Application of Soft Computing Tools for Wave Prediction at Specific Locations in the Arabian Sea using Moored Buoy observations

J. Vimala¹, G. Latha² and R. Venkatesan²
¹Research scholar, Sathyabama University, Chennai.
²National Institute Of Ocean Technology, Tamil Nadu, India.
Email: vimalaneru@yahoo.co.in

Abstract

The knowledge of design and operational values of significant wave heights is perhaps the single most important input needed in ocean engineering studies. Conventionally such information is obtained using classical statistical analysis and stochastic methods. As the causative variables are innumerable and underlying physics is too complicated, the results obtained from the numerical models may not always be very satisfactory. Soft computing tools like Artificial Neural Network (ANN) and Adaptive Network based Fuzzy Inference System (ANFIS) may therefore be useful to predict significant wave heights in some situations. The study is aimed at forecasting of significant wave height values in real time over a period of 24hrs at certain locations in Indian seas using the models of ANN and ANFIS. The data for the work were collected by National Institute of Ocean Technology, Chennai. It was found that the predictions of wave heights can be done by both methods with equal efficiency and satisfaction.

Keywords: ANN, ANFIS, Forecasting, Correlation coefficient

1. INTRODUCTION

The veracious predictions of the significant wave height (Hs) are of immense importance in ocean and coastal engineering applications. The significant wave height is calculated as the average of the highest one-third of all of the wave heights during a given sampling period. The phenomenon of generation of ocean waves depends on a number of atmospheric and meteorological factors and hence is a very complex process. Although there exists a number of wave height estimation models, it is known that they do not consider all causative factors at the same time and consequently their results are more or less a general approximation of the overall dynamic behaviour (Røst, 1981; Schwab et al., 1984; Komen et al., 1994; Tsi and Lee, 1999; Tucker and Pitt, 2001; Liu et al., 2002).

To predict a future value of Hs in real time mode a historical sequence of measured significant wave heights is normally considered and a statistical autoregressive method is then employed. When point forecasts at a specified location are needed such time series based models like Auto Regressive (AR), Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA) and soft computing tools can be preferred as these are advantageous when compared with the numerical models due to relatively simple modelling techniques and less requirement of computer memory and time. Wave forecasting using ANN has been developed by Londhe and Deo (2003) at specific locations in Indian seas. This paper makes use of another soft computing tool namely Adaptive Neuro Fuzzy Inference System (ANFIS) to analyse see if both work with equal efficiency.
2. METHODOLOGY

Deep-sea and shallow water moored buoys have been functional in the Indian Ocean seas since 1997 under the National Data Buoy Program (NDBP), National Institute of Ocean Technology (NIOT) of India. In the present study, wave parameters measured by the deep-sea buoys operational in the Arabian Sea at 15°53'20"N/60°18'30"E (DS1) and 10°37'19"N/72°55'29"E (DS2) were used. The data used were significant wave heights for the period 2001-2003.

The environment of MATLAB was used for implementing ANN. In all the cases, trainlm was adopted as the network training function. It is a network training algorithm that updates weight and bias values according to the Levenberg-Marquardt optimization. The traditional training method is the standard back-propagation, although numerous training schemes are available to impart better training with the same set of data, as indicated by Londhe and Desai (2003) in their harbour tranquility studies. All input and output values were normalized within the range -1 to 1. All weights and bias values were initialized to a value 1. The transfer function used was 'logarithmic sigmoid', uniformly for first hidden and output nodes and 'purelin' for the second hidden layer and output layer nodes. The developed model was verified in each case and the details are given in results section. The structure of ANN is given in Figure 1. Details of the network types and learning mathematics can be found in standard books such as Kosko (1992), Wu (1994), Bose and Liang (1998).

![Structure of Artificial Neural Network](image)

Figure 1. Structure of Artificial Neural Network

The structure of ANFIS is given in Figure 2. In the case of ANFIS the first layer is the fuzzifying layer in which Ai and Bi are the linguistic labels. The output of the layer is the membership function of these linguistic labels. The second layer calculates the firing strength for each rule quantifying the extent to which any input data belong to IF-THEN rule. The output of the layer is the algebraic product of the input signals. The third layer is the normalization layer. Every node in this layer calculates the ratio of the i\textsuperscript{th} rule’s firing strength to the sum of the firing strength of all the rules. The fourth layer is the

International Journal of Ocean and Climate Systems
output of every node. The fifth layer computes the overall output as the summation of all the incoming signals, which represents the results of wave height.

![ANFIS Structure](image)

**Figure 2. Structure of ANFIS**

In ANFIS, the premise and consequent parameters are optimized using a hybrid learning algorithm. In this way, a two-step process is used for the learning or adjustment of the network parameters. In the first step, the premise parameters are kept fixed and the information is propagated forward in the network to Layer 4, where the consequent parameters are identified by a least-squares estimator. In the second step, the backward pass, the consequent parameters are held fixed while the error is propagated and the premise parameters are modified using a gradient descent algorithm. The only user specified information is the number of membership functions for each input and the input-output training information.

3. DATA AND LOCATION

The measurements of significant wave height time series considered in this work were made at locations DS1 and DS2 in Arabian Sea shown in Fig. 3. The location details are given in Table 1. The observations in the form of 3 hourly Hs values were used for the period 2001-2003. Out of this, the data of period Jun 2001-Jun 2002 were used for training and the data for the period July 2002-Dec 2003 were used for testing.

<table>
<thead>
<tr>
<th>Buoy ID</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Depth(m)</th>
<th>Distance from Shore(Km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>15.818°N</td>
<td>69.368°E</td>
<td>3800</td>
<td>463.338</td>
</tr>
<tr>
<td>DS2</td>
<td>10.622°N</td>
<td>72.923°E</td>
<td>1850</td>
<td>372.15</td>
</tr>
</tbody>
</table>

Table 1. Latitude and longitude of selected buoy locations
4. RESULTS AND DISCUSSION

The aim of the present study is to predict the significant wave height ahead at the required time step. Predictions of the wave heights at intervals of 3 hours, 6 hours, 12 hours and 24 hours, have been carried out.

The ANFIS model results were compared with that of ANN model. Comparisons between the model network output and the observed values have been carried out using both qualitative as well as quantitative measures. The qualitative assessment is made with the help of time history plots, where deviation from the ideal fit line across the entire range of predictions is immediately seen (Fig.4). A variety of error measures serve the quantitative requirement and they include the correlation coefficient (CC), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) (Table 2).

The results of modeling in terms of comparisons made using scatter plots and time series plots are given in the figure 4, 5, 6 and 7.

The most common feed forward back propagation network was used. The number of input nodes was 5 and there was a single output node. The number of hidden nodes was found by trial and error for each case. All input and output values were normalized within the range of -1 to 1. All weights and bias values were initialized to a value of 1. The transfer function used was logarithmic sigmoid and purelin for hidden and output nodes.

Models were tested for lead times 3, 6, 12, and 24 hours. Different input combinations were tried for significant wave heights. The significant wave height values up to previous 24 hours were taken into consideration as predictor variables. For all the models, test data were selected from the same portion of entire data. It is particularly important to ensure that accurate forecasts with lead times of 3 and 6 hours are obtained.

The correlation coefficient (CC) and mean absolute error (MAE) were used as scores to evaluate the model performance. The scatter and time series plots between the forecasted and observed significant wave heights at DS1 and DS2 locations are given in Figures: 4, 5, 6 and 7.

The results show different levels of performance of each of the model in terms of RMSE and CC. The behavior changed according to the parameter being predicted and the period of forecasting. In Table 2, the values of CC, RMSE and MAE for the corresponding hours are given.

The best ANN models, had a relatively simple architecture. The three layer network and four neurons for a hidden layer were sufficient to produce results for all lead times. The ANN models

International Journal of Ocean and Climate Systems
developed in this study resulted in the correlation coefficient of 0.99 for 3 hour lead time. The mean absolute error for 3 and 24 hour lead times appears to be 0.001 and 0.002 m respectively at location DS1. The correlation coefficient decreases from 0.99 to 0.94 when the lead time increases from 3 hours to 24 hour at DS1. As the prediction interval increases, CC decreases and, RMSE and MAE increase.

The ANFIS models were tested for all the scenarios. It can be seen from Table 2 that CC changes with respect to lead times. The correlation coefficient decreases from 0.99 to 0.96 when the lead time increases from 3 hours to 24 hours at DS2. The RMSE values from ANN are less compared to that of ANFIS for both DS1 and DS2 locations.

Figure 4. Time series plot between Observed value and Neural Network Model output at DS1 and DS2 Location
Figure 5. Scatter plot between Observed value and Fuzzy Logic Model output at DS1 and DS2 Location

Table 2. Verification statistics of Hs simulations with different lead times 3, 6, 12 and 24hrs.

<table>
<thead>
<tr>
<th>Buoy ID</th>
<th>Statistics</th>
<th>03^th hr</th>
<th>06^th hr</th>
<th>12^th hr</th>
<th>24^th hr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANFIS</td>
<td>ANN</td>
<td>ANFIS</td>
<td>ANN</td>
<td>ANFIS</td>
</tr>
<tr>
<td>DS1</td>
<td>Correlation Coefficient</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>RMSE (m)</td>
<td>0.021</td>
<td>0.018</td>
<td>0.032</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>MAE (m)</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>DS2</td>
<td>Correlation Coefficient</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>RMSE (m)</td>
<td>0.011</td>
<td>0.009</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>MAE (m)</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>

International Journal of Ocean and Climate Systems
Figure 6. Time series plot between Observed value and Forecasting Significant wave height at DS1 Location
Figure 7. Time series plot between Observed value and Forecasting Significant wave height at DS2 Location

5. CONCLUSION
The wave measurements made by the wave buoys deployed at two locations along the west coast of India have been analyzed using auto-regressive types of ANN’s and ANFIS. The significant wave heights obtained with the ANFIS were compared with that of ANN. The predictions of Hs were made for lead times of 3, 6, 12 and 24 hr. It was found that both the adopted methods, namely, ANN and ANFIS showed satisfactory performance in this prediction exercise as revealed by high values of CC and low magnitudes of MAE and RMSE, although decrease in the performance to a certain level was noticed with increasing prediction intervals. Incorporation of fuzzy logic in ANN did not yield any special dividend in this work. The choice in between ANN and ANFIS is thus left to the confidence level of the user to use these tools.
6. ACKNOWLEDGEMENTS
The authors thank Ocean Observation Systems team of NIOT for their efforts in deploying and maintaining the buoys. The authors gratefully acknowledge the support and encouragement by Director, NIOT in carrying out this work.

APPENDIX I

Correlation Coefficient (CC):

\[ r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \]

Where \( r = (X - \bar{X}) \), \( y = (Y - \bar{Y}) \), \( X = \) Observed values, \( \bar{X} = \) Mean of \( X \), \( Y = \) Predicted value, \( \bar{Y} = \) Mean of \( Y \). The summation in the above equation as well as in the following two equations is carried out over all \( n \) number of testing patterns.

Mean Absolute Error (MAE)
The Mean Absolute Error is (MAE) the sum of the absolute values of the differences between corresponding forecasted and observed values, divided by the total number of events. The MAE is individually calculated for each forecast category.

\[ MAE = \frac{\sum |X - Y|}{n} \]

Root Mean Square Error (RMSE)
The Root Mean Square Error (RMSE) is the square root of the sum of the squares of the differences between corresponding forecasted and observed values, divided by the total number of events. The RMSE is individually calculated for each forecast category.
REFERENCES


