NEURAL NETWORK FORECASTING OF STORM SURGES ALONG THE GULF OF MEXICO

Philippe E. Tissot[1], Daniel T. Cox[2], M.ASCE, and Patrick Michaud[3]

Abstract: Accurate water level forecasts are of vital importance along the Gulf of Mexico as its waterways play a critical economic role for a number of industries including shipping, oil and gas, tourism, and fisheries. While astronomical forcing (tides) is well tabulated, water level changes along the Gulf Coast are frequently dominated by meteorological factors. Their impact is often larger than the tidal range itself and unaccounted for in present forecasts. We have taken advantage of the increasing availability of real time data for the Texas Gulf Coast and have developed neural network models to forecast future water levels. The selected inputs to the model include water levels, wind stress, barometric pressures as well as tidal forecasts and wind forecasts. A very simple neural network structure is found to be optimal for this problem. The performance of the model is computed for forecasting times between 1 and 48 hours and compared with the tide tables. The model is alternatively trained and tested using three-month data sets from the 1997, 1998 and 1999 records of the Pleasure Pier Station located on Galveston Island near Houston, Texas. Models including wind forecasts outperform other models and are considerably more accurate than the tide tables for the forecasting time range tested, demonstrating the viability of neural network based models for the forecasting of water levels along the Gulf Coast.

INTRODUCTION

Accurate water level forecasts along the Gulf of Mexico coast, estuaries, and intracoastal waterways are of great importance to federal, state, and local agencies, industries such as ports, fisheries, construction and coastal communities. The overall economic importance of the Gulf of Mexico for the U.S. economy is high: nine out of the twelve largest U.S. ports with tonnage greater than 50 million tons are located along

1 Assistant Professor of Physics, Physical and Life Sciences Department, and Research Faculty, Conrad Blucher Institute, Texas A&M University-Corpus Christi, Corpus Christi, Texas 78411, ptissot@cbi.tamucc.edu

2 Associate Professor, Coastal and Ocean Engineering Division, Civil Engineering Department, Texas A&M University, College Station, Texas 77843-3136, dtc@tamu.edu

3 Professor of Computer Sciences, Computing and Mathematical Sciences Department, and Director, Division of Near Shore Research, Conrad Blucher Institute, Texas A&M University-Corpus Christi, Corpus Christi, Texas 78411, pmichaud@cbi.tamucc.edu

the Gulf coast and account for 52% of the U.S. tonnage (NOAA, 1999). For ports and waterways along the Atlantic and Pacific coasts water level forecasts are obtained by consulting the tide tables. In the Gulf of Mexico, water level changes are often dominated by meteorological factors the impact of which are often larger than the tidal range itself and unaccounted for in present forecasts (Cox et al., 2002). A one-month comparison between measured water levels and tidal forecasts is presented in figure 1 for the Pleasure Pier tide station located on Galveston Island near Houston, Texas, for the spring of 1997. As can be observed the difference between tidal forecasts and actual water levels can be larger than one foot for several consecutive days corresponding to the passage of frontal systems. The passage of frontal systems does in fact represent one of the major forcing for water levels and one of the main reasons for the inadequacy of the tide tables along the Gulf Coast. The passage of frontal systems across Texas and the Gulf of Mexico takes place approximately with a weekly frequency between early October and late May. In 1994 the National Oceanic and Atmospheric Administration (NOAA) conducted a current assessment program in Aransas Pass, Texas, and Corpus Christi Bay, Texas, and indicated that for typical weather conditions and for current predictions (a closely related parameter) the "presently published predictions do not meet working standards" (NOAA, 1991; NOAA 1994). Differences between observed and predicted current velocities were up to 100%, and wind was identified as the main cause with density variations, morphology and fresh water runoff playing a secondary role. To remedy the problem, both reports recommended forecasting based on real-time data including wind data. As present models are based on harmonic analysis they are fundamentally unable to account for these aperiodic forcing functions and new modeling techniques relying on real-time data must be introduced.

Sophisticated models based on finite elements and finite differences are ideal to understand the physical processes of coastal and estuarine dynamics and for simulating storm surges during hurricanes events. These models provide highly accurate solutions

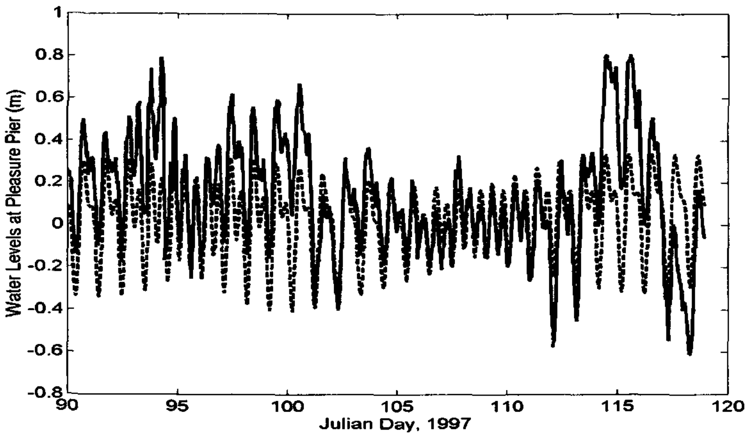

Fig. 1. Comparison between measured water levels (-) and tidal forecasts (...) at Galveston Pleasure Pier during the spring of 1997

to the governing equations of motion when the boundary conditions and time histories of the forcing functions are well prescribed. Their ability as forecasting tools, however, is limited to the accuracy of the forecasted time histories of the forcing conditions such as wind stress. On the other hand, empirical models such as those constructed using neural network techniques can yield accurate forecasts of water levels by incorporating data for established weather patterns. The prediction of water level by a neural network model at an entrance to a harbor channel, for example, can be used either directly as an aid to navigation or indirectly as a seaward boundary for a finite element model of the detailed flow in the harbor channel. In the second case, the neural network model provides two advantages: it reduces the requirements of having a large computation grid outside the entrance channel, and it can easily incorporate information from land-based stations (e.g., information about frontal passages) which may be difficult to incorporate into a finite element model discretized over only the water. The purpose of this paper is to show that a relatively simple neural network model can be constructed and trained using a fairly modest data set (three months) to provide accurate forecasts of water levels on a time scale of 1 to 48 hours. Direct comparisons or integration with finite element models is beyond the scope of the present work and will be pursued in the future.

The modeling philosophy applied in this work is to include and scale data streams such as observational data and forecasts to account for the main forcing of a problem and train a neural network to establish relationships between forcing functions and future water level changes. Of course, large amounts of real-time observational data are required to apply this modeling technique. Over the past ten years Texas has seen a dramatic increase in the availability of real-time observational data along the Gulf coast including parameters such as water levels, wind speeds, wind directions, barometric pressures, water temperatures and air temperatures. The Texas Coastal Ocean Observation Network (TCOON) is one of the main sources of such data and consists of 60 platforms from Brownsville to the Louisiana border (Michaud et al., 2001). TCOON station locations are illustrated in figure 2. The location of the Pleasure Pier tide station located on Galveston Island near Houston is highlighted as data from this station is used to test the model. Increases in the performance and decreases in the cost of sensors, telecommunication and overall information processing equipment should continue for the foreseeable future and make the deployment of data intensive models possible for most coastlines. The present work takes advantage of the real-time data available through the TCOON network and the modeling capabilities of neural networks to predict water levels in real-time and alleviates the present limitations of the tide tables.

NEURAL NETWORK MODELING OF WATER LEVEL CHANGES

The concept of neural networks emerged in the sixties as scientists aimed at emulating the functioning of the brain. The main advantages and key characteristics of neural networks for water level forecasting are their non-linear modeling capability, their generic modeling capacity, their robustness to noisy data, and their ability to deal with high dimensional data (Rumelhart et al., 1995). At the heart of a neural network is the assignment of judicious weights and biases to the elemental neurons of the network. This learning process must be based on a large set of prerecorded observations such as the TCOON database. Rumelheart et al. (1986) developed a type of learning algorithm to assign such a set of optimum weights and biases called backpropagation.

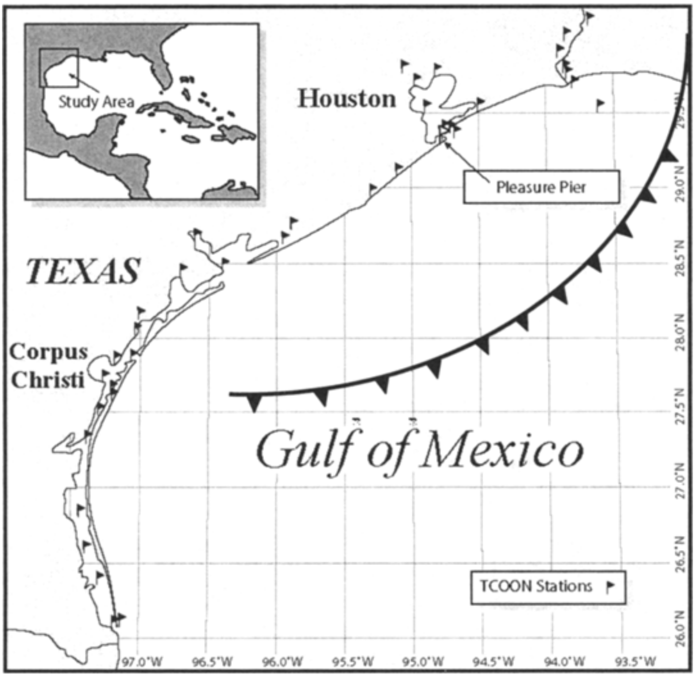

Fig. 2. Gulf of Mexico with the locations of the TCOON stations, the Pleasure Pier Station on Galveston Island and schematic of a typical frontal boundary

Backpropagation neural networks use the repeated comparison between the output of a neural network for a given input and an associated set of target vectors and optimize the neuronal weights and biases by backpropagating a function of this error through the network. The use of error backpropagation has been a key for the application of neural network to a growing number of practical cases including environmental, financial, and engineering problems (Swingler, 1996, Zirili, 1997). During the past five to ten years, neural networks have also been applied successfully to a growing number of coastal and riverine cases such as the forecasting of physical or water quality parameters (Mase et al., 1995, Recknagel et al., 1997, Mase and Kitano, 1999, Moatar et al., 1999, Tsai and Lee, 1999). Neural networks are increasingly tested for the forecasting of flooding along rivers (Campolo et al., 1997, Kim and Barros, 2001) a related application. The application of neural networks to water level forecasting consists in designing and training a network that, given a time series of water levels weather observations and forecasts (wind and tidal forecasts), accurately predicts the next water levels for a period of one, six, twelve, twenty four hours or more. The typical structure of the neural networks used in this work is illustrated in figure 3 and is relatively simple with one hidden layer and one output layer. Neural networks with additional hidden layers were also tested but did not improve on the performance of the two layer models. The elements of the input decks are chosen to track the variation of the main forcing

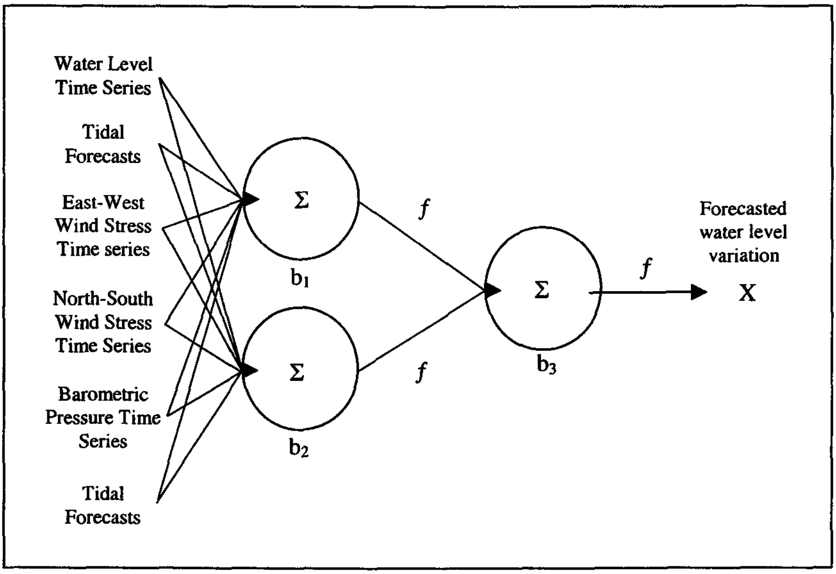

Fig. 3. Schematic of the type of neural network model applied to the problem of water level forecasting including outputs, inputs, and neural network

functions to the problem. They consist of time series of previous water levels, barometric pressures, wind speeds, and wind directions. Also included as part of the input decks are tidal forecasts computed using Xtide 2 (Hopper, 2000). The tidal forcing is included in the model by using water level differences between measured and forecasted water levels and the water levels predicted by the tide tables. The changes in the resulting water level differences are then a direct function of the meteorological forcing. Finally the model forecasts changes in water level differences rather than absolute water level differences. Focusing the model on changes in water level differences allows a more direct relationship between short-term forcing and changes in water levels. Also this allows inclusion of long-term effects such as steric effects as part of the input to each short-term forecasts. The models were tested with and without wind forecasts. When tested with wind forecasts the models initially included exact wind predictions with actual future measurements used as the forecasts. The influence of the accuracy of the forecasts was then evaluated by adding an error to the wind forecasts proportional to the wind and forecasting time. A discussion of the influence of the wind forecasts on the model accuracy as well as discussions on the characteristics of the optimum input deck and neural network structure are included as part of the results and discussion section. 'Tansig' transfer functions are used for the hidden and output layers while the input decks are scaled to a [-1,1] range. The neural network models are trained using a backpropagation algorithm and all computations are performed within the MATLAB 5.3/version 3 of the Neural Network Toolbox (The Math Works Inc., 1998) computational environment running on a Pentium III PC.

RESULTS AND DISCUSSION

The model was tested for the Pleasure Pier tide station on the Gulf Coast side of Galveston Island near Houston, Texas, (http://dnr.cbi.tamucc.edu/overview/022). The station is located near the ship channel leading to the port of Houston--one of the largest ports in United States. To test and optimize the models, the neural networks were successively trained over three data sets, spring 1997, spring 1998, and spring 1999. The approximately three-month data sets were chosen as they are representative of the type of conditions that can be encountered during frontal passages. The model was trained over one of the data sets and then applied to the two other data sets to assess the model performance. Figure 4 compares the performance of the model when trained over the 1999 data set and applied to make 24-hour forecasts for the 1997 data set. As can be observed the neural network outperforms significantly the tide tables during the frontal passages when wind forcing becomes the primary forcing driving water level changes. Figure 5 compares the performance of the model trained during 1997 and applied to the 1999 data set. The 24-hour water level forecasts displayed in figure 5 improve considerably on the tide tables matching closely the measured water levels and demonstrate the ability of the neural network to predict water levels during frontal passages. The input deck of the neural network used for the examples in figures 4 and 5 consists of 5-hour time series of previous water levels, tidal forecasts, wind speeds and wind directions, 20-hour times series of barometric pressures and an exact wind forecast for the time of forecast, i.e. a 24-hour wind forecast. The structure of the neural network is very simple with only one neuron in both the hidden and output layers. The optimization of the model is discussed later in this section. The fact that a very simple neural network is capable of making relatively accurate water level predictions

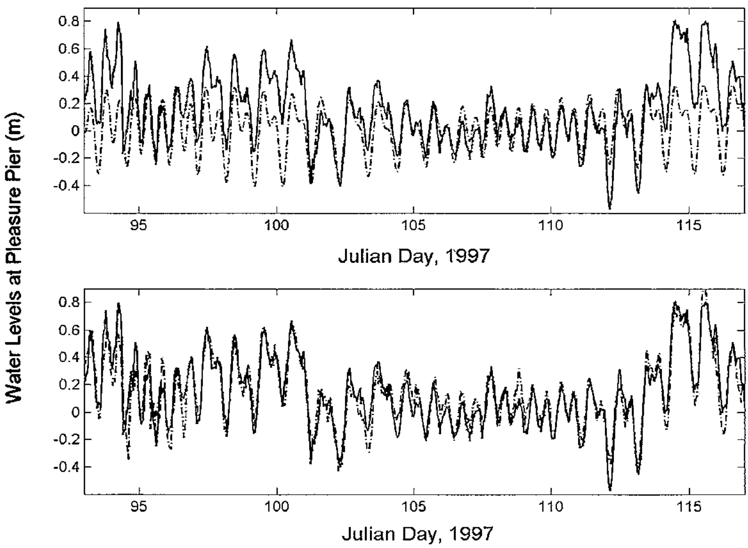

Fig. 4. Comparison of water levels measured at Galveston's Pleasure (-) with tide tables (top, ...) and neural-network forecasts (bottom chart, ...)

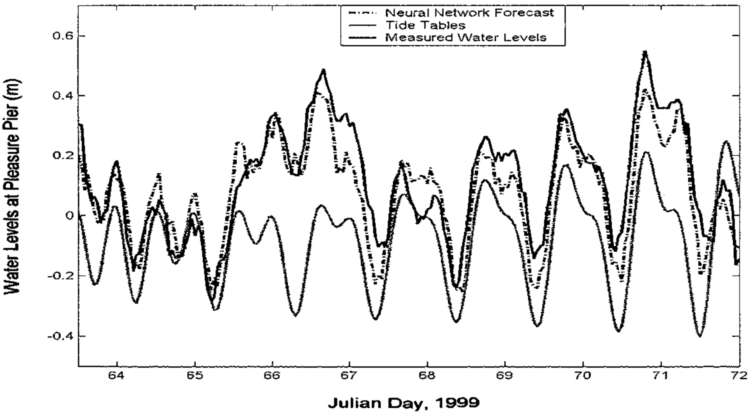

Fig. 5. Performance of the neural network model when trained over the 1997 data set and applied to a strong frontal passage during the spring of 1999

should not be surprising as Cox et al. (2002) have shown that a simple linear model based on the wind stress and a time delay could already lead to significant improvements of water level forecasts as compared to the tide tables.

To compare different versions of the model quantitatively over the complete data sets, models were evaluated by computing an error index including all forecasted water levels. The error index is detailed below and is the relative standard deviation of the water level differences or the ratio of the root mean square of the differences between forecasted and measured water levels and the root mean square of the measured water levels (Cox et al., 2002). The error index is zero for a perfect forecast and approaches one when the accuracy of the forecasts is of the same order as the water level variation.

$$E = \frac{\left[\frac{1}{N}\sum_{i=1}^{N}(H_i - X_i)^2\right]^{1/2}}{\left[\frac{1}{N}\sum_{i=1}^{N}(H_i)^2\right]^{1/2}} \qquad (1)$$

where N = number of water level measurements and forecasts; H_i = observed water levels (m); and X_i = forecasted water levels (m). To assess the model performance over the three data sets the models are first trained over one of the data sets and then applied to the two other data sets. The operation is repeated by rotating the training data sets and the data sets over which the models are tested. An average model performance is then computed by averaging the performance of the models for the six resulting time series of forecasted water levels. The performance variability of the models is estimated by computing the standard deviation over the six different cases.

The error index was used to optimize the network structure and the input deck. The network parameters were optimized for 24-hour forecasts and verified by comparing with the optimum parameters obtained for 9-hour forecasts as well. There were no substantial differences between the optimized neural network parameters optimized for 9 and 24-hour forecasts. The extent of the past wind speeds and wind directions time series was varied between 0 and 30 hours. The optimal time series length was found to be 5 hours with relatively small changes in performance when increasing or decreasing the length of the time series in the 1 to 20-hour range. Input time series longer than 20 hours led to decrease in performance. It should be noted that including longer time series in the input deck does not necessarily mean that the additional input data will be taken into account by the model. As the weights of the neural network are optimized during the training procedure the weights of the additional inputs can be zero. A more in-depth study of the variation of neural network weights and biases when optimizing the model is ongoing. Although the neural network should optimize itself and not take into account unnecessary inputs, larger input decks will affect the training times and possibly the final weights and biases. The length of the input deck was therefore chosen as the smallest time series leading to the best model performance. The optimum past water level measurements and tidal levels time series was 5 hours while the optimum barometric pressure time series length was 20 hours. The barometric pressure time series had a relatively small impact on the model performance on the order of 5%. This is likely due to the fact that winds are in large part a result of the pressure differences across the frontal boundaries and that therefore the barometric pressure effect is in large part already included through the wind inputs. As will be discussed later in this section, the addition of a wind prediction for the time of the forecast has a significant effect on the performance of the model for prediction times longer than 6 hours. The optimized model for this study includes 1 hidden and 1 output neuron, 5 hour time series of previous wind speeds, wind directions, water level measurements and tidal predictions, 20 hour time series of barometric pressures and a wind forecast for the time of the forecast. The performance of the model is displayed for a range of 1 to 48 hours forecasts in figure 6. As can be observed, the model outperforms significantly the tide tables indicating that neural network models can indeed factor in meteorological forcing and lead to more accurate water level forecasts along the Gulf Coast.

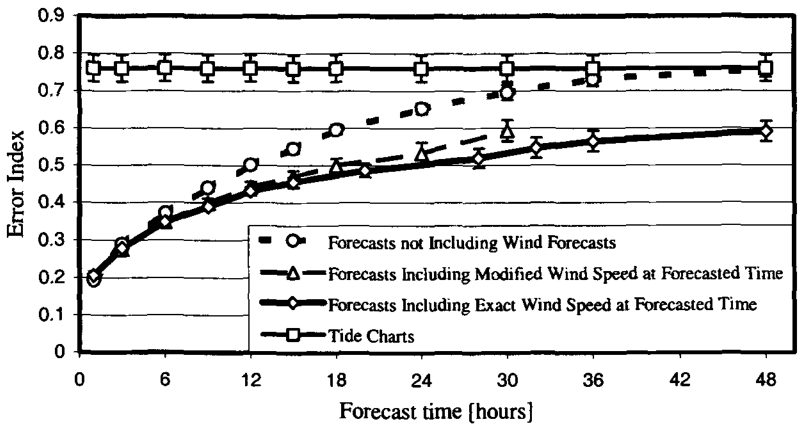

Fig. 6. Comparison of the performance of neural network models forecasting future water levels at Galveston, Texas, Pleasure Pier Station.

To assess the sensitivity of the model to the accuracy of the wind forecasts, an error proportional to the magnitude of the wind forecast and the forecasting time was added to the exact wind forecasts (Cox et al., 2002). The model was then trained on the exact wind forecasts and tested with the modified wind forecast. As can be observed in figure 6, the model performance is not affected for forecasting times below 15 hours and not affected significantly for longer forecasting times. The model was also tested while excluding wind forecasts. The results are displayed in figure 6 and show that a model without wind forecasts performs well for up to 6-hour forecasts and does not provide significant improvements over the tide tables for forecasting times longer than 24 hours. The accuracy of the model without wind forecasts will likely vary depend on the location as it depends on the natural time lags between the onset of frontal driven winds and the water level response (Cox et al., 2002). As the wind forecasts are the primary factor for accuracy of the model past 12 to 15 hours, the model is presently being modified to include wind forecasts for up to 60-hour predictions extracted in real time from the National Center for Environmental Predictions Meso Eta model. The real time inclusion of the wind forecasts is conducted as part of a collaboration with the National Weather Service. The Meso Eta forecasts will allow access not only to accurate wind forecasts for the station for which the model is trained but also wind forecasts for large portions of the Gulf of Mexico allowing for the development of more sophisticated models.

Present plans include the application of the model to other locations along the coast of Texas and its real time access through the World Wide Web. Depending on the location the models could require the addition of other inputs such as precipitation and riverine inflows. As the availability of real-time data is constantly improving along the Texas Gulf Coast this should not be a significant limiting factor. This technique will be more difficult to apply for the case of predicting storm surges during strong tropical storms and hurricanes for several reasons. First, compared to frontal events which occur almost weekly in the fall and spring months, tropical storms are more episodic and the available data base is small. Second, frontal events cover a large spatial area with conditions (wind speed, direction) relatively constant over that area compared to tropical storms which are more localized. Different types of neural network modeling techniques may have to be adapted to address the dynamic and localized nature of tropical storms.

CONCLUSIONS

A new data intensive forecasting method based on neural network modeling was developed to predict water levels along the Gulf Coast. The neural networks are trained to forecast water levels by establishing a relationship between time series of previous water levels, wind stress, and barometric pressures and future water levels. The technique was tested over a period of three years for a tide station along Galveston Island near Houston, Texas. Models were tested with and without including wind forecasts. Models including wind forecasts significantly outperformed the tide tables for the tested forecasting range of 1 hour to 48 hours. The models not including wind forecasts performed well up to 12 to 18 hours. Further work includes the addition of more sophisticated wind forecasts extracted in real time from the National Center for

Environmental Predictions Meso Eta model and expanding the model to other locations along the Texas coast.

REFERENCES

Campolo, M., Andreussi, P., and Soldati, A. 1997. River Flood Forecasting with a Neural Network Model. *Water Resources Research*, 35 (4), 1191-1197.

Cox, D., Tissot, P, and Michaud, P. 2002. Water Level Observations and Short Term Predictions Including Meteorological Events for the Entrance of Galveston Bay, Texas. *Journal of Waterways, Port, Coastal and Ocean Engineering*, accepted for publication, January 2002, ASCE.

Hopper, M. 2000. WXTide32 version 2.6, January 20, 2000, Copyright © 1998-2000 Michael Hopper, http:www.GeoCities.com/SiliconValley/Horizon/1195.html.

Kim, G., and Barros, A. 2001. Quantitative Flood Forecasting Using Multisensor Data and Neural Networks. *Journal of Hydrology*, 246, 45-62.

Mase, H. and Kitano, T. 1999. Prediction Model for Occurrence of Impact Wave Force. *Ocean Engineering*, 26 (10), 949-961.

Mase, H., Sakamoto, M., and Sakai, T. 1995. Neural Network For Stability Analysis of Rubble-Mound Breakwaters. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 121 (6), ASCE, 294-299.

Michaud, P., Jeffress, G., Dannelly, R., and Steidly, C. 2001. Real-Time Data Collection and the Texas Coastal Ocean Observation Network. *Proceedings of INTERMAC '01 Joint Technical Conference*, Tokyo, Japan, accepted.

Moatar F., Fessant, F., and Poirel, A. 1999. pH Modelling by Neural Networks. Application of Control and Validation Data Series in the Middle Loire River. *Ecological Modeling*, 120, 141-156.

NOAA 1999. Assessment of the National Ocean Service's tidal current program. NOAA Tech. Rpt. NOS Co-OPS 022, United States Department of Commerce, 71p.

NOAA 1994. Special 1994 tidal current predictions for Aransas Pass, Corpus Christi, Texas. NOAA Technical Memorandum NOS OES 8, United States Department of Commerce, 26p.

NOAA 1991. Corpus Christi Bay Current Prediction Quality Assurance Miniproject. *NOAA Technical Memorandum NOS OMA 60*, United States Department of Commerce, 46p.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K-I. 1997. Artificial Neural Network Approach for Modeling and Prediction of Algal Blooms. *Ecological Modeling*, 96, 11-28.

Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. 1995. Backpropagation: The Basic Theory. *Backpropagation: Theory, Architectures, and Applications*, Rumelhart, D. E., Chauvin, Y., eds, Lawrence Erlbaum Associates, Publishers, Hillsdale, 1-34.

Rumelhart, D. E., Hinton, G. E., and Williams, R.J. 1986. Learning Representations by Back-Propagating Errors. *Nature*, 323, 533-534.

Swingler, K. 1996. *Applying Neural Networks, A Practical Guide*, Academic Press Limited, London, 303 p.

The MathWorks, Inc. 1998. *Neural Network Toolbox for use with Matlab 5.3/version 3*, The MathWorks, Natick, MA.

Tsai, C-P., and Lee, T-L. 1999. Back-Propagation Neural Network in Tidal-Level Forecasting. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 125 (4), ASCE, 195-202.

Zirili, S.J. 1997. *Financial Predictions Using Neural Networks*, International Thomson Computer Press, London, 135 p.