Interpolation of wave heights

M.C. Deo, N. Kiran Kumar

Department of Civil Engineering, Indian Institute of Technology, Bombay, Mumbai 400 0076, India

Abstract

Remote sensing of waves often necessitates presentation of data in the form of wave height values grouped over large time intervals. This restricts their use to long-term applications only. This paper describes how such data can be made suitable for short-term usage in the field. Weekly mean significant wave heights were derived from their monthly mean observations with the help of different alternative techniques. These include model-free neural network schemes as well as model-based statistical and numerical methods. Superiority of neural networks was noted when the estimations were compared with corresponding observations. The network was trained using three different training algorithms, viz., error back propagation, conjugate gradient and cascade correlation. The technique of cascade correlation took minimum training time and showed better coefficient of correlation between observations and network output.

Keywords: Wave heights; Neural networks; Satellite data; Ocean waves; Disaggregation models

1. Introduction

Knowledge of the heights of ocean waves forms an essential prerequisite for any engineering or transportation-related activity in the ocean. This information can be obtained either by gathering wind data and then using wind–wave relationships or by analyzing the instrumental wave data itself. The later technique is generally more accurate and reliable and involves, commonly, the use of a floating wave rider buoy or a submerged pressure gage or a radar altimeter mounted overboard a satellite.
The satellite-based wave observations have the advantage that they are cheaply available and that they can give information at very large water depths and over longer distances offshore where other ‘in situ’ equipment is not deployable. Satellites release vast amounts of information worldwide and are increasingly regarded as extremely powerful tools for acquiring wave data (World Meteorological Organization, 1988).

Due to the typical sensing process of satellites the corresponding information is generally made available to users in the form of average values over large space and time intervals. Such information then becomes useful for long-term applications like knowing general wave climatology of the area concerned. If the vast database created by the satellite is to be employed for short-term usage, conversion of the long duration wave information into a small-duration one becomes necessary. To the authors’ knowledge there is no technique currently available to do so. This paper discusses use of neural networks for this purpose and evaluates the neural output by comparing it with possible statistical and numerical options.

In March 1985 a satellite named GEOSAT was launched by the USA. After a period of 18 months, it was made to undergo a 17-day repeat cycle or Exact Repeat Mission. The satellite had a radar altimeter onboard which collected a variety of environmental data. This included wave height observations, among other parameters. These were sensed by transmitting a narrow pulse of electromagnetic radiation to the ocean surface and then analyzing the characteristics of the return pulse. Average values of significant wave height, $H_s$, were collected over footprint circles of 5 km radius along the satellite pass with a repeat period of 17.05 days at a resolution of about 150 km at the equator. Sarkar et al. (1990) synthesized such $H_s$ values around the Indian coastline and derived charts showing their monthly mean values over various grids of size $2.5^\circ \times 2.5^\circ$. The validity of these observations against ship-reported and buoy data was also confirmed by Sarkar et al. (1990).

For the present work, GEOSAT data of 32 months, from November 1986 to June 1989, were retrieved to obtain monthly mean $H_s$ values over 12 grids of size $1^\circ \times 1^\circ$ off the west coast of India as shown in Fig. 1. This was done after averaging all observations available over each grid as per the algorithms developed by Kumar et al. (1991). These grids were chosen considering the presence of certain offshore structures in grids 6 and 7.

As stated earlier, the objective of the present studies was to interpolate these monthly mean wave heights so that they become useful for short-term applications like determining clear weather windows to carry out some construction or repair activity. A stepwise application of the interpolation technique can then be made to obtain waves over still shorter intervals.

2. Use of neural networks

The technique of neural networks was first employed to carry out the interpolation. Occurrence of waves is basically a random and unpredictable phenomenon and hence neural networks seem suitable as they provide a non-deterministic and model-free mapping between a given set of input and output values. Built-in dynamism in fore-
casting, data error tolerance and lack of requirements of any exogenous input further make the network suitable for the present application.

2.1. The network and its training

The common three-layered feed forward type of network shown in Fig. 2 was employed. In such a network the input nodes receive the data value (of say monthly mean significant wave height) and pass them on to the hidden layer nodes. These nodes individually sum up the received values after multiplying each input value by a weight, attach a bias to this sum and pass on the result through a non-linearity like a sigmoid transfer function. This forms input to the output layer that operates identically to the hidden layer nodes. The resulting transformed output from each output node forms the network output. If the desired output is weekly mean wave heights, there will be four output nodes.
The training of the network was done using three different algorithms to see which one works better for the application under consideration. These are the error back-propagation, as followed by Rabelo (1990); Yeh et al. (1993), the conjugate gradient, developed by Fletcher and Reeves (1964), and the cascade correlation presented by Karunanithi et al. (1994).

The objective of the training is to reduce the global error, $E$, defined below:

$$E = \frac{1}{P} \sum_{p=1}^{P} E_p$$

(1)

where $P =$ total number of training patterns, $E_p =$ error for $p$th training pattern

$$E_p = \frac{1}{2} \sum_{k=0}^{N} (o_k - t_k)^2$$

(2)

$N =$ total number of output nodes, $o_k =$ network output at the $k$th output node, and $t_k =$ target output at the $k$th output node.

In all training schemes, attempt is made to reduce the global error to minimum.

### 2.2. Error back-propagation algorithm

The scheme of error back-propagation involves minimization of the error between the realized output and the actual one for each training pattern supplied, using steepest descent or gradient descent approach. The network weights and biases are adjusted by moving a small step in the direction of negative gradient of the error function during each iteration. The iterations are repeated until a specified convergence or number of iterations is achieved. Mathematically, the gradient descent is given by

$$\tilde{X}_k + 1 = \tilde{X}_k - n g(\tilde{X}_k + 1 - 1)$$

(3)

where $\tilde{X}_k =$ vector of weights at $k$th iteration index, $\tilde{X}_k + 1 =$ vector of weights at $(k + 1)$th iteration index, $n =$ step size supplied by the user, $g =$ gradient vector

$$= \nabla f(\tilde{X})$$

(4)

and $f(\tilde{X}) =$ error function for a general weight vector $\tilde{X}$.

The above error gradient approach is simple to use. However, it converges slowly and may exhibit oscillatory behaviour due to fixed step size. These difficulties can be removed by adopting a rather complex algorithm of conjugate gradient. Further, the network architecture of the back-propagation is required to be prefixed by trial. If the resulting size of the network is too small, it gives rise to under-learning of the problem. On the contrary, if it is too big, lack of generalization and convergence difficulties can arise. The training algorithm of cascade correlation is directed towards solving these problems.
2.3. Cascade correlation algorithm

The algorithm of cascade correlation starts its training without any hidden node. If the error between the realized output and the targeted one is not low, it adds one hidden node (Fig. 3). This node is connected to all other nodes except the output nodes. Weights associated with hidden nodes are optimized by a gradient ascent method in which the correlation between the hidden unit’s output and the residual error of the network is maximized. If \( S \) is an overall sum of such correlations then

\[
S = \sum_o \left| \sum_p (V_p - \bar{V})(E_{p,o} - \bar{E}_o) \right|
\]

where \( o \) = \( o \)th hidden unit, \( p \) = \( p \)th training pattern, \( V_p \) = average of the hidden unit at pattern \( p \), \( \bar{V} \) = value of all \( V_p \) values over all patterns, \( E_{p,o} \) = residual error for \( o \)th hidden unit at \( p \)th pattern, and \( \bar{E}_o \) = average of all \( E_{p,o} \) over all patterns.

The sum \( S \) is maximized by differentiating it with respect to weights and then equating the result to zero.

When a given hidden unit is trained, weights connecting it to the output layer are kept constant. When a new hidden unit is added, its incoming weights are kept unchanged for the remaining training period, during which weights of all links directly connected to output nodes are updated. The addition of hidden units continues until the desired learning is over.

2.4. Conjugate gradient algorithm

This technique differs from the back-propagation one in gradient calculations and subsequent corrections to weights and bias (Wasserman, 1993). Here, the gradient descent does not move down the error gradient but along a direction which is conju-
gate to the previous step. The change in gradient is taken as orthogonal to the previous step with the advantage that the function minimization carried out in each step is fully preserved due to lack of any interference from subsequent steps. A search direction \( d_k \) is computed at each training iteration, \( k \), and the error function is minimized along it using a line search. The five-step iteration process is as follows (Fitch et al., 1991):

1. Initialize weight vector, \( \bar{X} \), by using uniform random numbers from the interval \((-0.5, 0.5)\). Calculate error gradient \( \bar{g}_o \) at this point. Select initial search direction \( \bar{d}_o = -\bar{g}_o \).

2. Determine constant \( \alpha_k \), for iteration \( k \), that minimizes the error function \( f(\bar{X}_k + \alpha_k \bar{d}_k) \) by line search, \( \bar{d}_k \) is the search direction at iteration \( k \). Update the weight vector \( \bar{X}_k \) to \( \bar{X}_{k+1} \) as below:
   \[
   \bar{X}_{k+1} = \bar{X}_k + \alpha_k \bar{d}_k
   \]  
   \( \text{(6)} \)

3. If the error at this iteration point is acceptable or if specified number of computations of the function and gradients is reached, terminate the algorithm.

4. Otherwise obtain new direction vector \( \bar{d}_{k+1} \).
   \[
   \bar{d}_{k+1} = -\bar{g}_{k+1}
   \]  
   \( \text{(7)} \)

   if \( k + 1 \) is an integral multiple of \( N \), where \( N \) is the dimension of \( X \)

   \[
   \bar{d}_{k+1} = -\bar{g}_{k+1} + \beta_k \bar{d}_k \text{ otherwise}
   \]  
   \( \text{(8)} \)

   where

   \[
   \beta_k = \frac{(\bar{g}_o \bar{g}_o^T)^{-1}}{(\bar{g}_o \bar{g}_o^T)}
   \]  
   \( \text{(9)} \)

5. Go to step 2 for next iteration.

2.5. Training patterns

As stated earlier the objective of the studies was to determine weekly mean significant wave heights from their given monthly mean values. For actual application of the network, the data of monthly mean values only is necessary. However, to train the network, both monthly as well as weekly \( H_s \) values are required. In the present case the weekly mean \( H_s \) values were retrieved from satellite observations only. This was done with much difficulty since very few observations were available within a one-week period for a given grid owing to typical data sensing. In other situations, when even some amount of reliable satellite wave data over smaller intervals are not available, use may be made of wind data obtainable form meteorological organizations. The wind observations can be input in the wind–wave relationships to derive small interval waves for network training. It is, however, recognised that inaccuracy in the small interval data made available for training will affect the network learning accuracy.
The 32 months sample size of individual grid data was found to be too small to obtain the required network output at desired grids 6 and 7 on the basis of grid-wise training. Hence, all available data surrounding grid number 7, viz., those pertaining to grids 2, 3, 4, 8, 12, 11, 10 and 6 were used to train the network so as to obtain the desired output at central grid 7 while data around grids 1, 2, 3, 7, 11, 10, 9 and 5 were used to train the network to yield the output at grid 6. The total number of training patterns was 288. Verification of the network outcome was done with the help of 32 input–output patterns each available at grids 6 and 7. These testing patterns were not involved in the training of the network.

2.6. Verification of the network

The network trained using the three algorithms mentioned earlier was used to get weekly mean wave heights from their monthly mean values. Corresponding results were compared with the actual observations. Figs. 4 and 5 show examples of such verifications where neural computations, made using the typical cascade correlation algorithm, can be seen along with corresponding actual observations in respect of grids 7 and 6, respectively. The same comparison made using scatter plots are shown.
in Figs. 6 and 7 for grids 7 and 6, respectively. It may be seen that the rising and falling trends in the observations appear to have been captured fairly well by the network. A quantitative evaluation of the results was made by deriving the correlation coefficient, $r$, for these two grids, using the following relation:

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

(10)

where $x = X - X'$, $X$ = actual observation of weekly mean $H_s$, $X'$ = average of $X$, $y = Y - Y'$, $Y$ = computed value of weekly mean $H_s$, and $Y'$ = average of $Y$.

In respect of Figs. 6 and 7 the value of $r$ was 0.81 and 0.79, respectively, indicating a fairly high amount of covariation of the computed values with the observed ones.

A comparison of the training and testing results for grids 7 and 6 is given in Tables 1 and 2, respectively, where the number of iterations for training, and the training time (on a PC-486) are given along with the correlation coefficient resulting from network testing. It is seen that the cascade correlation algorithm takes a minimum amount of time for training and gives a relatively higher correlation coefficient. This may be due to the optimum configuration of the network selected by the algorithm. The training was continued until the overall mean square error between the target output and the network output converged to 0.01%. The number of iterations required to achieve this tolerance was the least for the conjugate gradient method. However, each iteration cycle in it took a very long time to complete when compared with back-propagation and cascade correlation schemes.

Since the underlying phenomena are purely random and uncertain a very high

![Fig. 6. Weekly mean wave heights at grid 7 (algorithm: cascade correlation).](image)
Table 1
Training and testing results for grid 7

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Network configuration</th>
<th>Iterations</th>
<th>Time (s)</th>
<th>Time/iteration (s)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-propagation</td>
<td>1-5-4</td>
<td>56 200</td>
<td>10 730</td>
<td>0.191</td>
<td>0.75</td>
</tr>
<tr>
<td>Cascade correlation</td>
<td>1-19-4</td>
<td>4 140</td>
<td>180</td>
<td>0.043</td>
<td>0.79</td>
</tr>
<tr>
<td>Conjugate gradient</td>
<td>1-2-4</td>
<td>50</td>
<td>326</td>
<td>6.520</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note: Under the column ‘Network configuration’, 1-5-4, etc., indicate 1 input node, 4 output nodes and 5 hidden nodes.

Table 2
Training and testing results for grid 6

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Network configuration</th>
<th>Iterations</th>
<th>Time (s)</th>
<th>Time/iteration (s)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-propagation</td>
<td>1-5-4</td>
<td>56 100</td>
<td>8204</td>
<td>0.146</td>
<td>0.76</td>
</tr>
<tr>
<td>Cascade correlation</td>
<td>1-18-4</td>
<td>4 120</td>
<td>180</td>
<td>0.044</td>
<td>0.81</td>
</tr>
<tr>
<td>Conjugate gradient</td>
<td>1-2-4</td>
<td>50</td>
<td>634</td>
<td>12.680</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Note: Under the column ‘Network configuration’, 1-5-4, etc., indicate 1 input node, 4 output nodes and 5 hidden nodes.
accuracy in the weekly wave height estimation cannot be expected. Further, such an accuracy may not be that necessary for taking certain operational decisions, like continuation of offshore drilling or repair work and operations of crew boats, etc., say over a couple of weeks in a given months. It is felt that for such field applications the wave estimations, as reported above, should be useful. Further processing of the yielded wave height information to derive percentage exceedance charts and rose diagrams could also provide a useful input in planning operational works in the ocean. The objective of these studies was restricted to operate on the available maps of monthly mean wave heights and come up with information useful in taking certain short-term operational decisions.

It was decided to see how the neural network performed when compared with other possible options like statistical disaggregation and numerical interpolations.

3. Use of statistical disaggregation models

The technique of disaggregation modelling is traditionally used by hydrologists to disaggregate or break apart a long-duration series, like annual stream flows, into a short-duration one, like monthly stream flows. A stepwise disaggregation is also applied to obtain further short-interval series. For the present study a series of weekly wave heights was derived from their monthly values using one such disaggregation model called condensed disaggregation.

In disaggregation modelling the small interval or the subseries is generated from a given large interval or key series in such a way that certain statistics like probability distributions are maintained individually in both synthetic key series and the generated subseries as well as in between them (Valencia and Schaake, 1973).

A wide variety of disaggregation models are available. This includes both linear and non-linear as well as temporal and spatial. Considering the scope of the present work only those models that depend on the linear dependency structure of the data and applicable at a single site were used. These are the basic, extended and condensed models (Valencia and Schaake, 1973; Mejia and Rouselle, 1976; Lane, 1979). The basic model is:

\[
\{Y\} = [A]\{X\} + [B]\{\epsilon\} \tag{11}
\]

where \(\{Y\}\) is a vector of current values of the small interval subseries being generated, and \(\{X\}\) denotes a vector of given large interval key series. \(\{\epsilon\}\) is a vector of current values from a totally random series while \([A]\) and \([B]\) are parameter matrices. The extended model is:

\[
\{Y\} = [A]\{X\} + [B]\{\epsilon\} + [C]\{Z\} \tag{12}
\]

where \(\{Z\}\) is a vector of all small interval values lying within the given previous large interval and \([C]\) is another parameter matrix. A computationally advantageous version of the extended model involves calculation of the small interval values on a one-at-a-time basis. This forms the condensed model indicated below:
\[ Y_j = A_jX + B_j\epsilon + C_jY_{j-1} \]  
\hspace{1cm} (13)

where subscript \( j \) indicates the current small interval value being generated and \( j - 1 \) denotes the earlier one. The parameter matrices \( A, B \) and \( C \) are evaluated based upon the auto- and cross-covariances of the short and large interval series (Salas et al., 1980).

Application of the above models to the present sets of data showed that the condensed model produces more satisfactory results than basic and extended versions. Fig. 8 shows how the weekly wave height series obtained by using the condensed model compares with the observed one at grid 6. It may be noted that the match between the two series is not as good as the one observed with the corresponding neural network computations shown in Fig. 5. The correlation coefficient in the case of the disaggregation model was also as low as 0.40. Similar discrepancies were noted in the case of grid 7. It thus appears that variations in the weekly wave height are difficult to express using the fixed form of statistical models.

4. Use of numerical interpolation

Attempts were also made to derive the weekly wave height series from its monthly observations using common numerical interpolation methods based upon polynomial fits. These included Lagrangian, Newton’s divided difference, Lagrangian piece-wise and cubic spline. The underlying details can be seen in any standard book of mathematics, e.g., Kreyszig (1985). As expected, such simple methods failed to produce any satisfactory estimates of the weekly wave heights over the range of observations. The variations of the weekly wave heights remained totally unexplained by these methods. Fig. 9 illustrates this with respect to the data at grid 7.

5. Conclusions

The foregoing sections described the application of different alternative techniques to derive information of wave heights over shorter time intervals from corresponding

Fig. 8. Weekly mean wave heights at grid 6 (method: disaggregation model).
long interval data. Both model-free as well as model-based methods were involved. It was found that neural networks, which do not assume any particular type of model, produce better results than the model-based statistical disaggregation or numerical interpolation.

The neural network was trained using three alternative algorithms, viz., error back-propagation, conjugate gradient and cascade correlation. The later two schemes yielded a slightly higher coefficient of correlation between network forecasts and actual observations than the back-propagation scheme. The cascade correlation algorithm took a minimum amount of time to train the data by virtue of its choice of optimum network configuration. The number of iterations required to complete the desired training was minimal in the case of the conjugate gradient method. However, each iteration in it took a very long time when compared with the other two schemes. It thus appears that the network trained using the cascade correlation algorithm can provide the best possible estimates of short interval waves.

Acknowledgements

The authors express their thanks to Mr. Konda Thirumalaiah, Research Scholar, for computational help and to Dr. A. Sarkar of SAC, Indian Space Research Organization, Ahmedabad, India for all help in data retrieval and related technical discussions.

References


