



Maximum wave height hindcasting using ensemble linear-nonlinear models

Rifat Tür¹

Received: 20 March 2020 / Accepted: 13 May 2020 / Published online: 28 May 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

A comparative study between classic linear and intelligent nonlinear time series approaches for short-term maximum wave height forecasting is presented in this study. The applied models to accomplish a use case for onshore measurements from the Mediterranean Sea include ordinary linear regression (LR), autoregressive integrated moving average (ARIMA), artificial neural networks (ANN), and genetic programming (GP). The study also introduces a new evolutionary ensemble model called ensemble GP, which integrates effective models' forecasts through an evolutionary procedure. The results from standalone models showed that both linear and nonlinear models provide the same accuracy for short-term maximum wave height hindcasting on a seasonal scale. The proposed ensemble model can enhance the forecasting accuracy of standalone models markedly. The new model can forecast maximum wave heights with the root mean squared errors less than 5 cm and Nash-Sutcliffe efficiency more than 0.97. It is explicit and secures parsimony conditions, thus it is proposed to be used in practice.

Keywords Wave height · Hindcasting · Time series modeling · Genetic programming

1 Introduction

Variations in oceanic/sea waves are usually characterized as stochastic. It is therefore not surprising that the development and application of emerging data mining techniques have received noteworthy attention in the ocean-/sea wave-analyzing community. From a coastal operation/engineering perspective, data mining techniques were typically applied to develop wave parameters forecasting models using long-term measurements of oceanic/climatic variables (e.g., Tsai et al. 2002; Kazeminezhad et al. 2005; Özger 2011; Vouterakos et al. 2012; Akpınar et al. 2014; Aydoğan et al. 2013; Hadadpour et al. 2014; Duan et al. 2016; Tsai et al. 2018; Law et al. 2020; Zubier 2020).

Artificial neural networks (ANN) are of the earliest data mining techniques that were applied for wave parameter prediction. In a seminal paper by Deo and Naidu (1998), ANNs were used for real-time wave forecasting in Yanam, along the east coast of India. The authors demonstrated that ANNs exhibited higher covariation between the predicted and observed

significant wave height than the classic autoregressive (AR) models. In a similar study, the promising role of ANNs for tide prediction was also reported by Deo and Chaudhari (1998). Agrawal and Deo (2002) compared the efficiency of different ANN models with classic linear time series modeling approaches of autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) and showed that ANNs produce more accurate predictions of wave heights than the linear time series schemes when shorter intervals of predictions (3 h and 6 h) were involved. For long-range predictions (12 h and 24 h), both the classic and ANN models showed similar performance. Makarynskyy et al. (2005) utilized ANNs in wave predictions on the west coast of Portugal and showed that ANNs can be successfully applied to predict significant wave heights and zero-up-crossing wave periods with short-term warning times. Mandal and Prabaharan (2006) reported the potential use of recurrent neural networks (RNNs) through a comprehensive case study in Marmugao, west coast of India. The results indicated that RNNs with exogenous inputs recurrent algorithms can be used for significant wave heights forecasting directly from measured waves. Zamani et al. (2008) showed that ANNs provide higher accuracy than instance-based learning when they applied for significant wave heights forecasting in the Caspian Sea. Londhe and Panchang (2007) and Londhe

✉ Rifat Tür
rifattur@akdeniz.edu.tr

¹ Civil Engineering Department, Akdeniz University, Antalya, Turkey

(2008) proved that ANNs can recognize the nonlinear correlation between buoy networks that could be beneficial in the estimation of missing wave parameters via correlating oceanographic data from multiple stations. Kamranzad et al. (2011) used wind speed, direction, and wave height as input parameters for an ANN model for significant wave height prediction for the Persian Gulf. More recently, Zubier (2020) demonstrated that ANN can be satisfactorily employed for the significant wave heights prediction in Eastern Central Red Sea for up to 12 h in advance.

For several years, genetic programming (GP) has been successfully used to solve forecasting and classification problems in hydro-climatological applications (Danandeh Mehr et al. 2018). Regarding ocean engineering, some examples include prediction of significant wave height and average zero-cross wave period (Kambekar and Deo 2012), wave runup (Power et al. 2019), and sea level (Ghorbani et al. 2010) as well as filling up gaps in wave data (Ustoorikar and Deo 2008), and optimizing pumping strategies for coastal aquifer management (Sreekanth and Datta 2010). Owing to the explicit structure of GP-based models, recent studies have also used GP to develop new formulae to optimize the design of coastal structures. For example, Formentin and Zanuttigh (2019) presented a new GP-based formula for parametrization of the reductive effects induced by crown walls and bullnoses on the average wave overtopping discharge at coastal structures. Most recently, the multigene GP technique was used by Lee and Suh (2020) to develop stability formula for rock armor and Tetrapods. The authors demonstrated that the multigene GP model is more accurate than previous empirical formulae as well as ANN-based models.

To cope with uncertainties available in wave-borne data, hybrid data mining models were suggested in the recent studies (Balas et al. 2010; Koç and Balas 2012). The hybrid models either integrate the capabilities of various data mining models or created through linking a data-preprocessing approach with a data mining technique (Danandeh Mehr 2018). For instance, an adaptive neuro-fuzzy inference system (ANFIS) is a hybrid technique that was suggested and applied by Mahjoobi et al. (2008) for wave parameters hindcasting in Lake Ontario. The authors compared ANFIS with ANN and FIS models and demonstrated that ANFIS is marginally more accurate than its counterparts. ANFIS was also used by Tür and Balas (2010) for accurate forecasting of daily significant wave height in the Black Sea. The authors used daily average wave height and wave period data recorded at different time intervals and compared the ANFIS results with those of ANN (Deo and Naidu 1998) and FIS (Kazeminezhad et al. 2005). The results showed that ANFIS is generally superior to the standalone models. A similar study conducted by Akpınar et al. (2014) indicated the promising role of ANFIS for significant wave heights and period forecasting in the Black Sea. Zanaganeh et al. (2009) introduced a hybrid genetic

algorithm–ANFIS model in which both clustering and rule base parameters in the prediction of significant wave height and peak spectral period are simultaneously optimized using genetic algorithm and ANNs. The results showed that the hybrid model is superior to ANFIS and Shore Protection Manual methods in terms of their prediction accuracy. Combining wavelet decomposition with ANN, a hybrid neuro-wavelet model was suggested by Dixit et al. (2015) and Dixit and Londhe (2016) which was able to satisfactorily predict extreme wave heights and remove timing error in standalone ANN, respectively. Duan et al. (2016) demonstrated that AR models are adaptive for wave forecasting. However, they have limitations in forecasting nonlinear and non-stationary waves. Inspired by the capability of empirical mode decomposition (EMD), the authors suggested a hybrid EMD-AR model for significant wave height forecasting using the data from the National Data Buoy Center, USA. More recently, Buyukyildiz and Tezel (2017) showed that hybridization of particle swarm optimization algorithm with ANN and ANFIS may increase sea level forecasting accuracy in Lake Beysehir Turkey.

The review showed that the prediction of wave parameters is of great importance for planning of many operation-related activities in the oceans/seas and the design of coastal structures. The issue has been investigated by numerous researchers utilizing various data mining techniques. While the previous studies have mostly been focused on the prediction of significant wave height for different lead times, the maximum wave height (H_{max}) hindcasting for concurrent data using GP techniques has not been investigated yet. Therefore, the main objective of this study is to explore a novel hybrid GP-based model that may increase the accuracy of standalone GP to forecast H_{max} 1 h in advance. The method presented in this paper follows the ideas of the ensemble GP models proposed by Rahmani-Rezaeieh et al. (2019) and Khozani et al. (2020) for streamflow and incipient sediment motion modeling, respectively. However, instead of different types of GP, the proposed model integrates simple linear models with a classic GP variant which may yield in hybrid models simpler than those of previous studies that guarantee its novelty.

2 Study area, wave parameter monitoring, and data preparation

The direct wave height measurement has been started in Turkey in recent years. In the earlier studies, long-term wind data was typically used to estimate wave parameters (height and period) that yielded in Turkish deep-sea wave-wind atlas (Aydoğan et al. 2013). The data used in the study is from a buoy station located in Antalya Bay (36°43'00" N - 31°01'00" E), approximately 20 km offshore and 350 m deep,

Mediterranean Sea, Turkey (Fig. 1). Atmospheric data (instantaneous wind velocity and direction, maximum wind velocity and direction, temperature, humidity, and pressure) and marine data (significant wave height and period, maximum wave height, wave direction, current velocity, current direction, salinity, and conductivity) are measured continuously at the station. The station is of tower type anchored to the seabed (Fig. 1). The station provides transmission of measurement data via GPS (Global Positioning System) and GPRS (General Packet Radio Service). Considering the sea depth, the station is in the deep sea and the waves are not transformed.

The hourly data from April 1, 2015 till March 31, 2016 were collected and used in this study. Because of a pronounced variation in the variance of measured wave heights, we divided the entire measurements into four subsamples of AMJ (April-May-June), JAS (July-August-September), OND (October-November-December), and JFM (January, February, March). Statistical characteristics of observed *Hmax* values were tabulated in Table 1. There are 2185 observations in each season. The samples at each season are also divided into two subsamples of training (70 days, 1640 observations) and holdout testing (20–21 days, 545 observations) periods. After detecting the best models, the testing data is used to control forecasting accuracy. Timeseries plot of the observations given in Fig. 2 provides a general view of the variation of the *Hmax* in different seasons. Presence of extreme waves with heights greater than 3 m is pronounced in the OND period. Due to short period (3-month) of each series, the plots reveal no seasonal component.

Figure 3 presents the directional wave rose showing the frequency of *Hmax*. It is observed from the figure that during the AMJ season, the waves smaller than 1 m are approached to the coast from S-SSW directions, and the higher waves come from SSW direction. In the JAS season, the waves mostly come from S-SSW directions. In the OND season, the waves below 1 m are seen from the SSW-S-SSE and SE directions, and the higher waves come predominantly from the SE

direction. Finally, in the JFM season, the waves smaller than 1 m mostly appear from the S direction, and waves with maximum height in the range 1–2 m come from the SSW, SE, and WNW, and the higher ones come from the SSW and SE directions.

Considering the annual wave directions, it is implied that waves smaller than 1 m are predominant and come from S and SSW directions. Larger waves appear to come mainly from the SSW and SE directions during the JFM and OND seasons. It is worth to mention that the wave directions obtained in the study agree with the wind-blowing directions in the region. Historical observations showed that most of the coastal damages were occurred by the waves coming from these directions during JFM and OND seasons.

3 Methodology

Successive observations of stochastic wave events are typically correlated, so future values may be predicted from past values. The discrete stochastic *Hmax* time series is analyzed in this study using simple descriptive techniques integrated with symbolic regression tools. The aim is to construct mathematical model which explains the observed variability in the data so that we can predict, with some confidence, future observations. As previously mentioned, two linear schemes including a simple linear regression (LR) and ARIMA model together with two nonlinear schemes including ANN and GP are used to model *Hmax* in each season. Besides, a new ensemble approach is suggested in this study that is described in this section.

3.1 The benchmark linear regression (LR) model

The ordinary LR model is the most basic type of statistical technique that is widely used to establish a linear relationship between two variables and forecast new observations. In the

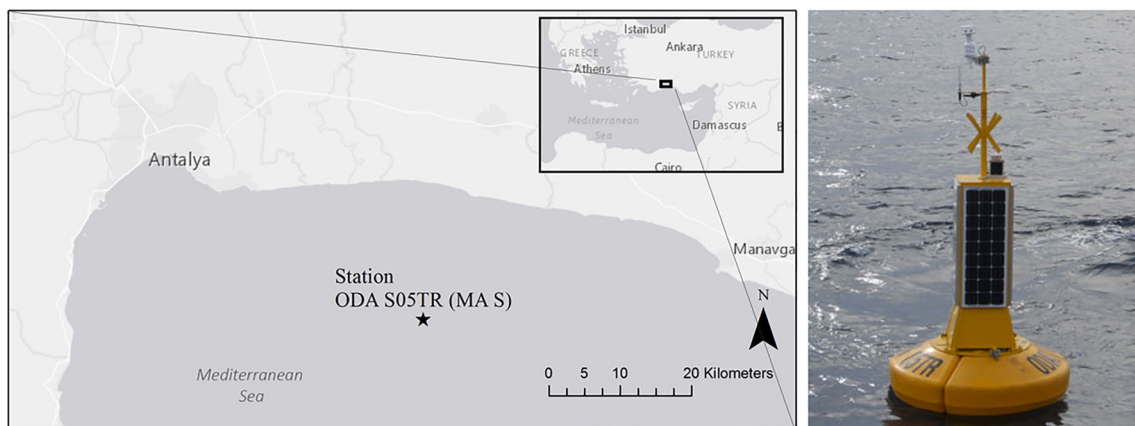


Fig. 1 Location of the buoy station (star) at Antalya coast

Table 1 Statistical characteristics of maximum wave height measured at different seasons

Season	Mean	Min	Max	Median	Mode	Skewness	Standard deviation
AMJ	0.58	0.06	2.87	0.48	0.26	1.66	0.40
JAS	0.43	0.11	2.29	0.37	0.29	2.69	0.23
OND	0.66	0.11	4.33	0.50	0.23	1.45	0.52
JFM	1.135	0.06	7.20	0.87	0.36	2.09	0.97

latter, the existing relationship about the observations is used to forecast unobserved values. It is known that wave height impacts on each other over the time and therefore, the previous values ($Hmax_{t-1}$) can be used to forecast the current value ($Hmax_t$). On average, they might follow a linear pattern and, therefore, the linear equation to represent them could be expressed as:

$$Hmax_t = \beta_0 + \beta_1 Hmax_{t-1} + \varepsilon \quad (1)$$

where β_1 is the slope coefficient of the linear line plot between the current and previous $Hmax$ values, β_0 is the constant term or intercept, and ε is the error term which is going to be minimized to get the best linear fit.

3.2 Overview of ARIMA model

ARIMA process is of the most general time series modeling approach that allows the modeler to model nonstationary time series. In the ARIMA process, instead of predicting the time series itself, differences of the time series from one time step to the previous time step are predicted. It basically means that instead of predicting $Hmax_t$, a transformation is carried out so that $Z_t = Hmax_{t+1} - Hmax_t$ as difference between consecutive values of $Hmax$ is modeled and predicted. The basic form of ARIMA (p, d, q) is presented as follows:

$$Z_t = \sum_{i=1}^p (\varphi_i \cdot Z_{t-i}) - \sum_{j=1}^q (\theta_j \cdot \varepsilon_{t-j}) + \varepsilon_t \quad (2)$$

where φ_i is the coefficient of AR model, θ_j is the coefficient of moving average (MA) model, and ε_t denotes the error in the current time step. The parameters p , q , and d represent the orders of AR, MA, and integrated part, respectively.

To identify the best ARIMA (p, d, q) models, first, the model parameters are determined. To determine order of integration, the parameter d , Augmented Dicky-Fuller (ADF) unit root test, is used in this study which indicates whether the time series does need to be differenced to make it stationary. For example, $d = 1$ means it needs to be differenced one time to make it stationary. The null hypothesis for an ADF unit root test is that the series contains a unit root against the alternative that the series does not contain a unit root because the series appears to be stationary around a constant. If the null hypothesis is rejected, then d equals to zero which means the data series does not need to be differenced to be made stationary, and therefore, the ARIMA (p, d, q) becomes the ARMA (p, q). Once Z_t is predicted, the $Hmax_{t+1}$ that the modeling process originally started with can be recovered using the following equation:

$$Hmax_{t+1} = Z_t + Hmax_t. \quad (3)$$

Fig. 2 Hourly values of maximum wave height measured at different seasons

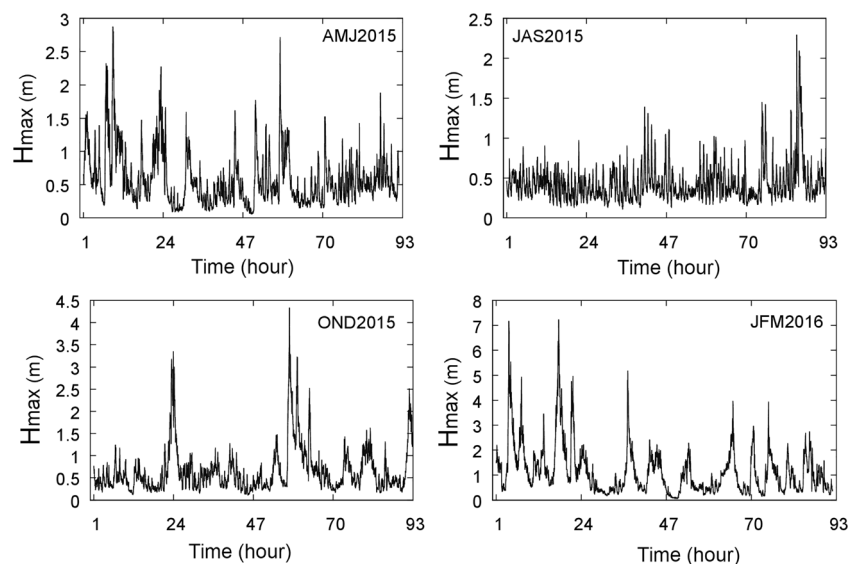
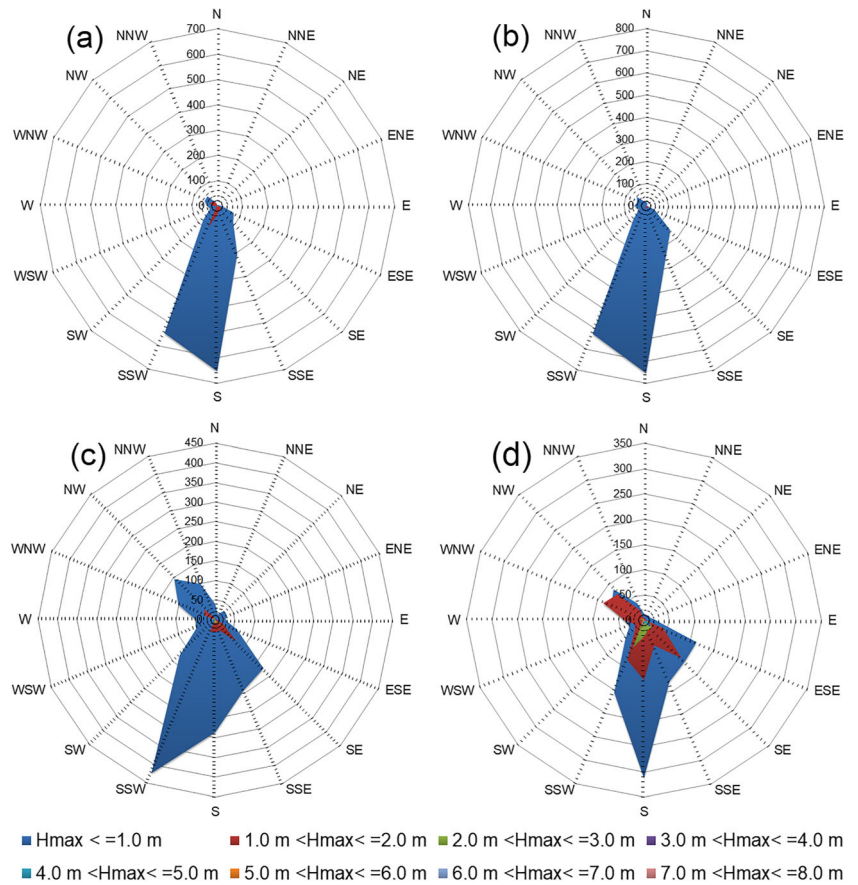


Fig. 3 Seasonal wave rose at the buoy station; **a** AMJ, **b** JAS, **c** OND, and **d** JFM seasons



Comparing to the classic Dickey-Fuller test, the ADF allows for testing the higher-order autoregressive processes. The other alternatives include (but are not limited to) ADF-GLS (Elliott et al. 1996) and KPSS (Kwiatkowski et al. 1992) tests.

3.3 Overview of ANN

An ANN model imitates human brain activity and it consists of neurons and some layers which are processing units. A feedforward neural network is a type of ANN algorithm that consists of two or more layers with neurons in each layer. The network response is:

$$y_i = f(\sum_{i=1}^n w_i x_i - b_j) \tag{4}$$

where n denotes the total number of input values (i.e., observed H_{max}) and their weights, x_i parameter are the input values of a perceptron, parameter $w_i w_i$ are weights for each input, b_j value is the bias of the perceptron, and y_j refers to output of a perceptron that states activation function values of summation.

There are hidden layers between the input and the output layers. Each neuron layer receives input from the previous layer and gets in contact with the next layer by sending an

output. The output y of a perceptron for hyperbolic tangent, or tanh, function in the hidden layers is denoted as:

$$y = f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{5}$$

where parameter z denotes the output of a perceptron summation.

3.4 Overview of genetic programming

GP is a domain-independent, problem-solving approach in which computer programs are evolved to find solutions to

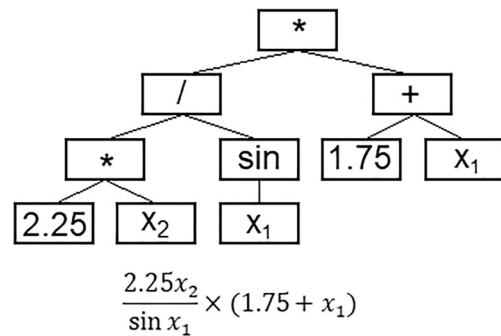


Fig. 4 A tree-shaped genomes and corresponding mathematical representations

problems. It is based on the Darwinian principle of “survival of the fittest.” The programs are typically characterized by a tree structure known as genome. Figure 4 illustrates a genome and corresponding mathematical equation using a root node (multiplication), inner nodes of addition, multiplication, and subtraction as well as terminal nodes of x_1 , x_2 , and random numbers of 2.25 and 1.75. Each node in a GP tree can adopt a function or terminal variables. Some of the main issues in the GP-based modeling is the selection of a set of appropriate functions, input variables, and maximum depth (also referred to as height) of GP trees.

No matter what the problem is, GP algorithm begins with creation of initial population of programs called potential solutions. Then, the solutions that show higher performance during training phase survive to the next generation of population where they are considered as parents to create offspring. To this end, typically three evolutionary operators are used. The operators that act on the genomes include reproduction, crossover, and mutation. Reproduction is transferring the best single solution into the new population set of offspring without any morph. The process commences with the goodness of fit assessment of initial programs and ends after the identification of the best fitted one. Crossover is an operation that needs two of the best solutions as parents. The operation yields two offspring by replacing of genetic material of the parents. The offspring are solutions that possess genetic materials of their parents. Many studies have shown that offspring fits to the training set better than their parents. Mutation is the third genetic operation in which genetic materials of a single parent is replaced with new genetic materials (a new subtree) at the mutation point (Karimi et al. 2019). These operations are repeated using the population of offspring as the new set of parents till an individual shows a desired level of fit to the training data set. If the individual shows favorite accuracy for the testing data, then it is called the best solution and the model is not overfitted. Different variants of GP and their applications in water engineering were reviewed by Danandeh Mehr et al. (2018). The reader is referred to this paper for the relevant details.

3.5 Ensemble GP (EGP) model

To overcome the problem of the use of single input in both linear and nonlinear forecasting models, a new ensemble learning method that combines the multiple models results to improve the accuracy of the forecasts is suggested and verified in this study. In the ensemble model, which is called EGP, a new GP tree is created so that the evolutionary algorithm tries to minimize error between new model outputs and observed data at training period, however, the best model is selected among those of potential solutions that provide the highest accuracy in validation dataset. The idea came from the unique feature of GP that can provide a population of models

for a single problem having more or less the same accuracy at training data, but different at validation set. In this way, the modeler would be avoided from overfitting problem. The other advantage of the proposed model is the automatic sensitivity analysis of standalone models using GP engine. When ensemble inputs are used, GP would be able to give more weights to the more effective inputs and reject the less effective inputs which may yield not only more accurate forecast but also a parsimonious structure. According to several researches in ensemble modeling, it is confirmed that combining the results of different models could enhance the prediction accuracy (Rahmani-Rezaeieh et al. 2019; Nourani et al. 2020).

3.6 Performance evaluation of the models

The models' performances are compared using two statistical indicators, Nash-Sutcliff Efficiency (NSE) and Root Mean Square Error (RMSE) as expressed below.

$$NSE = 1 - \frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{\sum_{i=1}^n (X_i^{obs} - X_{mean}^{obs})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{n}} \quad (7)$$

where X_i^{obs} is the wave height measured in the buoy, X_i^{pre} is the height calculated by the models, and n is the number of measurements.

4 Results and discussion

In this section, at first, the outcomes of standalone modeling via LR, ARIMA, ANN, and GP techniques are presented. Then the results of the new ensemble model and discussion on different models are presented.

4.1 Evolution of standalone linear and nonlinear models

The best LR models attained for AMJ, JAS, OND, and JFM seasons were presented in Eqs. 8 to 10, respectively.

$$Hmax_t = 0.032 + 0.9454Hmax_{t-1} + \varepsilon \quad (8)$$

$$Hmax_t = 0.0674 + 0.834Hmax_{t-1} + \varepsilon \quad (9)$$

$$Hmax_t = 0.018 + 0.9722Hmax_{t-1} + \varepsilon \quad (10)$$

$$Hmax_t = 0.041 + 0.9659Hmax_{t-1} + \varepsilon \quad (11)$$

The constant term in all models are close to 0.0 and coefficient for the $Hmax_{t-1}$ is close to 1.0. These equations show that for every meter increase in $Hmax$, the next hour it will rise by 0.9 on average.

As previously described, to develop an ARIMA model, the relevant parameters must be identified first. Here, the ADF test was applied for each time series to figure out the order of integration d . To make sure to include enough lags, the max lag of ADF test was selected equal to 24 which is two times more than the cube root of the sample size (~ 11.8). Therefore, the test is started with 24 lags and test down using a 10% significant level to determine number of lags to include in the ADF unit root test. As the mean value of each series is not zero (see Table 1), the test was done with a constant. The p values of AMJ, JAS, OND, and JFM seasons were calculated equal to $7.445e-10$, 0.001166 , 0.0006641 , and $1.749e-5$ which imply that the null hypothesis of a unit root can be rejected. This means that the $Hmax$ observations at each season does not have to be differenced to be made stationary and thus, $d = 0$.

To determine the parameters p and q , the correlogram of the series were plotted and the patterns of autocorrelation function (ACF) and partial ACF (PACF) visually inspected as depicted in Fig. 5. Both are decaying but an oscillation pattern is seen for JAS season. As expected, there is no significant partial correlation after the first lag. These patterns point to p and q being equal to one in all the series. Thus, the model would be ARIMA (1,0,1) which is also called ARMA (1,1). Therefore, the evolved models have the form given in Eq. (12). The associated parameters attained for each season at training period were presented in Table 2.

$$Hmax_{t+1} = \varphi.Hmax_t + \varepsilon_t - \theta.\varepsilon_{t-1} \tag{12}$$

In the evolution of nonlinear data mining models (here ANN and GP), dimensions of the inputs together with the types of transfer functions within the computation units of the models (if any) must be taken into consideration before training the models. To attain dimensionally correct ANN and GP models in this study, the observed $Hmax$ series were scaled to the range between 0.1 and 0.9 as follows.

$$Hmax_{nt} = 0.8 \times \frac{Hmax_t - Hmax_{min}}{Hmax_{max} - Hmax_{min}} + 0.1 \tag{13}$$

To develop ANN models, the well-known Levenberg-Marquardt algorithm existing in neural network toolbox of MATLAB was used. Several three-layer feed-forward networks with hyperbolic tangent and pure linear transfer functions, as the activation functions in the hidden and output layers, respectively, were trained and tested. To achieve an appropriate topology at each season, different numbers of hidden neurons were used in the training process. The results from training process at all the seasons showed that the best performance at testing period is produced when only two hidden neurons are used. Although increasing the number of hidden neurons may slightly intensify the models' performance at training sets, it may not increase the performance at testing set, which means the models are overfitted to the training data set.

To develop GP models, GPdotNET software packages (Hrnjica and Danandeh Mehr 2019) were used in this study. The package evolves hundreds of potential models via the combination of input parameters, user defined functions, and some random constants. Like the ANN models, RMSE was used as the fitness fiction to be minimized during the training process. Rank selection method was utilized to select the successful individuals among the potential models during the evolutionary process (model generation), and the best model was considered as the one which provides the least RMSE at the training data sets after 500 generation. The probability of crossover, mutation, and reproduction operations were equal to 0.8, 0.2, and 0.1, respectively. The mathematical expressions of the best attained GP models for AMJ, JAS, OND, and JFM seasons were presented in Eqs. 14 to 17, respectively.

$$Hmax_{nt} = Hmax_{nt-1} + (((cos(Hmax_{nt-1})) - 0.97607) \times 0.00326) \tag{14}$$

$$Hmax_{nt} = Hmax_{nt-1} - ((sin((0.1996 \times (sin((Hmax_{nt-1} - 0.1996)))))) \tag{15}$$

$$Hmax_{nt} = Hmax_{nt-1} \times (cos((0.67758 (Hmax_{nt-1})^2 \times (Hmax_{nt-1} \times (cos((Hmax_{nt-1})^2)))))) \tag{16}$$

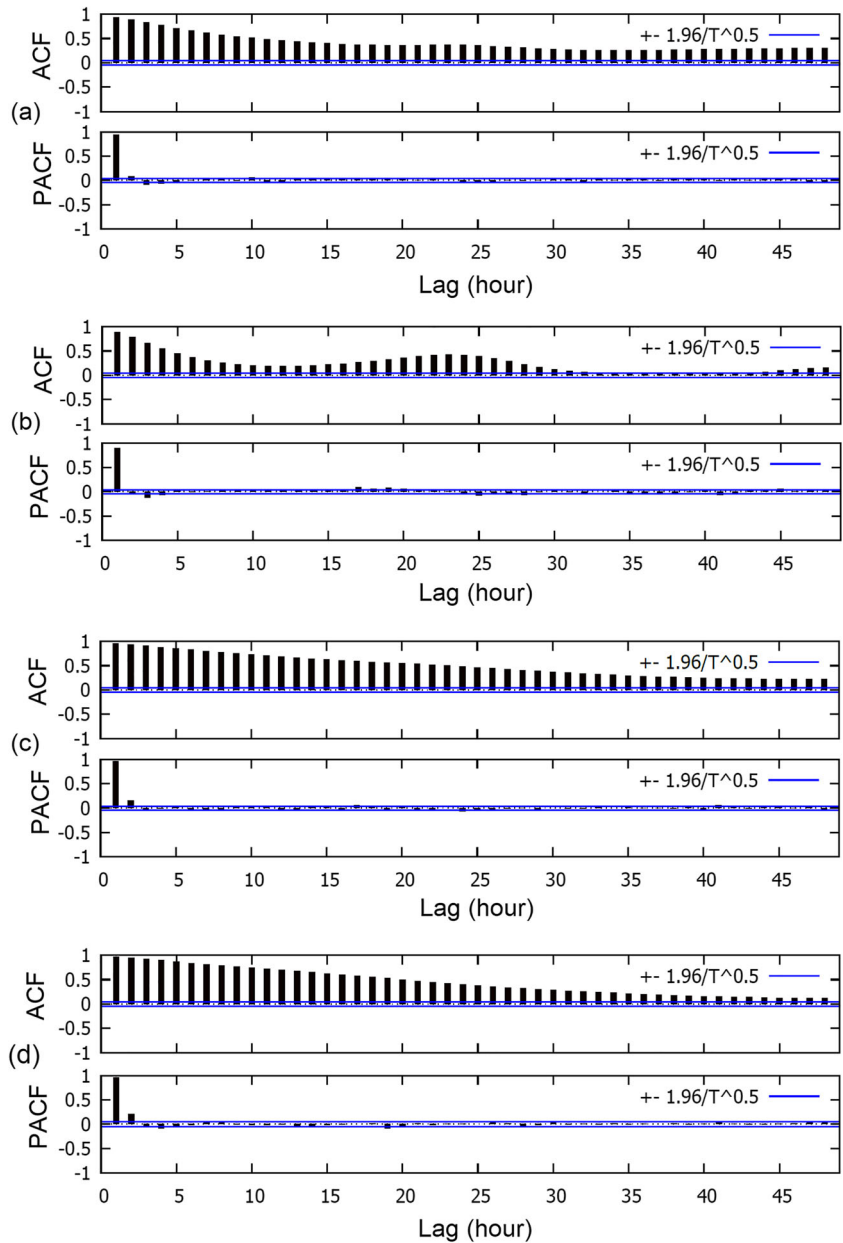
$$Hmax_{nt} = Hmax_{nt-1} \times (cos((((sin(Hmax_{nt-1})) \times (Hmax_{nt-1})) \times ((cos(Hmax_{nt-1})) \times (cos(Hmax_{nt-1})))))) \tag{17}$$

4.2 Comparison of classic linear and data mining nonlinear models

For further assessment of the standalone models, the performance measures of each model during training and testing

periods were summarized in Table 3 and the scatter plots of the forecasts vs observations were presented in Fig. 6. Since the actual data for the testing period is available, the static method, aka one step ahead or rolling forecast, was used instead of dynamic forecasts. In other words, the forecasts in

Fig. 5 Correlogram of the maximum wave height time series in **a** AMJ, **b** JAS, **c** OND, and **d** JFM seasons



testing periods are based on actual hold out series rather than the forecasted value.

According to Table 3, both linear and nonlinear models show more or less same accuracy at each season. While they are not well enough in AMJ, their forecasting accuracy is

acceptable ($NSE > 0.9$) in the OND season. However, the maximum forecasting error during training period is 0.15 m which belongs to the ARMA model in OND season. The reason behind is perhaps the extreme wave height ($H_{max} = 4.33m$) observed in this season.

Table 2 Parameter values obtained for ARMA (1,1) models at each season

Season	AR component (φ)	MA component (θ)	constant	Sample size
AMJ	0.955852	-0.103415	0.590405	1656
JAS	0.830995	0.0777106	0.404869	1656
OND	0.972633	-0.138914	0.649897	1656
JFM	0.977553	-0.185120	1.20279	1656

Table 3 The summary of standalone models for one-step ahead maximum wave height forecasting

Method	Season	Training		Testing	
		RMSE (m)	NSE	RMSE (m)	NSE
LR	AMJ	0.14	0.894	0.13	0.729
	JAS	0.08	0.695	0.14	0.827
	OND	0.12	0.945	0.12	0.922
	JFM	0.27	0.933	0.21	0.900
ARMA (1, 1)	AMJ	0.14	0.895	0.13	0.723
	JAS	0.09	0.728	0.14	0.828
	OND	0.15	0.930	0.12	0.925
	JFM	0.27	0.932	0.21	0.898
ANN (1-2-1)	AMJ	0.14	0.892	0.13	0.730
	JAS	0.08	0.690	0.15	0.807
	OND	0.13	0.942	0.13	0.916
	JFM	0.27	0.933	0.21	0.902
GP	AMJ	0.14	0.894	0.13	0.728
	JAS	0.08	0.693	0.14	0.824
	OND	0.12	0.945	0.13	0.921
	JFM	0.27	0.934	0.21	0.899

Table 4 Performance indices obtained for the ensemble models at each season

Models	Season	Training		Testing	
		RMSE	NSE	RMSE	NSE
EM	AMJ	0.14	0.893	0.13	0.705
	JAS	0.08	0.694	0.15	0.819
	OND	0.12	0.946	0.12	0.925
	JFM	0.27	0.936	0.21	0.900
EGP	AMJ	0.02	0.997	0.01	0.998
	JAS	0.03	0.969	0.05	0.978
	OND	0.03	0.997	0.02	0.997
	JFM	0.04	0.998	0.03	0.998

4.3 Evolution of ensemble models

As previously mentioned in the ensemble modeling, the results of the best LR, ARMA, ANN, and GP models are recombined through a nonlinear scheme. Thus, the ensemble results are not sensitive to the limit of a single input used in standalone models. To attain an optimum ensemble model in this stage, the GP engine was executed again to minimize mean absolute error as the secondary fitness function, and the maximum depth of genes (up to seven to secure parsimony condition) as well as the optimum rate of evolutionary operations were obtained via a trial-error process. Given an extra attempt, a simple averaging (called ensemble mean; EM) model was also considered as another benchmark in which arithmetic mean of the standalone models' (i.e., LR,

According to Fig. 6, the GP and LR forecasts have distributed closer to the 1:1 line in AMJ season. The ARMA exhibited the worst performance in this case. Considering the distribution of the results in JAS season, it is seen that ARMA still scattered far away from diagonal line and tends to underestimate *Hmax*. As already mentioned, the best forecasts were obtained in OND season with slight superiority of GP to its counterparts.

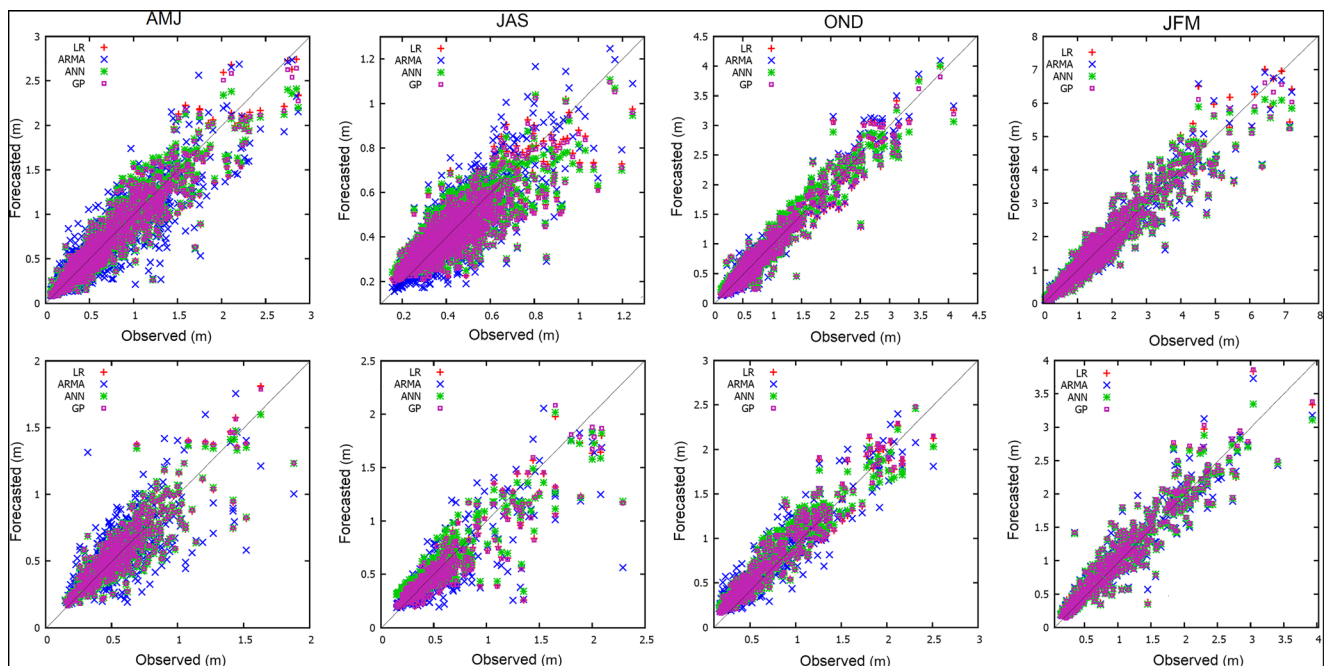


Fig. 6 Scatter plots of the standalone model estimation vs. observed experimental training (top panel) and testing (bottom panel) data sets

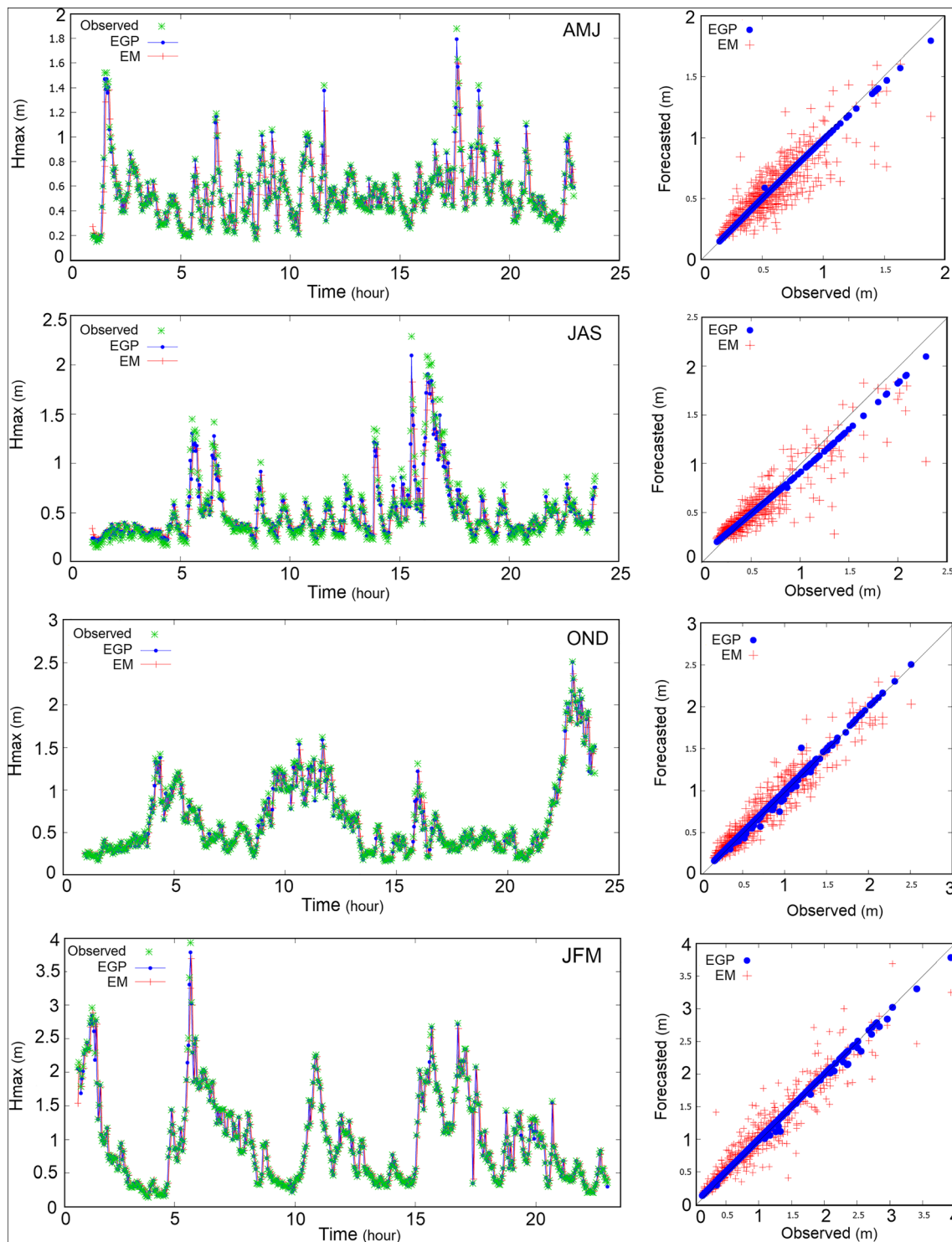


Fig. 7 Observed and ensemble model forecasts of maximum wave height hydrograph and their scatter plot during testing period

ARIMA, ANN, GP) forecasts were calculated as the new forecasts. Table 4 compares the efficiency results of the best evolved ensemble models at different seasons.

The results indicated that the EM scheme did not necessarily increase forecasting accuracy of the standalone models. However, the EGP significantly improved the forecasting

results. The scheme decreases the forecasting errors to 0.03 m on average. The reason behind such promising results is further postprocessing of the forecasts of linear/nonlinear models so that the results of the effective models were recombined to achieve higher accuracy in testing period. Figure 7 compares ensemble models' forecasts with

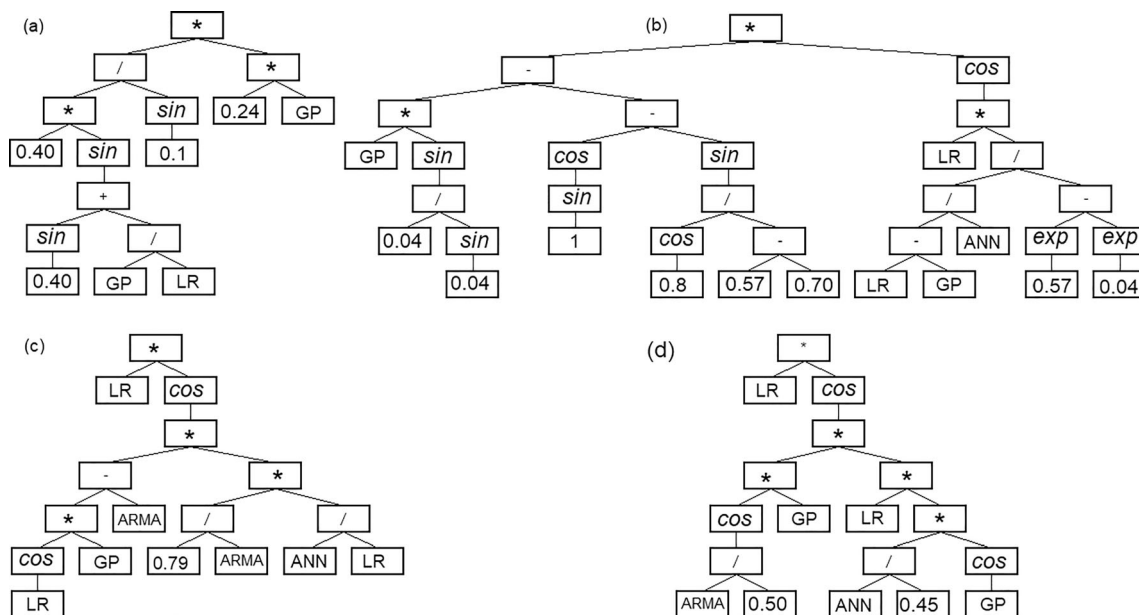


Fig. 8 The best EGP tree evolved for maximum wave height forecasting in **a** AMJ, **b** JAS, **c** OND, and **d** JFM seasons

corresponding observations during the testing period. Although both ensemble models can capture the stochastic feature of the highly fluctuating, they slightly underestimate *Hmax* values higher than 1 m in AMJ and JAS seasons. The EGP is superior to EM and can capture the global and local maxima better than EM and all the standalone models.

As previously mentioned, the EGP model is explicit and can be expressed in the tree genome. The corresponding genome of the best evolved EGP models were illustrated as depicted in Fig. 8. The GP and LR models appeared in all the seasons more frequently which implies higher efficiency of the associated models in the accuracy of final solutions. The ARMA model appeared only in the EGP model of OND season.

5 Conclusions

The tremendous growth in offshore operational activities demands enhanced wave forecasting models. In this study, a comparative study between classic linear and data mining nonlinear methods was accomplished and a new ensemble model was proposed for *Hmax* forecasting. The models were trained and verified using hourly wave measurements from a buoy station located in the route of intensive maritime trade in Mediterranean Sea, Antalya region, Turkey.

The results showed that both linear and nonlinear models are generally qualified to 1-h ahead forecast of *Hmax* in the study region. As the forecasts obtained in seasonal scale, it can be concluded that the use of ordinary linear regression models, with a given confidence, would be applicable in practice.

While a higher degree of accuracy is needed, the proposed ensemble modeling methodology can be implemented.

From a modeling perspective, despite being hybrid, the new EGP model is explicit and allows the modeler to improve the accuracy of standalone models through a secondary fitness function. The other important feature of the proposed ensemble model which highlights its superiority over the existing ensemble models (e.g., Nourani et al. 2020; Tehrani et al. 2019, among others) is that it meets simplicity conditions via an internal sensitivity analysis. These features encourage the EGP model to be applied by practitioners. This study was limited to 1-h ahead forecast of *Hmax* for a single station in the Mediterranean Sea. Future studies may explore the performance of EGP approach for *Hmax* prediction with higher lead times. Increasing the number of stations along the sea gives more accurate results on the modeling of the entire coast. Moreover, efficiency of the proposed approach for wind-wave models could be investigated, however, their complexity and inclusion of errors due to sudden changes in wind measurements must be carefully addressed.

Acknowledgments The data used in this study was provided by the Turkish State Meteorology Service (MGM). The author appreciates the fruitful comments from three anonymous reviewers that caused a significant improvement in this paper.

References

Agrawal JD, Deo MC (2002) On-line wave prediction. *Mar Struct* 15(1): 57–74. [https://doi.org/10.1016/s0951-8339\(01\)00014-4](https://doi.org/10.1016/s0951-8339(01)00014-4)
 Akpınar A, Özger M, Kömürçü Mİ (2014) Prediction of wave parameters by using fuzzy inference system and the parametric models along

- the south coasts of the Black Sea. *J Mar Sci Technol* 19(1):1–14. <https://doi.org/10.1007/s00773-013-0226-1>
- Aydoğan B, Ayat B, Yüksel Y (2013) Black Sea wave energy atlas from 13 years hindcasted wave data. *Renew Energy* 57:436–447. <https://doi.org/10.1016/j.renene.2013.01.047>
- Balas CE, Koç ML, Tür R (2010) Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. *Appl Ocean Res* 32:425–433. <https://doi.org/10.1016/j.apor.2010.09.005>
- Buyukyildiz M, Tezel G (2017) Utilization of PSO algorithm in estimation of water level change of Lake Beysehir. *Theor Appl Climatol* 128(1–2):181–191
- Danandeh Mehr A (2018) An improved gene expression programming model for streamflow forecasting in intermittent streams. *J Hydrol* 563:669–678
- Danandeh Mehr A, Nourani V, Kahya E, Hrnjica B, Sattar AM, Yaseen ZM (2018) Genetic programming in water resources engineering: a state-of-the-art review. *J Hydrol* 566:643–667
- Deo MC, Chaudhari G (1998) Tide prediction using neural networks. *Comp Aided Civil Infrastruct Eng* 13(2):113–120. <https://doi.org/10.1111/0885-9507.00091>
- Deo MC, Naidu CS (1998) Real time wave forecasting using neural networks. *Ocean Eng* 16:191–203. [https://doi.org/10.1016/s0029-8018\(97\)10025-7](https://doi.org/10.1016/s0029-8018(97)10025-7)
- Dixit P, Londhe S (2016) Prediction of extreme wave heights using neuro wavelet technique. *Appl Ocean Res* 58:241–252. <https://doi.org/10.1016/j.apor.2016.04.011>
- Dixit P, Londhe S, Dandawate Y (2015) Removing prediction lag in wave height forecasting using neuro-wavelet modeling technique. *Ocean Eng* 93:74–83. <https://doi.org/10.1016/j.oceaneng.2014.10.009>
- Duan WY, Huang LM, Han Y, Huang DT (2016) A hybrid EMD-AR model for nonlinear and non-stationary wave forecasting. *J Zhejiang Univ-Sci A* 17(2):115–129
- Elliott G, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836
- Formentin SM, Zanuttigh B (2019) A genetic programming based formula for wave overtopping by crown walls and bullnoses. *Coast Eng* 152:103529
- Ghorbani MA, Makarynsky O, Shiri J, Makarynska D (2010) Genetic programming for sea level predictions in an island environment. *Int J Ocean Climate Syst* 1(1):27–35
- Hadadpour S, Etemad-Shahidi A, Kamranzad B (2014) Wave energy forecasting using artificial neural networks in the Caspian Sea. In *Proc Instit Civil Eng-Marit Eng* 167(1):42–52. <https://doi.org/10.1680/maen.13.00004>
- Hrnjica B, Danandeh Mehr A (2019) Optimized genetic programming applications: emerging research and opportunities. IGI global, PA
- Kambekar AR, Deo MC (2012) Wave prediction using genetic programming and model trees. *J Coast Res* 28(1):43–50
- Kamranzad B, Etemad-Shahidi A, Kazeminehad MH (2011) Wave height forecasting in Dayyer, the Persian Gulf. *Ocean Eng* 38(1):248–255. <https://doi.org/10.1016/j.oceaneng.2010.10.004>
- Karimi B, Safari MJS, Danandeh Mehr A, Mohammadi MA (2019) Monthly rainfall prediction using ARIMA and gene expression programming: a case study in Urmia, Iran. *Online J Eng Sci Technol* 2(3):8–14
- Kazeminehad MH, Etemad-Shahidi A, Mousavi SJ (2005) Application of fuzzy inference system in the prediction of wave parameters. *Ocean Eng* 32(14–15):1709–1725. <https://doi.org/10.1016/j.oceaneng.2005.02.001>
- Koç ML, Balas CE (2012) Genetic algorithms based logic-driven fuzzy neural networks for stability assessment of rubble-mound breakwaters. *Appl Ocean Res* 37:211–219. <https://doi.org/10.1016/j.apor.2012.04.005>
- Khozani ZS, Safari MJS, Mehr AD, Mohtar WHMW (2020) An ensemble genetic programming approach to develop incipient sediment motion models in rectangular channels. *J Hydrol* 124753
- Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J Econ* 54:159–178
- Law YZ, Santo H, Lim KY, Chan ES (2020) Deterministic wave prediction for unidirectional sea-states in real-time using artificial neural network. *Ocean Eng* 195:106722
- Lee, J. S., & Suh, K. D. (2020). Development of stability formulas for rock armor and tetrapods using multigene genetic programming. *J Waterway Port Coastal Ocean Eng*, 146(1), 04019027
- Londhe SN (2008) Soft computing approach for real-time estimation of missing wave heights. *Ocean Eng* 35:1080–1089. <https://doi.org/10.1016/j.oceaneng.2008.05.003>
- Londhe SN, Panchang V (2007) Correlation of wave data from buoy networks. *Estuar Coast Shelf Sci* 74:481–492. <https://doi.org/10.1016/j.ecss.2007.05.003>
- Mahjoobi J, Etemad-Shahidi A, Kazeminehad MH (2008) Hindcasting of wave parameters using different soft computing methods. *Appl Ocean Res* 30(1):28–36. <https://doi.org/10.1016/j.apor.2008.03.002>
- Makarynsky O, Pires-Silva AA, Makarynska D, Ventura-Soares C (2005) Artificial neural networks in wave predictions at the west coast of Portugal. *Comput Geosci* 31(4):415–424. <https://doi.org/10.1016/j.cageo.2004.10.005>
- Mandal S, Prabakaran N (2006) Ocean wave forecasting using recurrent neural networks. *Ocean Eng* 33(10):1401–1410. <https://doi.org/10.1016/j.oceaneng.2005.08.007>
- Nourani V, Behfar N, Uzelaltinbulat S, Sadikoglu F (2020) Spatiotemporal precipitation modeling by artificial intelligence-based ensemble approach. *Environ Earth Sci* 79(1):6
- Özger M (2011) Prediction of ocean wave energy from meteorological variables by fuzzy logic modeling. *Expert Syst Appl* 38:6269–6274. <https://doi.org/10.1016/j.eswa.2010.11.090>
- Power HE, Gharabaghi B, Bonakdari H, Robertson B, Atkinson AL, Baldock TE (2019) Prediction of wave runup on beaches using gene-expression programming and empirical relationships. *Coast Eng* 144:47–61
- Rahmani-Rezaei A, Mohammadi M, Mehr AD (2019) Ensemble gene expression programming: a new approach for evolution of parsimonious streamflow forecasting model. *Theor Appl Climatol* 139(1–2):549–564
- Sreekanth J, Datta B (2010) Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *J Hydrol* 393(3–4):245–256
- Tsai CC, Wei CC, Hou TH, Hsu TW (2018) Artificial neural network for forecasting wave heights along a ship's route during hurricanes. *J Waterw Port Coast Ocean Eng* 144(2):04017042
- Tsai CP, Lin C, Shen JN (2002) Neural network for wave forecasting among multi-stations. *Ocean Eng* 29(13):1683–1695
- Tehrany MS, Jones S, Shabani F, Martínez-Álvarez F, Bui DT (2019) A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor Appl Climatol* 137(1–2):637–653
- Tür R, Balas CE (2010) Neuro-fuzzy approximation for prediction of significant wave heights: the case of Filyos region. *J Fac Eng Archit Gazi Univ* 25(3):505–510
- Ustoorikar K, Deo MC (2008) Filling up gaps in wave data with genetic programming. *Mar Struct* 21(2–3):177–195
- Vouterakos PA, Moustiris KP, Bartzokas A, Ziomas IC, Nastos PT, Paliatatos AG (2012) Forecasting the discomfort levels within the greater Athens area, Greece using artificial neural networks and multiple criteria analysis. *Theor Appl Climatol* 110:329–343. <https://doi.org/10.1007/s00704-012-0626-x>

- Zamani A, Solomatine D, Azimian A, Heemink A (2008) Learning from data for wind-wave forecasting. *Ocean Eng* 35(10):953–962
- Zanaganeh M, Mousavi SJ, Shahidi AFE (2009) A hybrid genetic algorithm–adaptive network-based fuzzy inference system in prediction of wave parameters. *Eng Appl Artif Intell* 22(8):1194–1202
- Zubier KM (2020) Using an artificial neural network for wave height forecasting in the Red Sea. *Indian J Geo Marine Sci* 49(02):184–191

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.