is randomly hunting for food, the remaining members of the flock can share their experience of food search and discovery, allowing the entire flock to get the best hunt faster.

While we can emulate a flock of birds' movement, we can also suppose that each bird is aiding us in identifying the best solution in a high-dimensional solution space, with the flock's best solution being the best solution in the space. This is a method based on metaheuristics. We can never be sure whether or not the genuine global optimum solution exists, and it rarely does. We commonly find, however, that the PSO solution is quite close to the global optimum.

Assume we have p particles and we denote the positions of the particles at i^{th} iteration as $X_i = (x_{1i}, x_{2i}, \ldots, x_{pi},)$ and the velocities of the particles are denoted as $V_i = (v_{1i}, v_{2i}, \ldots, v_{pi},)$, now we can update the positions in the next iterations given as follows:

$$X_{i+1} = X_i + V_i \tag{16}$$

and also at the same instant, the updated velocity will be given as:

$$V_{i+1} = w * r * V_i + c1 * r1 * (pbest_i - X_i) + c2 * r2 * (gbest - X_i) \tag{17}$$

In the equation (25), r, $r1$ and $r2$ are the random numbers ranging from 0 to 1, $c1$ and $c2$ are the acceleration co-efficients whose values are kept generally equal to 2 and w is the inertial weight

which is equal to 1 generally. The flow of the process can be studied and visualized from the
Figure 5.
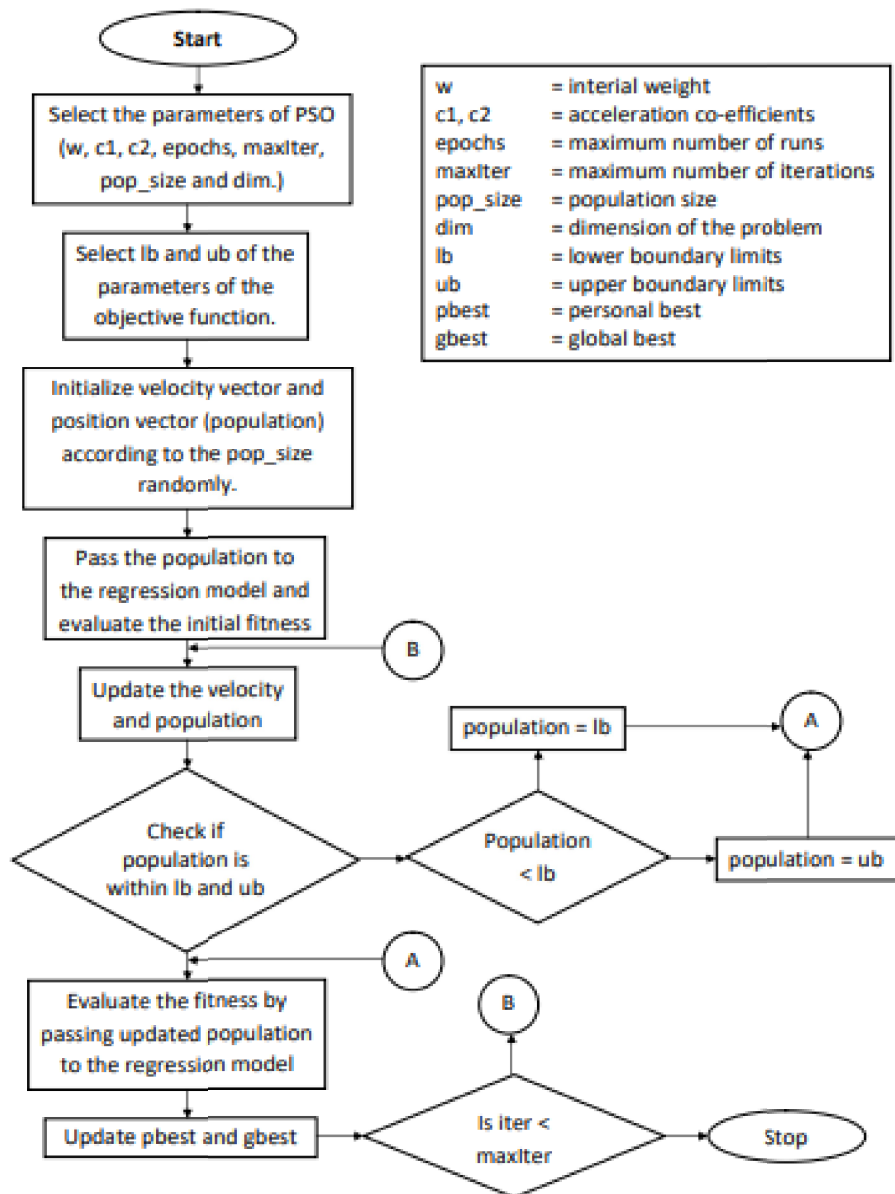
Figure 5: Flowchart of PSO

Chapter 3

Decomposition Methods

We have used the decomposition methods to generate informative and robust features for machine learning models in our developed wave hybrid models. In this chapter, we have briefly described the decomposition methods used in our wave hybrid models.

3.1 Wavelet Decomposition

A Wavelet is a time-localized wave-like oscillation. Scale and location are the two most basic features of wavelets. The scale (or dilation) of a wavelet determines how "stretched" or "squished" it is. This feature has to do with the frequency of waves. The wavelet's position in time is defined by its location (or space). In this study, discrete wavelet transform (DWT) is used to compute decompositions since the datasets are in the form of the discrete time sequence. The discrete wavelet transform can be expressed as,

$$w_\phi(a,b) = \frac{1}{\sqrt{n}} \sum_m x(m)\phi_{a.b}(m) \tag{18}$$

where, *x is function to be decomposed or transformed*, ϕ is *mother(base) wavelet function*, *A is the number of decompositions required*, *a is the scaling factor* $= 0,1,2,\dots,A-1$, $n = 2^A$, $m = 0,1,2,\dots,n-1$ *and b is the shifting factor* $= 0,1,2,\dots,2^a-1$

In other words, we select a wavelet of a specific scale. Then we slide this wavelet through the full signal, varying its position, and multiply the wavelet and signal at each time step. This multiplication yields a coefficient for that wavelet scale at that time step. The procedure is then repeated by increasing the wavelet scale. Basically, this decomposition is repeated to improve frequency resolution, and the approximation coefficients are decomposed with high-pass and

low-pass filters before being down-sampled. The binary tree, with each node indicating a distinct time-frequency localization of a sub-space is known as a filter bank. Figure 6 represents the decomposition workflow in the filter bank.
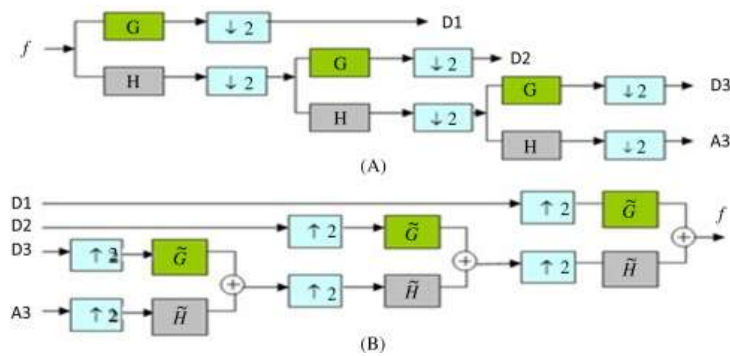
Figure 6: Decomposition workflow in the filter bank

3.2   Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) [5] is a powerful signal processing method, particularly for non-linear and non-stationary data. In case of wavelet decomposition (WD), the prior/initial information or knowledge for mother wavelet is required which drives the algorithm for the appropriate decomposition whereas EMD is absolutely data driven where no prior information is required and decomposes the signal recursively until desired specifications are met. We shall now discuss the EMD decomposition technique. The EMD approach divides an arbitrary signal $x(t)$ into many Intrinsic Mode Functions (IMFs) plus a residual function. There are two requirements that the IMFs must meet:

- The number of zero-crossings and extrema should be identical or differ by at most one.

- The mean of the upper and lower envelopes, which are functions of local maxima and minima respectively, should be zero.
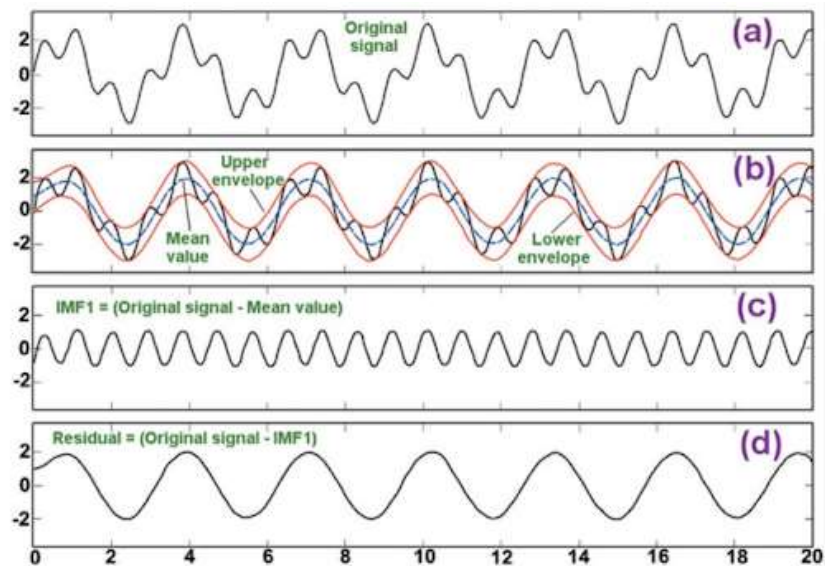
Figure 7: Internal work flow of EMD

The EMD approach iteratively detects and estimates the lower and upper envelopes by detecting local minima and maxima. The average of envelopes is removed as 'low pass' center line. The high-frequency components are segregated as 'mode' in this fashion as shown in Figure 7. The process is repeated by decomposing the remaining low pass center line iteratively. The flow of EMD decomposition is depicted in the following Figure 8 below.
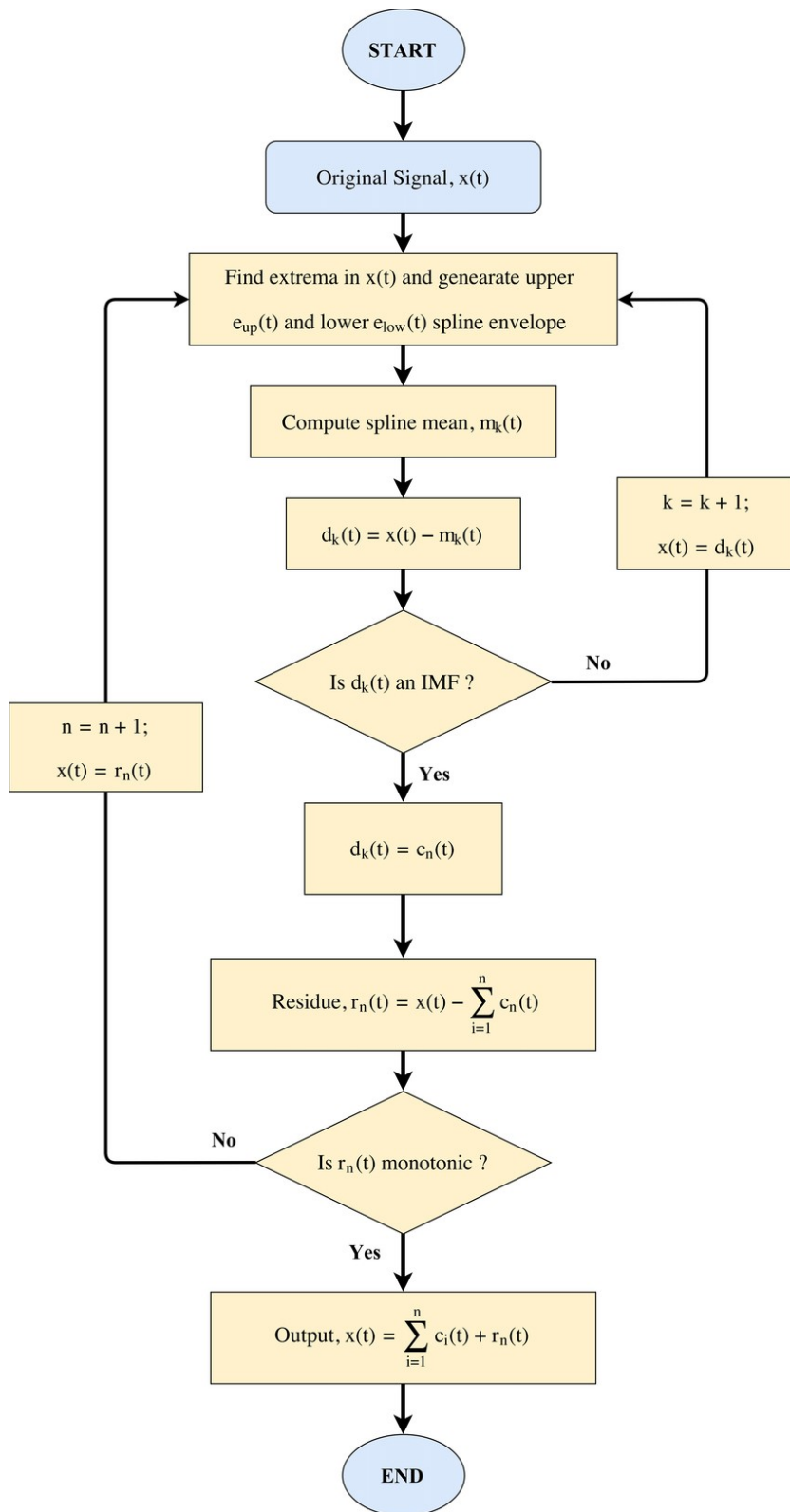
Figure 8: Flowchart of EMD

3.3 Variational Mode Decomposition (VMD)

The EMD approach is vulnerable to noise and sampling effects, according to researchers [5]. The Variational Mode Decomposition (VMD) approach is an adaptive, non-recursive method for obtaining several modes at the same time. The model seeks for an ensemble of modes and their corresponding centre frequencies that collectively reconstruct the input signal while each getting smooth after demodulation into base band.

The input signal f is decomposed into different frequency components (a modal component function) $u_k(t)$ using VMD. VMD decomposed the daily stream-line sequence X into $n + 1$ intrinsic mode functions (IMFs), i.e., $u_1, u_2, \ldots, u_n$. Each mode $u_k(t)$ had a finite bandwidth with different centre frequencies in which the minimum computed bandwidth's summation of each mode was required. The more details on VMD can be studied from [6]. Figure 9 gives an idea of flow of VMD.
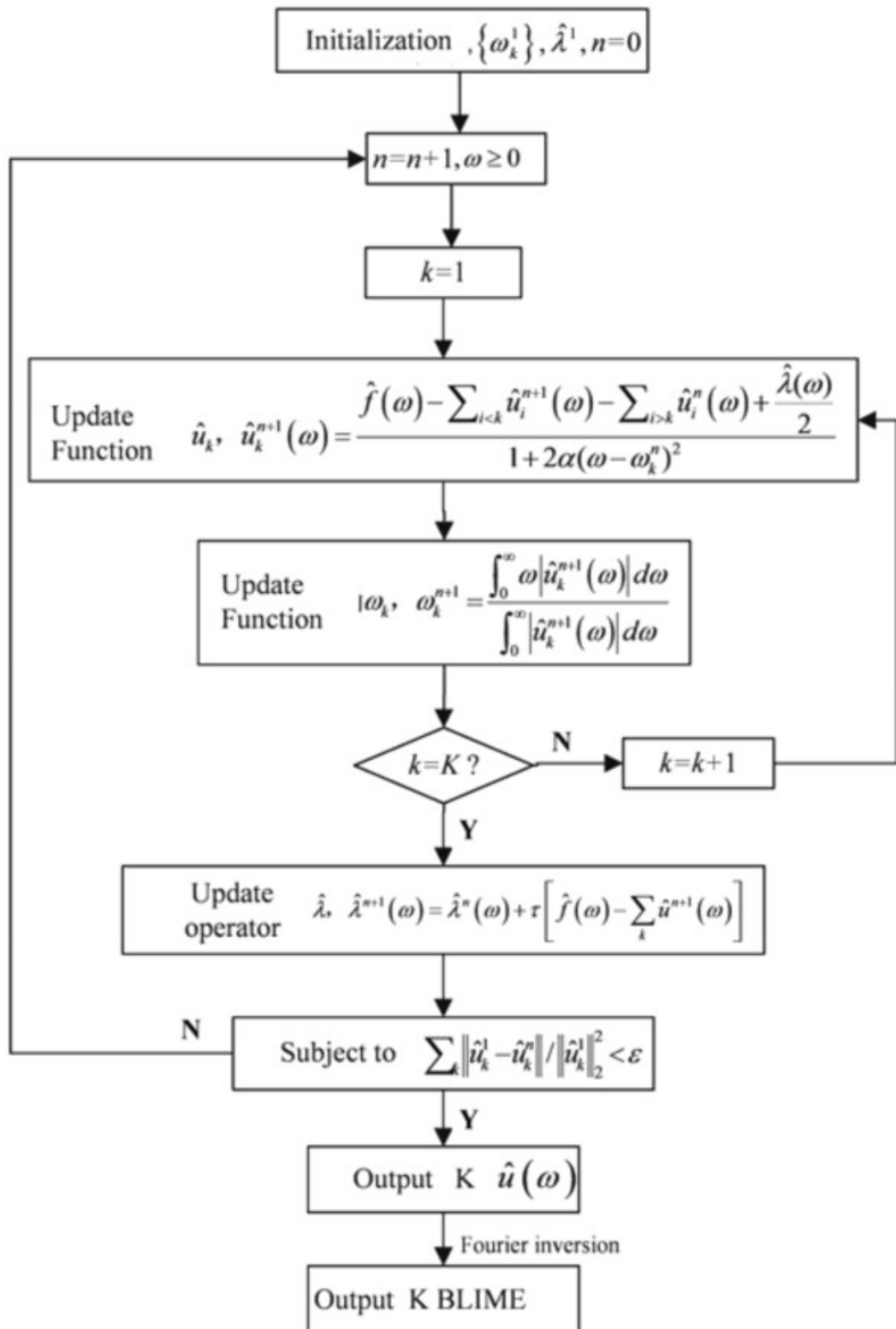
Figure 9: Flowchart of VMD

Image reference: https://www.researchgate.net/publication/324917202/69865/

Chapter 4

Methodology of Proposed Wave Hybrid Models for SWH Prediction and their Implementation Details

In this chapter, we shall briefly describe our methodology used by our developed wave hybrid models along with their implantation details.

4.1 Methodology of Proposed Wave Hybrid Models

Our wave hybrid model involves different successive steps which we have presented in the Figure 10.

We describe these steps as below:

1. **Decomposition of signal:** After receiving the time series SWH data, we need to decompose it into more informative data signals. The decomposition of the signal is important as it helps to obtain the more informative features for our machine learning method. In our developed wave hybrid models, we have used the WD, EMD or VMD decomposition method to decompose the time series SWH $x(t)$ into $[x_1(t), x_2(t), \dots, x_n(t)]$ component data signals.

2. **Attributes/features composition:** After decomposing the SWH signal, we need to prepare the dataset for machine learning model. For this, we need to compose the feature set and corresponding response values. For i^{th} signal component, we compose the p-dimensional $[x_i(t), x_i(t-1), x_i(t-2), \dots, x_i(t-(p-1))]$ features for the prediction of $x_i(t+1)$.

3. **Parameters Tuning:** We need to train the n machine learning models to obtain the prediction for each $x_i(t+1)$ for $i = 1,2,..n$, using our composed feature sets. But, before training a machine learning model, we need to tune its parameters. We have used the PSO algorithm to tune the parameters of our machine learning models.

4. **Prediction:** After selecting the appropriate values of parameters for our machine learning model, we finally train them for obtaining the prediction of $x_i(t+1)$ for $i = 1,2,..n$.

5. **Aggregation:** After obtaining the prediction of $x_i(t+1)$ for $i = 1,2,..n$, we aggregate them to obtain the final prediction for $x(t+1)$.
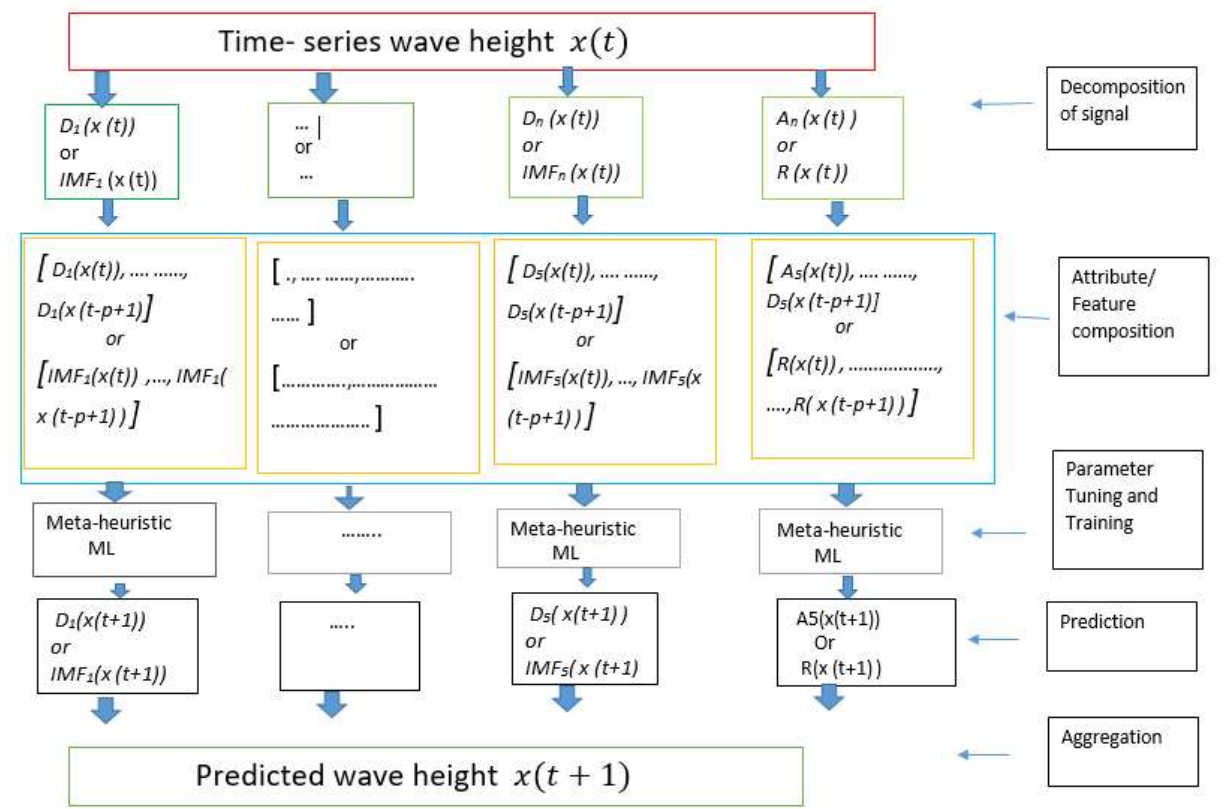
Figure 10: Flowchart of wave hybrid model

4.2 Implementation Details

We will be systematically describing the implementation details of our developed wave hybrid models. We have used the WD, EMD and VMD methods to develop the different wave hybrid models.

In Wavelet decomposition based hybrid models, we have decomposed the SWH data signal into five high frequency detail signals (D_1, D_2, D_3, D_4, D_5) and a low frequency approximate signal A_5. For this, we have used the Daubechies 4 discrete wavelet filter available in MATLAB (www.mathworks.com). In Figure 12, we have shown the wavelet decomposition for the dataset A (refer Table 1). For our EMD and VMD decomposition based wave hybrid models, we have decomposed the initial data signal into six IMFs using their respective algorithms. In Figure 13, we have shown the VMD decomposition for the dataset A (refer Table 1). After decomposing the original signal into six different components, we have composed the feature sets and response values in each case and divided the obtained datasets into training set and testing set. The 80% of a dataset was used as training set and remaining was used as testing set.

We have used the SVR, LS-SVR, LDMR and LSTM models for the prediction of SWH data. We have implemented the SVR, LS-SVR and LDMR models in MALTAB environment by writing an appropriate function in MATLAB. The SVR and LDMR model require the solution of the dual QPPs (5) and (15) respectively. The solution of these QPPs of SVR and LDMR models have been obtained using the 'quadprog' function available in MATLAB by using the 'interior-point convex' algorithm. The LS-SVR model only requires the solution of a system of equations to obtain its solution. In SVR, LDMR and LS-SVR models, we have used the RBF kernel of the form $k(x,y) = e^{-q\|x-y\|}$, where q is the kernel parameter.

We need to tune and obtain the appropriate parameters of our machine learning model for a given dataset before using them for obtaining the prediction. We have used the PSO algorithm to tune the parameters of our used machine learning models. We have illustrated the flow chart of used PSO algorithm in our hybrid model in Figure 5. The LDMR model involves four parameters namely ϵ, C, c and kernel parameter q. The ϵ-SVR involves three parameters namely ϵ, C and kernel parameter q and LS-SVR models involve two parameters C and kernel parameter q only. The ranges of parameters ϵ, C, c and q have been fixed with [0, 2], [0, 1024], [0, 1024], [0, 1024] respectively for all considered SVR models. For the evaluation of fitness function in our PSO algorithm, we need to obtain the testing error by solving the dual problem (5) and (15) for ϵ-SVR and LDMR model respectively. For the LS-SVR model, we need to obtain the solution of system of equations (11). In LSTM, the number of hidden unit used were 100 units, optimizer used was 'adam' optimizer, maximum number of epochs taken was 1000, the gradient threshold was kept 0.01 and initial learning learning rate was 0.0001.

Chapter 5

Experimental Results

In this chapter, we have compared the performances of developed wave hybrid models along with different existing wave hybrid models on different ocean significant wave height datasets. Further we have briefly analyzed our obtained numerical results.

5.1 Dataset Description

We have collected the time-series hourly SWH data from the different ocean buoys available on the National Data Buoy Center (NDBC) (https://www.ndbc.noaa.gov/). At first, we have listed brief details of chosen ocean buoys datasets at Table 1. The location of the considered buoy stations has also been mapped at Figure 11. We have shown the VMD and WD decomposition of the time-series signal for dataset A (refer Table 1) at Figure 12 and Figure 13.
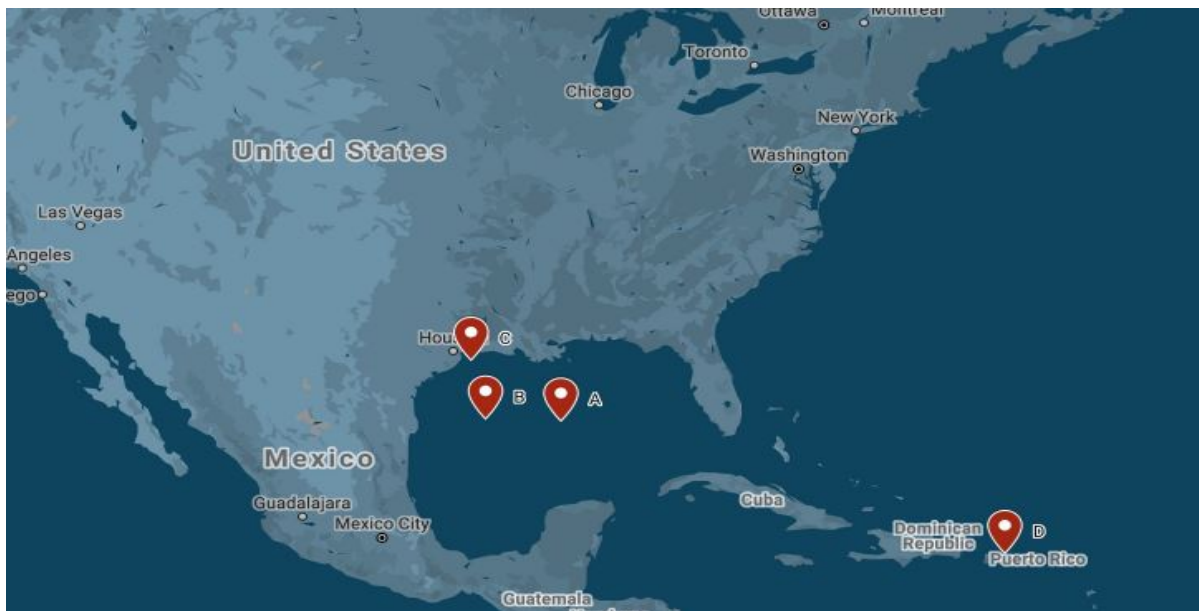
Figure 11: Pin locations of the buoy stations

Dataset	Station ID	Station Name	Co-ordinates	Time	No. of samples	Mean, STD	Payload	Type of Buoy
A	410532	San Juan, PR	66°5'58" W, 18°28'27" N	Feb'21 - Mar'21	571	1.5452, 0.2812	Scoop	2.3-meter foam discus buoy
B	42001	MID GULF - 180 nm South of Southwest Pass, LA	89°39'25" W, 25°56'31" N	Jan'15 - Apr'15	380	1.1354, 0.6125	Scoop	3-meter discus buoy
C	42002	WEST GULF - 207 NM East of Brownsville, TX	93°38'46" W, 26°3'18" N	Jan'15 - Apr'15	379	1.1785, 0.6754	Scoop	3-meter foam buoy
D	42035	GALVESTON,TX - 22 NM East of Galveston, TX	94°24'45" W, 29°13'54" N	Jan'15 - Apr'15	378	0.7570, 0.4047	Scoop	Moored Buoy

Table 1: Description of datasets
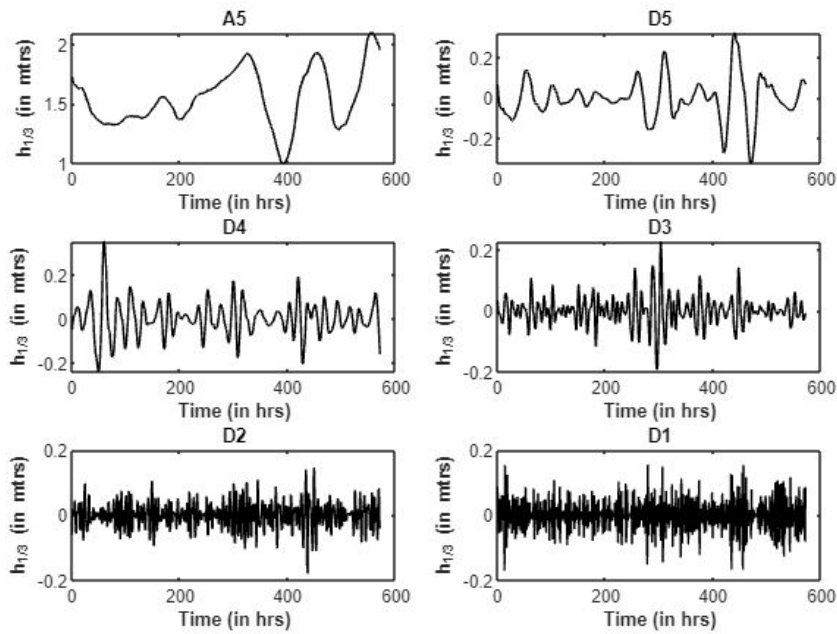
Figure 12: WD decomposition plot of dataset A
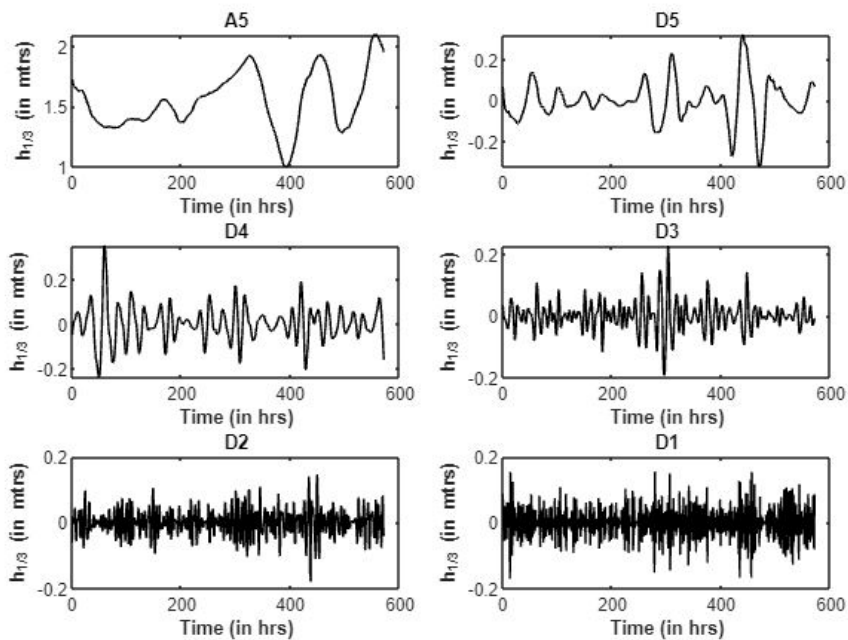
Figure 13: VMD decomposition plot of dataset A

5.2 Evaluation Criteria

For evaluating the different considered wave hybrid models, we need to fix the evaluation criterion first. We have used the commonly used evaluation criterion for measuring the performance of 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:

- Root Mean Square of Errors (RMSE):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- Mean of Absolute Deviations (MAD):

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- Mean Absolute Percentage Error (MAPE):

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

5.3 Results and Discussion

After fixing the evaluation criteria, we present our numerical results from the extensive experiments. We have listed the performance of proposed PSO-LDMR, PSO-VMD-LDMR, PSO-VMD-LS-SVR, PSO-VMD-SVR, PSO-VMD-LSTM, PSO-EMD-LDMR and PSO-WD-LDMR models along with PSO-WD-SVR, PSO-WD-LS-SVR, PSO-WD-LSTM, PSO-EMD-SVR, PSO-EMD-LS-SVR, PSO-EMD-LSTM, PSO-SVR, PSO-LS-SVR and PSO-LSTM models on our selected SWH datasets using RMSE, MAD and MAPE at Table 2 and Table 3.

Dataset	Method	RMSE	MAD	MAPE	Dataset	Method	RMSE	MAD	MAPE
	PSO-LDMR	**0.1037**	**0.0866**	**5.11**		**PSO-LDMR**	**0.2965**	**0.23**	**17.49**
	PSO-SVR	0.112	0.0888	5.15		PSO-SVR	0.3195	0.2379	17.41
	PSO-LS-SVR	0.1039	0.0876	5.18		PSO-LS-SVR	0.3305	0.2554	19.48
	LSTM	0.224	0.1832	10.41		LSTM	0.7271	0.627	47.22
	WD-PSO-LDMR	**0.0381**	**0.0307**	**1.91**		**WD-PSO-LDMR**	**0.1226**	**0.0908**	**7.17**
	WD-PSO-SVR	0.6061	0.5135	4.78		WD-PSO-SVR	0.1168	0.0911	7.22
	WD-PSO-LS-SVR	0.049	0.0391	2.43		WD-PSO-LS-SVR	0.1873	0.1443	12.01
4105322021	WD-PSO-LSTM	0.1219	0.1032	6.26	42001h2015	WD-LSTM	0.1802	0.2047	25.24
A	**EMD-PSO-LDMR**	**0.1896**	**0.1669**	**10.56**	B	**EMD-PSO-LDMR**	**0.3189**	**0.2614**	**19.72**
	EMD-PSO-SVR	0.177	0.1559	9.81		EMD-PSO-SVR	0.3916	0.3048	21.41
	EMD-PSO-LS-SVR	0.1768	0.1556	9.69		EMD-PSO-LS-SVR	0.4153	0.3407	25.10
	EMD-PSO-LSTM	0.357	0.3039	17.05		EMD-PSO-LSTM	0.2342	0.3331	45.92
	VMD-PSO-LDMR	**0.0344**	**0.0283**	**2.52**		**VMD-PSO-LDMR**	**0.0661**	**0.0538**	**4.23**
	VMD-PSO-SVR	0.0349	0.0285	1.72		VMD-PSO-SVR	0.2238	0.1702	13.51
	VMD-PSO-LS-SVR	0.042	0.0348	2.08		VMD-PSO-LS-SVR	0.0616	0.0584	4.57
	VMD-LSTM	0.1491	0.132	8.12		VMD-LSTM	0.2022	0.1603	13.30

Table 2: Numerical Results of datasets A and B

Dataset	Method	RMSE	MAD	MAPE	Dataset	Method	RMSE	MAD	MAPE
	PSO-LDMR	**0.2812**	**0.2233**	**20.95**		**PSO-LDMR**	**0.241**	**0.0876**	**5.18**
	PSO-SVR	1.0469	0.8815	70.74		PSO-SVR	0.2475	0.1919	26.94
	PSO-LS-SVR	0.3055	0.2323	21.45		PSO-LS-SVR	0.2502	0.1941	29.25
	LSTM	0.6587	0.4807	39.44		LSTM	0.5597	0.4031	47.63
	WD-PSO-LDMR	**0.1137**	**0.081**	**7.24**		**WD-PSO-LDMR**	**0.1**	**0.0671**	**9.28**
	WD-PSO-SVR	0.1072	0.0817	7.75		WD-PSO-SVR	0.0955	0.0642	9.43
	WD-PSO-LS-SVR	0.1344	0.1074	10.75		WD-PSO-LS-SVR	0.1859	0.1418	18.99
42002h2015	WD-LSTM	0.32	0.2802	27.23	42035h2015	WD-LSTM	0.2832	0.2305	34.42
C	**EMD-PSO-LDMR**	**0.3123**	**0.2425**	**21.57**	D	**EMD-PSO-LLDMR**	**0.1983**	**0.1531**	**19.31**
	EMD-PSO-SVR	0.2991	0.2273	20.28		EMD-PSO-SVR	0.2068	0.1528	21.12
	EMD-PSO-LS-SVR	0.3202	0.2531	22.46		EMD-PSO-LS-SVR	0.2542	0.1776	21.45
	EMD-LSTM	0.3589	0.285	27.04		EMD-LSTM	0.295	0.2385	31.73
	VMD-PSO-LDMR	**0.0672**	**0.05**	**4.97**		**VMD-PSO-LDMR**	**0.0592**	**0.0443**	**6.77**
	VMD-PSO-SVR	0.0667	0.0527	5.40		VMD-PSO-SVR	0.0613	0.0486	7.55
	VMD-PSO-LS-SVR	0.0711	0.0536	5.54		VMD-PSO-LS-SVR	0.083	0.0625	9.62
	VMD-LSTM	0.2378	0.2735	21.33		VMD-LSTM	0.306	0.2735	34.21

Table 3: Numerical Results of datasets C and D

Now, we shall briefly analyze the numerical results presented in the Table 2 and Table 3 and

Figure 14 attempts to verify our claims. At first, we plot the accuracy obtained by PSO-LDMR,

PSO-SVR, PSO-LS-SVR and PSO-LSTM regression models when used with different decomposition methods in Figure 14 on our considered SWH datasets A, B, C and D.

Top-left chart:

	410532	42001	42002	42035
PSO-LLDMR	94.89	82.59	79.05	74.84
WD-PSO-LLDMR	98.09	92.83	92.76	90.72
EMD-PSO-LLDMR	89.44	80.28	78.43	80.69
VMD-PSO-LLDMR	98.32	95.77	95.03	93.23

Top-right chart:

	410532	42001	42002	42035
PSO-SVR	94.85	82.51	29.26	73.06
WD-PSO-SVR	95.22	92.78	92.25	90.57
EMD-PSO-SVR	90.19	78.59	79.72	78.88
VMD-PSO-SVR	98.28	86.49	94.6	92.45

Bottom-left chart:

	410532	42001	42002	42035
PSO-LSTM	89.59	52.78	60.56	52.37
WD-LSTM	93.74	74.76	72.77	65.58
EMD-LSTM	82.95	54.08	72.96	68.27
VMD-LSTM	91.88	86.7	78.67	65.79

Bottom-right chart:

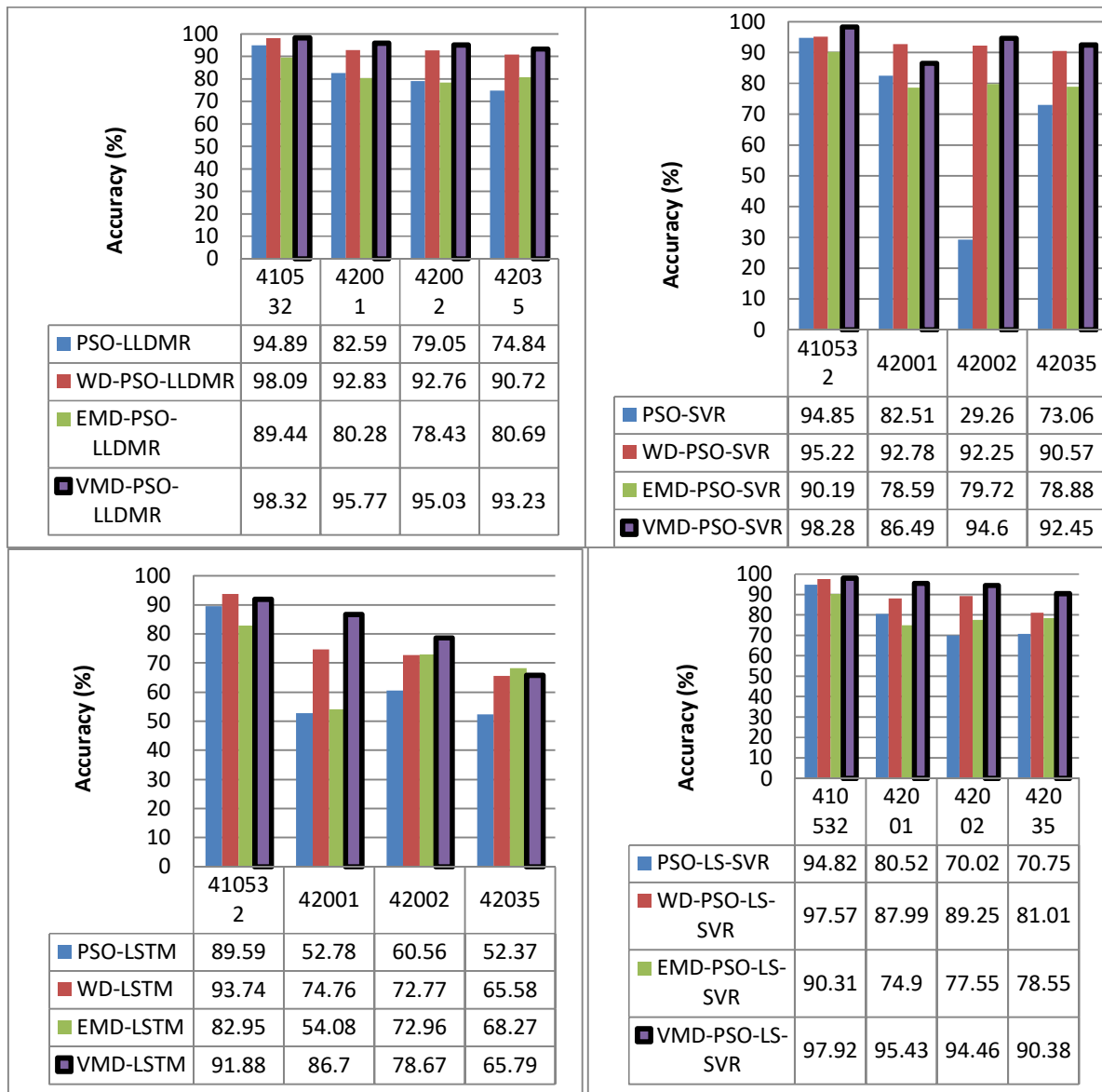|	410532	42001	42002	42035
PSO-LS-SVR	94.82	80.52	70.02	70.75
WD-PSO-LS-SVR	97.57	87.99	89.25	81.01
EMD-PSO-LS-SVR	90.31	74.9	77.55	78.55
VMD-PSO-LS-SVR	97.92	95.43	94.46	90.38

Figure 14: Comparison of the accuracies of the different machine learning models

We can observe from the Table 2, Table 3 and Figure 14 that incase of VMD and WD decomposition, the PSO-LDMR model obtains the best accuracy as compared to PSO-SVR, PSO-LS-SVR and LSTM models for all considered datasets. Also, the PSO-LDMR model

obtains the best accuracy among all other used regression models for dataset A, C and D when no any decomposition technique is used. In case of EMD decomposition also, the PSO-LDMR model obtains the best accuracy for the dataset B and D.

42002

	PSO	WD-PSO	EMD-PSO	VMD-PSO
SVR	29.26	92.25	79.72	94.6
LLDMR	79.05	92.76	78.43	95.03
LS-SVR	70.02	89.25	77.55	94.46
LSTM	60.56	72.77	72.96	78.67

42001

	PSO	WD-PSO	EMD-PSO	VMD-PSO
SVR	82.51	92.78	78.59	86.49
LLDMR	82.59	92.83	80.28	95.77
LS-SVR	80.52	87.99	74.9	95.43
LSTM	52.78	74.76	54.08	86.7

410532

	PSO	WD-PSO	EMD-PSO	VMD-PSO
SVR	94.85	95.22	90.19	98.28
LLDMR	94.89	98.09	89.44	98.32
LS-SVR	94.82	97.57	90.31	97.92
LSTM	89.59	93.74	82.95	91.88

42035

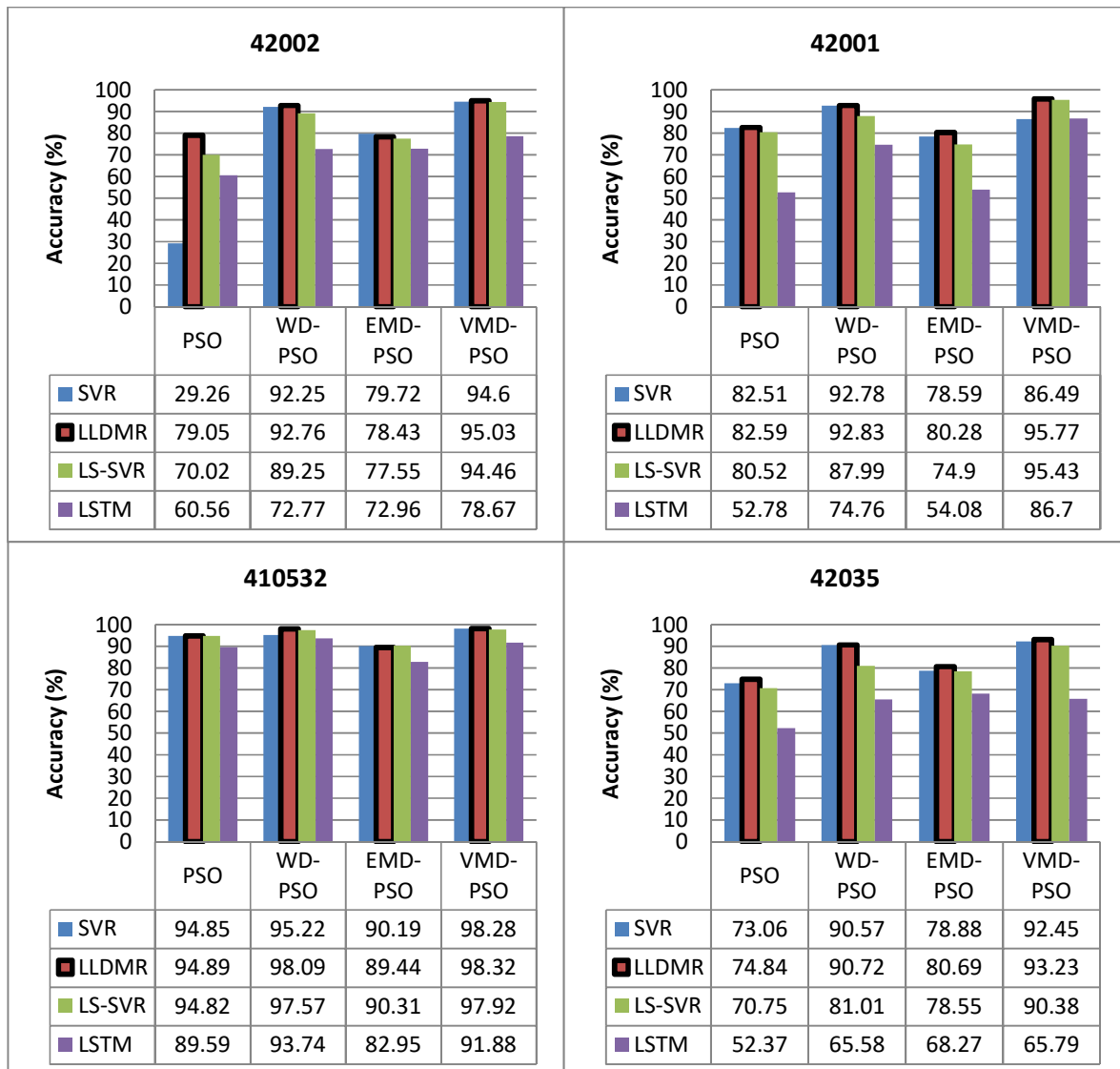|	PSO	WD-PSO	EMD-PSO	VMD-PSO
SVR	73.06	90.57	78.88	92.45
LLDMR	74.84	90.72	80.69	93.23
LS-SVR	70.75	81.01	78.55	90.38
LSTM	52.37	65.58	68.27	65.79

Figure 15: Comparison of the accuracies of the different decomposition methods

These observations make us infer that the LDMR model was the most promising regression model for SWH prediction in our experiments. Further, we compare the heights of histograms which appear in Figure 14. We can observe that all regression models when used with any

decomposition method perform better than the case where no decomposition is used except in case of dataset B and D (42001h2015 and 42035h2015) in which PSO-LS-SVR performs better than EMD-PSO-LS-SVR. Also, the VMD decomposition based hybrid models obtains the best performance among other decomposition based hybrid models in case of all regression models.

We have shown the prediction and scatter plot obtained by different VMD and WD based hybrid models on test set of dataset A and D at Figure 16-18. We can observe that in case of both decomposition methods the PSO-LDMR based hybrid models obtain the best performance.
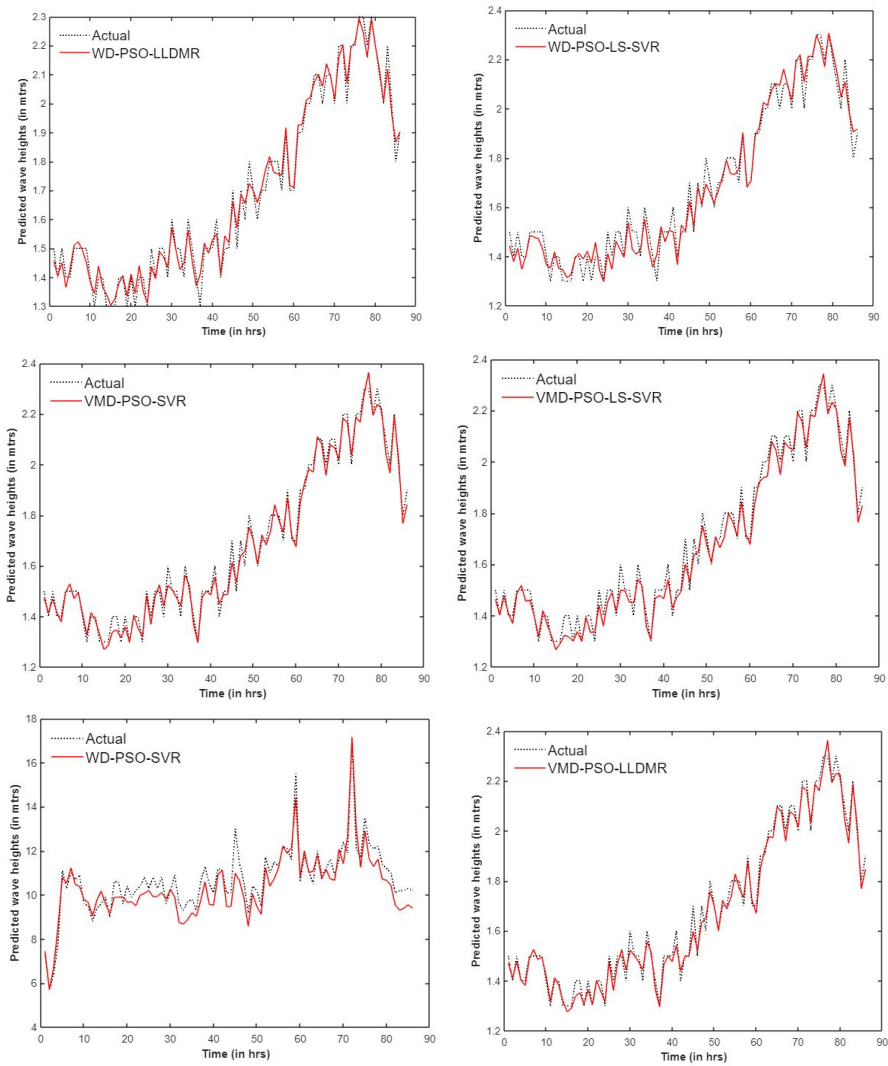
Figure 16: Prediction plots of dataset A using WD and VMD based hybrid models
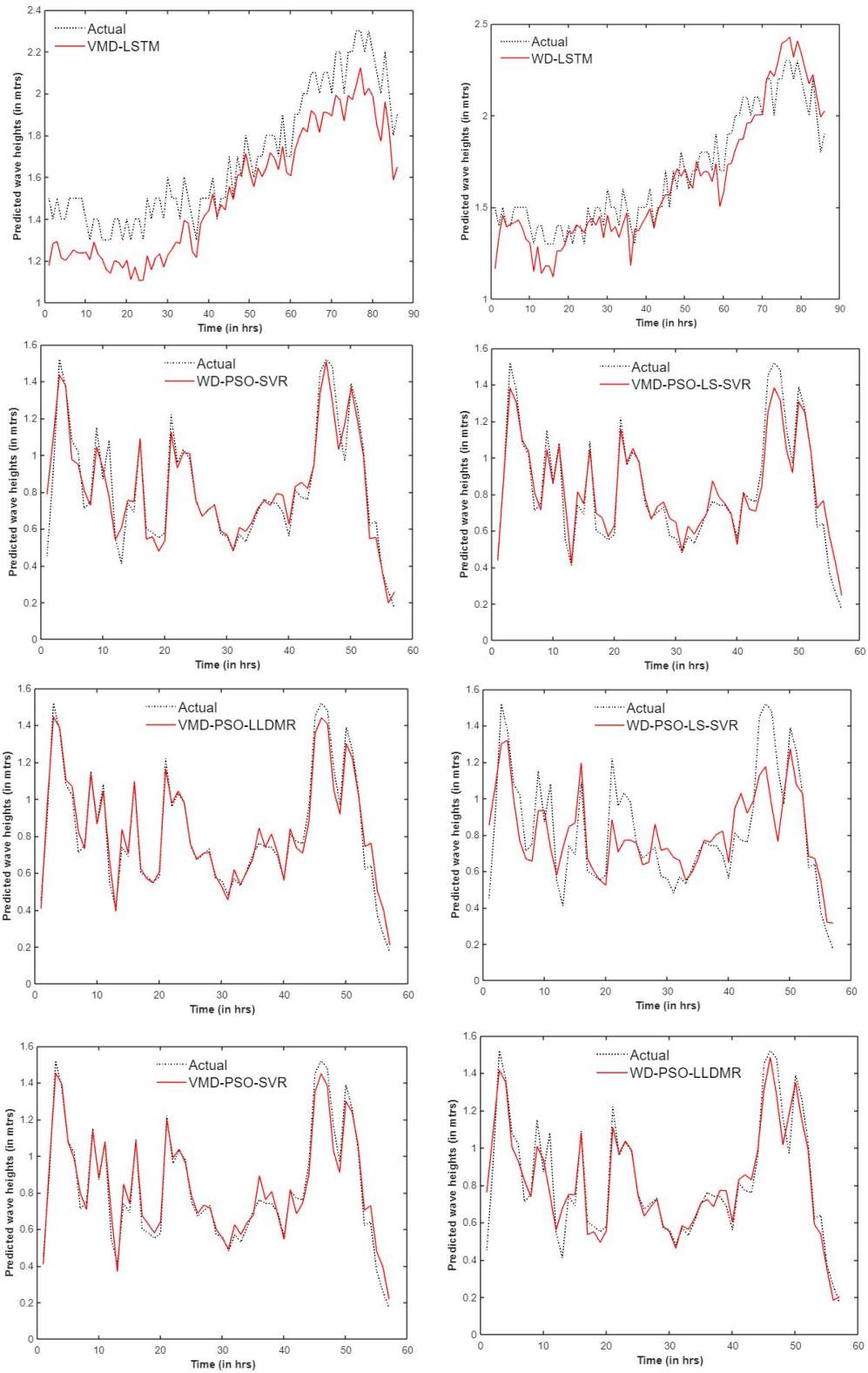
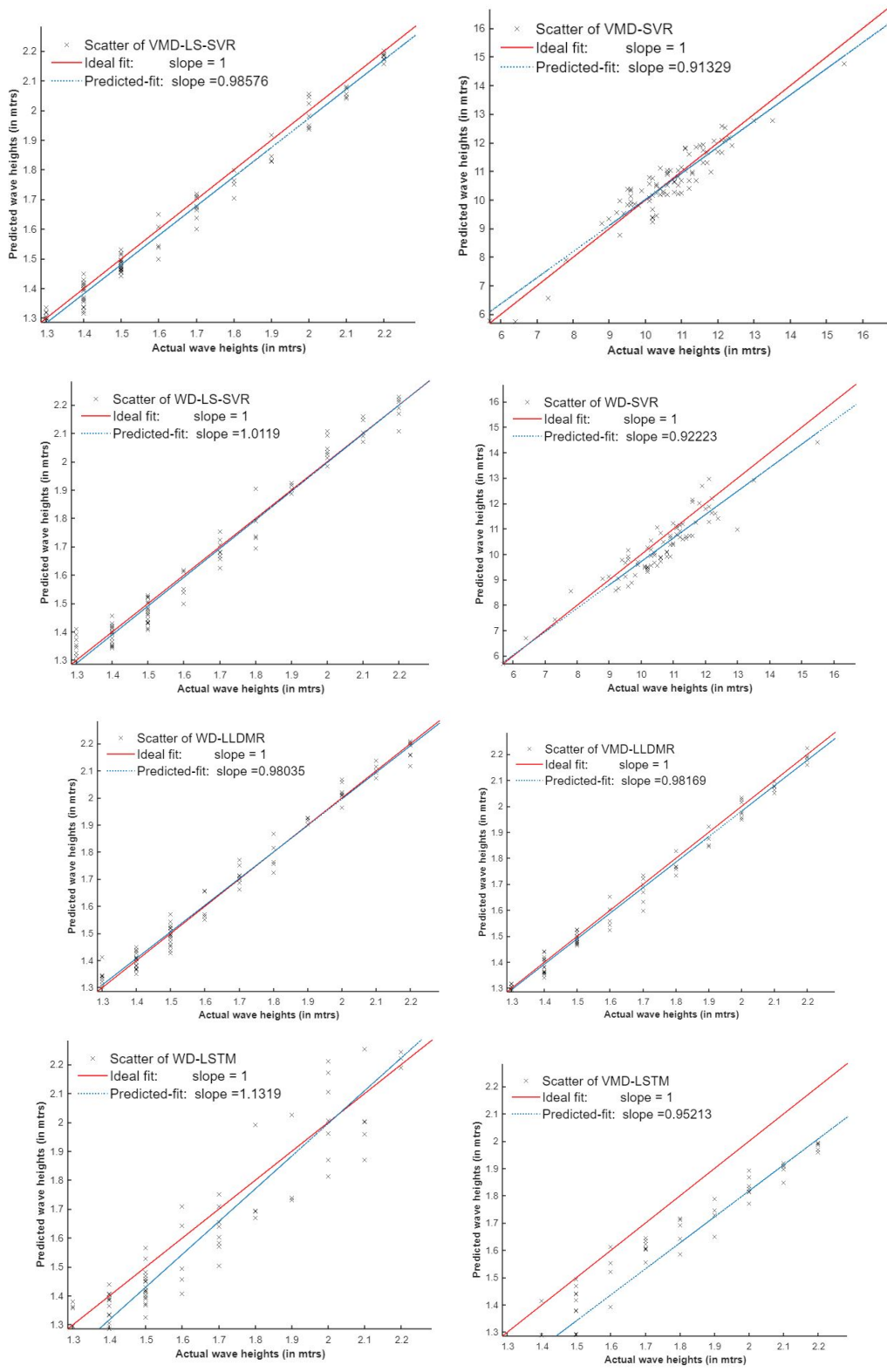Figure 17: Prediction plots of dataset D using WD and VMD based hybrid models

Figure 18: Scatter plots of dataset A using WD and VMD based hybrid models

Now, we shall attempt to quantitatively analyze our numerical results presented in the Table 2 and Table 3 to obtain the clearer picture. At first, we have computed the average of the accuracy obtained by our different regression models when used with different decomposition methods in Table 4. In Figure 15, we have plotted the average accuracy obtained by different regression models in case of different used decomposition methods in two different ways. In Table 4, we have also computed the total average accuracy obtained by different regression models and decomposition methods.

	No Decomp.	WD	EMD	VMD	AVG
PSO-LDMR	82.82 ± 8.63	93.6 ± 3.15	82.21 ± 4.92	95.59 ± 2.11	88.56
PSO-SVR	69.94 ± 28.55	92.7 ± 1.92	81.85 ± 5.58	92.96 ± 4.94	84.36
PSO-LS-SVR	79.03 ± 11.57	88.95±6.79	80.33 ± 6.83	94.55 ± 3.14	85.71
PSO-LSTM	63.83 ± 17.59	76.71 ± 12.01	71.34 ± 14.82	80.76 ± 11.36	73.16
TOTAL AVG	73.90	88.00	78.93	90.96	

Table 4: Average accuracy obtained by different wave hybrid models

We can observe that the LDMR model based hybrid models manages to obtain the best rank in average accuracy followed by LS-SVR model based hybrid models. The SVR and LSTM based wave hybrid models obtain the third and fourth rank respectively. On average, the LDMR model based hybrid models are 4.26% more accurate than SVR based hybrid models and 2.85% more accurate than LS-SVR based hybrid model. Also, we can find that the VMD decomposition based wave hybrid models obtains the best average prediction followed by WD and EMD based hybrid models. On average, the VMD based hybrid models obtain the 2.96% and 12.03% more accuracy than WD and EMD based hybrid models. Also, every decomposition based hybrid

models obtains a far better accuracy than hybrid models which do not make use of any decomposition technique.

	No Decomp.	WD	EMD	VMD	Average
PSO-LDMR	0.2306+0.0878	0.0936+0.0381	0.2548+0.0704	0.0567+0.0153	0.1589
PSO-SVR	0.4315+0.4192	0.2314+0.25	0.2686+0.0971	0.0967+0.0859	0.2570
PSO-LS-SVR	0.2475+0.1015	0.1392+0.0649	0.2916+0.1012	0.0644+0.0173	0.1857
PSO-LSTM	0.5424+0.2231	0.2263+0.0914	0.3113+0.0593	0.2238+0.0658	0.3259
Average	0.3630	0.1726	0.2816	0.1104	

Table 5: Average RMSE obtained by different wave hybrid models

We have also computed the average RMSE obtained by considered regression models when used with different decomposition methods in Table 5. We can conclude that the LDMR based hybrid models obtain lower RMSE than SVR, LS-SVR and LSTM based hybrid models. Also, the VMD decomposition method based hybrid models obtain lower RMSE than EMD, WD and no decomposition based hybrid models.

5.4 ANOVA Analysis

ANOVA, is the abbreviation of Analysis Of Variance, is a statistical method used to determine if the means of two or more groups differ from one another substantially. ANOVA compares the means of various samples to examine the influence of one or more factors. The formula for ANOVA analysis is as follows:

$$F = \frac{MST}{MSE}$$

where, *F* is ANOVA coefficient, *MST* is mean of squares due to treatment and *MSE* is mean of squares due to error.

Using the ANOVA test, you may compare more than two groups at once to see whether there is a correlation between them. The F statistic, or F-ratio, which is the outcome of the ANOVA formula enables the examination of several sets of data to ascertain the variability within and across samples.

The outcome of the ANOVA's F-ratio statistic will be near to 1 if there is no discernible difference between the tested groups, which is known as the null hypothesis. The F-distribution is the distribution of all potential F statistic values. The numerator degrees of freedom and the denominator degrees of freedom are two characteristic numbers that describe this collection of distribution functions.

Now, we shall be using the two way ANOVA analysis to obtain the answer of the following questions.

- Do performances obtained by different regression models in Table 2 and Table 3 differ significantly?
- Do performances obtained by different decomposition methods in Table 2 and Table 3 differ significantly?
- Is there any significant interaction between the used decomposition methods and regression models in Table 2 and Table 3?

For this, we shall be testing the following hypothesis:

a) H_1 : All the regression models have equal mean accuracy

b) H_2: All the decomposition methods have equal mean accuracy.

c) H_3: These factors are independent i,e. the interaction effects do not exist.

Now, we perform the two way ANOVA analysis in Microsoft Excel with 5% of level of significance using the numerical results of Table 6 and listed the final output table in Table V. It lists the Sum of Squares of variation (SS), Degree of Freedom (DF), Mean Square of variation (MS), F statistics, p statistics and F critical values also. We can observe at last section of Table V that the obtained F-statistics are greater than F-critical value for the first two rows. Also, the obtained p-values are lesser than our chosen level of significance 0.05. So, we must reject the null hypothesis H_1 and H_2. It means that the obtained accuracy by different used regression models and decomposition methods are significantly different. For the row 'interaction' listed at the last section of Table 6, the obtained F statistics is lower than the critical value. Also, the obtained p-value is greater than our chosen level of significance. So, we cannot reject the null hypothesis H_3.

SUMMARY	NO Decomp.	WD	EMD	VMD	Total
PSO-LDMR					
Count	4	4	4	4	16
Sum	331.2937	374.3983	328.8378	382.3633	1416.893
Average	82.82343	93.59958	82.20945	95.59083	88.55582
Variance	74.54703	9.93402	24.19148	4.455487	62.11028
PSO-SVR					
Count	4	4	4	4	16
Sum	279.7628	370.8126	327.3838	371.83	1349.789
Average	69.9407	92.70315	81.84595	92.9575	84.36183
Variance	814.9628	3.697973	31.17603	24.38557	270.2476
PSO-LS-SVR					
Count	4	4	4	4	16
Sum	316.1181	355.8187	321.3146	378.1818	1371.433
Average	79.02953	88.95468	80.32865	94.54545	85.71458
Variance	133.7891	46.10736	46.68546	9.829498	90.53046
PSO-LSTM					
Count	4	4	4	4	16
Sum	255.3034	306.8469	285.3405	323.0309	1170.522
Average	63.82585	76.71173	71.33513	80.75773	73.15761
Variance	309.329	144.3531	219.7669	129.099	203.3887
Total					
Count	16	16	16	16	
Sum	1182.478	1407.877	1262.877	1455.406	
Average	73.90488	87.99228	78.92979	90.96288	
Variance	326.02	89.30147	85.40289	71.52094	

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	2191.236	3	730.4121	5.767427	0.001883	2.798061
Columns	3001.713	3	1000.571	7.900637	0.00022	2.798061
Interaction	313.5134	9	34.83482	0.27506	0.978482	2.08173
Within	6078.929	48	126.6444			

Table 6: Two way ANOVA analysis for obtained numerical results

From the presented two way ANOVA analysis, we can statistically infer that the use of different regression model in wave hybrid models significantly affect their performance. Also, the use of different decomposition method in wave hybrid model significantly affects their performance.

But, the effect of the used decomposition method and regression model on the performance of the resulting wave hybrid model is independent i,e. these two factors do not interact among themselves. Therefore, we cannot claim that a certain combination of regression model and decomposition method can obtain any special effect on the performance of wave hybrid model.

Chapter 6

Conclusions and Future Work

In this work, we have considered the different combinations of decomposition methods, regression models and metaheuristic algorithm to develop the different variants of wave hybrid models for hourly significant wave height prediction. We have compared the performance of developed sixteen different wave hybrid models namely PSO-LDMR, PSO-SVR, PSO- LS-SVR, LSTM, WD-PSO-LDMR, WD-PSO-SVR, WD-PSO-LS-SVR, WD-LSTM, EMD-PSO-LDMR, EMD-PSO-SVR, EMD-PSO-LS-SVR, EMD-LSTM, VMD-PSO-LDMR, VMD-PSO-SVR, VMD-PSO-LS-SVR and VMD-LSTM models using different evaluation criteria on real world ocean wave height datasets collected from four different ocean buoys. The VMD [6] method and LDMR [4] model have not been exploited in any wave hybrid models before this. After the brief analysis of the numerical results, we conclude that the VMD [6] method is a superior decomposition method over Wavelet Decomposition and EMD [5] method for hourly significant wave height forecasting. Also, the LDMR [4] based hybrid models have outperformed the ε-SVR, LS-SVR and LSTM based wave hybrid models. Further, we have also carried out the two-way ANOVA analysis for obtained numerical results. It statistically infers that the used decomposition method and regression model affect the performance of the wave hybrid models significantly but, their effects are independent.

In this work, the wave hybrid models can forecast the ocean wave height of immediate next hour. We have planned to extend our developed wave hybrid models for long hour forecasts. Also the rate of convergence of the used metaheuristic method is quite low due to which the computational time complexity is considerably expensive. We would also like to develop or

explore the metaheuristic algorithms which has higher rate of convergence when used in our wave hybrid models in future. Further, we also feel the need of developing an appropriate machine learning model for wave height forecasting which can properly incorporate the prior information available.

References

[1] Drucker, H., Burges, C.J., Kaufman, L., Smola, A. and Vapnik, V., 1996. Support vector regression machines. *Advances in neural information processing systems*, *9*.

[2] Suykens, J.A. and Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), pp.293-300.

[3] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.

[4] Rastogi, R., Anand, P. and Chandra, S., 2020. Large-margin distribution machine-based regression. *Neural Computing and Applications*, *32*(8), pp.3633-3648.

[5] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry 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, 1998.

[6] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE transactions on signal processing*, *62(3)*:531–544, 2013.

[7] James Kennedy and Russell Eberhart. Particle swarm optimization. *In Proceedings of ICNN'95-international conference on neural networks*, *volume 4*, pages 1942–1948. IEEE, 1995.

[8] Marcus Lehmann, Farid Karimpour, Clifford A Goudey, Paul T Jacobson, and Mohammad-Reza Alam. Ocean wave energy in the United States: Current status and future perspectives. *Renewable and Sustainable Energy Reviews*, 74:1300–1313, 2017.

[9] Shaobo Yang, Zegui Deng, Xingfei Li, Chongwei Zheng, Lintong Xi, Jucheng Zhuang, Zhenquan Zhang, and Zhiyou Zhang. A novel hybrid model based on stl decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast. *Renewable Energy*, 173:531–543, 2021.

[10] WY Duan, Y Han, LM Huang, BB Zhao, and MH Wang. A hybrid emd-svr model for the short-term prediction of significant wave height. *Ocean Engineering*, 124:54–73, 2016.

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

[12] C Guedes Soares and AM Ferreira. Representation of non-stationary time series of significant wave height with autoregressive models. *Probabilistic Engineering Mechanics*, *11(3)*:139–148, 1996.

[13] Ming Ge and Eric C Kerrigan. Short-term ocean wave forecasting using an autoregressive moving average model. In 2016 UKACC 11th *International Conference on Control (CONTROL)*, pages 1–6. IEEE, 2016.

[14] Erik Vanem and Sam-Erik Walker. Identifying trends in the ocean wave climate by time series analyses of significant wave heightdata. *Ocean engineering*, 61:148–160, 2013.

[15] Mc C Deo, Abhay Jha, AS Chaphekar, and K Ravikant. Neural networks for wave forecasting. *Ocean engineering*, *28(7)*:889–898, 2001.

[16] Makarynskyy. Improving wave predictions with artificial neural networks. *Ocean Engineering*, *31(5-6)*:709–724, 2004.

[17] L Cornejo-Bueno, JC 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*:380–389, 2016.

[18] N Krishna Kumar, R Savitha, and Abdullah Al Mamun. Ocean wave height prediction using ensemble of extreme learning machine. *Neurocomputing*, *277*:12–20, 2018.

[19] Can Elmar Balas, Levent Koc¸ and Lale Balas. Predictions of missing wave data by recurrent neuronets. *Journal of waterway, port, coastal, and ocean engineering*, *130(5)*:256–265, 2004.

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

[21] Xinwei Chen and Weimin 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*, 2021.

[22] Shuntao Fan, Nianhao Xiao, and Sheng Dong. A novel model to predict significant wave height based on long short-term memory network. *Ocean Engineering*, *205*:107298, 2020.

[23] Felix Enigo VS et. al. Forecasting significant wave height using rnn-lstm models. In 2020 4th International Conference on *Intelligent Computing and Control Systems (ICICCS)*, pages 1141–1146. IEEE, 2020.

[24] Shuang Li, Peng Hao, Chengcheng Yu, and Gengkun Wu. Clts-net: A more accurate and universal method for the long-term prediction of significant wave height. *Journal of Marine Science and Engineering*, *9(12)*:1464, 2021.

[25] Nawin Raj and Jason Brown. An eemd-bilstm algorithm integrated with boruta random forest optimiser for significant wave height forecasting along coastal areas of Queensland, Australia. *Remote Sensing*, *13(8)*:1456, 2021.

[26] Vladimir Vapnik. The nature of statistical learning theory. *Springer science & business media*, 1999.

[27] Jadran Berbi'c, Eva Ocvirk, Dalibor Carevi'c, and Goran Lonˇcar. Application of neural networks and support vector machine for significant wave height prediction. *Oceanologia*, *59(3)*:331–349, 2017.

[28] Dong Gao, Yongxin Liu, Junmin Meng, Yongjun Jia, and Chenqing Fan. Estimating significant wave height from sar imagery based on an svm regression model. *Acta Oceanologica Sinica*, *37(3)*:103–110, 2018.

[29] Kaixin Zhao and Jichao Wang. Significant wave height forecasting based on the hybrid emd-svm method. 2019.

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

[31] R Prahlada and Paresh Chandra Deka. Forecasting of time series significant wave height using wavelet decomposed neural network. *Aquatic Procedia*, *4*:540–547, 2015.

[32] Mosbeh R Kaloop, Deepak Kumar, Fawzi Zarzoura, Bishwajit Roy, and Jong Wan Hu. A wavelet-particle swarm optimization-extreme learning machine hybrid modeling for significant wave height prediction. *Ocean Engineering*, *213*:107777, 2020.

[33] Pradnya Dixit, Shreenivas Londhe, and Yogesh Dandawate. Removing prediction lag in wave height forecasting using neuro-wavelet modeling technique. *Ocean Engineering*, *93*:74–83, 2015.

[34] Mumtaz Ali and Ramendra Prasad. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*, *104*:281–295, 2019.

[35] Shuyi Zhou, Brandon J Bethel, Wenjin Sun, Yang Zhao, Wenhong Xie, and Changming Dong. Improving significant wave height forecasts using a joint empirical mode decomposition–long short-term memory network. *Journal of Marine Science and Engineering*, *9(7)*:744, 2021.

[36] James Mercer. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London. Series A*, containing papers of a mathematical or physical character, 209(441-458):415 446, 1909.

[37] Ozger, M., 2010. Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Engineering, 37(16)*, pp.1443-1451.

[38] Wolfe, P., 1961. A duality theorem for non-linear programming. *Quarterly of applied mathematics*, *19*(3), pp.239-244.