KSCE Journal of Civil Engineering (2017) 21(1):1-10

DOI 10.1007/s12205-016-0770-4

Coastal and Harbor Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

# Significant Wave Height Modelling using a Hybrid Wavelet-Genetic Programming Approach

**Sajad Shahabi*, Mohammad-Javad Khanjani**, and Masoud-Reza Hessami Kermani*****

Received September 17, 2015/Revised February 21, 2016/Accepted March 4, 2016/Published Online April 29, 2016

## Abstract

In this paper, Genetic Programming (GP) based wavelet transform (WGP) was developed to forecast Significant Wave Height (SWH) in different lead times. The hourly SWH values for two buoy stations located in the North Atlantic Ocean were applied to train and validate the WGP model. For this purpose, the SWH main time series was decomposed into some subseries using wavelet transform and then decomposed time series were imported to GP model to forecast the SWH. Furthermore, GP approach was independently used to the same data set for comparison purposes. Performance of the WGP model was evaluated using correlation coefficient (R), Root Mean Square Error (RMSE), index of agreement ($I_a$) and Mean Absolute Error (MAE). The analysis proved that the model accuracy is highly depended on the decomposition levels. The obtained results showed that WGP model is able to forecast the SWH with a high reliability.

Keywords: *genetic programming, lead time, significant wave height, time series, wavelet*

## 1. Introduction

Knowledge of Significant Wave Height (SWH) is vital for most of maritime activities consisting of planning, operation and maintenance. In this study, SWH is shown by $H_s$. It is traditionally performed by converting wind-related information to waves (Nitsure *et al.*, 2012). In 1960s and 1970s empirical models were tried for modelling of wave height (US Army, 1984). After that, modelling of SWH based on historical data was considered. Forecasting of SWH using a time history is fed with an input including a sequence of previous observations, so that it distinguishes a hidden pattern in such a sequence and accordingly forecasts the future values in continuation (Gaur and Deo, 2008). Therefore, many SWH time series based models like Auto Regressive (AR), Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA) and Kalman filter have been developed (Soares and Ferreira, 1996; Soares *et al*., 1996; Soares and Cunha, 2000; Scotto and Soares, 2000; Ozger, 2010; Altunkaynak, 2013).

Over the last decades, the different sorts of soft computing approaches have been widely used in time series forecasting. One of the most popular of these approaches is artificial neural network (ANN). Presumably, Deo and Naidu (1998) have initiated investigations on the application of ANN into wave forecasting. They employed ANN to forecast wave height in east coast of India. Deo *et al.* (2001) applied a simple 3-layer feed forward type of network to obtain the output of SWH and the mean wave periods from the input of generating wind speeds. The model provided satisfactory results in open wider areas in deep water. Later Asma *et al.* (2012) compared the ANN results with those derived from Multiple Linear Regressions (MLR) and found that non-linear models at the same time step produced to better significant wave height models. In addition, a good many investigations on the SWH forecasting can be found in literature Markarynskyy *et al*. (2005), Jain and Deo (2007), Günaydýn (2008), Kamranzad *et al.* (2011) and Nitsure *et al.* (2014).

In the recent years, some other artificial intelligence techniques such as Genetic Algorithm (GA), Genetic Programming (GP) and Fuzzy Logic (FL) made of use to forecast SWH parameter (Ozger and Sen, 2007; Canellas *et al*., 2010; Nitsure *et al*., 2012; Altunkaynak, 2013). Gaur and Deo (2008) explored dependency between input and output data sets using GP in the Gulf of Mexico. The forecasted SWH over lead times of 3, 6, 12 and 24h. The performance of GP indicated more precise prediction for small-interval forecasts (3 and 6h) than those obtained for large-interval (12 and 24h). Conspicuously, they concluded that GP can be taken into account as a promising tool for its future applications into coastal and ocean problems

The above-mentioned approaches suffer from restricted capability in forecasting the non-stationary time series. Ergo, some improved Artificial Intelligence (AI) models were employed to forecast SWH and gave accurate outperformances (Ozger 2010,

*Ph.D, Dept. of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran (Corresponding Author, E-mail: sajad.shahaaby@gmail.com)
**Professor, Dept. of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran (E-mail: khangani @yahoo.com)
***Assistant Professor, Dept. of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran (E-mail: masoud.hessami@gmail.com)

Deka and Prahlda, 2012, Shahabi and Khanjani, 2015). Furthermore, Nourani *et al.* (2014) presented an extensive review on the application of hybrid wavelet-AI approaches. Recently, Prahlada and Deka (2015) developed a hybrid model of wavelet and artificial neural network (WLNN) to forecast time series significant wave height for lead times up to 48h in two stations located in Indian Ocean and North Pacific Ocean. The three hourly significant wave height data for a period of one year was decomposed through discrete wavelet and used as inputs into Levenberg Marquardt artificial neural network models to forecast time series SWH at multistep lead time. To recap, results of proposed model indicated good predictions for two stations at lower lead times but slight deviation observed at higher lead times.

There is an extensive literature available on the application of wavelet technique to forecast time series, few are them are Cheng *et al.* (2006), Dixit *et al.* (2015), Seo *et al.* (2015).

Some application of models such as fuzzy interface system, Bayesian inference, modified Weibull distribution, regressive support vector machine and hybrid Wavelet Group Method Of Data Handling (WGMDH) were described by Kazeminezhad *et al.* (2005), Scotto and Soares(2007), Muraleedharan (2007), Mahjoobi and Mosabeb (2009), Shahabi and Khanjani (2015) and Shahbi *et al.* (2016) for forecasting of SWH.

In the present research, the possibility of forecast SWH in two different locations in North Atlantic Ocean using a hybrid Wavelet Genetic Programming (WGP) approach are investigated. For this purpose, WGP algorithm has been introduced and employed to develop an SWH forecasting model when has an ability to make forecast up to 48 h lead time using hourly wave height observed data. Therefore, current research is initiated with a data preprocessing, i.e. de-noising of predictor time series using discrete wavelet transform technique, and followed by a GP-based mode. The decomposed time series can be presented as input parameter to Genetic Programming (GP) which can handle non-linearity capability and higher forecasting accuracy can be achieved. Forecasts are more accurate compared to those yield by original signals due to the fact that the features of the subseries are vivid. This is why the hybrid method of wavelet transform and GP have better performances than single GP or similar models. Finally, the proposed wavelet GP (WGP) model are evaluated to assess the model efficiency in the higher lead times along with different decompositions level using three different efficiency indices. The results were compared with the GP model ones for the same SWH data set.

## 2. Wavelet Transform

Here, a summary of wavelet main concepts are presented. For more information some excellent supplementary text such as Mallat (1998) and Bogges and Narcowich (2005) are recommended.

A wavelet transform presents a powerful tool for non-stationary data analysis. It is especially beneficial in selecting characteristic variations at different resolution or scales and it is similar to Fourier transform. In other word, it is the derivative from of Fourier transform. The wavelet transform decomposes a signal into its subseries in time and frequency domains. It has been used for studying non-stationary time series unlike Fourier transform. This point is the most important benefit of wavelet transform. In traditional transformation methods such as Fourier transform, production of both time and frequency information with a higher resolution is not possible, but wavelet transform resolved this shortage. There are two types of Wavelet Transforms (WT): Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). (Mallat, 1998; Missiti *et al*., 2000; Boggess and Narcowich, 2009; Ozger, 2010; Nourani *et al.*, 2012; Danandeh Mehr *et al.*, 2014)

The continuous wavelet transform of a time series, $f(t)$, is defined as:

$$T(s,b) = \frac{1}{\sqrt{s}}\int_{-\infty}^{+\infty} \varphi^*\left(\frac{t-b}{s}\right) f(t)dt \tag{1}$$

where $\varphi(t)$, $s$, $b$ and $t$ are the mother wavelet function, dilation factor (or contraction coefficient), scale parameter and time, respectively. As well as * which corresponds to the complex conjugate.

The mother wavelet function is described by:

$$\varphi(b,s) = \frac{1}{s}\varphi\left(\frac{t-b}{s}\right) \tag{2}$$

The mother wavelet function is used for both wavelet decomposition and composition transforms.

In empirical applications, discretization of Eq. (1) was used. This transform produces $N^2$ coefficients from a data set of length N. DWT computes the wavelet coefficients at discrete intervals of time and scale. The signal is passed through a series of high pass filters and low pass filters to analyze the high frequencies and low frequencies, respectively. The low-frequency value is the most important part for many signals. Frequently is addressed of approximations and details. The approximation is the high-scale and low-frequency component of the signal. The detail is the low-scale and high-frequency component. The detail coefficients (cD) are small and consist mainly of a high frequency noise, while, the approximation coefficients (cA) consist much less noise than the original signal (mallat, 1998; Missiti *et al*., 2000; Deka and Prahlada, 2012).

The pick out of mother wavelet is an important part of wavelet analysis and depends on both the properties of the signal under investigation and what the researchers are looking for. In this study, Haar and Daubechies wavelet functions were used to analyze their similarity to nature of data.

Decomposition of signal process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This process is called the wavelet decomposition tree. Looking at a signal's wavelet decomposition tree can provide valuable information. Despite the analysis process can theoretically be continued indefinitely, in practice, it can be proceed only until the individual details consist of a single sample or pixel. The researcher can select an appropriate number of levels based on the nature of the signal, or on an appropriate criterion (Missiti *et al.*, 2000).

## 3. Genetic Programming

Here, a brief overview of the Genetic Programming (GP) is presented for motivation. GP is an evolutionary computing technique and proposed by Koza (1992). The GP nature authorizes to gain additional information on how the system performs, i.e., gives insight into the relationship between input and output time series of data set (Nourani *et al*., 2012). The tree based GP was used in this study.

In GP a random population of individuals is created, the fitness of individuals is evaluated and then parents are designated out of these individuals. The parents are then made to achieve offspring's by following the process of reproduction, crossover and mutation. The production of offspring's continues till a determined number of offspring's in a generation are produced and further till another determined number of generations are created. The resulting offspring's at the end of all this process (equation or computer program) is the solution of the problem. In other words, the GP transforms one population of individuals into another one in an iterative manner by employing some operators. In GP computations, it can distinguish between three different types of operators which are called mutation, reproduction and crossover (Gaur and Deo, 2008; Nourani *et al*., 2012).

## 4. Wavelet-Genetic Programming

The discrete wavelet transform combine to genetic programming to obtain a powerful nonlinear ability that call Wavelet-genetic Programming (WGP) model in this paper. In other word, the WGP is a hybrid model that combines the DWT with GP to improve the performance and ability of the GP. In the WGP model, the original series of SWH decompose to some subseries by DWT and then these time series are imposed as inputs to the GP model to forecast SWH in different lead times. The schematic diagram of the proposed WGP model is illustrated in Fig. 1.

The proposed model comprises two main stages. In the first stage (pre-processing stage), the main SWH time series are decomposed into some sub-series using DWT. For this purpose, Haar and db3 were selected as a mother wavelet. In the second stage (simulation stage), the decomposed time series analyze with the GP to obtain a nonlinear approximation formulation. Thus, the SWH signal decomposes into some sub signals with different level of decomposition. The decomposed signal at level n consisted of $n+1$ sub signals are included those of an approximation and n details. After that, the decomposed signals were considered as input parameters for GP model. Finally, the WGP model obtained to GP formulations to predict the SWH.

In the proposed approach, approximation and details play key role in the performance of model. Here, $(n+1)$ variables (an approximation and n details) were taken into account as input parameters for GP model so as to yield the SWH at the time $(t+m)$, where $m$ is lead time.

## 5. Study Area and Data

In this study, a buoy station (41013, Latitude 33°26'11" and longitude 77°44'35") located in coastal areas and another one (41048, Latitude 31°57'00" and longitude 69°269'48") located in deep waters in North Atlantic Ocean were employed. The water depths of these two stations are 23.5 m and 5261 m, respectively. The data used in this study are $H_s$ time series and can be downloading from website of National Data Buoy Center (NDBC) (http://www.ndbc.noaa.gov). These two stations were selected based on continuity in the reported values in recent years. Fig. 2 shows the location map for two current stations.

Table 1 presents the statistical properties of wave height time series. 75% of the wave height data was devoted to perform training and remaining 25%, for testing. Also, the statistical characteristics of training and testing dataset were separately given in Table 2.

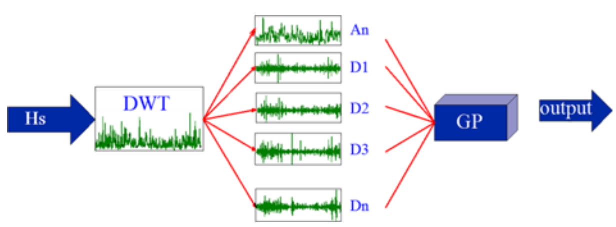


Fig. 1. The Schematic Diagram of WGP Model

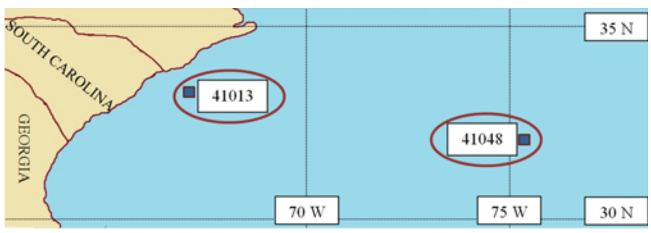


Fig. 2. Location Map for the Study Area

Table 1. Statistical Properties of All Data

| Station ID | Water depth (m) | Period | Significant wave height | | | |
|---|---|---|---|---|---|---|
| | | | Min. (m) | Max. (m) | Mean (m) | Standard deviation (m) |
| 41013 | 23.5 | 2013-1-1 to 2013-12-31 | 0.33 | 5.12 | 1.35 | 0.637 |
| 41048 | 5261 | 2013-1-1 to 2013-12-31 | 0.50 | 9.00 | 1.81 | 1.006 |

Table 2. Statistical Properties of Training and Testing Data

| Station ID | Data | Period | Min. (m) | Mean (m) | Max. (m) | Standard deviation (m) |
|---|---|---|---|---|---|---|
| 41013 | Training dataset | 2013-1-1 to 2013-9-30 | 0.33 | 1.32 | 5.12 | 0.61 |
| | Testing dataset | 2013-10-1 to 2013-12-31 | 0.37 | 1.43 | 4.63 | 0.697 |
| 41048 | Training dataset | 2013-1-1 to 2013-9-30 | 0.51 | 1.81 | 9.00 | 1.079 |
| | Testing dataset | 2013-10-1 to 2013-12-31 | 0.50 | 1.81 | 5.83 | 0.742 |

## 6. Analysis and Results

Data sets from two different stations located in the North Atlantic Ocean were used to develop the proposed models. In this section, the results of proposed WGP model were compared with GP model results. These models were employed to make forecast for various lead times. The lead times were fixed as 3, 6, 12, 24 and 48h. Different combinations of data sets were imposed as input variable to forecast SWH. In addition, three various scenarios which have various predictor configurations were employed. These scenarios are: (a) $h_s(t)$, (b) $h_s(t-1)$, $h_s(t)$ and (c) $h_s(t-2)$, $h_s(t-1)$, $h_s(t)$. While $h_s(t)$, $h_s(t-1)$ and $h_s(t-2)$ are current SWH, one time step and two time step past wave height, respectively. The results of the scenarios demonstrated that enhancement in the number of lagged values have the ability to improve the model performance any more. In this way, three lagged values were sufficiently determined to make trustworthy predictions. The predicted value is shown by $h_s(t+m)$ where m is the lead time.

In this study, four different performance indices were employed to evaluate the model performances. These indices are: correlation coefficient (CC or R), index of agreement ($I_a$), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) according to Eq. (3) to (6).

$$CC = R = \sqrt{1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \bar{x})^2}} \tag{3}$$

$$I_a = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (|y_i - \bar{y}_i| + |x_i - \bar{x}|)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{n}} \tag{5}$$

$$MAE = \frac{1}{n}\sum_i |x_i - y_i| \tag{6}$$

where $x_i$, $y_i$, $\bar{x}$, $\bar{y}$ and $n$ are observed wave height, predicted wave height, mean of observed wave height, mean of predicted wave height and number of observations, respectively.

R and $I_a$ range from 0 to 1, closer values to 1 demonstrate outperform agreement between the observed and predicted SWH. The GP model for each scenario without SWH time series preprocessing was employed to forecast SWH time series. This approach was carried out for modelling of SWH time series at stations. Table 3 presents the best GP model testing results of analysis in terms of R, RMSE and MAE for predicted SWH.

Table 3. Evaluation Results of the GP Model Performances for Station 41048

| Station ID | Lead time | R | | RMSE (m) | | MAE (m) | |
|---|---|---|---|---|---|---|---|
| | | Train | Test | Train | Test | Train | Test |
| 41013 | 3h | **0.941** | **0.941** | **0.208** | **0.237** | **0.141** | **0.154** |
| | 6h | 0.846 | 0.825 | 0.534 | 0.395 | 0.225 | 0.263 |
| | 12h | 0.657 | 0.658 | 0.463 | 0.528 | 0.329 | 0.368 |
| | 24h | 0.351 | 0.195 | 0.576 | 0.696 | 0.426 | 0.526 |
| | 48h | 0.089 | 0.007 | 0.614 | 0.706 | 0.470 | 0.539 |
| 41048 | 3h | **0.973** | **0.951** | **0.247** | **0.230** | **0.220** | **0.230** |
| | 6h | 0.941 | 0.885 | 0.366 | 0.348 | 0.220 | 0.230 |
| | 12h | 0.861 | 0.730 | 0.551 | 0.523 | 0.339 | 0.352 |
| | 24h | 0.713 | 0.481 | 0.756 | 0.673 | 0.489 | 0.490 |
| | 48h | 0.535 | 0.247 | 0.913 | 0.761 | 0.613 | 0.566 |

According to Table 3, R values vary with respect to lead times. The correlation coefficient values for 3 and 48h lead times were declined from 0.941 to 0.007 in station 41013 for scenario 2. It seems to be satisfactory for 3 and 6h lead times but for higher lead times, predictions with an acceptable accuracy was not met. The GP capability in SWH modelling decreases drastically as lead times progresses. The root mean square error was increased from 0.237 for 3h to 0.706 for 48h lead times in station 41013. On the other hand, while the R values decreases from 0.951 to 0.247, the RMSE values range between 0.230 and 0.761 m at 48h lead time (Table 3).

Data pre-processing is the first stage of the proposed WGP model. In this stage, the main time series were decomposed into an approximation and some details. Next, these decomposed subseries were determined as input parameters for GP model to enhance the level accuracy of model. Here, discrete wavelet transform (DWT) was employed for processing of SWH time series in the form of approximation and details at different decomposition levels. These subseries were used as input variable for GP model which is applied in the proposed hybrid model to forecast SWH time series. The mother wavelet type selection is one of the most significant parameters in WGP modelling. On the basis of two time series formation and structure of the Haar (db1) and db3 wavelet, these mother wavelets were selected to decompose and analyze the SWH time series.

This approach was performed with different decomposition levels from 3 to 8. No significant improvement in the model

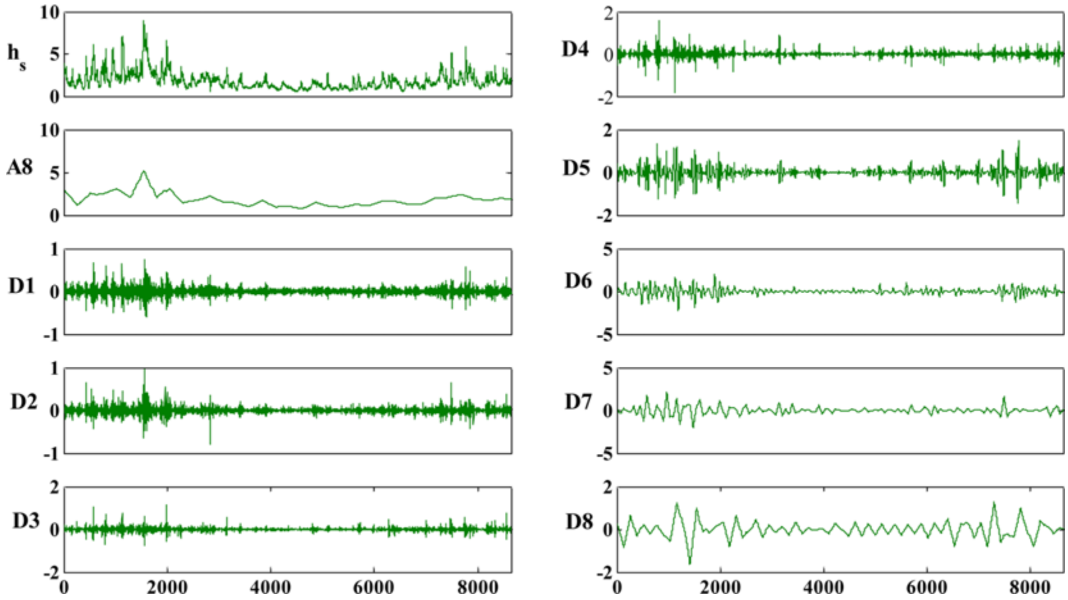


Fig. 3. Approximation and Detail Subseries of Significant Wave Height Time Series for Station 41048

Table 4. Parameter Setting for the GP Model

| Parameter | Value |
|---|---|
| Initial populations (programs) | 500 |
| Mutation frequency | 50% |
| Crossover frequency | 50% |
| Generation without improvement | 300 |
| Generation since start | 1500 |

performance was found when the number of decomposition levels was increased from a threshold, which is similar to the experience reported by Deka and Prahlda (2012).

The output subseries from the DWT in the form of approximation and details at station 41048 are shown in Fig. 3. The decomposed subseries were considered as input parameter to forecast the SWH in different lead times. Then, the trained model was tested by testing data. The data obtained from both stations in different

Table 5. Evaluation Results of the WGP Model Performances

| Station ID | Mother Wavelet | D.L* | Lead time | Training data R | Training data RMSE (m) | Training data $I_a$ | Training data MAE (m) | Testing data R | Testing data RMSE (m) | Testing data $I_a$ | Testing data MAE (m) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 41013 | Haar | 4 | 3 | 0.953 | 0.189 | 0.976 | 0.123 | 0.949 | 0.231 | 0.973 | 0.146 |
| | Haar | 5 | 6 | 0.883 | 0.290 | 0.937 | 0.189 | 0.867 | 0.350 | 0.927 | 0.223 |
| | Haar | 5 | 12 | 0.746 | 0.410 | 0.848 | 0.289 | 0.716 | 0.491 | 0.827 | 0.350 |
| | Haar | 7 | 24 | 0.586 | 0.498 | 0.717 | 0.363 | 0.559 | 0.584 | 0.714 | 0.435 |
| | Haar | 8 | 48 | 0.454 | 0.550 | 0.612 | 0.406 | 0.471 | 0.617 | 0.620 | 0.466 |
| | db3 | 6 | 3 | **0.957** | **0.179** | **0.977** | **0.122** | **0.952** | **0.213** | **0.975** | **0.142** |
| | db3 | 5 | 6 | 0.899 | 0.269 | 0.944 | 0.228 | 0.874 | 0.339 | 0.930 | 0.186 |
| | db3 | 7 | 12 | 0.813 | 0.357 | 0.891 | 0.254 | 0.787 | 0.432 | 0.874 | 0.312 |
| | db3 | 7 | 24 | 0.685 | 0.448 | 0.802 | 0.338 | 0.638 | 0.541 | 0.773 | 0.411 |
| | db3 | 8 | 48 | 0.587 | 0.498 | 0.691 | 0.372 | 0.462 | 0.620 | 0.620 | 0.473 |
| 41048 | Haar | 4 | 3 | **0.981** | **0.214** | **0.990** | **0.128** | **0.964** | **0.200** | **0.982** | **0.130** |
| | Haar | 5 | 6 | 0.953 | 0.328 | 0.976 | 0.191 | 0.903 | 0.320 | 0.949 | 0.204 |
| | Haar | 6 | 12 | 0.911 | 0.449 | 0.953 | 0.287 | 0.777 | 0.470 | 0.869 | 0.316 |
| | Haar | 7 | 24 | 0.847 | 0.575 | 0.915 | 0.373 | 0.574 | 0.616 | 0.726 | 0.420 |
| | Haar | 7 | 48 | 0.778 | 0.636 | 0.872 | 0.468 | 0.520 | 0.682 | 0.672 | 0.434 |
| | Db3 | 5 | 3 | 0.963 | 0.291 | 0.981 | 0.183 | 0.909 | 0.309 | 0.952 | 0.207 |
| | Db3 | 5 | 6 | 0.957 | 0.315 | 0.978 | 0.206 | 0.901 | 0.324 | 0.947 | 0.224 |
| | Db3 | 7 | 12 | 0.932 | 0.390 | 0.964 | 0.253 | 0.875 | 0.361 | 0.933 | 0.264 |
| | Db3 | 8 | 24 | 0.851 | 0.567 | 0.917 | 0.376 | 0.730 | 0.509 | 0.835 | 0.364 |
| | Db3 | 8 | 48 | 0.809 | 0.634 | 0.889 | 0.433 | 0.641 | 0.568 | 0.755 | 0.435 |

*Decomposition Level

lead times are summarized in Table 5. The best results in terms of CC, $I_a$, RMSE and MAE are highlighted for both stations. When these subseries are used as input, as the number of decomposition level increases accordingly number of input layer also increases. The results of proposed model for various decomposition levels evidently illustrated the performance of the WGP model for both low and high lead times compared to the GP (Table 5). For both stations, the WGP model demonstrated better prediction than GP model for all cases with an exception in 3h lead time that indicated satisfactorily results. The CC (R) value is about zero resulted in GP modelling was increased up to 0.471 by WGP at 48h lead time in station 41013. This considerable increment showed the capability of the WGP model. This noticeable improvement can be explained by deletion of noisy data from original time series.

Next, the WGP model formulations were explored for both stations. Nourani *et al.* (2012) reported that it is inevitable to acquire more input variables in evolutionary computing methods lead to more complex formulations. The best ranked WGP formulations were expressed for 3h lead time for the stations 4013 Eq. (7) and 41048 Eq. (8) as,

$$SWH^{db3}_{41013}(t+3) = d1(t)+2\times d3(t)+d4(t)+d5(t)+d7(t)+ d3(t)\times(d2(t)+d3(t))-((0.23999\times d4(t))+Sin(d6(t))) \tag{7}$$

$$SWH^{Haar}_{41048}(t+3) = d1(t)+d2(t)+d3(t)+d4(t)+d5(t)+ \left(\frac{d2(t)\times Sin(7.436249+d2(t))}{d1(t)^{\frac{1}{3}}}\right)^2 \tag{8}$$

Furthermore, Eqs. (9) and (10) at 48h lead time for the stations 41013 and 41048 were respectively presented as follows:

$$SWH^{Haar}_{41013}(t+48) = d1(t)+Sin((Cos(d1+d7(t))) (Sin(4.576904)\times d7(t)))+((d3(t)\times d3(t)\times d8(t))\times (8.465607\times d5(t)-8.465607\times d1(t))) \tag{9}$$

$$SWH^{db3}_{41048}(t+48) = d3(t)+Sin(Arctan(Sin(d4(t)+4.639923)^2- d9(t)\times Arctan(d9(t)+d1+d2(t)+d7(t)) \tag{10}$$

where $SWH^{db3}_{41013}(t+3)$, $SWH^{Haar}_{41013}(t+48)$, $SWH^{Haar}_{41048}(t+3)$ and $SWH^{db3}_{41048}(t+48)$ are the forecasted SWH at 3 and 48h lead times in stations 41013 and 41048, respectively. Also d1(t) to d9(t) are the approximation and the details time series that obtained by DWT at time t.

As shown in Eqs. (7) to (10), only four operations including addition, subtraction, multiplication and division and seven functions consisting of square root, cube root, sin, cosine, arctangent and power (2 and 3) were set in constructing the formulations. The GP provided these equations via an evolutionary process. The GP characteristics that used in this study are presented in Table 4.

The model efficiency criterion showed the high performance of model for the small lead times. For low lead time, performances

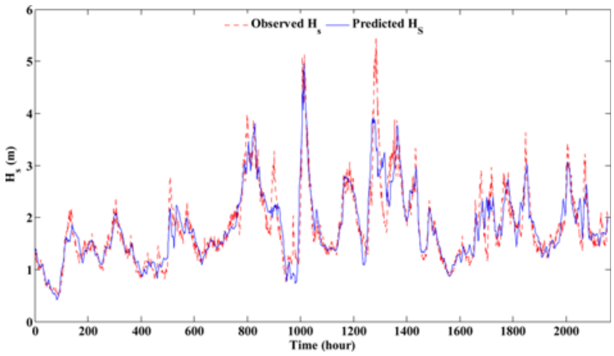


Fig. 4. Observed and Predicted SWH by WGP and db3 Mother Wavelet at 3h Lead Time for Station 41048 (testing data)

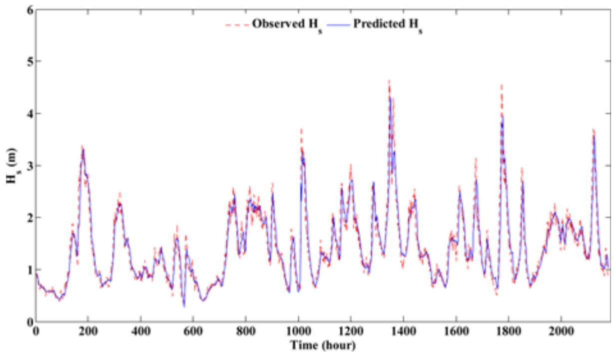


Fig. 5. Observed and Predicted SWH by WGP Model and db3 Mother Wavelet at 3h Lead Time for Station 41013 (testing data)

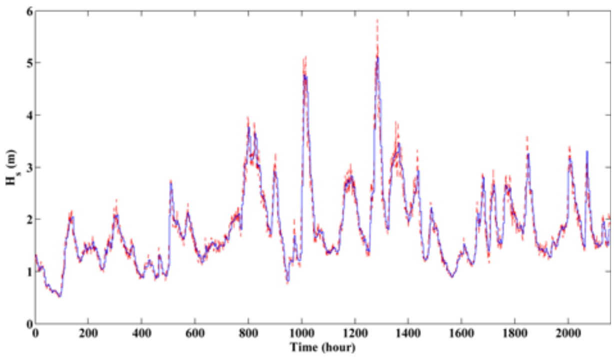


Fig. 6. Observed and Predicted SWH by WGP Model and Haar Mother Wavelet at 3h Lead Time for Station 41048 (testing data)

of both models have approximately same performances whereas accuracy of the GP model was dramatically on the decline for the higher lead times. Evidently, shorter lead times performance is better than higher lead times. The observed and forecasted SWH time series for 3h lead time were shown Fig. 4 to Fig. 7. From figures, it is crystal clear that WGP model forecasts the general behavior of the observed data. The period related to the testing period results indicated which proposed WGP provided slightly higher accuracy than GP model. As lead time increases, the GP model performance decreases dramatically. But the WGP model

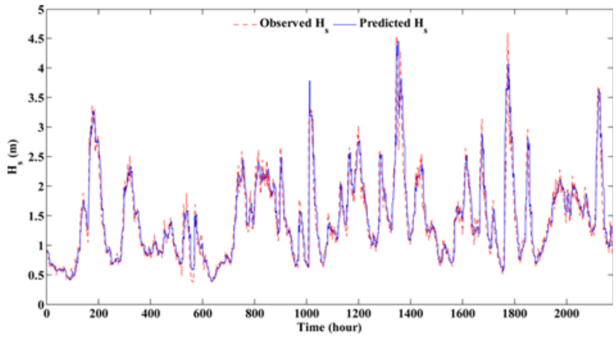


Fig. 7. Observed and Predicted SWH by WGP Model and Haar Mother Wavelet at 3h Lead Time for Station 41013 (testing data)

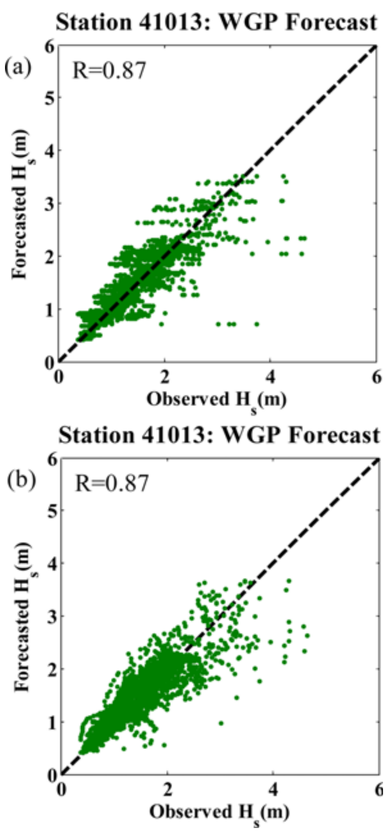


Fig. 8. Scatter Plot of Observed and Predicted SWH by WGP Model and (a)Haar and (b)db3 Mother Wavelet at 6h Lead Time (testing data)

performance decreases gradually as shown in Fig. 10 to 11 in the scatter diagram. For instance, the $I_a$ decreases from 0.975 to 0.620 in station 41013 for 3h and 48h lead times, respectively. The R varies between 0.952 and 0.462 for 3 to 48h lead times. The RMSE increases from 0.213 for 3h to 0.620 m for 48h lead times.

Figure 11 illustrate good accuracy of forecasting by the WGP model at station 41048 for which $0.64 < R < 0.91$ and $0.75 < I_a < 0.96$ are yielded. The scatter plots indicate that the proposed model capability in deep waters is remarkably better than shallow waters. In this way, the proposed model cannot predict

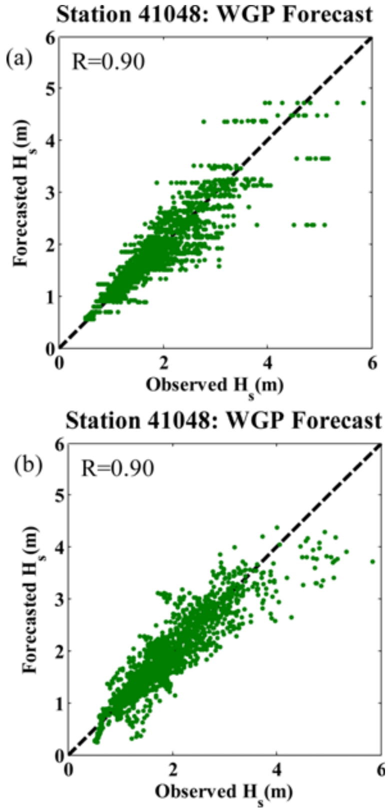


Fig. 9. Scatter Plot of Observed and Predicted SWH by WGP Model and (a) Haar and (b) db3 at 6h Lead Time (testing data)

SWH satisfactorily in shallow waters (station 41013) in comparison with deep waters (41048). This one is an special property of single parameter models. Some other parameters such as wind speed are necessary to improve the model capability in shallow waters.

Throughout the WGP model, the results obtained for 5 different lead times had undergone various decomposition levels starting from 3 to 8 using two different mother wavelets at two stations. For higher lead times (≥ 12h), the optimum accuracy was obtained in higher decomposition levels (7-8). In shorter lead times (≤ 6h), a low decomposition level (5-6) produces satisfactorily accuracy. The optimum decomposition level can be obtained by trial and error process (Table 5).

A possible reason for WGP model performance improvement is that WGP extract the characteristics of SWH variation process through decomposing the non-stationary SWH time series into several stationary SWH time series.

For better comparison purposes the error indices are plotted versus lead time in Fig. 12 and Fig. 13 for station 41013 and 41048, respectively. These plots evidently demonstrates that shorter lead times (3 and 6h) are more accurate than larger lead times (24 and 48h). These results are in agreement with other studies (Gaur and Deo, 2008; Ozger, 2010; Karmanzad et al., 2011; Deka and Phralda, 2012). Fig. 14 illustrated SWH values forecasted by WGP at various lead times in comparison with

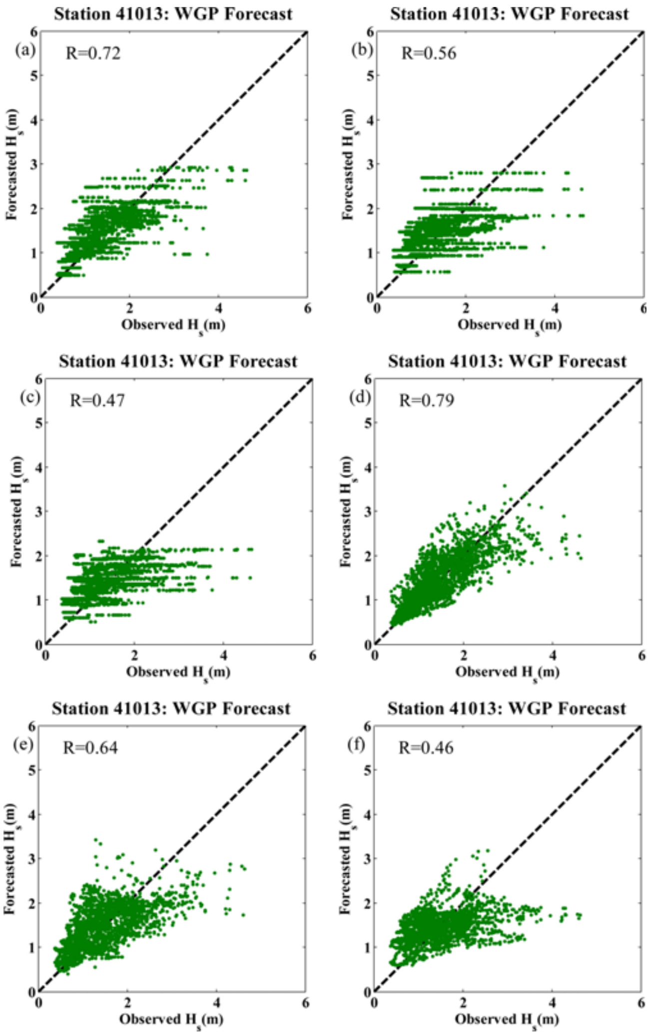


Fig. 10. Scatter Plot of Observed and Predicted SWH by WGP Model and (a) Haar at 12h, (b) Haar at 24h, (c) Haar at 48h, (d) db3 at 12h, (e) db3 at 24h., (f) db3 at 48h (testing data)

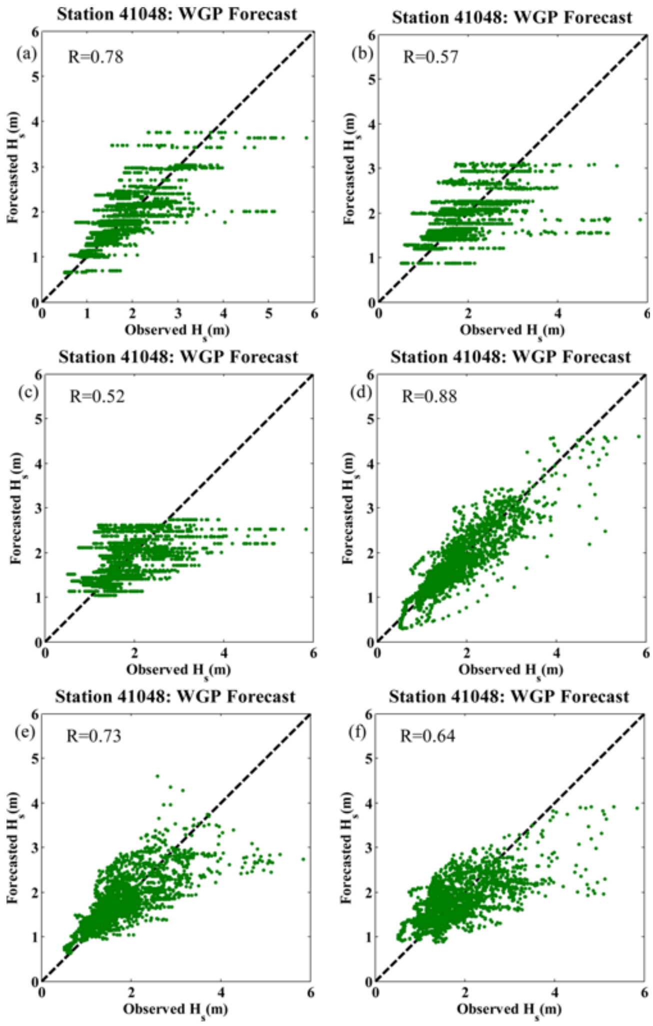


Fig. 11. Scatter Plot of Observed and Predicted SWH by WGP Model and (a) Haar at 12h, (b) Haar at 24h, (c) Haar at 48h, (d)db3 at 12h, (e) db3 at 24h., (f) db3 at 48 h (testing data)

observed values. According to Fig. 14, it is obviously clear that the time series preprocessing by the wavelet can heighten the accuracy of model particularly in higher lead times. At station 41013, the model efficiency criterions showed the higher error of forecasting in comparison with results obtained for station 41048. It is presumably due to the significant fluctuation of time series data around the mean value, so that the regression between data is reduced (Table 1). In contrast, the performance of model for training and testing data sets is closed to each other in station 41013 compare to the results extracted from station 41048. A possible reason is the statistical properties of both stations as shown in Table 2. From Table 2, the statistical properties of training and testing data are closed to each other for station 41013.

As a whole, by comparing the performance of WGP with other model that proposed and examined by some researchers (Soares *et al.*, 1996; Gaur and Deo, 2008; Ozger, 2010; Karmanzad *et al.*, 2011; Deka and Phralda, 2012), it can be controlled that model has comparable accuracy.

## 7. Conclusions

The oceanic phenomena generally and the significant wave height time series in particular are characterized by high non-stationary and non-linearity. A wavelet was introduced to present useful decomposition of the original time series, so that the data preprocessing enhance the forecasting model ability. The wavelet decomposition of non-stationary time series leads to produce bunches of stationary time series that can be applied in analyzing by obtaining useful information on various time resolution levels. Here, the wavelet transform was linked with the Genetic Programming (GP) in order to provide the Wavelet Genetic Programming (WGP) model for SWH forecasting of two stations located in North Atlantic Ocean. Furthermore, the impact of decomposition level on the model efficiency was investigated by considering various decomposition levels. The outperformance of proposed model represented the high merit of both Haar and db3 mother wavelet. Moreover, the db3 mother wavelet yielded better performance than Haar in most cases. The

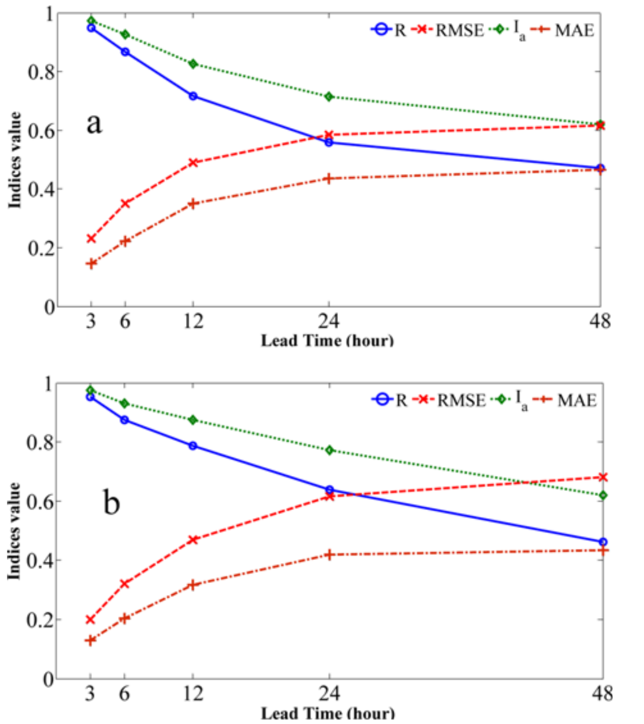


Fig. 12. Performance Indices of WGP Model by (a) Haar and (b) db3 vs. Lead Times for Station 41013

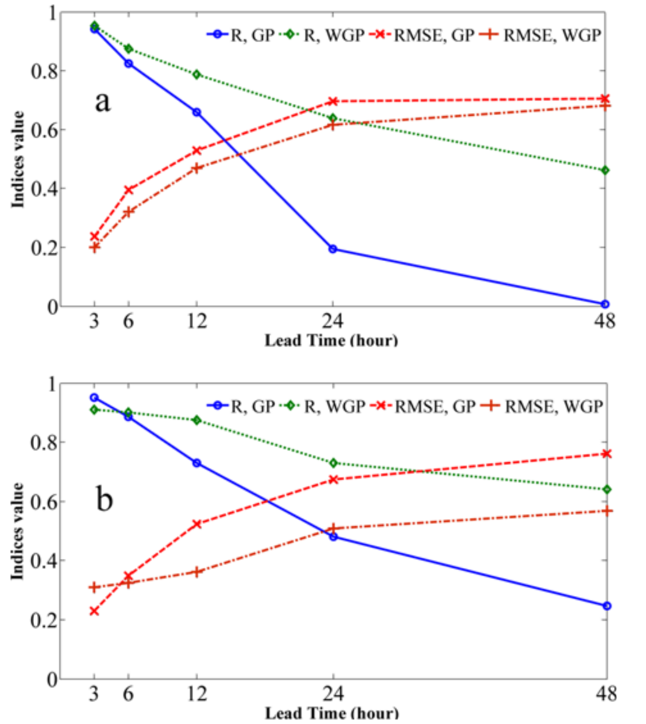


Fig. 14. R and $I_a$ Indices of GP and WGP Model vs. Lead Times for Station (a) 41013 and (b) 41048

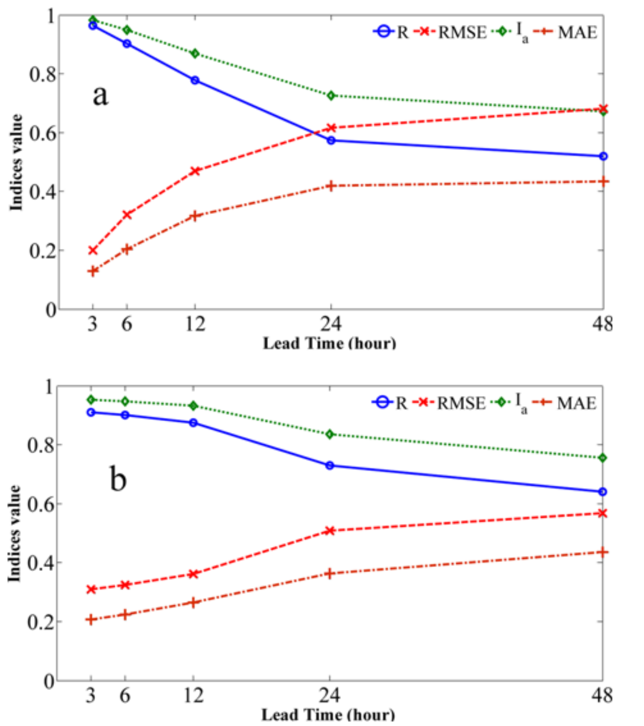


Fig. 13. Performance Indices of WGP Model by (a) Haar and (b) db3 vs. Lead Times for Station 41048

WGP demonstrated prediction with the high level of validation compared to the classic GP especially in higher lead times. The reason for this high efficiency is the utilizing of a wavelet as a preprocessing in the WGP. The enhancement of model performance provided by WGP over GP model was seen, particularly in 48h lead time. The superiority of WGP to GP model can be evidently observed particularly for 48h lead time. From the results, the correlation coefficient values were increased from 0.007 to 0.471 and 0.247 to 0.641 by the WGP model for stations 41013 and 41048, respectively. As a whole, WGP model can be distinguished as a promising tool in the SWH forecasting.

## Acknowledgements

The authors are grateful to the editor and reviewer for their valuable and constructive comments, which help enhance the quality of the paper.

## References

Altunkaynak, A. (2013). "Prediction of significant wave height using geno-multilayer perceptron." *Ocean Engineering*, Vol. 58, pp. 144-153, DOI: 10.1016/j.oceaneng.2012.08.005.

Altunkaynak, A. and Wang, K. H. (2012). "Estimation of significant wave height in shallow lakes using the expert system techniques." *Expert Systems with Applications*, Vol. 39, No. 3, pp. 2549-2559, DOI: 10.1016/j.eswa.2011.08.106.

Asma, S., Sezer, A., and Ozdemir, O. (2012). "MLR and ANN models of significant wave height on the west coast of India." *Computers & Geosciences*, Vol. 49, pp. 231-237, DOI: 10.1016/j.cageo.2012.05.032.

Boggess, A. and Narcowich, F. J. (2009). *A First Course in wavelets with Fourier analysis*, John Wiley Publications, New Jersey.

Canellas, B., Balle, S., Tintore, J., and Orfila, A. (2010). "Wave height prediction in the western Mediterranean using genetic algorithms." *Ocean Engineering*, Vol. 37, Nos. 8-9, pp. 742-748, DOI: 10.1016/j.oceaneng.2010.02.006.

Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P., and Du, H. N. (2006). "Approaches for automated detection and classification of masses in mammograms." *Pattern Recognition*, Vol. 39, No. 4, pp. 646-668, DOI:10.1016/j.patcog.2005.07.006.

Danandeh Mehr, A., Kahya, E., and Ozger, M. (2014). "A gene-wavelet model for long lead time drought forecasting." *Journal of Hydrology*, Vol. 517, pp. 691-699, DOI: 10.1016/j.jhydrol.2014.06.012.

Deka, P. C. and Prahlada, R. (2012). "Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time." *Ocean Engineering*, Vol. 43, pp. 32-42, DOI: 10.1016/j.oceaneng.2012.01.017.

Deo, M. C. and Naida, C. S. (1998). "Real time wave forecasting using neural network." *Ocean Engineering*, Vol. 26, No. 3, pp. 191-203, DOI: 10.1016/S0029-8018(97)10025-7.

Deo, M. C., Jha, A., Chaphekar, A. S., and Ravikant, K. (2001). "Neural networks for wave forecasting." *Ocean Engineering*, Vol. 28, No. 7, pp. 889-898, DOI: 10.1016/S0029-8018(00)00027-5.

Dixit, P., Londhe, S., and Dandawate, Y. (2015). "Removing prediction lag in wave height forecasting using Neuro-wavelet modelling technique." *Ocean Engineering*, Vol. 93, pp. 74-83, DOI: 10.1016/j.oceaneng.2014.10.009.

Gaur, S. and Deo, M. C. (2008). "Real-time wave forecasting using genetic programming." *Ocean Engineering*, Vol. 35, No. 11-12, pp. 1166-1172, DOI: 10.1016/j.oceaneng.2008.04.007.

Günaydýn, K. (2008). "The estimation of monthly mean significant wave heights by using artificial neural network and regression methods." *Ocean Engineering,* Vol. 35, Nos. 14-15, pp. 1406-1415, DOI: 10.1016/j.oceaneng.2008.07.008.

Jain, P. and Deo, M. C. (2007). "Real-time wave forecasts off the western Indian coast." *Applied Ocean Research*, Vol. 29, Nos. 1-2, pp. 72-79, DOI: 10.1016/j.apor.2007.05.003.

Kamranzad, B., Shahidi, A. E., and Kazeminezhad, M. H. (2011). "Wave Height forecasting in Dayyer, the Persian Gulf." *Ocean Engineering,* Vol. 38, No. 1, pp. 248-255, DOI: 10.1016/j.oceaneng.2010.10.004.

Kazeminezhad, M. H., Etemad-Shahidi, A., and Mousavi. (2005). "Application of fuzzy inference system in the prediction of wave parameters." *Ocean Engineering*, Vol. 32, Nos. 14-15, pp. 1709-1725, DOI: 10.1016/j.oceaneng.2005.02.001.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*, MIT press, Cambridge, MA, USA.

Mahjoobi, J. and Mosabbeb, E. A. (2009). "Prediction of significant wave height using regressive support vector machines." *Ocean Engineering,* Vol. 36, No. 5, pp. 339-347, DOI: 10.1016/j.oceaneng.2009.01.001.

Makarynskyy, O., Pires-Silvab, A. A., Makarynska, D., and Ventura-Soaresc, S. (2005). "Artificial neural networks in wave predictions at the west coast of Portugal." *Computers & Geosciences*, Vol. 31, No. 4. pp. 415-424, DOI: 10.1016/j.cageo.2004.10.005.

Mallat, S. (1998). *A wavelet tour of signal processing*, CA: Academic press, San Diego.

Misiti, M., Misiti, Y., Oppenheim, C., and Poggi, J. (2000). *Wavelet Toolbox: For use with MATLAB*, The MathWorks. Natick, Mass.

Muraleedharan, G, Rao, A. D., Kurup, P. G, Unnikrishnan, N. N., and Sinha, M. (2007). "Modified Weibull distribution for maximum and significant wave height simulation and prediction." *Coastal Engineering,* Vol. 54, No. 8, pp. 630-638, DOI: 10.1016/j.coastaleng.2007.05.001.

Nitsure, S. P., Londhe, S. N., and Khare, K. C. (2012). "wave forecasts using wind information and genetic programming." *Ocean Engineering*, Vol. 54, pp. 61-69, DOI: 10.1016/j.oceaneng.2012.07.017.

Nitsure, S. P., Londhe, S. N., and Khare, K. C. (2014). "Prediction of sea water levels using wind information and soft computing techniques." *Applied ocean Research*, Vol. 47, pp. 344-351, DOI: 10.1016/j.apor.2014.07.003.

Nourani, V., Hosseini, B., Adamowski, J., and Kisi, O. (2014). "Application of hybrid artificial-intelligence models in hydrology. A review." *Journal of Hydrology*, Vol. 514, pp. 358-377, DOI: 10.1016/j.jhydrol.2014.03.057.

Nourani, V., Komasi, M., and Alami, M. T. (2012). "Hybrid wavelet-genetic programming approach to optimize ANN modelling of rainfall-runoff process." *Journal of Hydrologic Engineering*, Vol. 17, No. 6, pp. 724-741, DOI: 10.1061/(ASCE)HE.1943-5584.0000506.

Ozger, M. (2010). "Significant wave height forecasting using wavelet fuzzy logic approach." *Ocean Engineering*, Vol. 37, No. 16, pp. 1443-1451, DOI: 10.1016/j.oceaneng.2010.07.009.

Ozger, M. and Sen, Z. (2007). "Prediction of wave parameters by using fuzzy logic approach." *Ocean Engineering*, Vol. 34, Nos. 3-4, pp. 460-469, DOI: 10.1016/j.oceaneng.2006.03.003.

Prahlad, R. and Deka, P. C. (2015). "Forecasting of time series significant wave height using a wavelet decomposed neural network." *Int. conference on water resources, coastal and ocean engineering (ICWRCOE 2015), Aquatic Proc*., Vol. 4, pp. 540-547, DOI: 10.1016/j.aqpro.2015.02.070.

Scotto, M. G. and Soares, C. G. (2000). "Modelling the long-term time series of significant wave height with non-linear threshold models." *Coastal Engineering*, Vol. 40, No. 4, pp. 313-327, DOI: 10.1016/S0378-3839(00)00016-8.

Scotto, M. G. and Soares C. G. (2007). "Bayesian inference for long-term prediction of significant wave height." *Coastal Engineering*, Vol. 54, No. 5, pp. 393-400, DOI: 10.1016/j.coastaleng.2006.11.003.

Seo, Y., Kim, S., and Singh, V. P. (2015). "Multistep-Ahead flood forecasting using wavelet and data-driven methods." *KSCE Journal of Civil Engineering*, Vol. 19, No. 2, pp. 401-417, DOI: 10.1007/s12205-015-1483-9.

Shahabi, S. and Khanjani, M. J. (2015). "Modelling of significant wave height using wavelet and GMDH." *Proc., 36th Int. IAHR World Congress.*, IAHR,Hague, Netherlands.

Shahabi, S., Khanjani, M. J., and Kermani, M. H. (2016). "Hybrid wavelet-GMDH model to forecast significant wave height." *Water Science and Technology: Water Supply*, Vol. 16, No. 2, pp. 453-459, DOI: 10.2166/ws.2015.151.

Soares, C. G. and Cunha, C. (2000). "Bivariate autoregressive models for the time series of significant wave height and mean period." *Coastal Engineering*, Vol. 40, No. 4, pp. 297-311, DOI: 10.1016/S0378-3839(00)00015-6.

Soares, C. G. and Ferreira, A. M. (1996). "Representation of non-stationary time series of significant wave height with autoregressive models." *Probabilistic Engineering Mechanics*, Vol. 11, No. 3, pp. 139-148, DOI: 10.1016/0266-8920(96)00004-5.

Soares, C. G., Ferreira, A. M., and Cunha, C. (1996). "Linear models of the time series of significant wave height on the southwest Coast of Portugal." *Coastal Engineering*, Vol. 29, Nos. 1-2, pp. 149-167, DOI: 10.1016/S0378-3839(96)00022-1.

US Army (1984). Shore Protection manual, Coastal Engineering Research Center. Washington, D.C: USA.