INVOLVING UNDERGRADUATE STUDENTS IN FUNDED INTERDISCIPLINARY RESEARCH

Carl Steidley, Alex Sadovski, Kelly Torres, Aimee Mostella, Ray Bachnak

Department of Computing and Mathematical Sciences
Texas A&M University Corpus Christi
6300 Ocean Blvd.
Corpus Christi, TX 78412

Abstract

The goal of our on-going research is to develop effective and reliable tools for modeling the environmental systems of the Gulf of Mexico. For example, our on-going research into methodologies for the prediction of water levels in the shallow waters of the bays and estuaries along the Texas gulf coast. Our modeling approaches are based on the real-time data collected by the Texas Coastal Ocean Observation Network (TCOON). TCOON is managed by the Division of Nearshore Research (DNR) in cooperation with the Department of Computing and Mathematical Sciences (CAMS) both of Texas A&M University-Corpus Christi. TCOON consists of approximately 50 data gathering stations located along the Texas Gulf coast from the Louisiana to Mexico borders.

TCOON, started in 1988, serves as the major environmental data acquisition system for our modeling efforts. TCOON stations automatically measure and archive various measurements such as water levels, wind speed and direction, temperature, salinity, and barometric pressure. TCOON follows U.S. federal standards for the installation of its stations and has a very useful real-time, online database.

Tide charts, based on harmonic analysis, are generally the method of choice for the forecast of water levels. However there are limitations to the use of tide charts. Tide charts are mostly based on astronomical forcing or the influence on water levels of the respective motions of the earth, the moon, and the sun. There are locations around the world, including the Gulf of Mexico, where other factors such meteorological forcing often dominate tidal forcing and limit significantly the application of tide charts. In such cases other models must be developed to accurately forecast water levels.

Different schemes that we are using for the prediction of water levels include harmonic analysis, multivariate statistical models, and neural networks. In addition to a short description of the major data acquisition system for our research efforts, this paper summarizes our interdisciplinary, NASA-funded, modeling efforts, which have had a great deal of student involvement.

Introduction

Due to the heavy dependence on water level forecasts of trade and industry along the Gulf of Mexico coast, accuracy in these forecasts is essential, but the current standard forecasting methodologies do not provide accurate predictions for this region. Tide charts, produced by harmonic analysis and published by the National Ocean Service, are the existing standard, but these charts only show the astronomical forces acting upon the water. While this proves to be an accurate predictor for major portions of the other coasts, water level changes along the Texas Coast are strongly effected by meteorological factors[1] and thus require a modified prediction model.

Texas Coastal Ocean Observation Network

The Texas Coastal Ocean Observation Network (TCOON) started in 1988 serves as the major environmental data acquisition system for our modeling efforts. TCOON consists of over 50 environmental data collection platforms along the Gulf Coast, from Mexico to Louisiana (Figure1). Primary project sponsors include the Texas General Land Office, Texas Water Development Board, U.S. Army Corps of Engineers, and NOAA National Ocean Service. TCOON stations[2] measure and archive various measurements such as water levels, wind speed and direction, temperature, salinity, and barometric pressure (Figure 2). TCOON follows U.S. federal standards for the installation of its stations and has a very useful real-time, online database.

Data sampled at these stations include: precise water levels, wind speed and direction, atmospheric and water temperatures, barometric pressure, and water currents. The measurements collected at these stations are often used in legal proceedings such as littoral boundary determinations; therefore data are collected according to National Ocean Service standards. Some stations of TCOON collect parameters such as turbidity, salinity, and other water quality parameters. All data are transmitted back to A&M-CC at multiples of six minutes via line-of-sight packet radio, cellular phone, or GOES satellite, where they are then processed and stored in a real-time, web-enabled database. TCOON has been in operation since 1988.

TCOON data are valuable for tidal datum, coastal boundaries, oil-spill response, navigation, storm preparation and response, as well as research. See Figure 3 for examples of TCOON web pages. The screen on the left depicts an illustration of graphical representations of TCOON measurements in near-real time. The screen depicted to the right contains the latest measurements taken at the selected station.

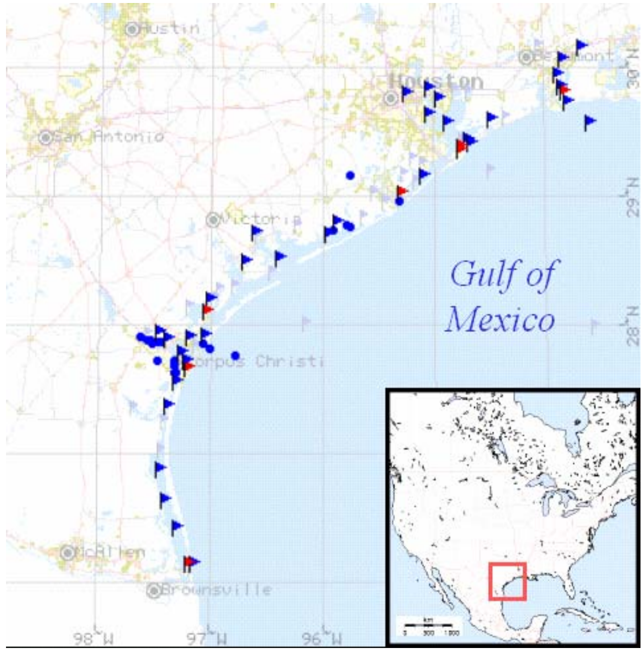

Figure 1: Map of TCOON Stations

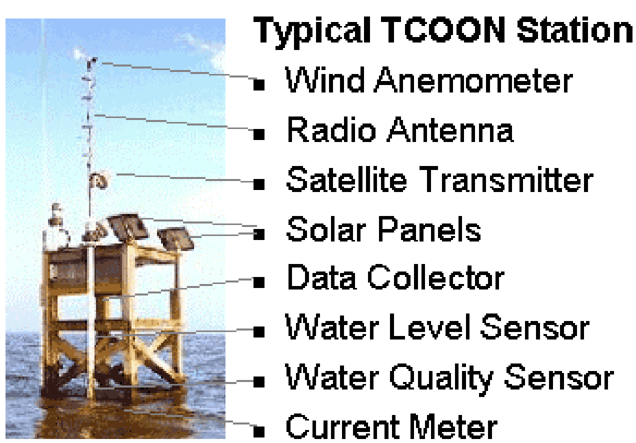

Figure 2: Example of a TCOON Station

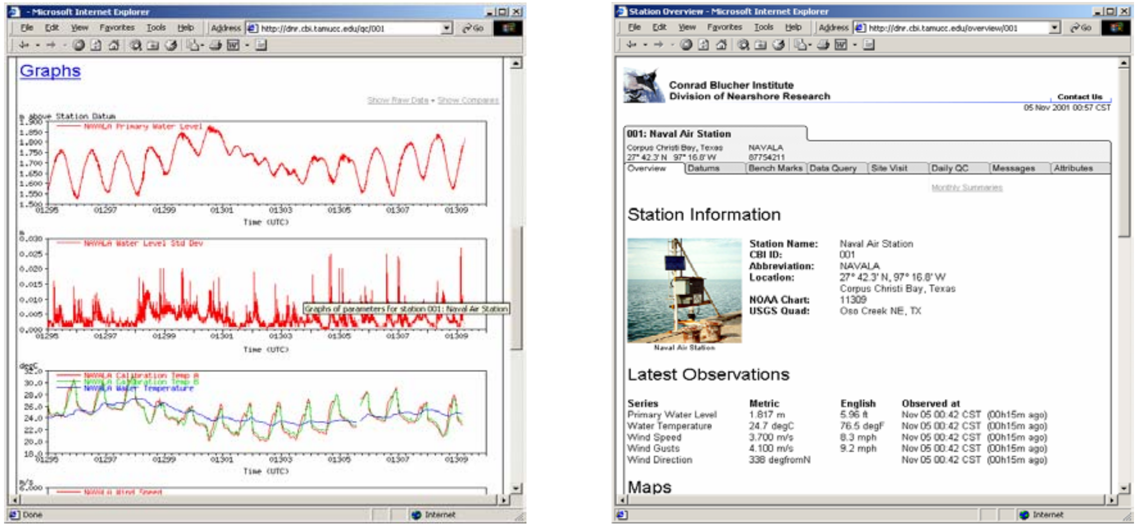

Figure 3: Typical Web Pages of the TCOON Web-Site

Tide (Water Level) Modeling

The goal of our on-going research is to develop effective and reliable tools for predicting water levels in the shallow waters of the Gulf of Mexico. Different schemes that we are using for the prediction of water levels include harmonic analysis, statistical models, and neural networks. Multivariate statistical based models of predictions of tides and neural network predictions are under development at the Division of Nearshore Research (DNR) of the Center for Coastal Studies in cooperation with the Department of Computing and Mathematical Sciences of Texas A&M University – Corpus Christi.

Statistical Modeling

Tide charts, based on harmonic analysis, are generally the method of choice for the forecast of water levels. However there are limitations to the use of tide charts. Tide charts are mostly based on astronomical forcing or the influence on water levels of the respective motions of the earth, the moon, and the sun. There are locations around the world, including the Gulf of Mexico, where other factors such meteorological forcing often dominate tidal forcing[12] and limit significantly the application of tide charts. In such cases other models must be developed to accurately forecast water levels.

We have considered three different models for "next-hour" predictions, and two of these produced quite reliable predictions. The first of these models is a multi-regression model in which the "next-hour" prediction is based on the levels of water, speeds and directions of wind for the previous 48 hours with a step of 2 hours. This model did not produce the expected results. The coefficient of correlation for these predictions was less than 0.5.

The second approach was another multi-regression model in which two-hour predictions of water level are based on the levels of water during the previous 48 hours, using 2-hour steps. Here we now believe that information about weather (pressure, wind, temperature, etc.), used in the model previously described, is hidden in the levels of water. Since this model excluding wind parameters worked remarkably well: R squared for all TCOON stations was greater than 0.95. To make further predictions we used the previously determined levels of water. Such a step by step approach produced quite good predictions. Table 1 below presents statistical data for the differences between predicted and real levels of water for 6, 12, 18, 24, 30, 36, 42, and 48 hours [12].

The third approach was also based on linear multi-regression of the levels of water, first differences, and second differences for such levels for the previous 48 hours with the step equal to two hours. This approach produces the same quality of water level prediction as the second approach, i.e. $R^2 > 0.95$. These results are quite understandable, since in both cases we have to deal with linear combinations of previous water levels. The difference in these two models is as follows: the third approach has between four (4) and eight (8) significant variables in a linear regression while in the second model of linear regression we use all twenty four (24) variables where these variables are the water levels for the previous 48 hours. Results of these predictions can be seen in Table 2 below.

We believe that this statistical model may also be useful as a means to fill gaps, due to equipment failure, in the observed water level data. To fill gaps in water level data, we will use the following procedure: First, we will find backward and forward linear regressions for the predicted water levels, and then we will evaluate lost data as a linear combination of forward and backward predictions with weights proportional to the distances from the edges of the gap.

Factor Analysis

After analyzing different regression models we faced the following question, “Why do models with only previous water levels work much better than models with all kinds of meteorological data provided by TCOON stations?” To answer this question we applied factor analysis to the water levels over the period of 48 hours with the interval of 2 hours. The conclusion is that no more than 5 factors explain over 90% of variance for water levels for all TCOON stations. Then we compared the results of the factor analysis for shallow waters with the results of the factor analysis for deep water stations.

Analyzing the for the different TCOON stations we have discovered the following:

- In shallow coastal waters and estuaries the principal or first of the major component is not periodical, and we call this component “weather”. Other main components are periodical and we call them “astronomical”.
- In off-shore deep waters, the first two or three components are astronomical components, while weather is a less dominant component.
- Our conclusion is that for estuaries and shallow waters, weather is the prime factor affecting the variations of water levels, while tidal forces are the major factors affecting the variations of water levels in deep waters.
- It has been observed also, that linear regression models for different locations have different coefficients for the same variables. We think that such differences may be explained by the geography of the place where the data is collected.

Integration of Regression and Harmonic Analysis

These conclusions assisted us in improving predictions in the shallow waters since the conclusion suggested integrating the regression approach with harmonic analysis. Namely, we use the idea that variations of water levels depend on two things – a harmonic component (which is called tides) and the weather component. Let us denote**:**

$$x_n = w_n - h_{n,}$$

where:

x_n is the difference between water level w_n and the harmonically predicted water level h_n at the moment n.

Then we can apply a technique, which is similar to that used for our statistical model described above. That is, we can predict the difference between water level and harmonic level for the next hour

$$x_1 = a_0x_0 + a_{-1}x_{-1} + \ldots + a_{-n}x_{-n}$$

and step by step

$$x_k = a_0\, x_{k-1} + a_{-1}\, x_{k-2} + \ldots + a_{-n}\, x_{k-n}$$

Now we can predict the water levels as follows:

$$p\, w_t = h_t + x_t$$

This approach to predictions of water levels proved to be very effective. In table 4 below we present comparisons of this approach with other approaches, thus, we were able to evaluate the effectiveness of this symbiosis of regression and harmonic analyses.

ANN Modeling and Predictions

The Artificial Neural Network (ANN) modeling approach is also based in forecasting future water level differences as a function of past water level differences. Other inputs to the ANN model have also been tested. For example, past wind squared is included in the model discussed below; as it has been recognized that wind forcing is well correlated with water anomalies. Other inputs, such as barometric pressure, have been tested but models which included past water level differences, past wind measurements and wind forecasts have been shown to be optimal[13]. It has also been shown that simple neural networks with one hidden layer and one output layer have the best performance[14]. With one input neuron with a tansig function and one output neuron with a purelin function and a number of total different inputs ranging from 10 to 30 the ANN forecast of a water level n hours beyond the time of forecast can be expressed as follows:

Model	RMSE	CF	POF	NOF	MDPO	MDNO
Harmonic	0.11+/-0.02	85.02+/-4.12	0.21+/- 0.19	1.90+/- 2.51	16+/-16	73+/-81
Pers 24hr	0.069+/- 0.006	95.75+/-1.19	0.249+/- 0.231	0.023+/- 0.029	14+/-19.17	0.6+/- 1.342
LR 24hr	0.106	97.18	0.261	0.027	9	1
NN-1 24hr	0.0588+/- 0.0085	97.848+/- 1.515	0.132+/- 0.114	0.104+/- 0.232	8.5+/-8.7	7.6+/- 18.5
NN-2 24hr	0.053+/- 0.0079	98.563+/- 1.284	0.124+/- 0.115	0.08+/- 0.204	8.4+/-8.7	6.2+/- 17.1
Pers 48hr	0.101+/- 0.009	87.18+/-2.22	0.785+/- 0.528	0.424+/- 0.255	25.4+/- 17.813	13.8+/- 10.232
LR 48hr	0.122	91.05	0.466	0.409	16	19
NN-1 48hr	0.0889+/- 0.0123	91.396+/- 2.768	0.199+/- 0.158	0.57+/- 0.937	9.6+/-7.6	26.4+/- 37.7
NN-2 48hr	0.0779+/- 0.0108	94.500+/- 2.616	0.123+/- 0.162	0.299+/- 0.575	6.8+/-9.6	16.3+/- 30.7

where:
RMSE - root mean square error
CF - central frequency; % of errors within the limits of -X and X
POF/NOF (2X) - positive/negative outlier frequency; % of errors greater than X
MDPO/MDNO (2X) -maximum duration of positive/negative outlier; an event is two or more consecutive occurrences of an error greater than X;
MDPO/MDNO is the length of the longest event

Table 2: Comparison of 24 and 48 hours predictions by different methods

$$x(t_o + n) = a + \left(\frac{2b}{\left(1 + e^{-\left(c + \sum d_i y_i \right)} \right)} \right) - 1$$

In the expression above, the additive parameters (a, c) are identified as the model biases and the multiplicative parameters (b, d_i) are referred to as the model weights. These parameters of the ANN are defined in the process of training of neural network over the known set of data. The y_i are the inputs to the model. The exponential terms in the ANN model provide a non-linear modeling capability.

The training of ANN models is different in nature as compared to the methods for our statistical model. There is typically no demonstrated method to identify a global optimum. The goal of the training process is therefore to find a suitable local optimum. To identify a good local optimum ANNs are trained over past data sets starting with a random guess of the model parameters and using the repeated comparison between the output of an ANN and an associated set of target vectors to optimize the weights of the neurons and biases of the model. All the ANNs discussed in this work were trained using the Levenberg-Marquardt back-propagation algorithm and implemented using version 4.0 of the Matlab Neural Network Toolbox and the MATLAB 6.0 Release 12 computational environment [15] running on a Pentium PC.

The performance of the ANN for the prediction of water levels was tested at the Bob Hall Pier, Texas, TCOON station. The model was trained and tested using three data sets composed of

3600 hourly measurements of water levels, wind speeds and wind directions. The data sets covered the spring seasons of 1998, 2000, and 2001 from Julian day 21 to Julian day 182. The model was successively trained on each data set and applied to the other two data sets. This procedure provided a set of six time series of predicted water levels to be used for validation. For each time series the average absolute error between predicted and measured water levels was computed.

Averages and standard deviations were then computed for the results of the six validation time series for these two parameters. The standard deviation gives an overall measure of the variability due to the differences between training sets as well as the differences resulting from the training process. The inputs to the model were selected as the previous12 hourly water level and wind measurements based on experience gathered during the modeling for other locations [16]. One model was trained without wind predictions while for the second case wind measurements were used to simulate wind forecasts. These wind forecasts consisted of future wind measurements at 3 hour intervals up to 36 hours. A database of wind forecasts is presently being constructed and models based on wind forecasts are expected to be more representative of future model performance. Figure 11 displays a comparison between a 36-hour water level hindcast, the tide tables, and TCOON measurements. As can be observed in the figure, the ANN model captures a large fraction of the water anomaly and improves significantly on the tide tables. The performance of the models with and without wind forecasts is compared with the performance of the tide tables for forecasting times ranging from 6 to 36 hours. Both ANN models improve significantly on the tide tables for forecasting times up to 24 hours. Improvements for 30-hours and 36-hours predictions are still measurable. The addition of wind forecasts improves the model performance although not significantly as compared to the improvement over the tide tables. Comparisons of ANN and Regression models maybe found in Table 2.

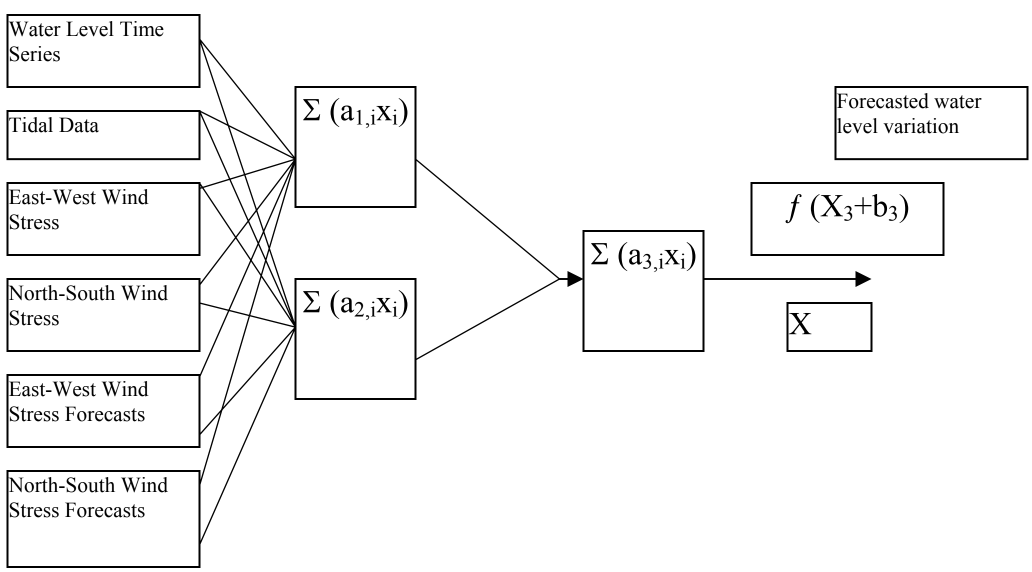

Figure10: Schematic of the type of neural network applied to the problem of water level

Student Participation

Students have been involved in this project from its inception. The first student to work on these projects was an undergraduate student in majoring in geology with minors in mathematics and computer science. He was recruited to the project because he performed well in his mathematics and computer science courses. He proved a “quick study” in the neural network modeling group. He was joined by two female students, both mathematics majors, one a graduate student, the other Hispanic undergraduate. One of these students worked on the statistical modeling, while the other worked with factor analysis.

As work progressed, this original group was joined by two additional mathematics graduate students, a male and a female. One of these students worked on a project utilizing partial differential equations to develop water level models. The other worked on the neural network modeling project.

Our original female undergraduate student graduated and was accepted to the graduate biomedical statistics program at North Carolina State University. The group has been joined by an exceptionally talented Hispanic male undergraduate mathematics major, who is working to develop entropy based criteria to evaluate learning by our artificial neural network models. The group has also been joined by a female undergraduate mathematics major, who is assisting us with the development of fuzzy systems for evaluating our models.

References

1. Sadovski, A. L., P. Tissot, P. Michaud, C. Steidley, Statistical and Neural Network Modeling and Predictions of Tides in the Shallow Waters of the Gulf of Mexico. In "WSEAS Transactions on Systems", Issue 2, vol. 2, WSEAS Press, pp.301-307, 2002.

2. Michaud, P., G. Jeffress, R. Dannelly, and C. Steidley 2001. Real Time Data Collection and the Texas Coastal Ocean Observation Network. *Proc. International Measurement and Control (InterMAC)*, Tokyo, Japan, 2002.

3. S.G. Hall, R.R. Price, and L. Wei, “Design of an Autonomous Bird Predation Reduction Device,” ASAE Annual International Meeting, Paper Number: 01-3131, 8 pages, 2001.

4. B. Ross, “A Robot Boat for Offshore Science,” The Robotics Institute-Carnegie Mellon University, United State, 2002.

5. R. Rocca, “Launching the Roboat,” Circuit Cellar Journal, Vol. 115, pp. 32-38, February 2000.

6. Woods Hole Oceanographic Institution, “Remote Environmental Monitoring Units,” http://adcp.whoi.edu/REMUS/index.html, September 2002.

7. R. Handerson, *Sea Sense*, International Marine Pub. Co., 1972.

8. Steidley, Carl, Ray Bachnak, Steve Dannelly, Patrick Michaud, Alex Sadovski, "Image and Data Logging Systems for Environmental Studies and Research", Proceedings of the ISCA 12th International Conference on Intelligent and Adaptive Systems and Software Engineering, pp 31-36, San Francisco, CA, July 2003.

9. Steidley, C., W. Lohachit, R. Bachnak, P. Michaud, S. Duff, and J. Adams, "A Remotely Operated Shallow-Water Boat for Wireless Data Logging," Proceedings of the 49th International Instrumentation Symposium of the Instrumentation, Systems and Automation Society, Orlando, FL, May 2003.

10. R. Bachnak, R.S. Dannelly, R. Kulkarni, S. Lyle, and C. Steidley, "Developing An Airborne Multi-Spectral Imaging System for GIS-based Environmental Studies," Proceedings of the ISCA 15th International Conference on Computer Applications in Industry and Engineering, Nov. 7-9, 2002, San Diego, CA.

11. R. Bachnak, R. Kulkarni, S. Lyle, and C. Steidley, "Airborne multi-camera system for geo-spatial applications," Proceedings of AeroSense 2003, April 21-55, 2003, Orlando, FL, SPIE Vol. 5097, pp. 100-105.

12. Sadovski, A. L., C. Steidley, P. Tissot, P. Michaud, "Developing a Goodness Criteria for Tide Predictions Based on Fuzzy Preference Ranking", Developments in Applied Artificial Intelligence, pp. 391-401, Lecture Notes in Artificial Intelligence 2718, Springer, June 2003.

13. Sadovski, A.L., C. Steidley, P. Tissot, G. Jeffress, "Neural Network and Statistics-Based Systems for Predictions in Coastal Studies", Intelligent Engineering Systems Through Artificial Neural Networks, pp.689-694, Vol.13, 2003.

14. Tissot, P.E D.T. Cox, and P.R. Michaud, "Optimization and Performance of a Neural Network Model Forecasting Water Levels for the Corpus Christi, Texas, Estuary", 3rd Conference on the Applications of Artificial Intelligence to Environmental Science, Long Beach, California, February 2003.

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

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

Biographical Information

CARL STEIDLEY
Carl Steidley is Professor of Computer Science and Chair of Computing and Mathematical Sciences. His interests are in the applications of artificial intelligence, real-time computing, and robotics. He has had research appointments at NASA Ames Research Center, Oak Ridge National Laboratories, and Electro Scientific Industries in Portland, Oregon. Steidley received his Ph.D. from the University of Oregon.

ALEXEY L SADOVSKI
Alexey L Sadovski, Professor of Mathematics, internationally recognized scientist. His research interests involve Mathematical Modeling, Applied Statistics, and Decision Making. His books and papers on applications of Mathematics to Economics, Sports, and Environment have been published in many countries around the world.

KELLY TORRES
Kelly Torres graduated from Texas A&M University Corpus Christi with a B.S. in Mathematics. She is currently a graduate student in Bio-Statistics at the University of North Carolina.

AIMEE MOSTELLA
Aimee Mostella is a Research Assistant at the Division of Nearshore Research at Texas A&M University Corpus Christi, where she recently graduated with an M.S. in Mathematics.

RAFIC BACHNAK

Rafic (Ray) Bachnak is a Professor in the Department of Computing and Mathematical Sciences (CAMS) at Texas A&M University-Corpus Christi (A&M-CC). He received his B.S., M.S., and Ph.D. degrees in Electrical and Computer Engineering from Ohio University in 1983, 1984, and 1989, respectively. Dr. Bachnak was previously on the faculty of Franklin University and Northwestern State University.