REAL TIME WEB AVAILABILITY OF STATISTICAL MODELS FOR WATER LEVELS ALONG THE TEXAS COASTLINE

Alex Sadovski
Department of Computing and Mathematical Sciences, Texas A&M University–Corpus Christi
6300 Ocean Dr. Corpus Christi, TX 78412

Carl Steidley
Department of Computing and Mathematical Sciences, Texas A&M University-Corpus Christi
6300 Ocean Dr. Corpus Christi, TX 78412

Philippe Tissot
Department of Physical and Life Sciences, Texas A&M University-Corpus Christi
6300 Ocean Dr. Corpus Christi, TX 78412

G. Beate Zimmer
Department of Computing and Mathematical Sciences, Texas A&M University-Corpus Christi
6300 Ocean Dr. Corpus Christi, TX 78412

Keywords: Harmonic analysis, prediction, tide chart, multivariate statistical modeling, linear regression, web, forecast, water level.

Abstract: Water level forecasts are essential to the success of trade and industry in the Gulf of Mexico, but present forecasting methodologies do not provide accurate predictions for the Gulf Coast region. Tide charts produced by harmonic analysis are the existing standard, but these charts only show the effect of astronomical forces acting upon the water. While this proves to be an accurate predictor for most of the Atlantic and Pacific Coasts, water level changes along the Texas Coast are strongly effected by meteorological factors and thus require a modified prediction model, rather than harmonic analysis alone. A web-based tool was created that combines harmonic analysis with multivariate statistical modeling to predict water levels along the Texas Gulf Coast. The result is a substantial improvement on the current model with forecasts available via the World Wide Web.

1 INTRODUCTION

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 and Department of Computing and Mathematical Sciences of Texas A&M University - Corpus Christi.

Due to the heavy dependence of trade and industry along the Gulf of Mexico coast on water level forecasts, 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 (Cox et al., 2002) and thus require a modified prediction model. A web-based tool was created that combines harmonic analysis with multivariate statistical modeling to predict water levels along the Texas Gulf Coast. The result is a substantial improvement on the tide charts with forecasts available via the web. Water level data used to make these predictions is gathered by the Texas Coastal Ocean Observation Network (TCOON), which is managed by the Division of Nearshore Research.

Sadovski A., Steidley C., Tissot P. and Beate Zimmer G. (2005).
REAL TIME WEB AVAILABILITY OF STATISTICAL MODELS FOR WATER LEVELS ALONG THE TEXAS COASTLINE.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Signal Processing, Systems Modeling and Control*, pages 218-223
DOI: 10.5220/0001157502180223

2 TEXAS COASTAL OCEAN OBSERVATION NETWORK

The Texas Coastal Ocean Observation Network (TCOON) started in 1989 and operates over 50 environmental data collection platforms along the Gulf Coast, from Mexico to Louisiana (Fig.1). 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 (Michaud et al., 2001) measure and archive various measurements such as water levels, wind speed and direction, temperature, salinity, and barometric pressure (Fig. 2). TCOON follows U.S. National Ocean Service standards for the installation of its stations and has a very useful real-time, online database. TCOON data is valuable for tidal datum, coastal boundaries, oil-spill response, navigation, storm preparation and response, as well as research.

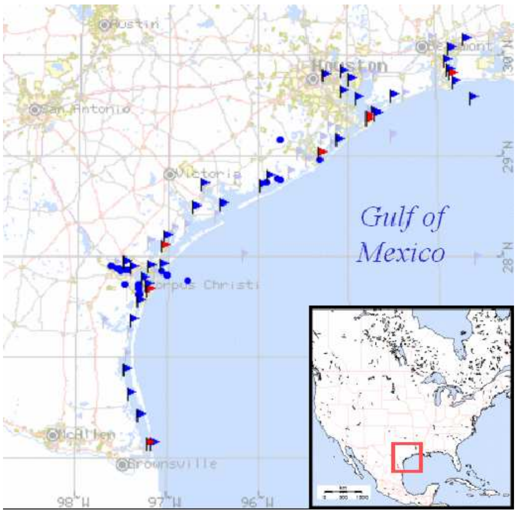


Figure 1: Map of TCOON Stations.

Figure 2: Typical TCOON station.

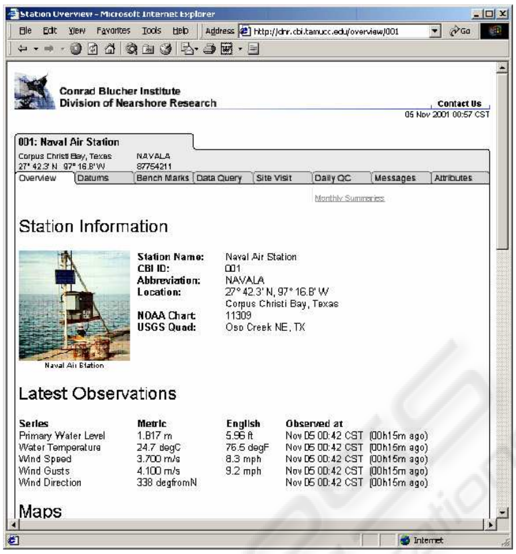


Figure 3: TCOON web site: top of the page about station 001: Naval Air Station.

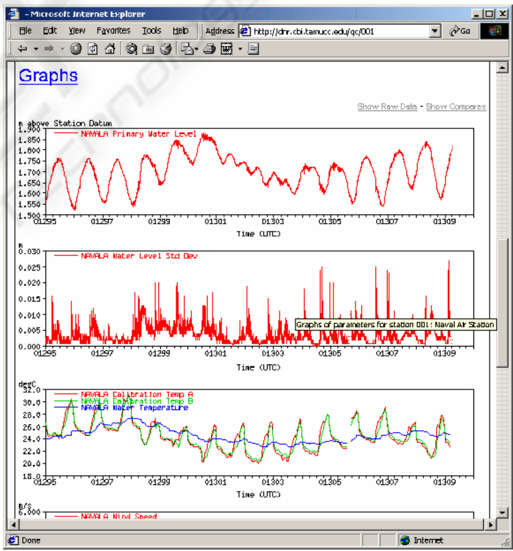


Figure 4: TCOON web site: graphs of water levels, air and water temperature data for Naval Air Station.

See Figures 3 and 4 for examples of TCOON web pages. The screen depicted in Figure 3 contains the latest measurements taken at the selected station. The screen in Figure 4 depicts an illustration of graphical representations of TCOON measurements in near-real time.

3 HARMONIC ANALYSIS

According to the Tide and Current Glossary issued by the National Ocean Service (NOS) in 2000, tide is defined as: *The periodic rise and fall of a body of water resulting from gravitational interactions between Sun, Moon, and Earth*

Thus, changes in water level from non-gravitational forces are not “tides” but rather “water levels” which can be defined as:

Astronomical+Meteorological Forcing+Other Effects

Present forecasting methodologies do not provide accurate predictions for the Gulf Coast region, because of shallow waters in this part of the Gulf of Mexico. The standard method for tide predictions is harmonic analysis. Harmonic analysis (HA) (Sadovski et al., 2003b) is represented by constituent cosine waves with known frequencies based on gravitational or periodic forces.

$$h(t) = H_O + \sum H_c f_{y,c} \cos(a_c t + e_{y,c} - k_c),$$

where
$h(t)$ = elevation of water at time t
a_c = frequency (speed) of constituent c
$f_{y,c}/e_{y,c}$ = node factors/equilibrium args
H_O = datum offset
H_c = amplitude of constituent c
k_c = phase offset for constituent c

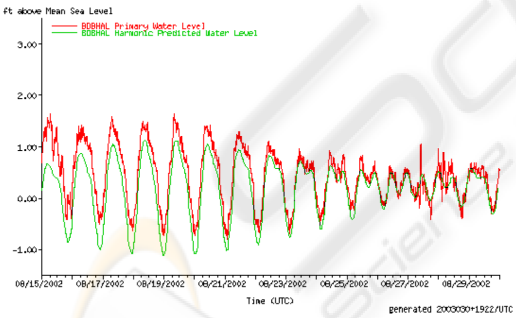


Figure 5: Comparison of water levels measured (red) and predicted by harmonic analysis (green) at the Bob Hall Pier Station for part of August 2002. It can be seen that tide tables provide accurate forecasts when the influence of meteorological and other factors is small.

4 IMPROVEMENT OF PREDICTIONS

Our goal is to develop and compare models forecasting the difference between observed water levels and

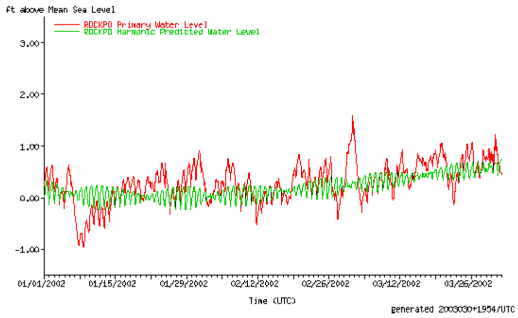


Figure 6: Comparison of water levels measured (red) and predicted by harmonic analysis (green) at the Rockport Station for January to March 2002. The large influence of meteorological forcing can be observed throughout the period.

the harmonic predictions. Approaches considered include persistence model, multivariate statistical modeling, and neural networks (Sadovski et al., 2003b), (Tissot et al., 2002). The models are built and tested based on the past observations and then applied to predict future water level differences. The methodology to develop and test the new model based on linear regression is illustrated in Fig 7.

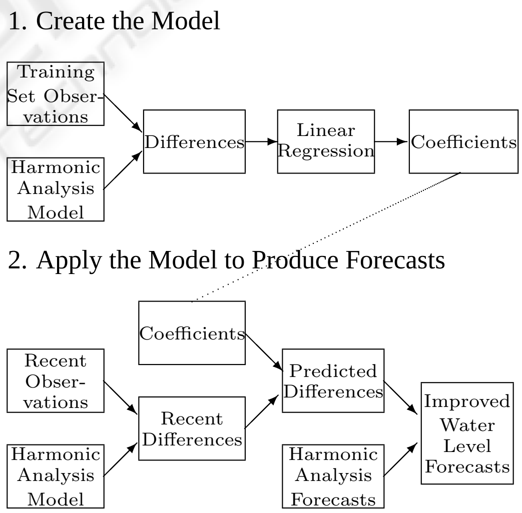


Figure 7: Schematics Illustrating the methodology to develop and test the new water level forecasting models.

We performed a factor analysis of water level data for over 20 of the TCOON stations, one station at a time, using water levels over a period of 48 hours with measurements for every other hour as variables. Data spanning nearly one year were used for each factor analysis. The stations included deep water stations such as Flower Garden near Houston and shallow water stations such as Bob Hall Pier near Corpus Christi. For Bob Hall Pier, four components extracted by Principal Component Analysis explained 96.2% of the total variance of the data. A look at the correlation matrix for the four main factors and the variables (Figure 8) shows all positive correlations in the column for the first component, indicating a non-periodic component, whereas for example the pattern of the correlation coefficients for the second component shows a periodic component.

time t	Component			
	1	2	3	4
0	.762	-.261	.452	-.276
-2	.777	-.437	.350	-.186
-4	.791	-.543	.160	-.120
-6	.804	-.551	-7.047E-02	-9.277E-02
-8	.815	-.455	-.284	-.113
-10	.825	-.278	-.426	-.174
-12	.833	-5.918E-02	-.462	-.253
-14	.840	.150	-.382	-.323
-16	.846	.303	-.205	-.354
-18	.851	.363	2.348E-02	-.332
-20	.854	.319	.247	-.256
-22	.856	.186	.410	-.139
-24	.857	8.896E-04	.470	4.746E-04
-26	.856	-.185	.411	.140
-28	.854	-.319	.248	.257
-30	.850	-.364	2.372E-02	.333
-32	.845	-.305	-.206	.354
-34	.840	-.153	-.383	.322
-36	.833	5.687E-02	-.464	.253
-38	.824	.276	-.429	.173
-40	.814	.455	-.287	.113
-42	.803	.551	-7.393E-02	9.279E-02
-44	.790	.545	.157	.120
-46	.776	.440	.348	.187
-48	.760	.265	.451	.277

Figure 8: Correlation matrix for the four main components for Bob Hall Pier.

For the deep water station Flower Garden, five main components explain 93.8% of the total variance of the data. The first two components for Flower Garden are periodic and the third component has all positive correlation coefficients, indicating that it is a non-periodic component. We found that at each station tested, no more than 5 factors explain over 90% of the variance for water levels. The periodic main components could be called “astronomical”, the non-periodic component that shows mainly in shallow waters could be called “weather”. For all stations in the coastal shallow waters and estuaries we found the first component to be not periodical; here weather is an important input for predictions. The other components are periodical. For the stations on off-shore deep waters the first two components are astronomical or periodical.

These conclusions assisted us in improving predictions in the shallow waters since the conclusion suggested integrating a regression approach with harmonic analysis. Namely, we use the idea that variations of water levels depend on two components - a harmonic component (tides) and another component which is affected by weather. The regression approach incorporates recent weather and the harmonic analysis parts accounts for the astronomical components. We separate out the weather part as follows: Let us denote

$$x_n = w_n - h_n$$

where x_n is the difference between the actual water level w_n and the harmonic forecast h_n at the moment n. Then we can apply a bootstrapping technique to find the next value of x from the water levels of the preceding n hours. That is, we can predict next hour difference between water level and harmonic level

$$x_1 = a_0 x_0 + a_{-1} x_{-1} + \ldots + a_{-n} x_{-n}$$

and step by step

$$x_k = a_0 x_{k-1} + a_1 x_{k-2} + \ldots + a_{-n} x_{k-n}.$$

Now we find the prediction for water levels as by adding the harmonic forecast h_t to the forecasted difference x_t

$$w_t = h_t + x_t.$$

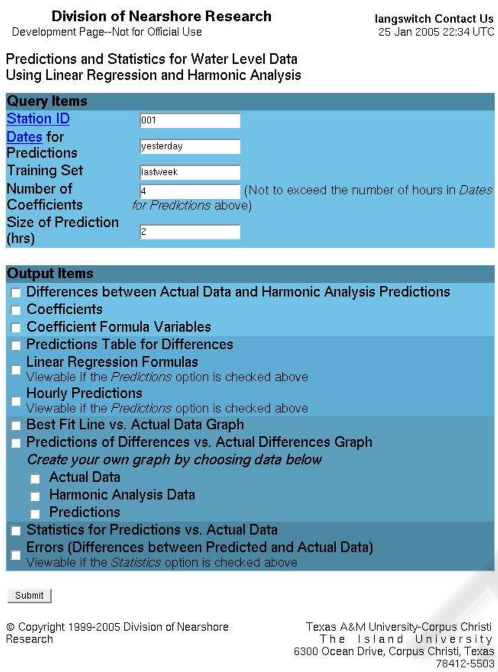
This symbiosis of regression and harmonic analysis approach to the predictions of water levels proved to be very effective.

5 MODEL PERFORMANCE ASSESSMENT

Once water level predictions are developed, they are evaluated based on a suite of National Ocean Service Skill Assessment Statistics.

Error	predicted value − observed value
SM	series mean
RMSE	root mean square error
SD	standard deviation
CF(X)	central frequency; % of errors between $-X$ and X
POF($2X$)	positive outlier frequency
NOF($2X$)	negative outlier frequency
MDPO($2X$)	max. duration of positive outlier
MDNO($2X$)	max. duration of negative outlier

A web-based tool (see Figure 9) was created that combines harmonic analysis with multivariate statistical modeling to predict water levels along the Texas Gulf Coast. Water level predictions, as well as skill assessment statistics, are dynamically generated based on a set of user given criteria including station identifier, dates, number of coefficients (that is, the number of hours) of linear regression, and a prediction range. The programming was done as CGI scripts in Perl using the Perl Data Language (PDL) and inputs from the TCOON database. The tool is available on the web and can be used for any TCOON station.


Division of Nearshore Research
Development Page--Not for Official Use
langswitch Contact Us
25 Jan 2005 22:34 UTC

Predictions and Statistics for Water Level Data
Using Linear Regression and Harmonic Analysis

Query Items

Station ID	001
Dates for Predictions	yesterday
Training Set	lastweek
Number of Coefficients	4 (Not to exceed the number of hours in *Dates for Predictions* above)
Size of Prediction (hrs)	2

Output Items

- Differences between Actual Data and Harmonic Analysis Predictions
- Coefficients
- Coefficient Formula Variables
- Predictions Table for Differences
- Linear Regression Formulas
  Viewable if the *Predictions* option is checked above
- Hourly Predictions
  Viewable if the *Predictions* option is checked above
- Best Fit Line vs. Actual Data Graph
- Predictions of Differences vs. Actual Differences Graph
  Create your own graph by choosing data below
  - Actual Data
  - Harmonic Analysis Data
  - Predictions
- Statistics for Predictions vs. Actual Data
- Errors (Differences between Predicted and Actual Data)
  Viewable if the *Statistics* option is checked above

Submit

© Copyright 1999-2005 Division of Nearshore Research

Texas A&M University-Corpus Christi
The Island University
6300 Ocean Drive, Corpus Christi, Texas
78412-5503

Figure 9: Web based tool to generate water level predictions at the TCOON stations using a symbiosis of harmonic forecasts and linear regression. Location: http://wip.cbi.tamucc.edu/∼jessica/pharosdb/cgi-bin/excel/sdiffcoeff.cgi (Development page)

6 MODEL PERFORMANCE

Using the previously described web based tool, forecasts were computed for the Packery Channel Station, near Corpus Christi, Texas. A graphical comparison between harmonic analysis forecasts, observed water levels and model predictions is presented in Figure 10. As can be observed in the figure model forecasts result in a substantial improvement over harmonic analysis. The performance improvement is further quantified using the NOS skill assessment criteria described above. Here are the National Ocean Service Skill Assessment Statistics for the given example:

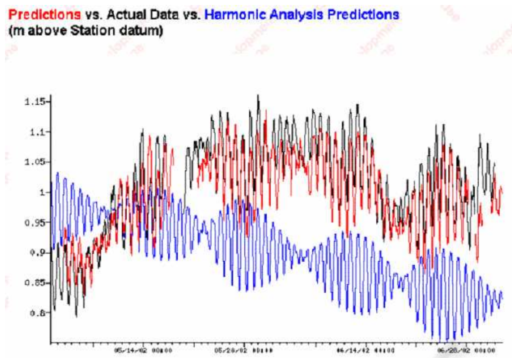


Figure 10: Comparison of Predictions with the Measured Water for Packery Channel 05/01/02-06/31/02

SM (predicted)	0.996
SM (actual)	0.995
RMSE	0.159
SD	0.159
CF (15 cm)	97.64%
POF (30 cm)	2.35%
NOF (30 cm)	0.00%
MDPO (30 cm)	24
MDNO (30 cm)	0

The advantage of this model is its stability; the quality of its forecasts are close to the quality of forecasts obtained with neural networks in (Sadovski et al., 2003b).

7 FURTHER MODEL DEVELOPMENT

Apart from being used as a forecast tool, this modeling technique has also been adapted to the problem of filling gaps. Gaps in the water level time series occur due to equipment failure, etc. Figure 11 depicts data collected at the Texas State Aquarium station with a gap spanning over 24 hours during January 2003. Linear interpolation of the data would not adequately fill the gap. Forward and backward linear regression are applied to complete the water level data sets. Other current work involves further development of the multivariate statistical model and comparisons at various locations along the Texas Gulf coast with harmonic analysis and other forecasting methodologies such as the persistence and neural network models (Sadovski et al., 2003a). Future work will generalize the forecast capabilities to points along the Texas coast that lie between TCOON stations.

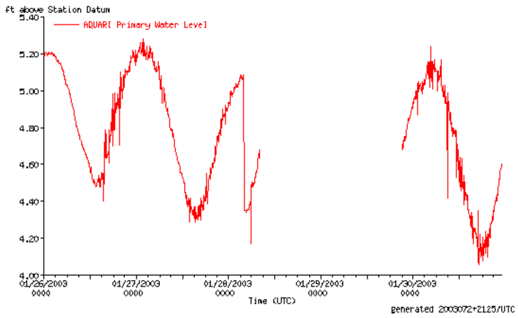


Figure 11: Example of a gap in a water level time series to be filled by the forward/backward linear regression method

8 SPONSORS AND RESOURCES

The work presented in this paper is funded in part by the following federal and state agencies of the USA
- National Aeronautic and Space Agency (NASA Grant #NCC5-517)
- National Oceanic and Atmospheric Administration (NOAA)
- Texas General Land Office - Coastal Management Program (CMP)

The following are the web based resources:

- Division of Nearshore Research Website
  http://dnr.cbi.tamucc.edu
- TCOON Data Query Page
  http://dnr.cbi.tamucc.edu/pquery
- Web-based Predictions Development Page
  http://wip.cbi.tamucc.edu/∼jessica/pharosdb
    /cgi-bin/excel/sdiffcoeff.cgi

REFERENCES

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*, 28(1):21–29.

Michaud, P., Jeffress, G., Dannelly, R., and Steidley, C. (2001). Real time data collection and the Texas Coastal Ocean Observation Network. In *Proc. International Measurement and Control (InterMAC)*.

Sadovski, A. L., C.Steidley, Tissot, P., and Michaud, P. (2003a). Developing a goodness criteria for tide predictions based on fuzzy preference ranking. In *Developments in Applied Artificial Intelligence*. Springer.

Sadovski, A. L., Tissot, P. E., Michaud, P., and Steidley, C. (2003b). Statistical and neural network modeling and predictions of tides in the shallow waters of the Gulf of Mexico. In *WSEAS Transactions on Systems*.

Tissot, P., Cox, D., and Michaud, P. (2002). Forecasting of storm surges along the Gulf of Mexico. In *Proceedings of the Fourth International Symposium on Ocean Wave Measurement and Analysis (Waves ‘01)*.