Enhanced Tide Model: Improving Tidal Predictions with Integration of Wind Data

Thomas P. Huff\textsuperscript{1}, Rusty A. Feagin\textsuperscript{1}, and Jens Figlus\textsuperscript{2}

\textsuperscript{1}Texas A&M University
Department of Ecosystem Science and Management
2138 TAMU College Station, Texas 77843
\texttt{thomas2013@tamu.edu}

\textsuperscript{2}Texas A&M University – Galveston Campus
Department of Ocean Engineering
200 Seawolf Pkwy., Galveston, TX, 77553

Keywords: Tides, Modeling, Wind, Wind Tides, R, Empirical Modeling, NOAA, NWS.
Publicly-available tidal predictions for coastlines are predominantly based on astronomical predictions. In shallow water basins, however, tides can deviate from these predictions by a factor of two or more due to wind-induced fluctuations from non-regional storms. To model and correct these wind-induced tidal deviations, a two-stage empirical model was created: the Enhanced Tidal Model (ETM). For any given NOAA tide gauge location, this model first measured the wind-induced deviation based on a compiled dataset, and then adjusted the astronomical predictions into the future to create a 144 hour forecast. The ETM, when incorporating wind data, had only 75.55% of the error of NOAA astronomical tidal predictions (e.g., if NOAA had 1.0 ft. of error, ETM had only 0.75 ft. error from observed water level). Certain ETM locations had over half (48.57%) as much prediction error as NOAA. The ETM can enhance navigation in shallow tidal waters and improve tidal forecasts. We envision the ETM as a resource for industry and the public to make informed decisions that impact their livelihood.
Introduction

Historically, tide estimation was built on lunar and solar cycles to give the timing and cycling of the tides. These models combine the gravitational pull placed on the earth by the moon, sun, and the centrifugal forces (NOAA 2008; Matsumoto, Takanezawa, and Ooe 2000). Tides have been measured and recorded since the early 1800’s in the United States, and today tide measurements and predictions are crucial for navigating coastal waterways (NOAA 2016). The National Oceanic and Atmospheric Administration (NOAA) provides a user-friendly and internet-accessible product that incorporates observational data with astronomical cycles to make tide predictions. These predictions are accurate at capturing the major physical forces that influence tides (Mukai et al. 2001). Predictions can be made years in advance as astronomical tide factors are highly regular (Price, Weller and Schudlich 1987; NOAA 2013). Several additional factors such as Ekman transport, wind direction, wind speed, storms, system location (bay or offshore), bay inlet dimensions and subsurface bathymetry, are known to affect the observed water level and these factors generally are not included in NOAA tide models (NOAA 2008; Price, Weller and Schudlich 1987; Cheng and Smith, 1998). In shallow water and micro-tidal coastlines, these factors have great influence on the overall tide level (NOAA 2008).

In a micro-tidal basin such as the Gulf of Mexico, it is not unusual to see the observed water levels exceed or fall below the NOAA estimates by a factor of two (Lefevre, Le Provost and Lyard, 2000; Zavala-Hidalgo, Morey, and O’Brien, 2003). A swing between
an accurate prediction and an inaccurate prediction of this magnitude can happen in only a few hours, and in some cases, have left ships stranded in shallow water until favorable tides occurred for their extrication (Coast Guard News 2015; NOAA 2001; NOAA 2016). Failing to correctly estimate both high tide (unexpected inundation events) and low tide conditions (unexpected problems with draft passage) can be equally detrimental. Moreover, as relative sea level rise continues, the margin between a safe water level and flooding becomes narrower (NOAA 2001), and the importance of an accurate prediction becomes greater.

Traditionally, numerical models have been created to address these problems (Zavala-Hidalgo, Morey, and O'Brien 2003; Tenorin-Fernandez, Valle-Levinson, and Gomez-Valdes, 2017) but they are often limited to local application. Alternately, coarse scale tidal models can focus on oceanic circulation (Padman et al. 2002), however these models lose their effectiveness closer to shore (Matsumoto, Takanezawa and Ooe 2000). To improve the predictive accuracy of tidal models in shallow basins, additional tide influencing variables need to be taken into consideration. To improve NOAA tide level predictions, we incorporated wind effects on water level into an Enhanced Tide Model (ETM). This model adjusts the NOAA water level predictions using wind data sets acquired from the US National Weather Service (NWS).

The ETM is not meant to predict storm surge from regional scale storms (i.e. hurricanes, tropical storms, or tropical depressions), and is currently limited to only predict water
levels with winds under 30 miles per hour (mph). The distinction between surge and wind-driven tides needs to be drawn as the physical factors driving both have parallels, however they remain distinct in their causes and behavior. Surge is the combined effect of (a) low atmospheric pressure allowing water to increase in elevation, this effect is significant in large storms (Feagin et al. 2010) (b) wave-induced set up, as driven by radiation stress (Longuet-Higgins and Stewart 1964). Wind tides are changes in water level that occur within enclosed basins, such that the wind drives currents which push water into or out of the basins. Surge is predominantly a regional-scale process – to model it requires knowledge of atmospheric pressure, waves, wind, and bathymetry. Wind tides are a local-scale process – to model it requires knowledge of wind alone. The computational approaches of the ETM versus storm surge models are also distinct, as surge models such as Sea, Lake, and Overland Surges from Hurricanes (SLOSH) are physics based and are highly location dependent (NOAA ND) (Mercado 1993; Glahn, Taylor, Kurkowski, and Shaffer 2009). The ETM is empirically based and does not require extensive parameterization and the fundamental parameters of the model are not adjusted based on location. This modeling approach allows the ETM to run in real time at hundreds of locations simultaneously updating tidal predictions every hour and is able to run on a single server. The same prediction intensity from a SLOSH or (Advanced Circulation Model) ADCIRC would be nearly impossible with a bank of supercomputers (Zhang, Li, Liu, Rhome, Forbes. 2013; Dietrich, Tanaka, Westerink, Dawson, Luettich Jr., Zijelema, Holthuijsen, Smith, Westerink, and Westerink 2012). Models such as the Extratropical Surge and Tide Operation Forecast System (ESTOFS) are produced at
higher temporal resolutions, but is run 2 times a day and at fewer locations (NOAA 2018). SLOSH and ADCIRC are extremely effective approaches to storm surge modeling and the ETM does not try to compete against these model, but instead fills a void in high precision, high temporal resolution tidal modeling via the integration of wind.

Methods

The ETM was broken into two separate modeling procedures. The first being the Base Model, which formed the framework for the second half of the model; the Predictive Model. The Base Model takes historical tide data, along with historical wind speed and wind direction, analyzes it, and creates a matrix. Second, the Predictive Model uses this matrix in combination with forecasted wind speed and direction data to create the tidal forecast. These forecasts are then graphed and linked to an html-based map.

Programming language

R programming was used for every aspect of both models along with the creation of the web Graphical User Interface GUI (R Core Team 2017).

Base Model

Data Gathering

The Base Model follows an empirical approach based on trends in historical hourly data sets, namely NOAA wind speed, NOAA wind direction, NOAA predicted water level,
and NOAA observed water level. These data sets were downloaded from NOAA (Tides and Currents 2017) websites and saved in .csv files by month and year. This procedure resulted in approximately 2.2 million hours of data collectively gathered from 89 individual NOAA gauges across the Gulf of Mexico and the east coast of Florida. Any data available for each gauge from 2008 to 2018 was collected and stored in a database (R Core Team 2017). Date formats were matched for each hour. If one of the four data sets (NOAA wind speed, NOAA wind direction, NOAA predicted water level, NOAA observed water level) was absent for a given hour, then that row of data was deleted from the compiled database.

**Conversion Matrix**

Once the data was gathered, a unique Conversion matrix was created for each individual NOAA tidal station with wind speed on the x axis and wind direction on the y axis. The matrix contained 50 columns and 360 rows. Wind speeds above 50 miles per hour (mph) were not included in the model. There was insufficient data for high wind events to be consistently considered, but moreover other processes such as regional scale storm surge begin to predominate at these speeds, thus limiting predictability. The difference (diff) between the NOAA predicted water level at a station (pred) and NOAA’s observed water level at the same station (obs) was taken for each hour (pred_{x+1} – obs_{x+1} = diff). An additional time step of one hour was used to account for the fact that wind speed and direction in hour x affect water level in the next hour (x+1) more than in the existing hour. Each occurrence of the same wind speed (ws) and wind direction (wd) for the
selected station was listed, and the median (diff) value for the next hour was entered in
the matrix for that (ws) and (wd) combination. This hour delay assumed that the wind
speed and direction had greater influence on hour x+1 than it did on hour x. The x+1
model reduced prediction error beyond an hour x model by approximately 62%.

Even with approximately ten years of hourly data for each station, a few cells were left
empty in each matrix, and thus interpolation was used to fill these missing values (the R
“interp” function from the R package “akima”) (Akima and Gebhardt 2016). The
interpolation method was tested against several data filling methods such as a moving
window, non-spatial spline interpolation, linear regression, kriging, and IDW. Each
method was evaluated independently and the most computationally efficient in terms of
time and reasonable method was chosen as “interp”. If this interpolation program was
unable to fill a gap in data, the missing value was left empty and averaging was used in
the next step (the Predictive Model) to adjust for these gaps.

Predictive Model

Data Gathering

To make hourly estimates for water level at each station for up to 144 hours ahead of
time, US National Weather Service (NWS) point forecasts of wind direction (wd) and
wind speed (ws) were obtained (NWS 2017). Wind direction was converted into degrees
(i.e. E wind corresponds to the range of 68 to 112 degrees wind direction). The values in
that range of that matrix for the given wind speed were averaged to create the final (diff) value.

NOAA tide station water level estimates (predictions only based on astronomical considerations) were also obtained. All three (NWS wind speed, NWS wind direction and NOAA predicted water level) data sets were gathered for each of the individual gauges.

Forecasting

The (diff) value in the selected cell from the Base Model’s Conversion Matrix was added to NOAA’s future predicted tide level (pred) for hour x+1 to adjust for wind influence and create the forecasted tide level (ETM_fore), thus (diff) + (pred) = (ETM_fore).

The (ETM_fore) value was then cleaned of data spikes using the smooth.spline function (“stats” package) in R with the (val) parameter set at 1 (a less aggressive level of smoothing that only removed the largest spikes) and with the weighting factor called nknots by the “stats” package set as the following: prediction length in hours/val (R Core Team 2017). The nknots act as “node” locations for the connection of separate portions of the spline arc and acting as a control point.
Finally, the (ETM_fore) value was further tuned by incorporating real time water level observations from each individual NOAA station tidal gauges (obs) at the hour at which the predictive model was first run. If (ETM_fore) was not equivalent to (obs) for that gauge, then the difference between (ETM_fore) and (obs) was calculated and all subsequent (ETM_fore) predictions were matched. The influence of the adjustment linearly decayed over each hour (h) out to 72 hours total (the decay rate was 1/72 of the difference between (ETM_fore) and (obs) per hour). For hours 72 to 144 of (ETM_fore), no further tuning was made. From here forward, we refer to the final forecasting output as sourcing from the ETM model, which is in effect a combination of the Base Model and Predictive Model.

Model Testing and Validation

Base Model

We conducted three types of tests to assess the validity of the ETM model. For the first, we sought to assess how accurately the Conversion Matrix predicted water level. In this test, the ETM was fed all the historic data for (ws), (wd) and NOAA’s predicted tide levels, for all stations in the Gulf of Mexico for the last 10 years. In this test, we were forced to utilize NOAA’s actual observed (ws) and (wd) recorded at each station in the past, because NWS point forecasts from past dates are not publicly-available. Thus, this test of validation uses verified wind data, and can only assess the ability of perfect knowledge of wind to accurately estimate the observed water level. This portion of the model did not incorporate a forecasting factor, rather only an instantaneous prediction.
Essentially the model was fed wind speed, wind direction, and NOAA predicted water level data for each station and a prediction for that corresponding timeframe was output from the model. This test essentially validates the Base Model alone. The ETM predictions were then compared to NOAA’s observed water level and the average deviation was recorded (ETM_dev). For comparison, NOAA’s own predicted values were compared with the NOAA observed water level and the average deviation recorded (NOAA_dev). The percent improvement of the ETM over NOAA was obtained by dividing (ETM_dev) by (NOAA_dev) then multiplying by 100. The resulting percent improvements were mapped by each gauge, using ArcMap 10.4.1 (NOAA 2008).

Predictive Model

For the second analysis, a forecast was made at time y from 0 to 144 hours in the future using NOAA predicted water level with (wd) and (ws) data being gathered from NWS point forecasts. The same x+1 time step was used for the Predictive Model the same as the Base Model. Meaning NWS (wd) and (ws) data created predictions for time x+1. The forecast’s accuracy was then assessed once measured data was available (after 144 hours). Thus, this test assesses the forecasting portion of the model, but any error could be both due to ETM-based errors, or errors in NWS wind forecasts. This test validates the Predictive Model. Similar to the first test, the forecasted values of ETM were then compared to NOAA’s observed water level and the average deviation was recorded, as well as the deviation between NOAA’s predicted and observed water level, and the deviations were compared.
Wind analysis

For the third analysis, the difference between NWS wind speed and NOAA recorded wind speed was calculated. Then, the difference between the ETM tide forecast values and the NOAA observed water level was taken for the same time frame as the above wind measurements. The two data sets were then regressed against each other to measure the contribution of wind forecasting error in the ETM water prediction. The accuracy of NWS wind direction (wd) prediction was not assessed at the current time.

Results

Base Model

For the first test, the ability of the base model to improve tidal water level estimates was dependent on location (Figure 1). In open water, the model performed very well as seen on the western coast of Florida where the ETM had only 52.43% of the error as NOAA. The model always improved over NOAA, except for a single outlier (Fernandina Beach, Florida where accuracy decreased 223%, likely because this location is at the confluence of several rivers and creeks.

ETM accuracy and number of hours of training data per gauge were regressed and found that beyond 100 hours no increase in accuracy was noticed \( (r^2 = 0.0653, y = 0.0003x + 62.63) \). The ETM performed satisfactorily with just more than 1000 hours of data.
Figure 1. ETM’s error from observed water level, as a percentage of NOAA error (e.g., if NOAA had 1.0 ft. of error, ETM had only 0.75 ft. error from observed water level then the ETM has 75% of NOAA’s error.). Red denotes the only outlier.
**Prediction Model**

For the second test, the ETM showed a 39.34% mean reduction in error over NOAA predictions from 0-36 hours (36-hour forecast) across all NOAA gauges in the Gulf of Mexico. For specific gauges however, these values should be considered much higher. At 72 hours, the ETM predictive mean accuracy improvement was 42.02%.

**Wind Analysis**

For the third test, there was no clear trend between forecast wind error and ETM model error with a correlation value of 0.305.

**Discussion**

**Run Time Considerations**

The data-gathering portion of the base model is costly in computing time and requires approximately three to five days to gather and calculate all the data for the 89 NOAA gauges across the Gulf of Mexico. Fortunately, this procedure only needs to be run once yearly to maintain the most recent data.
According to our testing, the most computationally costly procedure is filling and interpolation of the matrix. The time frame for the longest model to run, the kriging model, was approximately 90 days. The R interpolation method (interp) was faster and more stable than other methods and produced only 48.56 to 98.42 (excluding outlier) as much error as NOAA. The interpolation method highly influenced the accuracy of the model and this is likely due to the matrix structure of the data. While unconfirmed, it is likely that the data displays a high degree of autocorrelation and thus needs a method that does not treat individual columns or rows as discrete data sets. Other data interpolation methods were attempted (Kriging, IDW, moving window, and linear regression) but all performed poorly compared to the interp method.

For the real-time ETM Prediction Model, the computational time was substantially lower and runs on a modern eight core server in six minutes.

**Future Work and Potential Improvements to ETM**

One of the short-comings of the ETM is that by design it is based on weather forecasts. If the wind predictions are wrong the model will be wrong. Thus, this fundamental source of error must be accepted in the model to make predictions. A comprehensive study of the National Weather Service’s wind forecast for the gauge locations needs to be conducted to fully understand the error, but Hu and Skaggs (2009) show that weather
estimates can display strong regional variations. As weather forecasts increases in accuracy so will the ETM’s forecasting strength.

Further improvements can be made in the accuracy assessment of the Predictive Model. Currently, past wind data forecasts are not available for each station so a proper analysis of the error introduced by inaccurately forecasting wind is still necessary. The model accuracy assessment for the Predictive Model is done by recording the National Weather Services’ wind forecasts in a database and subsequently comparing these forecasts to wind measurements. Assessing NWS wind forecasting strength will require the accuracy assessment to run for a year or two to fully understand the impacts of weather prediction using point forecasts at each gauge.

An addition to the model that could help boost overall accuracy would be to incorporate time duration of wind speed and direction. Duration was considered and attempted to be added to the model but couldn’t be accomplished in the time allotted.

**Final Product**

A GIS based graphical user interface was created to represent the ETM forecasts in a spatially accurate map using (leaflet, shiny, and plotly) (Figure 2 and 3) (Cheng, Karambelkar and Xie 2017; Chang, Cheng, Allaire, Xie and McPherson 2017; Sievert, Parmer, Hocking, Chamberlain, Ram, Corvellec and Despouy 2017).
Each independent NOAA gauge location can be selected on the map which displays the ETM graphical forecast for that gauge. R programming, shiny, leaflet, and plotly were used to create this graphical plot that included NOAA’s prediction, the ETM prediction, and the observed tidal measurement (Cheng, Karambelkar and Xie 2017; Chang, Cheng, Allaire, Xie and McPherson 2017; Sievert, Parmer, Hocking, Chamberlain, Ram, Corvellec and Despouy 2017).

Figure 2. Graph of ETM tide prediction (black line), NOAA tide prediction (red line), and observed water level (green line). This figure shows a close up of the graphs that are created live in the GUI.
Figure 3. HTML GUI of gauge locations with popup graph of forecast. Black line is the Enhanced Tide Model, Red line indicates NOAA’s estimate, and the Green line is the observed water level. The Blue circles on the map show each forecasting location.
Conclusion

The ETM model possesses the potential to predict the tide level nearly universally, as it simply requires empirical seed data. By using commonly-available recorded data at each location, the model incorporates many stochastic factors such as inlet dynamics, gauge location, bathymetry, time, and local weather factors. The two-part structure of the ETM model (Base and Predictive) allows for its expansion to any location that has enough data to appropriately fill the base matrix.

Public use of the ETM can enable the transportation and shipping industry to operate more safely in shallow waters, allow outdoor enthusiasts access to recreational fishing and hunting locations at the appropriate portions of the tidal cycle, support the public use of boat ramps within more appropriate tolerance limits, enable infrastructure construction and ecological restoration operations to have advance notice of optimum water level working conditions, and improve the prediction of flood risk for coastal communities. The hope is that the ETM becomes a resource to aid in decision making for coastal navigation, and safety.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
References


Coast Guard News, 2015, Coast Guard oversees work to remove grounded tug, tank barge on Galveston Island, Texas. http://coastguardnews.com/coast-guard-oversees-work-to-remove-grounded-tug-tank-barage-on-galveston-island-texas/2015/10/26/


Sievert, C.; Parmer, C.; Hocking, T.; Chamberlain, C.; Ram, K.; Corvellec, M.;

Despouy, P. 2017. plotly: Create Interactive Web Graphics via 'plotly.js'. R
package version 4.7.1. https://CRAN.R-project.org/package=plotly

hydrodynamics in a tropical lagoon: A dimensionless numbers approach.
Estuarine, Coastal and Shelf Science

Zhang, K.; Li, Y.; Liu, Huiq; Rhome, J.; Forbes, C. 2013. Transition of the Coastal and
Estuarine Storm Tide Model to an Operational Storm Surge Forecast Model: A
Case Study of the Florida Coast. American Meteorological Society.

western shelf of the Gulf of Mexico using a high-resolution numerical model.