

Artificial neural networks in wave predictions at the west coast of Portugal

O. Makarynskyy^{a,*}, A.A. Pires-Silva^b, D. Makarynska^a, C. Ventura-Soares^c

^a Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia

^b Instituto Superior Técnico, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

^c Instituto Hidrográfico, Rua das Trinas, 49, 1249-093 Lisbon, Portugal

Abstract

In coastal and open ocean human activities, there is an increasing demand for accurate estimates of future sea state. In these activities, predictions of wave heights and periods are of particular importance. In this study, two different neural network strategies were employed to forecast significant wave heights and zero-up-crossing wave periods 3, 6, 12 and 24 h in advance. In the first approach, eight simple separate neural nets were implemented to simulate every wave parameter over each prediction interval. In the second approach, only two networks provided simultaneous forecasts of these wave parameters for the four prediction intervals. Two independent sets of measurements from a directional wave buoy moored off the Portuguese west coast were used to train and to validate the artificial neural nets. Saliency analysis of the results permitted an optimization of the networks' architectures. The optimal learning algorithm for each case was also determined. The short-term forecasts of the wave parameters verified by actual observations demonstrate the suitability of the artificial neural

* Corresponding author. Phone: +61 8 9266 7248. Fax: +61 8 9266 2703.
E-mail address: O.Makarynskyy@exchange.curtin.edu.au (O.Makarynskyy).

technique.

Keywords: Artificial intelligence; Wave simulations; Saliency analysis; Optimization.

1. Introduction

In recent decades, the use of numerical wind wave models has become essential for an adequate evaluation of wind wave characteristics. As a rule, modelling is based on deterministic equations or stochastic time domain approaches. However, these do not entirely account for the complexity and uncertainty of wave phenomena. Assessing the present status of spectral wind wave modelling, Liu et al. (2002) compared four prediction models and concluded that the models performed similarly reflecting the general trends and patterns presented in the observations, although the differences between the results of computations with different wave models themselves were of the same order of magnitude as the discrepancies between model results and observations. Therefore, there is still a requirement for new wave forecasting techniques.

The technique of artificial neural networks (ANNs) is a potential alternative methodology. The inspiration for the 60-year development of ANNs was found in the biological neural system. Though these artificial intelligence information-processing structures are still very primitive compared to biological ones (Fausett, 1994), there is a wide scope of problems being solved thanks to the ANNs' ability to approximate nonlinear behaviour without *a priori* knowledge of interrelations among the elements within a system (Haykin, 1999). Moreover, it has been proven that a feed-forward ANN with an arbitrary number of processing units is a

universal function approximator (Hornik, 1993). For a number of years, ANNs have been successfully used to solve geophysical problems in ocean, coastal and environmental engineering applications. Some examples include assessing the stability of rubble-mound breakwaters (Mase et al., 1995) and studying the storm-built beach profile predictability (Tsai et al., 2000), as well as predictions of the tide (Tsai and Lee, 1999), sea level (Roske, 1997; Makarynskyy et al., 2004), sea currents (Babovic, 1999), daily river stage (Thirumalaiah and Deo, 1998), wave parameters (Agrawal and Deo, 2002; Makarynskyy et al., 2002; Makarynskyy, 2004) and salinity variations in a tidal environment (Huang and Foo, 2002).

In the present paper, two different approaches to neural forecasting of wave parameters are investigated. The first approach consists of forecasting the significant wave height H_s and zero-up-crossing wave period T_{02} over warning times of 3, 6, 12 and 24 hours sequentially. Thus a forecast of each parameter over every warning interval is produced by a separate ANN. In the second approach, only two ANNs are used to concurrently simulate H_s and T_{02} over the aforementioned prediction intervals. The ANNs employed in the “sequential” forecasts were of relatively simple architecture, whereas much more sophisticated nets were developed for the “concurrent” simulations. In both cases, the initial architectures of the ANNs were optimised implementing saliency analysis (Abrahart et al., 2001). An investigation regarding the best training algorithm applicable to this case was also conducted.

This paper commences with an outline of ANNs and brief descriptions of the saliency analysis technique, the employed networks and learning algorithms in Section 2. Section 3 describes the data used. Several sets of both the sequential and concurrent ANN simulations with different numbers of neurons in the input, hidden and output layers are presented in Sections 4 and 5, respectively. These Sections also include comparisons of simulated wave parameters with estimated from measurements as well as some results of saliency analysis implementation. Section 6 presents final remarks.

2. Techniques and learning algorithms

The neuron is the basic concept underpinning ANNs. The neuron takes an argument, which can be formed as a sum of a weighted input and bias, and, by means of a transfer function (typically a step function or a sigmoid function), produces an output. Several neurons can be combined into a neural layer. Furthermore, a network can contain one or more interconnected neural layers. The method of determining the weights and biases is called learning. The learning process requires a set of patterns “input–target output” of proper network behaviour. During the learning process the weights and biases of the network are iteratively adjusted to minimize the network performance function (the averaged squared error between the network outputs and the target outputs is the default case for feed-forward networks) urging the entire network to perform in some expected way. Each representation of a training set to a net is called an epoch.

Saliency analysis is a technique revealing the relative importance of the input and/or the processing components of a network by intentional introduction of missing units (Abrahart et al., 2001). This technique was derived from the idea that an ANN, which is a parallel distributed processing system, should operate sufficiently well even in the case of incomplete input or when one of its internal components does not function properly.

In this study, common three-layer feed-forward ANNs, with a non-linear differentiable log-sigmoid activation function in the hidden layer and linear transfer function in the output layer, were used. The numbers of neurons in the input and output layers were determined according to the imposed tasks and saliency analysis of the obtained results. To avoid the problem of overfitting that may occur while an ANN is being trained, the number of nodes in

the hidden layer, h , of the nets employed in the first stages of the study was computed using one of the empirical expressions mentioned by Huang and Foo (2002)

$$h = 2z + 1, \quad (1)$$

where z is the number of input nodes. A short description of the training algorithms (Hagan et al., 1996) employed in this research follows.

The gradient descent backpropagation learning algorithm updates the weights and biases of artificial neural nets in the direction of the negative gradient of the performance function with a pre-set learning rate. This can cause slow training when the gradient is of a small magnitude.

The resilient backpropagation algorithm eliminates any harmful effects due to the magnitude of the partial derivatives as it accounts for the derivatives' signs. The sign is used to determine the direction of the weight update, which increases whenever the derivative of the performance function has the same sign for the two successive iterations, or decreases whenever the derivative changes the sign. When the derivative is zero, the update value remains the same.

Although the performance function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. The conjugate gradient algorithms perform a search along a conjugate direction to determine the step size that minimizes the performance function along that line. This generally produces faster convergence than steepest descent directions.

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the second derivatives (Hessian matrix) of the performance index at the current values of the weights and biases. Here the Hessian matrix and the gradient are approximated using the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases. In turn, the Jacobian matrix can be computed

through a standard backpropagation technique that is much less complex than computing the Hessian matrix itself.

3. Data employed

The data from a “WAVEC” buoy, which is moored offshore of Sines Harbor (Sines 1D buoy, $37^{\circ} 55' 16''$ N, $8^{\circ} 55' 44''$ W) in water 97 m deep (Fig.1) and is maintained by the

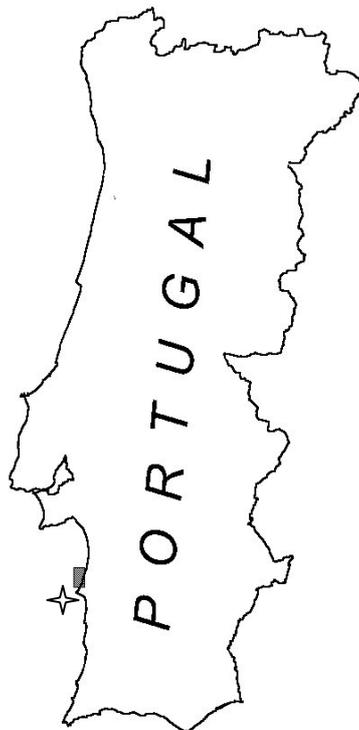


Fig. 1. Location of the Sines 1D buoy station (star) at the Portuguese coast

Hydrographic Institute of the Portuguese Navy, were used to train and test the ANNs employed. The 3-hourly values of the H_s and T_{02} for the period from December 4, 1999, 0000 UTC to January 31, 2000, 2100 UTC were divided into two independent data sets. Data set 1 was used to train the corresponding neural networks, while data set 2 served for validation

purposes. The validation of the ANNs was performed in terms of the root mean square error *RMSE*, correlation coefficient *R* and scatter index *SI*, computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - x_i)^2}{N}}, \quad (2)$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (3)$$

$$SI = \frac{RMSE}{\bar{x}}, \quad (4)$$

where x_i is the value observed at the i -th time step, y_i is the value simulated at the same moment of time, N is the number of time increments, \bar{x} is the mean value of observations, and \bar{y} is the mean value of simulations.

4. Sequential forecasts over four different time intervals

The natural processes determining the variability of the wave parameters in the time domain are of different temporal scales. Therefore, a separate ANN was used to simulate each wave parameter for every warning time. It was assumed that a 24 h wave history represented by 8 measurements contains all the necessary information to simulate the wave parameters one (+3 h), two (+6 h), four (+12 h) and eight (+24 h) time steps ahead. The resilient backpropagation training algorithm was implemented to train all of these nets. The number of training epochs in each simulation was 1000.

The nets used at the beginning with 8 input nodes, 17 processing neurons (see Eq. 1) and 1 output neuron (hereafter referred to as 8x17x1, and likewise for other nets) works as follows.

Each input layer of 8 nodes transfers the initial information to 17 hidden processing units, which further fire the result of simulation to the interested user through the only output neuron.

When short prediction intervals of three and six hours were concerned, the H_s was simulated by this net with high accuracy. The $RMSE$ were less than or equal to 0.20 m, the coefficients of correlation were higher than or equal to 0.94 and the scatter indexes of these forecasts were less than or equal to 0.18 (Table 1). The predictions of the H_s for 12 h were less reliable, while the predictions with leading time of 24 h exhibited neither a reliable correlation pattern nor reasonable values of $RMSE$ and SI .

Table 1. Verification statistics of H_s and T_{02} simulations with different lead times. Here $RMSE$ is root mean square error, R is correlation coefficient, SI is scatter index

Hours	H_s			T_{02}		
	$RMSE(m)$	R	SI	$RMSE(s)$	R	SI
8 x 17 x 1 neurons in the input x hidden x output layers						
+3	0.14	0.97	0.12	0.88	0.89	0.14
+6	0.20	0.94	0.18	1.13	0.81	0.18
+12	0.37	0.82	0.32	1.56	0.64	0.24
+24	0.77	0.40	0.66	1.93	0.41	0.30
4 x 9 x 1 neurons in the input x hidden x output layers						
+3	0.15	0.97	0.13	0.83	0.91	0.13
+6	0.21	0.94	0.18	1.11	0.83	0.17
+12	0.31	0.89	0.26	1.45	0.70	0.22
+24	0.53	0.74	0.45	1.71	0.55	0.26
2 x 5 x 1 neurons in the input x hidden x output layers						
+3	0.15	0.97	0.13	0.82	0.91	0.13
+6	0.21	0.94	0.17	1.11	0.83	0.17
+12	0.34	0.84	0.27	1.39	0.72	0.21
+24	0.53	0.69	0.45	1.69	0.54	0.26

The T_{02} was also simulated reasonably well for three- and six-hour periods of forecasting. The $RMSE$ were around 1s, R were higher than or equal to 0.81 and the SI were less than or

equal to 0.18. The predictions for 12 and 24 h were less successful showing unreliable R although with low values of $RMSE$ and SI (Table 1).

Table 1 also demonstrates an application of saliency analysis to the simulations of H_s and T_{02} . The numbers of neurons in the input and hidden layers were changed twice. The technique of saliency analysis is usually used to assess the role of the units in the input and the hidden layers of ANNs. An analysis of Table 1 shows that the use of half the input nodes (4 units spanning 12 h history) and corresponding to the expression (1) number of hidden neurons improved the statistics of forecast. A further simplification of the ANN architecture (the 2x5x1 net) produced a mixed effect: some statistics slightly improve (for instance, the predictions of T_{02} for six hours), while others degrade (the simulations of H_s with warning time of 12 and 24 h).

To illustrate these considerations, time series plots of the wave parameters estimated from observations and simulated by the best performing ANN (4x9x1) are displayed in Figures 2 and 3. From an analysis of these plots it appears that all rising and falling trends were properly caught by the methodology proposed, especially when the shorter prediction intervals of 3, 6 and 12 hours are concerned. As a result, scattering of the predicted values around the fit line were not significant (Figs. 4-5).

Thus, the implementation of saliency analysis revealed that the performance of the ANNs employed in the simulations of the H_s and T_{02} with leading times of 3, 6, 12 and 24 h can be, to some extent, improved with simplification of the initial architecture of the nets. The reduction of the number of units in the input and hidden layers also resulted in a faster training of the ANNs.

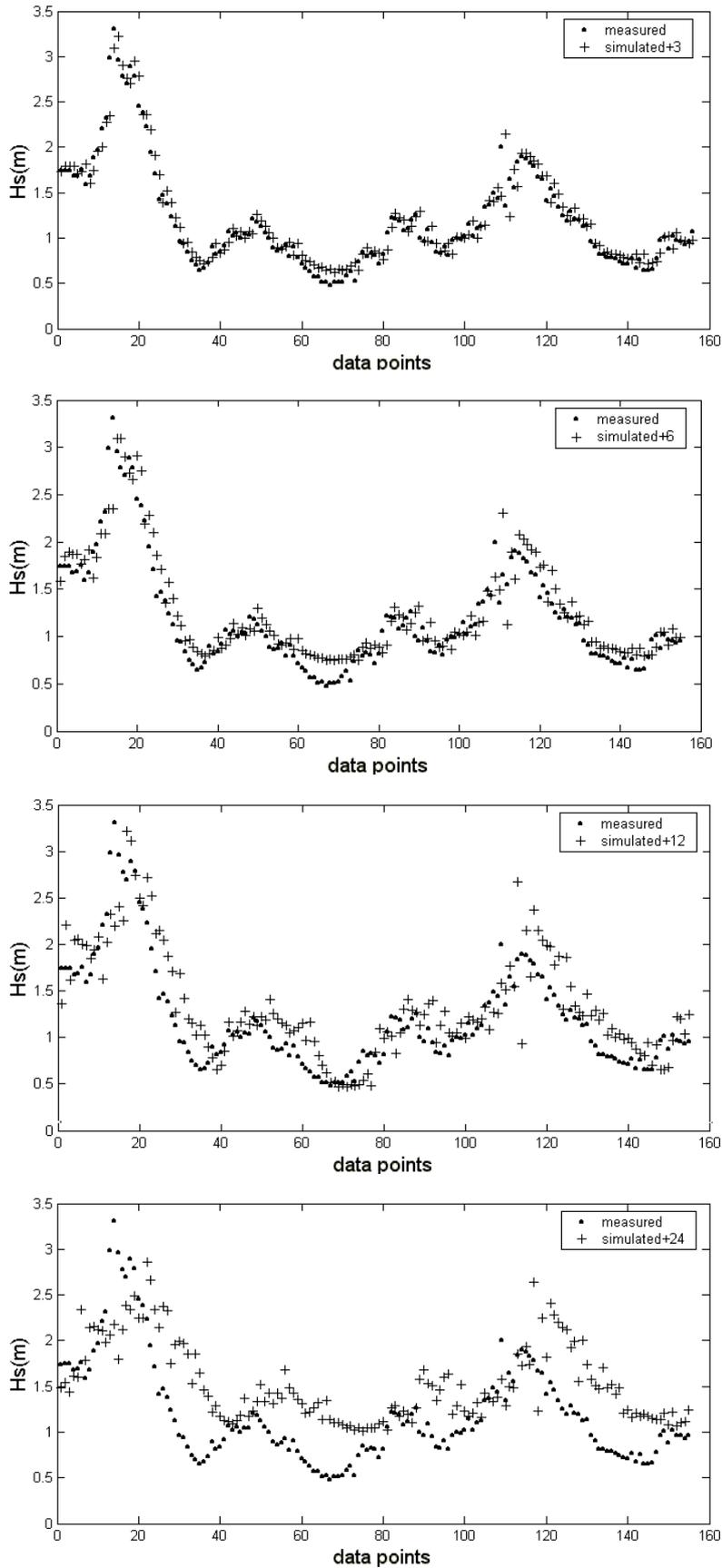


Fig. 2 Time series plots of significant wave heights measured by the Sines 1D buoy and simulated by the separate 4x9x1 ANNs for 4 different warning intervals, January 12 – 31, 2000.

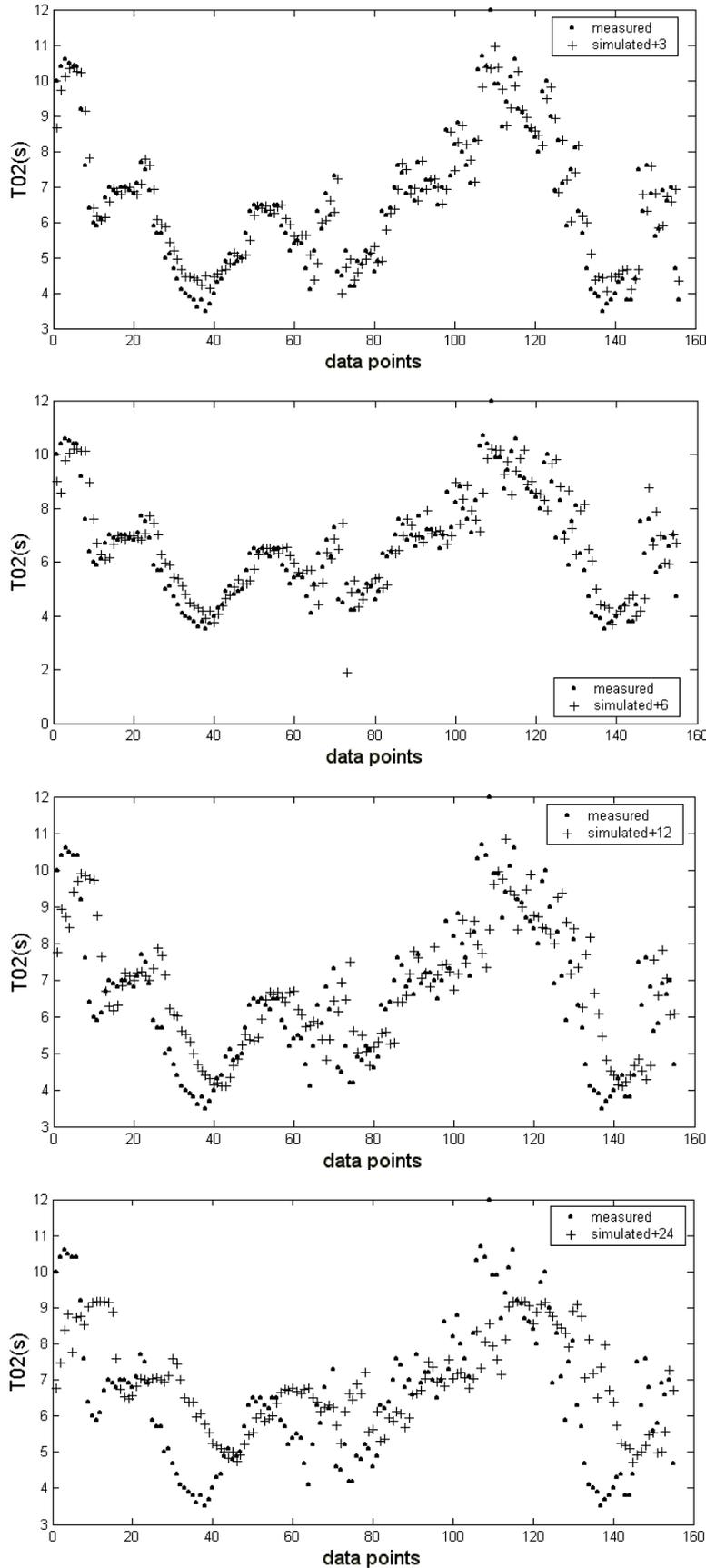


Fig. 3. Time series plots of zero-up-crossing wave periods measured by the Sines 1D buoy and simulated by the separate $4 \times 9 \times 1$ ANNs for 4 different warning intervals, January 12 – 31, 2000.

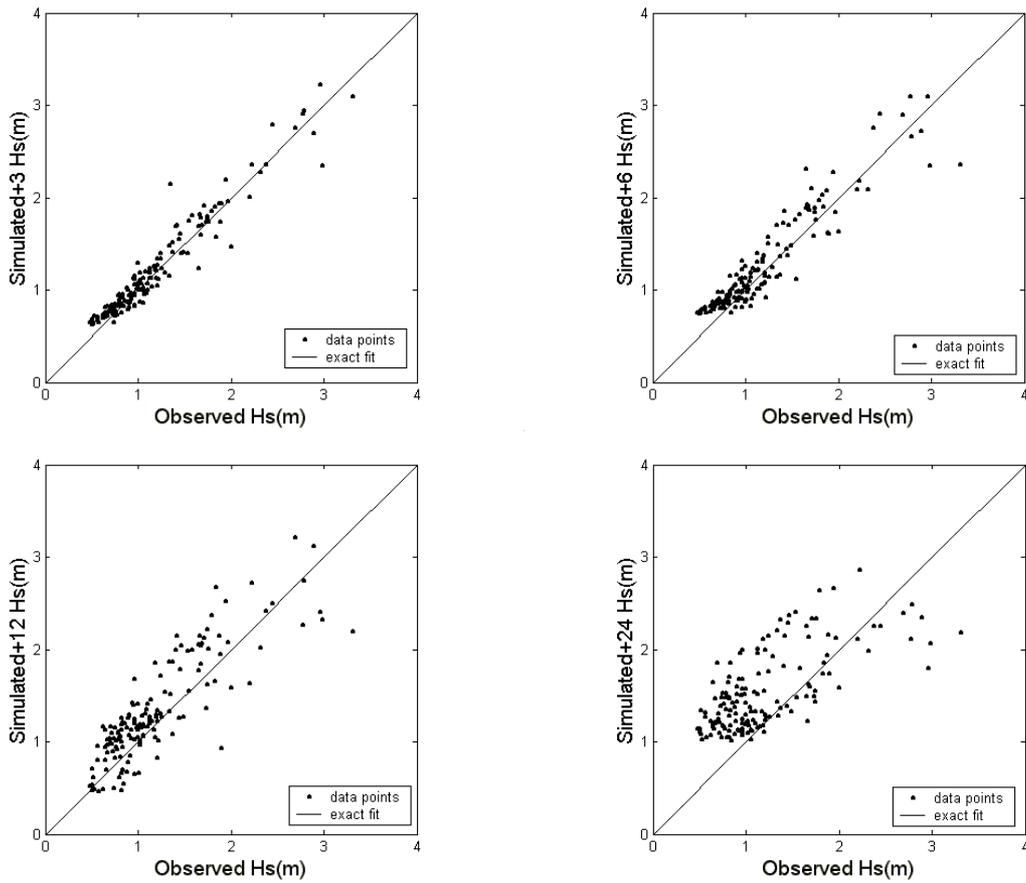


Fig. 4. Scatter diagrams of significant wave heights measured by the Sines 1D buoy versus simulated by the separate 4x9x1 ANNs for 4 different warning intervals, January 12 – 31, 2000

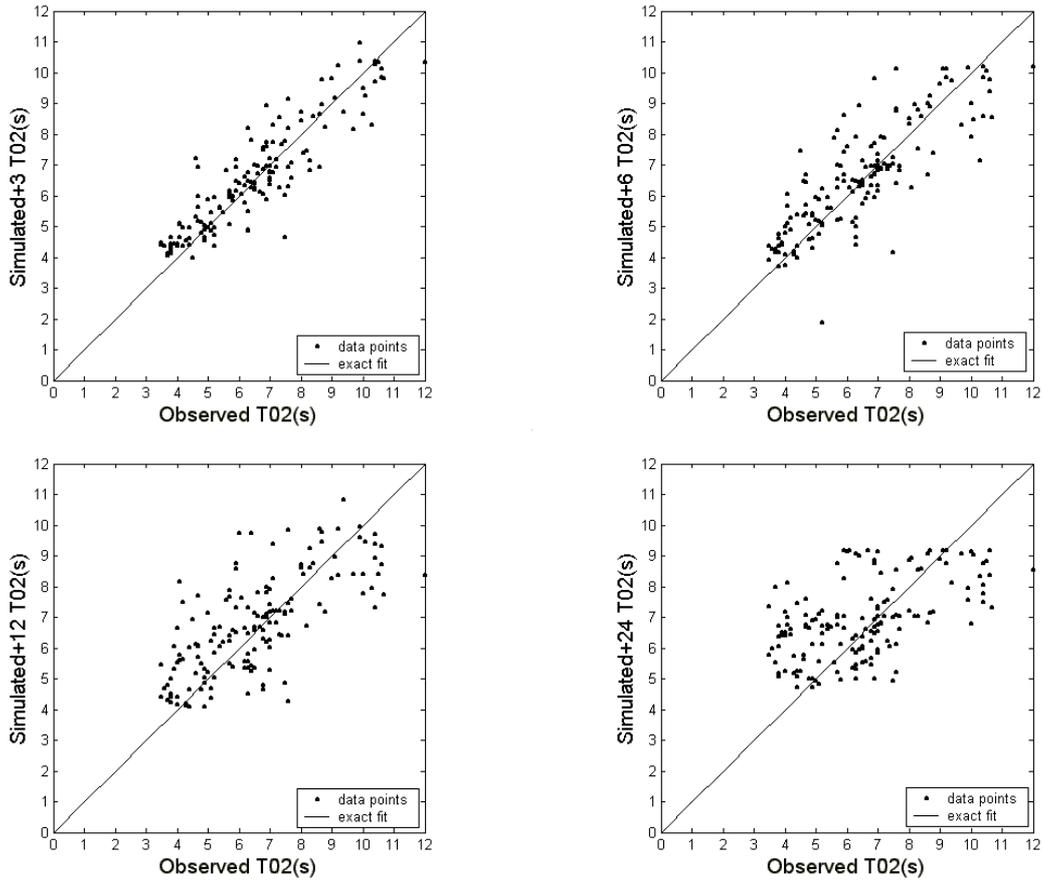


Fig. 5. Scatter diagrams of zero-up-crossing wave periods measured by the Sines 1D buoy versus simulated by the separate 4x9x1 ANNs for 4 different warning intervals, January 12 – 31, 2000.

5. Concurrent forecasts over four time intervals

In the series of experiments described in this section, it was initially assumed that doubling the history (48 h, or 16 measurements) could reveal consistent patterns for concurrent simulations of the wave characteristics with the same lead times as in the previous section. Thus, in this case, 16 input units presented 2-day wave records to 33 hidden processing neurons, which in turn transferred the results of processing to 4 (number of warning intervals) output neurons. Later, for this specific application, an adequate size of these ANNs was determined, again, involving the technique of saliency analysis.

In the first set of the experiments, the ANNs were trained with the basic gradient descent backpropagation algorithm in 3000 epochs. 18 sets of experiments were performed to assess the role of the input and hidden units of the H_s predicting net (Fig. 6). Another independent set of 16 similar simulations was carried out for the T_{02} (Fig. 7). This set had less experiments, because the ones with one neuron in the hidden layer equally as with one neuron in the input layer were skipped due to their lowest performance in the predictions of H_s (see Fig. 6). An analysis of the statistics in Figures 6 and 7 demonstrates that nets with fewer neurons in the hidden layer (16x17x4 for the significant wave height and 16x21x4 for the zero-up-crossing wave period) function equally well as the net 16x33x4 with 33 hidden neurons used initially. Similar results in terms of the $RMSE$, R and SI were obtained with the removal of units from the input layers: the ANNs 4x17x4 in the H_s simulations and 6x21x4 in the T_{02} simulations serve generally better than the other nets taking into consideration all 4 prediction intervals. Therefore, for the data series used, the technique of saliency analysis allowed a significant simplification of the ANNs' initial architecture without a loss in accuracy while decreasing the computational efforts needed.

An additional set of experiments with these nets of optimized architecture (4x17x4 for H_s

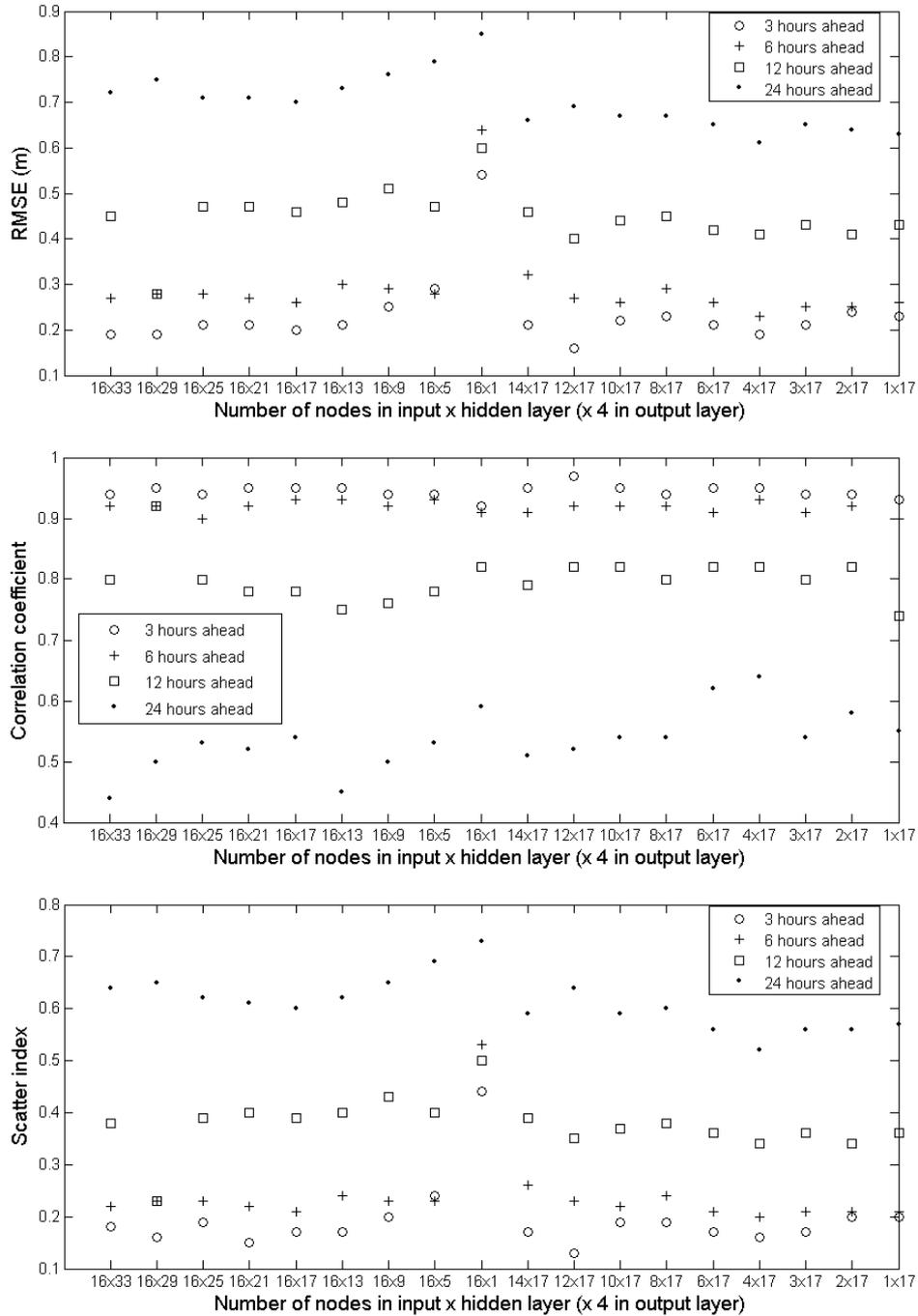


Fig. 6. The root mean square error, correlation coefficient and scatter index of the significant wave height simulations for 4 different leading times with indicated number of units in input x hidden x output layers.

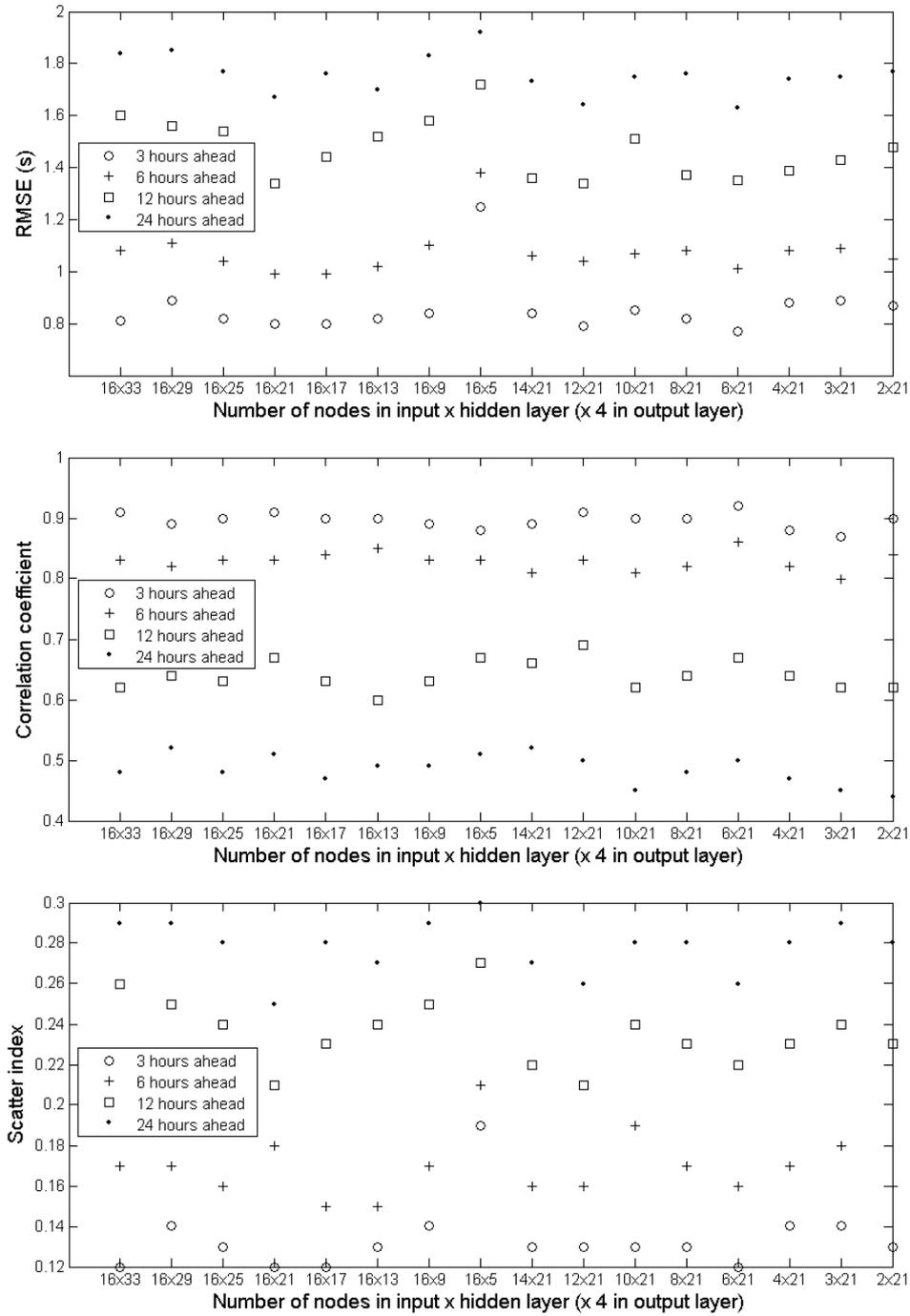


Fig. 7. The root mean square error, correlation coefficient and scatter index of the zero-up-crossing wave period simulations for 4 different leading times with indicated number of units in input x hidden x output layers.

and 6x21x4 for T_{02}) was carried out (Table 2). This was done in order to select a better training algorithm by comparing the verification statistics of the simulations using the resilient backpropagation, Fletcher-Reeves conjugate gradient and Levenberg-Marquardt algorithms with the previously employed gradient descent backpropagation algorithm. An analysis of the presented statistics reveals that, for the case under consideration, the former algorithms do not exhibit any essential quantitative differences from, or advantage over, the gradient descent backpropagation algorithm.

Table 2. Verification statistics of the significant wave height (6 x 17 x 4 net) and the zero-up-crossing wave period (12 x 21 x 4 net) simulations with different training algorithms. 3000 epochs. Here $RMSE$ is the root mean square error, R is the correlation coefficient, SI is the scatter index

Hours	H_s			T_{02}		
	$RMSE(m)$	R	SI	$RMSE(s)$	R	SI
Resilient backpropagation						
+3	0.20	0.95	0.16	1.03	0.87	0.16
+6	0.23	0.93	0.19	1.30	0.79	0.21
+12	0.66	0.64	0.56	2.01	0.49	0.33
+24	0.91	0.38	0.78	2.33	0.23	0.37
Conjugate gradient						
+3	0.17	0.96	0.14	0.89	0.89	0.14
+6	0.23	0.93	0.19	1.30	0.75	0.21
+12	0.46	0.79	0.37	2.16	0.35	0.34
+24	0.75	0.53	0.65	2.31	0.16	0.37
Levenberg-Marquardt						
+3	0.21	0.95	0.17	1.08	0.85	0.17
+6	0.29	0.90	0.24	1.50	0.73	0.24
+12	0.59	0.72	0.50	2.84	0.31	0.45
+24	0.91	0.50	0.78	3.51	0.17	0.55

6. Concluding remarks

The technique of artificial neural networks was used to predict significant wave heights and zero-up-crossing wave periods with warning times of 3, 6, 12 and 24 h sequentially and concurrently. The simulations were compared to time series of these wave parameters estimated off the west coast of Portugal.

The neural forecasts generally follow the rising and falling trends present in the observations. Different levels of accuracy in terms of the root mean square error, correlation coefficient and scatter index were achieved. The ANNs' performance changed according to the lead-time and the parameter being predicted. The wave parameters were better simulated for shorter lead times (3 and 6 h) than for longer prediction intervals (12 and 24 h). A higher accuracy was also noticed of the ANNs simulating the significant wave height than the ones forecasting zero-up-crossing wave period. The results obtained in simultaneous forecasts exhibited less accuracy than the predictions produced separately due to large numbers of input-processing-output nodes and thus more complicated interrelations among them.

An implementation of the saliency analysis technique allowed both an estimation of the relative importance of the input and processing units of the ANNs and simplification of the net's architectures. To some extent, the simplifications increased the accuracy of the simulations.

For the case of concurrent wave forecasts, the testing of three "faster" learning algorithms, namely, resilient backpropagation, Fletcher-Reeves conjugate gradient and Levenberg-Marquardt, did not exhibit any significant improvement of the statistics compared to the results obtained with the gradient descent backpropagation algorithm. However, some of these algorithms provide the same or similar accuracy in fewer training epochs requiring less computation.

The case studies presented clearly demonstrate that the proposed neural methodology could be successfully applied to short-term wave predictions. It also implies that, before employing it in other geographic areas, artificial neural networks must be trained on site-specific data sets.

Acknowledgments

The authors are grateful to the Hydrographic Institute of the Portuguese Navy for making the buoy data available. This work was part of the research project “Nearshore Wind Wave Prediction: Data Assimilation and Spectral Models” (PDCTM/MAR/15242/99), when Oleg Makarynsky was a post-doctoral fellow and Dina Makarynska was a doctoral fellow, all funded by the “Fundação para a Ciência e Tecnologia” of the Ministry of Science and Technology of Portugal.

References

- Abrahart, R.J., See, L., Kneal, P.E., 2001. Investigating the role of saliency analysis with neural network rainfall-runoff model. *Computers & Geosciences*, 27, 921-928, doi:10.1016/S0098-3004(00)00131-X.
- Agrawal, J.D., Deo, M.C., 2002. On-line wave prediction. *Marine Structures*, 15, 57-74, doi:10.1016/S0951-8339(01)00014-4.

- Babovic, V., 1999. Subsymbolic process description and forecasting using neural networks. In: Garcia-Navarro, P., Playan, E. (Eds.), Proceedings of the International Workshop: Numerical Modelling of Hydrodynamic Systems, 57-79.
- Fausett, L., 1994. Fundamentals of Neural Networks. Architectures, Algorithms, and Applications, Prentice-Hall, 462pp.
- Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing, Boston, MA, 628pp.
- Haykin, S., 1999. Neural Networks: a Comprehensive Foundation. Prentice-Hall, 842pp.
- Hornik, K., 1993. Some new results on neural network approximation. Neural Networks, 6, 1069-1072.
- Huang, W., Foo, S., 2002. Neural Network Modelling of Salinity Variation in Apalachicola River. Water Research, 36, 356-362, doi:10.1016/S0043-1354(01)00195-6.
- Liu, P.C., Schwab, D.J., Jensen, R.E., 2002. Has wind-wave modelling reached its limit? Ocean Engineering, 29, 81-98, doi:10.1016/S0029-8018(00)00074-3.
- Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean Engineering, 31 (5-6), 709-724, doi:10.1016/j.oceaneng.2003.05.003
- Makarynskyy, O., Makarynska, D., Kuhn, M., Featherstone, W.E., 2004. Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia. Estuarine, Coastal and Shelf Science, 61 (2), 351-360, doi:10.1016/j.ecss.2004.06.004.
- Makarynskyy, O., Pires-Silva, A.A., Makarynska, D., Ventura-Soares, C., 2002. Artificial neural networks in the forecasting of wave parameters. The 7th International Workshop on Wave Hindcasting and Forecasting, Banff, Alberta, Canada, October 21-25, pp. 514-522.
- Mase, H., Sakamoto, M., Sakai, T., 1995. Neural network for stability analysis of rubble-mound breakwaters. Journal Waterway, Port, Coastal and Ocean Engineering, ASCE, 121 (6), 294-299.

- Roske, F., 1997. Sea level forecasts using neural networks. *German Journal of Hydrography*, 49 (1), 71-99.
- Thirumalaiah, K., Deo, M.C., 1998. River stage forecasting using artificial neural networks. *Journal of Hydrologic Engineering*, ASCE, 3 (1), 26-32.
- Tsai, C.P., Lee, T.L., 1999. Back-propagation neural network in tidal-level forecasting. *Journal Waterway, Port, Coastal and Ocean Engineering*, ASCE, 125 (4), 195-202.
- Tsai, C.P., Hsu, J.R.-C., Pan, K.L., 2000. Prediction of storm-built beach profile parameters using neural network. *Proceedings of the 27th International Conference of Coastal Engineering*, ASCE, V.4, 3048-3061.