2007

Analysis and Predictions of Extreme Coastal Water Levels

Sudong Xu
THE FLORIDA STATE UNIVERSITY  
COLLEGE OF ENGINEERING

Analysis and Predictions of Extreme Coastal Water Levels

By

Sudong Xu

A Dissertation submitted to the
Department of Civil and Environmental Engineering
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2007

Copyright © 2007
Sudong Xu
All Rights Reserved
The members of the Committee approve the Dissertation of Sudong Xu
defended on April 06, 2007

__________________________________________
Wenrui Huang
Professor Directing Dissertation

__________________________________________
Xufeng Niu
Outside Committee Member

__________________________________________
Soronnadi Nnaji
Committee Member

__________________________________________
Tarek Abichou
Committee Member

Approved:

__________________________________________
Kamal Tawfiq, Chair
Department of Civil and Environmental Engineering

The office of Graduate Studies has verified and approved the above named committee members.
ACKNOWLEDGEMENTS

The author wishes to thank his committee members, Dr. Xufeng Niu, Dr. Soronnadi Nnaji and Dr. Tarek Abichou, for their friendly encouragement and patience during the extend time this study has taken. Special thanks go to Dr. Wenrui Huang, who, as major professor and friend, continued to encourage me to finish. His unfailing assistance and guidance, when requested, were invaluable. His inspiring, encouraging, and enthusiastic way of doing research is the best thing I ever learned. Without him, this research would not have been possible.

I extend many thanks to my colleagues and friends, especially Xiaohai Liu, Lei Yuan, Hongqing Wang, Bin Xu, Qipng Yang and Yi Zheng. I am also thankful to Maria Little in NOAA, for providing the data for this study.

Finally, and most importantly, I'd like to thank my family. My parents are a constant source of support and love. I'm grateful to my brother and sister in law for their encouragement and enthusiasm.
TABLE OF CONTENTS

LIST OF TABLES ...................................................................................................................... vii
LIST OF FIGURES ................................................................................................................... viii
ABSTRACT ............................................................................................................................... xi
INTRODUCTION ...................................................................................................................... 1
MANUSCRIPT 1 ......................................................................................................................... 3
Characteristics of Probability Distributions of Extreme Water Levels along the Coast of the
United States .............................................................................................................................. 3

1. Abstract ............................................................................................................................... 3
2. Introduction .......................................................................................................................... 2
3. General Extreme Value Distribution Method ................................................................... 5
   T-test for parameter c .......................................................................................................... 6
4. Annual Maximum Water Level Datasets .......................................................................... 7
5. GEV Parameters Determined by Maximum Likelihood Estimation (MLE) ................. 8
6. Characteristics of Probability Distributions of Annual Maximum Water Levels in
   Different Coastal Areas ..................................................................................................... 10
   6.1 Pacific Coast ................................................................................................................. 10
   6.2 Northeast Atlantic Coast ............................................................................................ 10
   6.3 Southeast Atlantic Coast ............................................................................................ 11
   6.4 Gulf of Mexico Coast .................................................................................................. 11
7. Performance of GEV Model For Different Coastal Areas .............................................. 12
   7.1 Pacific Coast ................................................................................................................. 12
   7.2 Northeast Atlantic Coast ............................................................................................ 13
Frequency Analysis for Predicting 1% Annual Maximum Water Levels In the Coast of Florida

1. Abstract .......................................................................................................................... 33
2. Introduction ................................................................................................................... 34
3. Review of Traditional Frequency Analysis Methods .................................................. 36
4. Annual Maximum Water Levels at Pensacola and Fernandina, Florida, USA .... 38
5. Evaluating the Performance of Traditional Frequency Analysis Methods ......... 39
   5.1 Traditional extreme value models ................................................................. 39
   5.2 Estimations of the Distribution Parameters ............................................... 40
   5.3 Comparison with Observations of Annual Maximum Water Levels .... 43
6. A Proposed Frequency Analysis Method ................................................................. 45
7. Comparison between Frequency Analysis and Simulation Method .................. 49
8. Conclusion ..................................................................................................................... 50
9. Acknowledgements ....................................................................................................... 51

Neural Network and Harmonic Analysis for Recovering Missing Extreme Water Level Data during Hurricanes

1. Abstract .......................................................................................................................... 61
2. Introduction ................................................................................................................... 62
3. Review of artificial neural network applications ..................................................... 63
4. Neural network methodology ...................................................................................... 65
   4.1 Multiple-layer ANN Model .............................................................................. 65
   4.2 Standard Network Training Using Gradient Descent Method ....................... 65
   4.3 Faster Training with Variable Learning Rate ................................................. 67
   4.4 Faster Training by Network Optimization with Conjugated Gradient Algorithm ........................................... 67
5. Data Analysis and Separating Tidal Signal from Storm Surges ..................... 69
LIST OF TABLES

Table 1-1 Characteristics GEV distributions of each station in different coastal areas 30

Table 1-2 Performance of GEV model in predicting 10% annual maximum water levels at different coastal areas 31

Table 1-3 Performance of GEV model in predicting maximum recorded extreme water level at different coastal areas 32

Table 2-1 Traditional Probability Distributions for Extreme Values 57

Table 2-2 Plotting positions of return years and corresponding annual maximum water levels 58

Table 2-3 Statistical summary of model performance 59

Table 2-4 Errors In Predicting Maximum Recorded Water Levels (m) 59

Table 2-5 Extreme water levels predicted by Xu-Huang’s method for different return years in Pensacola and Fernandina (in NGVD datum) 60

Table 3-1 Tidal Harmonic components in Panama City during Hurricane Dennis starting from local time 0.00 hour on 05/01/07 to 23:00 hour on 07/31/07 85

Table 3-2 Tidal Harmonic components in Panama City during Hurricane Ivan starting from local time at 0.00 hour on 07/01/04 to 23:00 hour on 09/30/04 85

Table 3-3 Tidal Harmonic components in Apalachicola during Hurricane Dennis starting from local time at 0.00 hour on 05/01/07 to 23:00 hour on 07/31/07 86

Table 3-4 Tidal Harmonic components in Panama City during Hurricane Ivan starting from local time at 0.00 hour on 07/01/04 to 0.00 hour on 09/30/04 86

Table 3-5 The extreme water level predictions combining storm surge with tides 87

Table 3-6 Correlation coefficients between models predicted and observed storm surge hydrograph. 87

Table 3-7 Extreme water levels (m) combining storm surge in spring and neap tides 87
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1.a</td>
<td>U.S. coastline and NOAA water level stations with long-term data over 70 years</td>
<td>19</td>
</tr>
<tr>
<td>1-1.b</td>
<td>Time series of annual maximum water levels at different coastal stations: Los Angeles, Portland, Sandy Hook and Pensacola</td>
<td>20</td>
</tr>
<tr>
<td>1-2.a</td>
<td>PDF of extreme water levels in Los Angeles, CA, USA</td>
<td>21</td>
</tr>
<tr>
<td>1-2.b</td>
<td>CDF of extreme water levels in Los Angeles, CA, USA</td>
<td>21</td>
</tr>
<tr>
<td>1-2.c</td>
<td>Extreme high water level in Los Angeles, CA, USA</td>
<td>21</td>
</tr>
<tr>
<td>1-3.a</td>
<td>PDF of extreme water levels in San Francisco, CA, USA</td>
<td>22</td>
</tr>
<tr>
<td>1-3.b</td>
<td>CDF of extreme water levels in San Francisco, CA, USA</td>
<td>22</td>
</tr>
<tr>
<td>1-3.c</td>
<td>Extreme high water level in San Francisco, CA, USA</td>
<td>22</td>
</tr>
<tr>
<td>1-4.a</td>
<td>PDF of extreme water levels in Portland, ME, USA</td>
<td>23</td>
</tr>
<tr>
<td>1-4.b</td>
<td>CDF of extreme water levels in Portland, ME, USA</td>
<td>23</td>
</tr>
<tr>
<td>1-4.c</td>
<td>Extreme high water level in Portland, ME, USA</td>
<td>23</td>
</tr>
<tr>
<td>1-5.a</td>
<td>PDF of extreme water levels in Sandy Hook, NJ, USA</td>
<td>24</td>
</tr>
<tr>
<td>1-5.b</td>
<td>CDF of extreme water levels in Sandy Hook, NJ, USA</td>
<td>24</td>
</tr>
<tr>
<td>1-5.c</td>
<td>Extreme high water level in Sandy Hook, NJ, USA</td>
<td>24</td>
</tr>
<tr>
<td>1-6.a</td>
<td>PDF of extreme water levels in Wilmington, NC, USA</td>
<td>25</td>
</tr>
<tr>
<td>1-6.b</td>
<td>CDF of extreme water levels in Wilmington, NC, USA</td>
<td>25</td>
</tr>
<tr>
<td>1-6.c</td>
<td>Extreme high water level in Wilmington, NC, USA</td>
<td>25</td>
</tr>
<tr>
<td>1-7.a</td>
<td>PDF of extreme water levels in Charleston, SC, USA</td>
<td>26</td>
</tr>
<tr>
<td>1-7.b</td>
<td>CDF of extreme water levels in Charleston, SC, USA</td>
<td>26</td>
</tr>
<tr>
<td>1-7.c</td>
<td>Extreme high water level in Charleston, SC, USA</td>
<td>26</td>
</tr>
<tr>
<td>1-8.a</td>
<td>PDF of extreme water levels in Fernandina, FL, USA</td>
<td>27</td>
</tr>
</tbody>
</table>
Fig. 1-8.b. CDF of extreme water levels in Fernandina, FL, USA  
Fig. 1-8.c. Extreme high water level in Fernandina, FL, USA  
Fig. 1-9.a. PDF of extreme water levels in Pensacola, FL, USA  
Fig. 1-9.b. CDF of extreme water levels in Pensacola, FL, USA  
Fig. 1-9.c. Extreme high water level in Pensacola, FL, USA  
Fig. 1-10.a. PDF of extreme water levels in Galveston Pier21, TX, USA  
Fig. 1-10.b. CDF of extreme water levels in Galveston Pier21, TX, USA  
Fig. 1-10.c. Extreme high water level in Galveston Pier21, TX, USA  
Fig. 2-1. Study sites: Pensacola and Fernandina Stations, Florida, USA  
Fig. 2-2. Time series of annual maximum water levels in Pensacola, USA  
Fig. 2-3. Time series of annual maximum water levels in Fernandina, USA  
Fig. 2-4. Comparison of annual maximum water levels for different return years between observations and predictions from traditional frequency analysis methods in Pensacola station  
Fig. 2-5. Comparison of annual maximum water levels for different return years between observations and predictions from traditional frequency analysis methods in Fernandina station  
Fig. 2-6. PDF and CDF distribution in Pensacola station  
Fig. 2-7. PDF and CDF distribution in Fernandina station  
Fig. 2-8. Comparison of annual maximum water levels in Pensacola between observations and predictions from the proposed frequency analysis method.  
Fig. 2-9. Comparison of annual maximum water levels in Fernandina between observations and predictions from the proposed frequency analysis method.  
Fig. 3-1. Study Areas in the coastal of North Florida  
Fig. 3-2. A three-layer feed-forward neural network for multivariate signal processing.  
Fig. 3-3. Harmonic analysis for Panama City during 05/01/05 to 07/31/05  
Fig. 3-4. Harmonic filtering of tidal signals from storm surge for Hurricane Dennis 2005 in Panama city Beach
Fig. 3-5. Harmonic filtering of tidal signals from storm surge for Hurricane Ivan 2004 in Panama city Beach

Fig. 3-6. Harmonic filtering of tidal signals from storm surge for Hurricane Dennis 2005 in Apalachicola

Fig. 3-7. Harmonic filtering of tidal signals from storm surge for Hurricane Ivan in Apalachicola

Fig. 3-8. Neural network Design for predicting storm surges using wind and atmospheric pressure after harmonic filtering tidal signal

Fig. 3-9. Panama City Beach for Hurricane Dennis: ANN model training using wind and atmospheric pressure to predict storm surges

Fig. 3-10. Panama City Beach for Hurricane Ivan: ANN model verification using wind and atmospheric pressure to predict storm surges

Fig. 3-11. Apalachicola for Hurricane Dennis: ANN model training using wind and atmospheric pressure to predict storm surges

Fig. 3-12. Apalachicola for Hurricane Ivan: ANN model verification using wind and atmospheric pressure to predict storm surges

Fig. 3-13. Compositions of water levels from storm surges and tidal variations in Panama City Beach during Hurricane Ivan.

Fig. 3-14. Combined water levels from storm surges and tidal variations in Apalachicola during Hurricane Ivan.

Fig. 3-15. Scenarios of extreme water levels composed by spring high tides and ANN predicted storm surges in Panama City Beach during Hurricane Ivan.

Fig. 3-16. Scenarios of extreme water levels composed by spring high tides and ANN predicted storm surges in Apalachicola. a) harmonic tides, b) storm surge, c) combined tides and storm surge.
ABSTRACT

Understanding the characteristics of probability distribution of extreme water levels is important for coastal flood mitigation and engineering design. In this study, frequency analysis has been conducted to investigate probability distributions along the coast of the U.S. by using three-parameter General Extreme Value (GEV) method. The GEV model combines three types of probability distributions (Type I for Gumbel distribution, Type II for Fretchet, or Type III for Weibull) into one expression. Types of distributions can be clarified by one of the three parameters of the GEV model for the corresponding studied stations. In this study, the whole U.S. coast was divided into four study areas: Pacific Coast, Northeast Atlantic Coast, Southeast Atlantic Coast and Gulf of Mexico Coast. Nine National Oceanic and Atmospheric Administration (NOAA) stations with a long history of data (more than 70 years) in the four study areas were chosen in this study. Parameters of the GEV model were estimated by using the annual maximum water level of studied stations based on the Maximum Likelihood Estimation (MLE) method. T-test was applied in this study to tell if the parameter, $c$, was greater than, less than or equal to 0, which was used to tell the type of the GEV model. Results show that different coastal areas have different probability distribution characteristics. The characteristics of probability distribution in Pacific Coast and Northeast Atlantic Coast are similar with extreme value I and III model. The Southeast Atlantic Coast and Gulf of Mexico Coast were found to have similar probability distribution characteristics. The probability distributions were found to be extreme value I and II model, which are different from those of the Pacific Coast and Northeast Atlantic Coast. The performance of the GEV model was also studied in the four coastal areas. GEV model works well in the five studied stations of both the Pacific Coast and the Northeast Atlantic Coast but does not work well in the Southeast Atlantic Coast and the Gulf of Mexico Coast.

Adequate predictions of extreme annual maximum coastal water levels (such as 100-year flood elevation) are also very important for flood hazard mitigation in coastal areas of Florida, USA. In this study, a frequency analysis method has been developed to provide more accurate predictions of 1% annual maximum water levels for the Florida coast waters. Using 82 and 94
years of water level data at Pensacola and Fernandina, performances of traditional frequency analysis methods, including advanced method of Generalized Extreme Value distribution method, have been evaluated. Comparison with observations of annual maximum water levels with 83 and 95 return years indicate that traditional methods are unable to provide satisfactory predictions of 1% annual maximum water levels to account for hurricane-induced extreme water levels. Based on the characteristics of annual maximum water level distribution Pensacola and Fernandina stations, a new probability distribution method has been developed in this study. Comparison with observations indicates that the method presented in this study significantly improves the accuracy of predictions of 1% annual maximum water levels. For Fernandina station, predictions of extreme water level match well with the general trend of observations. With a correlation coefficient of 0.98, the error for the maximum observed extreme water level of 3.11 m (NGVD datum) with 95 return years is 0.92 %. For Pensacola station, the prediction error for the maximum observed extreme water level with a return period of 83 years is 5.5 %, with a correlation value of 0.98.

In frequency analysis of 100 year coastal flood (FEMA 2005), annual extreme high water levels are often used. However, in many coastal areas, long history data of water levels are unavailable. In addition, some water level records may be missed due to the damage of measurement instruments during hurricanes. In this study, a method has been developed to employ artificial neural network and harmonic analysis for predicting extreme coastal water levels during hurricanes. The combined water levels were de-composed into tidal signals and storm surge. Tidal signal can be derived by harmonic analysis, while storm surge can be predicted by neural network modeling based on the observed wind speeds and atmospheric pressure. The neural network model employs three-layer feed-forward backpropagation structure with advanced scaled conjugate training algorithm. The method presented in this study has been successfully tested in Panama City Beach and Apalachicola located in Florida coast for Hurricane Dennis and Hurricane Ivan. In both stations, model predicted peak elevations match well with observations in both hurricane events. The decomposed storm surge hydrograph also make it possible for analysis potential extreme water levels if storm surge occurs during spring high tide.
INTRODUCTION

This study describes the analysis and prediction of extreme water levels in coastal waters. This dissertation has been prepared as a series of separate manuscripts, each focusing on a specific part of the study.

The first manuscript describes the characteristics of probability distribution of extreme water levels in the four studied coastal areas along the U.S. coastline by using three parameters GEV model. Understanding the characteristics of probability distribution of extreme water levels is important for coastal flood mitigation and engineering design. In this study, the whole USA coast was divided into four studied coastal areas: Pacific Coast, Northeast Atlantic Coast, Southeast Atlantic Coast and Gulf of Mexico Coast. Nine NOAA stations with a long history of data (more than 70 years) in the four studied coastal areas are chosen in this study. Parameters of the GEV model are estimated by using the annual maximum water level of studied stations based on the MLE method. The performance of GEV model was studied in the four coastal areas.

The second manuscript presents a development of frequency analysis for predicting 1% annual maximum water levels in the coast of Florida. Adequate predictions of extreme annual maximum coastal water levels (such as 100-year flood elevation) are also very important for flood hazard mitigation in coastal areas of Florida. Using 82 years and 94 years length of water level data at Pensacola and Fernandina, performances of traditional frequency analysis methods, including the advanced method of the GEV distribution method as recommended by FEMA (2005), were evaluated. Comparison with observations of annual maximum water levels with 83 and 95 return years indicate that traditional methods are unable to provide satisfactory predictions of 1% annual maximum water levels to account for hurricane-induced extreme water levels. A new probability distribution method was developed in this study. Comparison with observations indicates that the method presented in this study significantly improves the accuracy of predictions of 1% annual maximum water levels.

The third manuscript describes the neural network and harmonic analysis for recovering missing extreme water level by using wind vectors and air pressure as inputs to recover the
missing data for the studies in manuscript 1 and manuscript 2. In many coastal areas, a long history of data of water levels is unavailable. In addition, some water level records may be missed due to the damage of measurement instruments during hurricanes. In this study, a method was developed to employ an artificial neural network and harmonic analysis for predicting extreme coastal water levels during hurricanes. Tidal signals can be derived by harmonic analysis, while storm surge can be predicted by neural network modeling based on the observed wind speeds and atmospheric pressure. The neural network model employs three-layer feed-forward back propagation structure with advanced scaled conjugate training algorithm. The method presented in this study was tested at Panama City Beach and Apalachicola located on the Florida coast for Hurricane Dennis and Hurricane Ivan. In both stations, model predicted peak elevations match well with observations in both hurricane events. The decomposed storm surge hydrograph were also analyzed potential extreme water levels if a storm surge occurs during spring high tide.
Characteristics of Probability Distributions of Extreme Water Levels along the Coast of the United States

1. Abstract

Understanding the characteristics of probability distribution of extreme water levels is important for coastal flood mitigation and engineering design. In this study, frequency analysis has been conducted to investigate the characteristics of probability distributions of annual maximum water levels along the coast of the U.S. by using the three-parameter GEV model. The GEV model combines three types of probability distributions (Type I for Gumbel distribution, Type II for Fretchet, or Type III for Weibull) into one expression. Types of distributions can be clarified by one of the three parameters of the GEV model for the corresponding studied stations. In this study, the whole U.S. coast was divided into four studied coastal areas: Pacific Coast, Northeast Atlantic Coast, Southeast Atlantic Coast and Gulf of Mexico Coast. Nine NOAA stations with a long history of data (more than 70 years) in the four studied coastal areas were chosen in this study. Parameters of the GEV model were estimated by using the annual maximum water level of studied stations based on the MLE method. T-test was applied in this study to tell if the parameter, \( c \), is greater than, less than or equal to 0, which was further used to tell the type of the GEV model. Results show that different coastal areas have different probability distribution characteristics. The characteristics of probability distribution in Pacific Coast and Northeast Atlantic Coast are similar with extreme value I and III model. The Southeast Atlantic Coast and the Gulf of Mexico Coast were found to have similar probability distribution characteristics. The probability distributions are found to be extreme value I and II model, which are different from those of the Pacific Coast and Northeast Atlantic Coast. The performance of the GEV model is also studied in the four coastal areas. The GEV model works well in the five studied stations of both the Pacific Coast and the Northeast Atlantic Coast. The correlation coefficients between observed data and the GEV model prediction are 0.97-0.99 and the RMSE
are less than 0.046 meters. It also performs well in predicting maximum recorded water level with errors less than 0.196 meters. For the Southeast Atlantic coast and the Gulf of Mexico Coast, the coefficients of most stations are more than 0.9, the RMSE are all more than 0.058 meters in all four stations, and reach 0.126 meter in Pensacola. In predicting the maximum recorded water level, the errors of all stations are more than 0.379 meter and reach 0.915 meters in the Charleston station, which means the GEV model cannot work well in the Southeast Atlantic Coast and the Gulf of Mexico Coast.

2. Introduction

The accurate prediction of extreme high water levels is very important for Coastal flood mitigation because the coastal areas of the world are more and more populated. Also, coastal engineering planners have to include their designs with estimations of flooding risks and allow appropriate protection against expected extreme high water levels during the lifetime of the coastal infrastructure systems. Careful assessment of the probabilities of extreme water levels is playing a necessary role in designing modern coastal infrastructure systems. In the United States, the Coastal Flood Insurance Study (FIS) by the Federal Emergency Management Agency (FEMA, 2005) is based on the flood levels throughout the study area that have a 1% chance of being exceeded in any given year, which is called 100 year extreme water level. There are two main methods in predicting extreme water levels, one is the numerical model, and another is frequency analysis.

There are some applications in numerical model analysis of extreme water levels. Beach and shores resource center of Florida State University estimated the combined total storm tides for different return periods through numerical modeling. Realistic statistical representations of the important hurricanes parameters are used for the purpose of calculating the storm tides that occur during a long period. The chosen important forcing parameters are: \( p_o \) (central pressure), \( R \) (radius to maximum winds), \( v_F \) (forward speed), \( \theta \) (hurricane translation direction) and landfall or alongshore characteristics. The statistical parameters are based on historical hurricane data in the studied area. All of the parameters are considered to be independent. The empirical cumulative probability distributions are plotted for each of the parameters of interest and qualify \( p_o, R, v_F, \theta \) and hurricane track in accordance with the historical probability. The empirical cumulative probability distributions are then approximated by a series of straight line segments
for computer application. With the available statistics of hurricanes, a set of numerical models are applied on the side of interest with a random astronomical tide from the hurricane season as boundary condition. After the required number of storms have been simulated, the peak water levels are ranked and the return year $T$ is $T=500/M$, where $M$ is the rank of the combined total tide level (for example, if the simulation is intended to a 500 year period). However, we can see that this method needs a lot work for numerical simulations and choosing empirical probability distributions for each statistical parameter. Also, the numerical model calibration may not be reliable in studied areas.

Joint Probability Method (JPM) described by Tawn and Vassie (1991) and Pugh (1987, 2004) is a method including tides and other water level fluctuations in the probability estimates of coastal flooding. This method convolves the PDF of the background water levels at a given coastal location with the PDF of modeled storm surges to get the total exceedance probability. JPM has ever been applied on ocean waves and period. JPM has been applied to flood studies in two distinct forms. First, joint probability has been used in the context of an event selection approach to flood analysis. In this form, JPM means the joint probability of the parameters that define a particular event; for example, the joint probability of wave height and water level. In this approach, one seeks to select a small number of events thought to produce flooding approximating the 1% annual chance level. Engineering judgment plays an important role in this method. A second sort of JPM approach has also been adopted for hurricane surge modeling on the Atlantic and Gulf coasts, which is generally acceptable for the sites or processes for which the forcing function can be characterized by a small number of variables (such as storm size, intensity, and kinematics). With these variables, cumulative probability distribution functions for each of the several parameters can be estimated by using storm data obtained from the region surrounding the studied site. All of these distributions are approximated by a small number of discrete values. Combinations of these discrete parameter values represent all possible storms simulated with the chosen model. The rate of occurrence of each storm simulated in this way is the total rate of storm occurrence at the site, estimated from the record, multiplied by each of the discrete parameter probabilities. If the parameters are not independent, then a suitable computational adjustment must be made to account for this dependence. Actually, the numerical simulation method is only recommended when long history data are not available. If long history data are available, the frequency analysis on extreme water levels are recommended, which is
usually more reliable, because the analysis is based on the observed water level, which have already included all components. Frequency analysis is also a simpler and time saving method.

The frequency analysis is recommended when reasonably long term observed water level data are available, such as 30 years or more observed data. The process of prediction on extreme water levels is usually based on the regional history of water level data and use an appropriate model to estimate the parameters. After getting the CDF of the model, extreme water levels at different frequencies can be predicted. In this extreme value analysis approach, studied data sets are used to establish a probability distribution that is assumed to describe the observed flooding data, which can be evaluated using the data to determine the flood elevation at any frequency. FEMA (2005) recommends the use of GEV distributions for Pacific coastal water. The GEV model actually includes Extreme Value I, II, III model, which can be told by the one of the three parameters of GEV distribution. Many researchers predicted extreme water level at different sites by using Extreme Value models. A study in San Francisco (Sobey 2005) showed that the Extreme Value III (Weibull model) performs best among the three Extreme Value models. Studies show that different sites may have different distributions. In analysis, different sites often have different tidal, wind and current influence, also the hurricanes greatly influenced the extreme water level. It is important in coastal flood mitigation and engineering design to know the characteristics and performance of the GEV model of the studied sites. In this study, the whole U.S. coastal area were divide into four parts and apply the GEV model on different coastal areas of U.S. to compare and analyze different characteristics and different performance of the GEV model in different coastal areas. For the parameters estimation method, Maximum Likelihood Estimate (MLE) is chosen in this study because of its advantages compared with other methods.

Due to the different ocean areas nearby and different latitudes, characteristics of extreme water level distribution in the Northeast Atlantic coast, Southeast Atlantic coast and Gulf of Mexico Coast may be different from those of the Pacific Coast. In this paper, performances of GEV model in predicting extreme coastal water levels and the characteristics in studied areas are evaluated by using annual maximum water levels for a long-term period. The finding from this study may be very useful for coastal hazard mitigations and coastal engineering design.
3. General Extreme Value Distribution Method

There are many popular distributions that are used in engineering design; for instance, Pearson III and lognormal distributions are used in hydrology. There is a special family distribution mostly recognized in predicting extreme values, which are so-called extreme value distribution. Among the extreme value distributions, the most used distributions are Gumbel, Weibull and Frechet (Viessman, 1996). Both Gumbel and Weibull all have some successful applications and are candidates for FIS. There is a distribution know as the general extreme value distribution that can combine the distributions of the Gumbel, Weibull and Frechet distributions. It is a three parameters distribution. The characteristic of the studied data set can be told by one of the three parameters. Many investigators have suggested the GEV distribution as an improvement over some traditional models in flood flow frequency analysis (Vogel et al., 1993). GEV has been compared with other traditional method (such as Gumbel, Lognormal, log Pearson type 3 (LP3), exponential, and uniform distributions) in flood flow frequency analysis. Studies show that the GEV model can perform better than other traditional methods. Similar solutions were also obtained from the evaluation study of various distributions for flood frequency analysis (Haktanir, et al., 1993) of the Rhine Basin in Germany and two streams in Scotland. GEV is also recommended in the final guidelines of FEMA for Coastal Flood Hazard Analysis and Mapping for the Pacific Coast of the United States. The probability definition function is given by:

\[
f(x) = \frac{1}{b} \left\{1 + c \left(\frac{x-a}{b}\right)\right\}^{-\frac{-1}{c}} e^{-(1+c(x-a)/b)^{-1/c}}
\]

For \(-\infty < x \leq a - \frac{b}{c}\) with \(c < 0\) and \(a - \frac{b}{c} \leq x < \infty\) with \(c > 0\)

\[
f(x) = \frac{1}{b} e^{-(x-a)/b}\]

for \(-\infty \leq x < \infty\) with \(c = 0\)

The cumulative distribution is given by the expressions:

\[
F(x) = e^{-(1+c(x-a)/b)^{-1/c}}
\]

For \(-\infty < x \leq a - \frac{b}{c}\) with \(c < 0\) and \(a - \frac{b}{c} \leq x < \infty\) with \(c > 0\)
\[ F(x) = e^{-e^{-(x-a)/b}} \quad \text{for} \quad -\infty \leq x < \infty \quad \text{with} \ c = 0 \quad (4) \]

Where \(a\), \(b\) and \(c\) are the location, scale and shape factors, respectively. This distribution includes the Frechet (Type 2) distribution for \(c > 0\) and the Weibull (Type 3) distribution for \(c < 0\). If the limit of the exponent in equation 1 is taken as \(c\) goes to 0, then the simpler second forms are obtained, corresponding to the Gumbel (Type 1) distribution. Also, the extreme annual water level data was used to fit the parameters of GEV model and show the qualitative idea of goodness of results of fitting by showing the PDF and CDF of the distributions.

**T-test for parameter \(c\)**

GEV distribution includes the Frechet (Type 2) distribution for \(c > 0\) and the Weibull (Type 3) distribution for \(c < 0\). If the limit of the exponent in equation 1 is taken as \(c\) goes to 0, then the simpler second forms are obtained, corresponding to the Gumbel (Type 1) distribution. Using the extreme annual water level data sets to fit the parameters of the GEV model and show the characteristics of distributions. In some studied stations, the parameter \(c\) may be very close to 0; it is hard to tell whether the parameter \(c\) is greater than, less than or equal to 0. Thus, it is necessary to find a test of parameter to show \(c\) is positive, negative or equal to 0.

To test if the parameter, \(c\), is less than, greater than or equal to 0, the student’s t-test is applied in this study (Box et al, 2005; Montgomery, 1999), which is widely used in parameter test.

The t-test is

\[ t = \frac{c - \eta}{s} \quad (5) \]

Which has a known distribution called the student’s distribution. In this study, \(\eta = 0\), \(s\) is estimated standard deviation of \(c\). Therefore, the distribution of \(t\) depends on the number of degrees of freedom \((\nu = n - 1)\) of \(s\).

Because \(\nu\) is between 60~120 in this study (the length of data sets of all studied stations are between 60~120), at the 0.05 level of significance chosen in this study (Bross, 1971), the value of \(t_{0.05,120} = 1.658\). So, if \(c\) is bigger than 0 and corresponding \(t\) is greater than 1.658, it means \(c\) exceeds 0 significantly. Thus, the probability distribution of the corresponding studied station is type II (Frechet). If \(t\) is less than 1.658, it means that parameter \(c\) failures to exceed 0,
which mean it goes to 0 and the probability distribution is Type I (Gumbel). For the case of \( c \) is less than 0, if the corresponding \( t \) is less than -1.658, it means \( c \) is significantly less than 0, which shows that the probability distribution is type III (Weibull). If the corresponding \( t \) is bigger than -1.658, it means \( c \) is not significantly less than 0 and it goes to 0, which shows that the probability distribution is Type I (Gumbel). By using the t-test on parameter \( c \), we can tell the characteristic of each studied station, which is shown in Table 1-1.

4. Annual Maximum Water Level Datasets

In this study, due to the different ocean area latitudes, the coastal line of U.S. is divided into four areas: Pacific Coast, Northeast Atlantic Coast, Southeast Atlantic Coast and the Gulf of Mexico Coast. The Pacific coastal area includes the coasts of the whole Pacific coast. The Northeast Atlantic coast includes the Atlantic coast from Maine state to North Carolina state. Southeast Atlantic coast area includes the states from South Carolina to Florida. The Gulf of Mexico Coast includes the states from Texas to Florida along the Gulf of Mexico. The annual maximum water level data sets (FEMA 2005) of the four coastal areas are used to analyze the probability distribution characteristics. The data sets are all from NOAA stations based on the datum of NGVD in meters, the data sets are actually observed water level, which are combinations of harmonic tides, storm surge, waves and all other components. The annual maximum water level data sets are the maximum water levels in the corresponding years. In NOAA stations, only a few of them have long term data (close to 100 years), most stations only have history observed data for less than 30 years, which are not reliable in probability distribution characteristics analysis and frequency analysis. In probability distribution characteristics analysis and frequency analysis, long term data are quite necessary; more data means more reliability. The objective of analysis on the characteristics of probability distribution is to predict the 100 year extreme water level, with the dataset close to 100 years, the analysis is reliable, because the maximum recorded water level in the data sets can be used to test the prediction of 100 years extreme water level. On the contrary, short term data may not include all the incidents that happened in the past. In studied stations, the data sets with shorter records were found to distort the quantile estimate of floods (Panu and Roswalka, 1997). Thus, the long term data are quite important in this study.
Nine stations are chosen in this study because of the availability of long history data (shown in Fig 1-1). Each studied coastal area has at least two stations. On the Pacific coast, San Francisco, CA with 108 years data set and Los Angeles, CA with 82 years data set are chosen. On the Northeast Atlantic Coast, studied stations are Portland, ME with 82 years data set, Sandy Hook, NJ with 84 years data set and Wilmington, NC with 70 years data set. Charleston, SC with 81 years data set and Fernandina, FL with 95 years data set are chosen as studied stations of the Southeast Atlantic coast. For the Gulf of Mexico Coast, Galveston Pier21, TX with 98 years data set and Pensacola, FL with 83 years data set are chosen. Studied sites are stationed in Fig 1-1.a. The length of data sets in most stations is close to 100 years with longer than 80 years. Time series of annual maximum water levels at selected stations are presented in Fig 1-1.b.

Before plotting the observed data sets, the annual maximum observed water level need to be ranked from high to low. The plotting position for the observed data can be obtained by the following equation (Chow et al., 1988). For observed data, the return year \( T(x_m) \) is given by:

\[
T_m = \frac{(n+1)}{m}
\]

Where, \( n \) is the length of the data, \( m \) is the \( mth \) ranking of the water level data from high to low. Then we get data set \( [T_m, x_m] \) for plotting. Plots of \( T_i \) (in log scale) verse \( x_i \) and \( T_m \) verse \( x_m \) for both observed and predicted water level of each station are shown in Fig. 1-2 to Fig. 1-10 for all 9 studied stations respectively.

5. GEV Parameters Determined by Maximum Likelihood Estimation (MLE)

Many researchers (e.g., Eduardo and Jery, 2000; Hirose, 1996; Otten and Montfort, 1980) have studied methods on estimating parameters for generalized extreme value distribution. The method chosen in parameters estimation of the GEV model is MLE. Because MLE generally shows less bias than other methods and provides a consistent approach to parameter estimation problems, they are preferred. Consider an observation, \( x \), obtained from the density distribution \( f(x) \). The probability of obtaining a value close to \( x \), say within the small range \( dx \) around \( x \), is \( f(x)dx \), which is proportional to \( f(x) \). Then the posterior probability of having obtained the entire sample of \( N \) points is assumed to be proportional to the product of the individual
probabilities estimated in this way, in consequence of Equation 7. This product is called the likelihood of the sample, given the assumed distribution:

\[ L = \prod_{i=1}^{N} f(x_i) \]  

(7)

Usually, it is more common to work with the logarithm of this equation, which is the log-likelihood, \( LL \), given by:

\[ LL = \sum_{i=1}^{N} \ln f(x_i) \]  

(8)

The simple idea of the maximum likelihood method is to determine the distribution parameters that maximize the likelihood of the given sample. Because the logarithm is a momentonic function, this is equivalent to maximizing the log-likelihood. Note that because \( f(x) \) is always less than one, all terms of the sum for \( LL \) are negative; consequently, larger log-likelihoods are associated with smaller numerical values.

For the GEV model, the maximum likelihood method can be described in the following equations:

\[
LL(a,b,c) = \frac{1}{N} \sum_{i=1}^{N} \ln[f(x_i;a,b,c)]
\]  

(9)

To estimate the parameters, let

\[
\frac{\partial LL(a,b,c)}{\partial a} = 0; 
\]  

(10)

\[
\frac{\partial LL(a,b,c)}{\partial b} = 0; 
\]  

(11)

\[
\frac{\partial LL(a,b,c)}{\partial c} = 0 
\]  

(12)

Where \( f(x_i) \) is the PDF of the distributions of the observed annual maximum water level \( x_i \) and \( a \), \( b \) and \( c \) are the distribution parameters of the GEV model. Solving the above equations usually requires iterative calculations to locate the optimum parameters. The MLE method was used in this study to estimate parameters of the GEV distributions.

After estimating the parameters of the GEV distribution in each station by using MLE method, we can get CDF \( F(x_m) \) for each station for a given data set \( x_m \). For the extreme
models, the horizontal axis, $T$, is the return year of the extreme water level in terms of CDF of the distributions $F(x)$ as given below (Chow et al., 1988):

$$T_m = 1/(1 - F(x_m))$$

(13)

The vertical axis is extreme water level $x_m$. For each water level value, the corresponding return period $T_m$ can be obtained from Equation 13 for plotting $T_m \sim x_m$ graphs.

Plots of $T_i$ (in log scale) verses $x_i$ and $T_m$ verses $x_m$ for both observed and predicted water levels of each station are shown in Fig. 1-2 to Fig. 1-10 for all 9 studied stations, respectively.

6. Characteristics of Probability Distributions of Annual Maximum Water Levels in Different Coastal Areas

6.1 Pacific Coast

After obtaining the parameters $a$, $b$ and $c$ by using the annual maximum water level data sets of San Francisco and Los Angeles, the values of $c$ and the corresponding estimated standard errors are shown in Table 1-1. For San Francisco, $c$ is -0.142, with a standard error of 0.0677; $c/s$ is equal to -2.097<-1.66, thus $c < 0$ by the t-test. Therefore, the type of GEV model in San Francisco is type III. In Los Angeles, $c$ is -0.163, standard error is 0.0677, $c/s$ is equal to -2.932<-1.66, thus $c < 0$ by t-test, which means the type of GEV model in Los Angeles is type III. The characteristics of both Pacific coastal stations are the same type with Weibull distribution.

6.2 Northeast Atlantic Coast

The parameters $a$, $b$ and $c$ are obtained by using the annual maximum water level data sets of Portland, Sandy Hook and Wilmington. The values of $c$ and the corresponding estimated standard errors are also shown in Table 1-1. For Portland, $c$ is -0.157 with standard error of 0.0511; $c/s$ is equal to -3.072<-1.66, thus $c < 0$ by the t-test. Therefore, the type of GEV model in Portland is type III. In Sandy Hook, $c$ is -0.0196, standard error is 0.0818, $c/s$ is equal to -0.240>-1.66, thus $c = 0$ by t-test, which means the type of GEV model in Wilmington is type I. In Wilmington, $c$ is -0.0703, standard error is 0.0612, $c/s$ is equal to -1.149>-1.66, thus $c = 0$
by t-test, which means the type of GEV model in Wilmington is type I. The values of $c$ in all studied stations are less than 0, which means these stations have similar characteristics with the studied stations on the Pacific Coast. However, two of them are very close to 0 by the t-test, those of Sandy Hook and Wilmington should be treated as $c \to 0$. In conclusion, the characteristics of the studied stations on the Northeast Atlantic Coast are Weibull and Gumbel distributions.

6.3 Southeast Atlantic Coast

After obtaining the parameters $a$, $b$ and $c$ by using the annual maximum water level data sets of Charleston and Fernandina, the values of $c$ and the corresponding estimated standard errors are shown in Table 1-1. For Charleston, $c$ is 0.0663, standard error is 0.0582, $c/s$ is equal to 1.139 < 1.66, thus $c \to 0$ by t-test, the type of GEV model in Portland is type I. In Fernandina, $c$ is 0.205, standard error is 0.0907, $c/s$ is equal to 2.260 > 1.66, thus $c > 0$ by t-test, which means the type of GEV model in Wilmington is type II. The values of $c$ in all studied stations of the Southeast Atlantic Coast are more than 0, which means these stations have similar characteristics but quite different from those of the Pacific Coast. However, $c$ of Charleston is very close to 0, by t-test, $c$ of Charleston should be treated as $c \to 0$. In the Northeast Atlantic Coast studied stations, the characteristics are Gumbel and Frechet distributions.

6.4 Gulf of Mexico Coast

In Pensacola and Galveston Pier21, the values of $c$ and corresponding estimated standard errors are shown in Table 1-1 by using the annual maximum water level data sets. In Pensacola, $c$ is 0.344, standard error is 0.0889, $c/s$ is equal to 3.870 > 1.66, thus $c > 0$ by t-test, which means the type of GEV model in Pensacola is type II. In Galveston Pier21, $c$ is 0.0399, standard error is 0.0755, $c/s$ is equal to 0.528 < 1.66, thus $c = 0$ by t-test, which means the type of GEV model in Galveston is type I. The values of $c$ in all studied stations of the Gulf of Mexico Coast are more than 0, which means these stations have similar characteristics with those of the Southeast Atlantic Coast. However, parameter $c$ of Galveston Pier21 is very close to 0, by t-test, that of Galveston Pier21 should be treated as $c \to 0$. In the Gulf of Mexico Coastal studied stations, the characteristics are Gumbel and Frechet distributions.
7. Performance of GEV Model For Different Coastal Areas

7.1 Pacific Coast

To test the performance of the GEV model in the two studied stations, the PDF, CDF and extreme water level fitted line of San Francisco and Los Angeles are plotted in Fig 1-2 and Fig 1-3 respectively. From Fig 1-2.a and Fig 1-2.b, we can see the PDF and CDF of the San Francisco station are well fitted by the GEV model, the CDF and PDF of the GEV model quite reasonably match the trend of PDF and CDF for observed data. In Los Angeles, from Fig 1-3.a and Fig 1-3.b, we can see the fitted GEV line is also reasonably matched to the probability distributions of observed data.

As shown in Fig 1-2.c and Fig 1-3.c, the GEV model fitted lines of the both stations can match the observed annual extreme water level time series trend well in both the low end and high end. Correlation coefficients shown in Table 1-1 are all 0.99 for both stations, the Root Mean Square Errors (RMSE) is 0.019 and 0.009 meters for San Francisco and Los Angeles respectively, which show that the model prediction match the observed data trend very well. To study the performance of the GEV model in predicting different return year extreme water level, predictions on 10 return year extreme water levels of the studied stations are shown in Table 1-2. Results show the predictions can match the observations very well with only 0.004 m errors for both studied stations, which shows the GEV model can work well in predicting low return year events. Furthermore, the most important thing for the model to predict the maximum recorded water level (close to 100 return years) is also attained in these two stations. The comparisons of observed maximum recorded extreme water levels and GEV model predictions are shown in Table 1-3. In San Francisco the maximum recorded extreme water level is 1.689 m and the GEV prediction is 1.612 m, the error is 5.15% with 0.077 m difference. In Los Angeles, the prediction is even closer to the observed one, the maximum recorded water level is 1.588 m and the prediction is 1.562 m, the error is only 1.64% with 0.026 m difference.

In analysis, the Pacific Coast is not vulnerable to hurricanes because the water levels are mostly decided by astronomical tides. Therefore, no extremely high water level which is much higher than other annual extreme water level as shown in Fig 1-2 and Fig 1-3. The annual maximum extreme water level of San Francisco is in the range of 0.9 m to 1.6 m and the range of Los Angeles is between 1.0 m and 1.6 m. The fitted line can fit the data set in both the low end
and high end with this kind of more convergent distribution. With the satisfactory prediction of the maximum water level, the GEV model recommended by the FEMA guidelines in frequency analysis of the Pacific Coast can be proved to work well in predicting extreme high water levels of the Pacific Coast at different frequencies.

### 7.2 Northeast Atlantic Coast

For the Northeast Atlantic Coast, the PDF, CDF and extreme water level fitted line by GEV model of Portland, Sandy Hook and Wilmington are plotted in Fig 1-4, Fig 1-5 and Fig 1-6 respectively. Compared with those figures of the Pacific Coast, we can see the similar characteristics between the Pacific Coast and the Northeast Atlantic Coast. From Fig 1-4.a and Fig 1-4.b, we can see that the PDF and CDF of the Portland station are well fitted by the GEV model, the CDF and PDF of GEV model quite reasonably match the trend of PDF and CDF for the observed data. In Sandy Hook, from Fig 1-5.a and Fig 1-5.b, we can see that the fitted GEV line is also reasonably matched to the probability distributions of the observed data. For Wilmington, there is just a little difference at the highest frequency part between observed data and the prediction of the GEV. Most parts are well matched by GEV model.

As shown in Fig 1-2.c and Fig 1-3.c, the GEV model fitted lines of all three stations can match the observed annual extreme water level time series trend well in most observed data. The correlation coefficients shown in Table 1-1 are 0.99, 0.99 and 0.97 respectively, and the Root Mean Square Errors are 0.032, 0.030 and 0.046 meters for Portland, Sandy Hook and Wilmington respectively, which show the model predictions match the observed data trend very well. To study the performance of GEV model in predicting different return year extreme water level, predictions on 10 return year extreme water levels of the studied stations are shown in Table 1-2. Results show the predictions can match the observations well with only 0.029 m, 0.027 m and 0.079 m errors for Portland, Sandy Hook and Wilmington respectively. This shows that the GEV model can work well in predicting low return year events on the Northeast Atlantic Coast. Furthermore, the most important feature for the model to predict the maximum extreme water level is also attained in these three stations. The comparisons of the observed maximum recorded extreme water level and the GEV model prediction are shown in Table 1-2. In Portland, the maximum recorded extreme water level is 2.929 m and the GEV prediction is 2.772 m, which under predict by 5.36% with 0.157 m difference. In Sandy Hook, the maximum recorded
water level is 2.460 m and the GEV prediction is 2.483 m. The error is only 0.93% with 0.023 m difference. The biggest error of 9.74% with 0.196 m difference appeared in Wilmington, with a recorded level of 2.013 m and a predicted level of 1.817 m. For the prediction of the maximum recorded water level, the performance of the GEV model is not as good as those of the Pacific Coast, but the errors are still acceptable.

In analysis, the Northeast Atlantic Coast is vulnerable to some hurricanes but not as serious as those on the Southeast Atlantic Coast and the Gulf of Mexico Coast, so the GEV model under predict the maximum recorded water level in Portland and Wilmington with not much error. With the satisfactory prediction of the maximum water level, the GEV model recommended by FEMA guideline in frequency analysis of the Pacific Coast can also be proved to work well in predicting extreme high water levels of the Northeast Atlantic Coast at different frequencies.

7.3 Southeast Atlantic Coast

To test the performance of the GEV model in the studied stations, the PDF, CDF and extreme water level fitted line of Charleston and Fernandina are plotted in Fig 1-7 and Fig1-8 respectively. From Fig 1-7.a and Fig 1-7.b, we can see the PDF and CDF of Charleston station are fitted well in trend by the GEV model. From Fig 1-8.a and Fig 1-8.b, we can see the fitted GEV lines also reasonably match the probability distributions of the observed data. For Wilmington, there is just a little difference at the highest frequency part between the observed data and the prediction of the GEV, but most parts are matched well by the GEV model.

As shown in Fig 1-7.c and Fig 1-8.c, the GEV model fitted lines of all three stations can match the observed annual extreme water level data set trend well in the low end, however the GEV model can not get a good trend at the high end for both of the stations. Correlation coefficients shown in Table 1-1 are 0.89 and 0.98, the RMSE are 0.108 and 0.058 meters for Charleston and Fernandina respectively, which are not as good as those on the Pacific Coast and Northeast Atlantic Coast. To study the performance of GEV model in predicting different return year extreme water level, predictions on 10 return year extreme water levels of the studied stations are shown in Table 1-2. Results show that the predictions match the observations well with only 0.089 m and 0.023 m errors for Charleston and Fernandina respectively, which show the GEV model can work well in predicting low return years events on Southeast Atlantic Coast. Furthermore, the most important feature for the model to predict the maximum extreme water
level is not performed well in these two stations. The comparisons of the observed maximum recorded extreme water level and the GEV prediction are shown in Table 1-3. In Charleston, the maximum recorded extreme water level is 3.158 m and the GEV prediction is 2.220 m, which under predict 29.70% with 0.938 m difference. In Fernandina, the maximum recorded water level is 3.298 m and the GEV prediction is 2.848 m, the error is 13.65% with 0.45 m difference.

In analysis, the characteristics of the Southeast Atlantic Coast are quite different from those of the Pacific Coast and the Northeast Atlantic Coast. There is a much bigger range of the whole annual maximum water level data set as shown in Fig 1-7.c and Fig 1-8.c. Because the Southeast Atlantic Coast is vulnerable to hurricanes, the maximum recorded water level is much higher than other annual maximum water levels. The GEV model fitted line can not get a good trend at the high end and under predict the maximum recorded water level in both Charleston and Fernandina with a large error. With the unsatisfactory prediction on the maximum water level, the GEV model recommended by FEMA guidelines in frequency analysis of the Pacific Coast can not be proved to work well in predicting extreme high water levels of the Pacific Coast at different frequencies.

7.4 Gulf of Mexico Coast

On the Gulf of Mexico Coast, the PDF, CDF and extreme water level fitted line of Pensacola and Galveston Pier21 are plotted in Fig 1-9, and Fig 1-10 respectively. From Fig 1-9.a and Fig 1-9.b, we can see that the PDF and CDF of Galveston station are fitted well in trend by GEV model. From Fig 1-10.a and Fig 1-10.b, we can see that there is a difference at the highest frequency part between observed data and prediction of the GEV in Pensacola.

As shown in Fig1-9.c and Fig 1-10.c, the GEV model fitted lines for the extreme water levels of both stations match the observed annual extreme water level data sets trend well in the low end. However, the GEV model can get good trends at the high end for neither Pensacola nor Galveston Pier21. Correlation coefficients shown in Table 1-1 are 0.98 and 0.99 for Pensacola and Galveston Pier21 respectively, and the Root Mean Square Errors are 0.046 and 0.126 meters respectively, which are also not as good as those on the Pacific Coast and the Northeast Atlantic Coast. To study the performance of GEV model in predicting different return year extreme water level, predictions on 10 return year extreme water levels of the studied stations are shown in
Table 1-2. Results show that the predictions match the observations well with only 0.010 m and 0.027 m errors for Pensacola and Galveston Pier21 respectively, which show the GEV model work well in predicting low return year events on the Gulf of Mexico Coast. Furthermore, the most important feature for the model to predict the maximum extreme water level is not performed well in Pensacola and Galveston Pier21. The comparison of the observed maximum recorded extreme water level and GEV prediction is shown in Table 1-3. In Pensacola, the maximum recorded water level is 3.109 m and the GEV prediction is 2.182 m. The error is 29.82% with 0.927 m difference. In Galveston Pier21, the maximum recorded water level is 2.49 m and the GEV prediction is 1.953 m. The error is 21.57% with 0.537 m difference. In the studied stations, the errors are much more than acceptable in engineering design.

In analysis, the Gulf of Mexico Coast is very vulnerable to hurricanes, so the maximum recorded water level is much higher than other annual maximum water levels. The GEV model fitted line can not get a good trend at high end and under predicts the maximum recorded water level in both Pensacola and Galveston Pier21 with a large error. With the unsatisfactory prediction on maximum water level, the GEV model recommended by FEMA guidelines in frequency analysis of the Pacific Coast can not be proved to work well in predicting extreme high water level of the Gulf of Mexico Coast at different frequencies.

8. Conclusion

The GEV model is a widely used extreme value model in extreme water level frequency analysis. In this study, the GEV model is applied in the Pacific Coast, Northeast Atlantic Coast, Southeast Atlantic Coast and the Gulf of Mexico Coast to study the different characteristics and the performance of the GEV model in different coastal areas. Nine typical stations of four studied areas with a long history data are chosen in this study. The MLE method is chosen for parameter estimation because of its advantages. Studies show the following conclusions:

a) The characteristics of distribution types in different coastal areas are different. The type of GEV distribution is indicated by the parameter $c$. On the Pacific Coast, two studied stations are all type III with $c$ less than 0 by t-test. On Northeast Atlantic Coast areas, the distribution characteristics are similar to those of Pacific Coast with the values of $c$ being less than 0. However, by t-test, $c$ is considered to be 0 in Sandy Hook and Wilmington. Thus, on the Northeast Atlantic Coast, Portland is type III distribution,
while Sandy Hook and Wilmington are type I distribution. On Southeast Atlantic Coast, the c values are all greater than 0, which is different from those of the Pacific Coast and the Northeast Atlantic Coast. By t-test, c is considered to be 0 in Charleston and greater than 0 in Fernandina, thus, Charleston is type I distribution and Fernandina is type II distribution. On the Gulf of Mexico Coast, the characteristics of the studied stations are similar with those of the Southeast Atlantic Coast with type I in Galveston and type III distribution in Pensacola.

b) PDF and CDF distribution estimated by the GEV model of all studied coastal areas can match the observed data trend well in all stations, especially in the stations on the Pacific Coast and Northeast Atlantic Coast.

c) In general consideration, the GEV model works better on the Pacific Coast and the Northeast Atlantic Coast than it work on the Southeast Atlantic Coast and Gulf of Mexico Coast. The RMSE of the five stations on the Pacific Coast and the Northeast Atlantic Coast are 0.0092, 0.0193, 0.0318, 0.0297 and 0.0462, respectively, compared with 0.1087, 0.0578, 0.1263 and 0.0653 for the stations on the Southeast Atlantic Coast and Gulf of Mexico Coast. For Correlation coefficient, they are 0.99, 0.99, 0.99, 0.99 and 0.97 in the stations of Pacific coast and Northeast Atlantic coast, compared with 0.89, 0.98, 0.98 and 0.99 on the Southeast Atlantic Coast and Gulf of Mexico Coast.

d) The GEV model generally works well in predicting water levels for low return year events. It performs well in predicting 10 return years extreme water levels in all studied areas. For predicting maximum recorded extreme water levels with return period close to 100 years, the GEV seems to work well in San Francisco and Los Angeles on the Pacific Coast with 4.56% and 1.64% error respectively. It also seems work well in Portland and Sandy Hook of Northeast Atlantic coast with 5.36% and 0.94% errors respectively. But it shows a large error in Wilmington with a 9.74% error. The GEV does not seem to work not well on both the Southeast Atlantic Coast and Gulf of Mexico Coast areas. It shows large errors in the four stations with 28.97%, 11.49%, 29.82% and 20.60% for Charleston, Fernandina, Pensacola and Galveston Pier21, respectively. The finding from this study will be very useful for coastal hazard mitigations and engineering designs.
9. Acknowledgements

The authors would like to thank Maria Little in NOAA for providing the data for this study.
Fig. 1-1.a. USA coastline and NOAA water level stations with long-term data over 70 years.
Fig. 1-1.b. Time series of annual maximum water levels at different coastal stations: Los Angeles, Portland, Charleston and Pensacola
Fig. 1-2. Probability distributions of annual maximum water levels in San Francisco, CA: a) PDF, b) CDF

Fig. 1-2.c. Extreme water levels predicted by GEV model in San Francisco, CA
Fig. 1-3. Probability distributions of annual maximum water levels in Los Angeles, CA: 
a) PDF, b) CDF

Fig. 1-3.c. Extreme water levels predicted by GEV model in Los Angeles, CA, USA
Fig. 1-4. Probability distributions of annual maximum water levels in Portland, ME: a) PDF, b) CDF

Fig. 1-4.c. Extreme water levels predicted by GEV model in Portland, ME, USA
Fig. 1-5. Probability distributions of annual maximum water levels in Sandy Hook, NJ: a) PDF, b) CDF

Fig. 1-5.c. Extreme water levels predicted by GEV model in Sandy Hook, NJ, USA
Fig. 1-6. Probability distributions of annual maximum water levels in Wilmington, NC: a) PDF, b) CDF

Fig. 1-6.c. Extreme water levels predicted by GEV model in Wilmington, NC, USA
Fig. 1-7. Probability distributions of annual maximum water levels in Charleston, SC: a) PDF, b) CDF

Fig. 1-7.c. Extreme water levels predicted by GEV model in Charleston, SC, USA
Fig. 1-8. Probability distributions of annual maximum water levels in Fernandina, FL: a) PDF, b) CDF

Fig. 1-8.c. Extreme water levels predicted by GEV model in Fernandina, FL, USA
Fig. 1-9. Probability distributions of annual maximum water levels in Pensacola, FL: a) PDF, b) CDF

Fig. 1-9.c. Extreme water levels predicted by GEV model in Pensacola, FL, USA
Fig. 1-10. Probability distributions of annual maximum water levels in Galveston Pier 21, TX: a) PDF, b) CDF

Fig. 1-10.c. Extreme water levels predicted by GEV model in Galveston Pier 21, TX, USA
Table 1-1 Characteristics GEV distributions of each station in different coastal areas

<table>
<thead>
<tr>
<th>Coastal Areas</th>
<th>Stations</th>
<th>Number of years</th>
<th>c</th>
<th>Std</th>
<th>t</th>
<th>GEV type</th>
<th>RMSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific</td>
<td>Los Angeles, CA</td>
<td>82</td>
<td>-0.142</td>
<td>0.0677</td>
<td>-2.097</td>
<td>III</td>
<td>0.0092</td>
<td>0.9925</td>
</tr>
<tr>
<td></td>
<td>San Francisco, CA</td>
<td>108</td>
<td>-0.163</td>
<td>0.0556</td>
<td>-2.932</td>
<td>III</td>
<td>0.0193</td>
<td>0.9933</td>
</tr>
<tr>
<td>Northeast Atlantic</td>
<td>Portland, ME</td>
<td>82</td>
<td>-0.157</td>
<td>0.0511</td>
<td>-3.072</td>
<td>III</td>
<td>0.0318</td>
<td>0.9858</td>
</tr>
<tr>
<td></td>
<td>Sandy Hook, NJ</td>
<td>84</td>
<td>-0.0196</td>
<td>0.0818</td>
<td>-0.240</td>
<td>I</td>
<td>0.0297</td>
<td>0.9948</td>
</tr>
<tr>
<td></td>
<td>Wilmington, NC</td>
<td>70</td>
<td>-0.0703</td>
<td>0.0612</td>
<td>-1.149</td>
<td>I</td>
<td>0.0462</td>
<td>0.9720</td>
</tr>
<tr>
<td>Southeast Atlantic</td>
<td>Charleston, SC</td>
<td>81</td>
<td>0.0663</td>
<td>0.0582</td>
<td>1.139</td>
<td>I</td>
<td>0.1087</td>
<td>0.8889</td>
</tr>
<tr>
<td></td>
<td>Fernandina, FL</td>
<td>95</td>
<td>0.205</td>
<td>0.0907</td>
<td>2.260</td>
<td>II</td>
<td>0.0578</td>
<td>0.9791</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>Pensacola, FL</td>
<td>83</td>
<td>0.344</td>
<td>0.0889</td>
<td>3.870</td>
<td>II</td>
<td>0.1263</td>
<td>0.9826</td>
</tr>
<tr>
<td></td>
<td>Galveston Pier21, TX</td>
<td>98</td>
<td>0.0399</td>
<td>0.0755</td>
<td>0.528</td>
<td>I</td>
<td>0.0653</td>
<td>0.9876</td>
</tr>
</tbody>
</table>
Table 1-2 Performance of GEV model in predicting 10 return years extreme water levels at different coastal areas

<table>
<thead>
<tr>
<th>Coastal Areas</th>
<th>Stations</th>
<th>Observed 10 return year extreme water level: (Datum/NGVD)</th>
<th>GEV Prediction</th>
<th>Error</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific</td>
<td>San Francisco, CA</td>
<td>1.437</td>
<td>1.433</td>
<td>-0.004</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>Los Angeles, CA</td>
<td>1.485</td>
<td>1.481</td>
<td>-0.004</td>
<td>0.269</td>
</tr>
<tr>
<td>Northeast Atlantic</td>
<td>Portland, ME</td>
<td>2.595</td>
<td>2.566</td>
<td>-0.029</td>
<td>1.118</td>
</tr>
<tr>
<td></td>
<td>Sandy Hook, NJ</td>
<td>2.102</td>
<td>2.075</td>
<td>-0.027</td>
<td>1.284</td>
</tr>
<tr>
<td></td>
<td>Wilmington, NC</td>
<td>1.495</td>
<td>1.574</td>
<td>0.079</td>
<td>5.284</td>
</tr>
<tr>
<td>Southeast Atlantic</td>
<td>Charleston, SC</td>
<td>1.776</td>
<td>1.865</td>
<td>0.089</td>
<td>5.011</td>
</tr>
<tr>
<td></td>
<td>Fernandina, FL</td>
<td>2.212</td>
<td>2.189</td>
<td>-0.023</td>
<td>1.040</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>Pensacola, FL</td>
<td>1.127</td>
<td>1.137</td>
<td>0.010</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>Galveston Pier21, TX</td>
<td>1.212</td>
<td>1.239</td>
<td>0.027</td>
<td>2.228</td>
</tr>
</tbody>
</table>
Table 1-3 Performance of GEV model in predicting maximum recorded extreme water level at different coastal areas

<table>
<thead>
<tr>
<th>Coastal Areas</th>
<th>Stations</th>
<th>Maximum recorded extreme water level: (Datum/NGVD)</th>
<th>GEV Prediction</th>
<th>Error</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific</td>
<td>San Francisco, CA</td>
<td>1.689</td>
<td>1.612</td>
<td>-0.077</td>
<td>4.559</td>
</tr>
<tr>
<td></td>
<td>Los Angeles, CA</td>
<td>1.588</td>
<td>1.562</td>
<td>-0.026</td>
<td>1.637</td>
</tr>
<tr>
<td>Northeast Atlantic</td>
<td>Portland, ME</td>
<td>2.929</td>
<td>2.772</td>
<td>-0.157</td>
<td>5.360</td>
</tr>
<tr>
<td></td>
<td>Sandy Hook, NJ</td>
<td>2.46</td>
<td>2.483</td>
<td>0.023</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>Wilmington, NC</td>
<td>2.013</td>
<td>1.817</td>
<td>-0.196</td>
<td>9.737</td>
</tr>
<tr>
<td>Southeast Atlantic</td>
<td>Charleston, SC</td>
<td>3.158</td>
<td>2.243</td>
<td>-0.915</td>
<td>28.974</td>
</tr>
<tr>
<td></td>
<td>Fernandina, FL</td>
<td>3.298</td>
<td>2.919</td>
<td>-0.379</td>
<td>11.492</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>Pensacola, FL</td>
<td>3.109</td>
<td>2.182</td>
<td>-0.927</td>
<td>29.817</td>
</tr>
<tr>
<td></td>
<td>Galveston Pier21, TX</td>
<td>2.49</td>
<td>1.977</td>
<td>-0.513</td>
<td>20.602</td>
</tr>
</tbody>
</table>
1. Abstract

Adequate predictions of extreme annual maximum coastal water levels (such as 100-year flood elevation) are very important for flood hazard mitigation in coastal areas of Florida, USA. In this study, a frequency analysis method has been developed to provide more accurate predictions of 1% annual maximum water levels for the Florida coast waters. Using 82 and 94 years of water level data at Pensacola and Fernandina, performances of traditional frequency analysis methods, including advanced method of Generalized Extreme Value distribution method, have been evaluated. Comparison with observations of annual maximum water levels with 83 and 95 return years indicate that traditional methods are unable to provide satisfactory predictions of 1% annual maximum water levels to account for hurricane-induced extreme water levels. Based on the characteristics of annual maximum water level distribution Pensacola and Fernandina stations, a new probability distribution method has been developed in this study. Comparison with observations indicates that the method presented in this study significantly improves the accuracy of predictions of 1% annual maximum water levels. For Fernandina station, predictions of extreme water level match well with the general trend of observations. With a correlation coefficient of 0.98, the error for the maximum observed extreme water level of 3.11 m (NGVD datum) with 95 return year is 0.92 %. For Pensacola station, the prediction error for the maximum observed extreme water level with a return period of 83 years is 5.5 %, with a correlation value of 0.98.

KEY WORDS: Frequency analysis; annual maximum water level; 100 year flood; storm surge, Coastal flood.
2. Introduction

Coastal flooding from hurricanes historically causes significant damage and accounts for most of the deaths associated with a hurricane landfall. Although inland flooding has become the biggest killer in recent times, storm surge still holds the greatest potential for loss of life during a hurricane. The problem of storm surge is becoming of greater concern as coastal areas of the United States and the world become more populated. Although the accuracy of forecasts of tropical cyclones has improved in recent years, allowing more time for the evacuation of coastal areas, the population growth in coastal areas makes evacuation more difficult, time consuming, and costly. A greater percentage of the U.S. and world’s population is vulnerable to storm surge than ever before. In recent years, several strong hurricanes have hit the United States. In 2005, Hurricane Katrina, the third most intense hurricane with the fourth lowest central pressure brought 9.15 m storm surge to Biloxi, Mississippi, which is the highest ever observed in North America. As a Category 4 storm, Katrina’s storm surge soon breached the levee system that protected New Orleans from Lake Pontchartrain. Most of the city was subsequently flooded by the lake’s waters. This and other major damage to the coastal regions of Louisiana, Mississippi, and Alabama made Katrina the most destructive, costly natural disaster in the history of the United States. The storm killed at least 1,836 people, making it the deadliest U.S. hurricane since the 1928 Okeechobee Hurricane. Katrina caused damage of higher than $200 billion, topped Hurricane Andrew as the most expensive natural disaster in U.S. history. Over one million people were displaced - a humanitarian crisis on a scale unseen in the U.S. since the Great Depression.

Planning for coastal flood mitigation requires good estimations of extreme annual maximum water levels. The Coastal Flood Insurance Study by the Federal Emergency Management Agency (FEMA, 2005) of the United States is based on the 1% annual maximum water levels, or flood levels that have a 1% chance of being exceeded in any given year. On the average, this level is exceeded once in 100 years and is commonly called the 100-year flood. The 1% annual maximum water level might result from a single flood process or from a combination of processes. In other words, the 1% annual maximum level is an abstract concept based as much on the statistics of floods as on the physics of floods. Because coastal flooding is a product of
combined offshore, nearshore, and shoreline processes, the interrelationships of these processes are complex, and their relative effects vary significantly with coastal setting. These complexities present challenges in the determining the 1% annual chance flood (or 100-year flood) elevation for coastal hazard mapping purposes.

There are two major approaches in determining 100-year flood elevation, depending upon the availability of observed flood data for the site. The first approach, the numerical simulation approach, can be used when no adequate observational record of flood levels exists. In this case, it may be possible to simulate the flood process using hydrodynamic models driven by meteorological or other processes for which adequate data exist. That is, the hydrodynamic model (perhaps describing waves, tsunamis, or surge) provides the link between the known statistics of the generating forces, and the desired statistics of flood levels. These simulation methods, such as JPM and Monte Carlo methods as described in FEMA (2005), are relatively complex and costly and be used only when no acceptable, more economical alternative exists. Dean and Chiu (1986) conducted a combined total storm tide frequency analysis for Escambia County, Florida. Historical storms were constructed by randomly choosing a value for each parameter by generating a random value uniformly distributed between 0 and 1, and then entering the cumulative distribution at this value and selecting the corresponding parameter value. Each storm selected by this Monte Carlo procedure was simulated with the hydrodynamic model, and shoreline elevations were recorded. Simulating a large number of storms in this way is equivalent to simulating a long period of history, with the frequency connection established through the rate of storm occurrence estimated from a local storm sample. In Dean and Zhu’s study, numerical model was set up to cover large coastal waters. 500 simulations were conducted based on hurricane conditions derived from probability analysis of historical hurricanes using Monte Carlo method.

The second approach is frequency analysis of annual maximum water levels resulting from combined forcing factors. This method is recommended by FEMA (2005) when reasonably long observational observations are available, such as 30 years or more of flood or other data. The 1% annual chance flood might result from a single flood process or from a combination of processes such as astronomic tide and storm waves. In this frequency analysis method, traditional probability distribution methods include Gumbel, Weibull, Lognormal, and the more advanced Generalized Extreme Value (GEV) distributions. For estimating 1% annual maximum water
levels in Pacific coastal water, FEMA (2005) recommends the use of GEV distributions with Maximum-Likelihood Estimations (MLE) for probability parameter determinations. However, there is no guideline from FEMA for estimating 1% annual maximum extreme water levels on the coast of Florida and Gulf of Mexico. This is due mainly to the stronger climatic impacts from hurricanes on coastal storm surges that significantly affect annual maximum water levels.

In this paper, performances of traditional probability distributions in predicting 1% extreme value of annual maximum coastal water levels were evaluated. Annual maxima water levels selected from two NOAA stations consist of 82 years data at Pensacola Station and 94 years data at Fernandina Station. Other NOAA stations along the Florida coast cover only shorter period of 20-30 years, and are not selected in this study. Because the long-term data selected for this study cover the period close 100 year, 1% flood water levels can be reasonably evaluated by comparing model predictions and observations. In following sections, observed data of annual maximum water levels are described. Evaluations of traditional frequency analysis methods in predicting 1% coastal flood are presented. A new method is proposed to provide better predictions of 1% annual maximum water levels in the coast of Florida.

3. Review of Traditional Frequency Analysis Methods

There are some research publications in frequency analysis of extreme water levels. Sobey (2005) established a methodology for the prediction of low and high extremes in sustained water levels. He assumed that a long-duration observational record of about 100 years is available, and that monthly-extremes of high water can be extracted from that data set. Sobey further selected the highest monthly sustained water levels, and defined separate Annual Maximum Series (AMS). The abbreviation AMS has been adopted where both series are implied. Sobey investigated the performance of common probability distributions described in Benjamin and Cornell (1970), and Ang and Tang (1984): Extreme Value I (Gumbel), Extreme Value II (Frechet), Extreme Value III (Weibull), and Log Normal. His analysis steps follows the following steps: (1) Datum shift the raw AMS series as required by the probability model, (2) Estimate the distribution parameters based on Maximum likelihood estimate, (3) Estimate the confidence limits (95%), (4) Estimate event magnitudes for an average recurrence interval of $T_r$ years, (5) Reintroduce the analysis datum shift. His study indicates that the Extreme Value I
result is promising, but the model appears to predict high for the more extreme events. Extreme Value II does not appear a viable option. Much of the data at the high end is outside the 95% confidence limits, and the confidence limits are especially wide at this extreme. The 100-year event magnitude is also very high in comparison to the other models, and well outside the expectations from the observational data. Extreme Value III is another promising result, but the model appears to predict low for the more extreme events. The two highest observed events are beyond the 95% confidence limits. The Log Normal result appears to be the most satisfactory for this data set. The confidence band is relatively narrow and all the observational data, especially at the high end, falls within the confidence limits.

Many researchers have applied Generalized Extreme Value (GEV) in hydrology and other fields (De Michele and Salvadori, 2005; D'Onofrio et al, 1999; Morrison and Smith, 2002; Phien and Fang, 1989). Methods on estimating parameters for generalized extreme value distribution have also been studied by many researchers (e.g., Eduardo and Jery, 2000; Hirose, 1996; Otten and Montfort, 1980). D'Onofrio et al. (1999) obtained the surge probability density distribution function by applying the GEV distribution to a set of extreme surges generated by removing tides from a series of 89 year of annual maxima. It is shown that in this case, the probability distribution that best fits the surge data is that of Gumbel Type I. Gelder and Neykov (1998) applied regional frequency analysis to extreme sea level analysis along the Dutch North Sea coast, which is widely used in flood analysis. Parameter estimation is based on the theory of L-moments developed by Hosking and Wallis (1997) for assessing exceedence probabilities of extreme environmental events when data is available from more than one site. Gelder and Neykov (1998) chose the data set contains a total of 6818 water level observations with sample sizes at the 5 sites varying from 53 to 104 years. The sites from north to south are Delfzijl, Den Helder, Harlingen, Hoek van Holland and Vlissingen. In their study, the number of locations is artificially increased by splitting the data sets of a location in two or three subsets. For example, the Delfzijl location consists of data from 1882 until 1985. It is split up into data sets from these sets of years: 1882-1916, 1917-1951, 1952-1985. In testing the goodness of fit, five distributions (generalized logistic, generalized extreme value, generalized Pareto, lognormal (LN3), Pearson type III) were fitted to the region. Results show that the generalized Pareto is acceptable, according to the Hosking and Wallis (1997) goodness-of-fit measure. In their study, the L-
moment ratio diagram shows that the generalized Pareto distribution produces better results for the region.

Federal Emergency Management Agency (FEMA, 2005) of the United States has presented guidelines for extreme value analysis of annual maxima to study coastal Flood Insurance (FIS) for the Pacific coastal waters. It is recommended that extremal analysis of annual maxima be performed using the Generalized Extreme Value (GEV) Distribution with parameters estimated by the Method of Maximum Likelihood. Sobey (2005) compares the performances of different popular extreme value methods. Results indicate that the extreme Value II (Frechet) seems not to fit the water level data very well. Gumbel, Weibull, and Log Normal methods are promising, fitting the data well in some places. Log Normal is the one that can fit the data better at the high end of the data series. In general, errors between data and predictions from traditional frequency analysis method seem to rise with an increase of the return years of the extreme water levels. This has caused the concerns regarding the accuracy of predicting the 100-year flood elevation for coastal flood mitigation planning, which often falls beyond the available data range.

4. Annual Maximum Water Levels at Pensacola and Fernandina, Florida, USA

In order to evaluate the accuracy of predictions from frequency analysis for 1% annual maximum water levels, long term data set covering a period close to 100 years are desired. For most NOAA water level stations in Florida coast except Pensacola and Fernandina stations, records of observation are less than 50 years which are insufficient to validate predictions of 100 year annul maximum water levels. Therefore, data set from Pensacola and Fernandina stations (Fig 2-1) were selected for this study. At Pensacola station, eighty-two years of annual maximum water level data from 1923 to 2004 are available (Fig 2-2). At Fernanina station, ninity-four years of observations of annual maximum water levels from 1898-2004 (excluding data gaps in some years). The long-term data covers years with hurricane effects and years without hurricane effects that provide a long data series to evaluate the performances of various annual maximum value distributions in predicting 1% annual maximum flood (Fig 2-3). Annual maximum water levels shown in Fig 2-2 and Fig 2-3 are in NGVD datum. As described in
FEMA (2005), annual maximum water levels for frequency analysis of 1% event are results from mixed forcing mechanics that may be the combination of tides and storm surges.

Pensacola and Fernandina were chosen as the study sites because of their long histories of data records and their vulnerability to hurricanes. Several major hurricanes have hit Pensacola in recent years. On July 10, 2005, Hurricane Dennis hit the Panhandle coast with 120 mile per hour winds. Most of Pensacola had winds of 100 miles per hour as worse conditions hit Navarre Beach. Santa Rosa County, which took a straight hit from the eye of the hurricane, had the most damage: the roof of the county court house blew off, and the causeways at Santa Rosa and Navarre Beach flooded. Downtown Pensacola did not suffer substantial damage, compared to other areas. Several homes and businesses were damaged at Pensacola Beach, and numerous power lines were knocked down.

In 1926 Pensacola suffered a Category 3 hurricane with tides of 2.9 meter and winds of over 100 miles per hour that lasted over 24 hours. The barometric pressure measured 28.56. Almost every pier and boat was destroyed. In September of 1995, the second hurricane to affect the western Florida Panhandle that season, Opal, hit just east of Pensacola, with 125 miles-per-hour winds, down from 150 miles per hour just offshore. The barometer at landfall registered 27.86. On Pensacola Beach at least half the homes were destroyed; all the hotels were damaged. Early on the morning of September 16, 2004, Hurricane Ivan passed approximately 20 miles to the west of Pensacola, with winds of 130 miles per hour, moving NNE over eastern Alabama. There was heavy beach damage and heavy flooding all around Pensacola. In Santa Rosa County 1,064 homes were destroyed; Escambia County lost over 6,000 homes.

Fernandina also has a history of hurricane damage. In 1898, a hurricane hit near Brunswick, Georgia, as a Category 4 storm; the winds at Fernandina were 90 to 100 miles per hour, the barometer measuring 28.984. Moving WNW at 15 miles per hour, the hurricane inundated Fernandina Beach. Nearby Fort Clinch flooded, with up to 2.4 meters of water. Three other hurricanes also affected Fernandina and induced storm surges as a result: a Category 3 unnamed hurricane in 1906, a Category 3 unnamed hurricane in 1945, and Category 3 Hurricane Betsy in 1965.

5. Evaluating the Performance of Traditional Frequency Analysis Methods
5.1 Traditional extreme value models
Annual extreme water levels from Pensacola and Fernandina stations were used to evaluate the performance of traditional probability models for predicting 1% annual maximum water levels. As described in FEMA (2005), Viessman (1996) and Chow (1988); commonly used probability models for extreme value predictions are Gumbel, Weibull, and GEV models. The study by Sobey (2005) for Pacific coast of USA indicates that Lognormal model may perform better in annual water level prediction, the Frechet model had a very poor performance. FEMA (2005) recommends GEV method for coastal flood frequency analysis for Pacific coast of USA. Probability density function (PDF) and cumulative density function (CDF) are summarized in Table 2-1.

Annual maximum water levels at Pensacola and Fernandina stations were used to evaluate the performance of traditional frequency methods. Following the procedures as given in Sobey (2005), datum shift were done to the raw data series as required by the probability model. The Weibull and Log Normal definition for \( x \) to be positive is accommodated by an analysis datum shift to a data lower bound at \( x = x_0 \); for annual maximum series of water surface elevations, \( x_0 = MHHW \) (mean higher high water) would be physically appropriate. The Gumbel model and GEV require no datum shift, so that \( x_0 = 0 \). The net data series for extreme value analysis would be \( x - x_0 \).

5.2 Estimations of the Distribution Parameters

5.2.1 Parameters for Gumble Distribution

For the Gumbel model, coefficients can be directly estimated by the mean and standard deviation of data series using the following equations (Viessman, 1996).

\[
\alpha = 0.78 \cdot \sigma \quad (4.1)
\]
\[
\xi = \mu - 0.5772\alpha \quad (4.2)
\]

Where \( \alpha \) and \( \xi \) are coefficients, \( \mu \) is the mean of \( x \), \( \sigma \) is the variance of \( x \).

For Pensacola, \( \alpha = 0.3138 ; \ \xi = 0.6047 \).

For Fernandina, \( \alpha = 0.2072 ; \ \xi = 0.6662 \).

5.2.2 Maximum-Likelihood Estimation Method
Parameters for the Weibull, Log Normal, and GEV model distribution are commonly estimated by maximizing the sample likelihood, using the Maximum-Likelihood Estimation (MLE) method (Sobey, 2005). MLE is recommended by FEMA (2005) because it generally shows less bias than other methods and provides a consistent approach to parameter estimation problems. The basic idea of the maximum likelihood method is to determine the distribution parameters that maximize the likelihood of the given sample. Because the logarithm is a monotonic function, this is equivalent to maximizing the log-likelihood. The Maximum Likelihood method (Chow, 1998) can be described in the following equations:

\[
L(p_1, p_2) = \frac{1}{N} \sum_{n=1}^{N} \ln[f(x_n; p_1, p_2)] \tag{4.3}
\]

\[
\frac{\partial L(p_1, p_2)}{\partial p_1} = 0; \quad \frac{\partial L(p_1, p_2)}{\partial p_2} = 0. \tag{4.4}
\]

Where \( f(x_n) \) is the PDF of the distributions of the observation \( x_n \), \( p_1 \) and \( p_2 \) are the distribution parameters (for instance, Weibull are \( \alpha \) and \( k \)). Solving above equations usually requires iterative calculations to locate the optimum parameters. The MLE method was used in this study to estimate parameters in Weibull, Log Normal, and GEV distributions.

5.2.3 Parameters for Weibull Distribution:

Applying the MLE method to Weibull distribution, the following equations can be obtained:

\[
\ln(f(x)) = \ln \alpha + \ln \beta + (\alpha - 1)(\ln x - \ln \beta) - (x / \beta)^\alpha \tag{4.5}
\]

Where, \( \alpha \) and \( \beta \) are parameters to be estimated, \( x \) is the data set in studied sites.

So,

\[
L(\alpha, \beta) = \frac{1}{N} \sum_{i=1}^{N} \ln(f(x_i)) = \frac{1}{N} \sum_{i=1}^{N} \left[ \ln \alpha + \ln \beta + (\alpha - 1)(\ln x_i - \ln \beta) - (x_i / \beta)^\alpha \right] \tag{4.6}
\]

According to the process of MLE method, let

\[
\frac{\partial L(\alpha, \beta)}{\partial \alpha} = 0; \quad \frac{\partial L(\alpha, \beta)}{\partial \beta} = 0. \tag{4.7}
\]

Then we get,
\[
\frac{1}{\alpha} = \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{x_i}{\beta} \right)^{\alpha} \ln \left( \frac{x_i}{\beta} \right) + \ln \left( \frac{x_i}{\beta} \right) \right]; \quad (4.8)
\]

\[
\beta = \left( \frac{N}{\sum_{i=1}^{N} x_i^{\alpha}} \right)^{\frac{1}{\alpha}}; \quad (4.9)
\]

For the data set \( x \) obtained from the Pensacola station, the parameters obtained from MLE method for Weibull distribution for Pensacola, \( \alpha = 2.0316 \), \( k = 0.9044 \). Fernandina, \( \alpha = 2.999 \), \( k = 1.99 \).

5.2.4 Parameters for Log Normal Distribution

The MLE method was also applied to find parameters for Lognormal distribution to obtain the following equations:

\[
\ln f(y) = -\ln(x \cdot \sigma_{y} \cdot \sqrt{2\pi}) - \frac{(y - \mu_{y})^2}{2\sigma_{y}^2} \quad (4.10)
\]

Where, \( y = \ln(x) \).

Then,

\[
L(\mu_{y}, \sigma_{y}) = \frac{1}{N} \sum_{i=1}^{N} \left[ -\ln \sigma_{y} - \ln(x_{i} \cdot \sqrt{2\pi}) - \frac{(y_{i} - \mu_{y})^2}{2\sigma_{y}^2} \right] \quad (4.11)
\]

Follows the MLE process, we get

\[
\mu_{y} = \sum_{i=1}^{N} y_{i} ; \quad (4.12)
\]

\[
\sigma_{y}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_{i} - \mu_{y})^2 \quad (4.13)
\]

Solving Eq 4.16 and 4.17, \( \mu_{y} = -0.3025 \), \( \sigma_{y} = 0.3534 \) for Pensacola Station; and \( \mu_{y} = 0.6176 \), \( \sigma_{y} = 0.1277 \) Fernandina Station.
5.2.5 Parameters for GEV Distribution

The generalized extreme value (GEV) distribution has cumulative distribution function (Hirose 1996).

\[
F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}
\]

For \(1 + \xi (x - \mu)/\sigma > 0\), where \(\mu \in R\) is the location parameter, \(\sigma > 0\) the scale parameter and \(\xi \in R\) the shape parameter.

The density function is, consequently

\[
f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}
\]

Again, for \(1 + \xi (x - \mu)/\sigma > 0\).

The shape parameter \(\xi\) governs the tail behavior of the distribution, the sub-families defined by \(\xi \rightarrow 0\), \(\xi > 0\), and \(\xi < 0\) correspond, respectively, to the Gumbel, Frechet, and Weibull families. The MLE method was applied to obtain parameters for GEV distributions as given below: For Pensacola, \(\xi = 0.344\), \(\sigma = 0.149\), \(\mu = 0.630\). For Fernandina, \(\xi = 0.205\), \(\sigma = 0.156\), \(\mu = 1.744\). Because \(\xi > 0\), it is the Frechet type distribution.

5.3 Comparison with Observations of Annual Maximum Water Levels

Plotting position for observed data can be obtained by the following equation (Chow et al., 1988):

\[
T_m = (1 + n)/m
\]

(4.16)

Where, \(T_m\) is the return year corresponding to the \(m^{th}\) annual maximum water level \(x_m\), \(m\) is the ranking of the water level data from high to low, \(n\) is the length (or the total number) of the data set. The each annual maximum water level over the 82 year period for Pensacola
Station (Fig 2-2), the corresponding return year was calculated and given in Table 2-2. In the same way, plotting position of return year for Fernandina Station can be determined for the annual maximum water levels (Fig 2-3) over the 94 year period.

For plotting results from probability models, the return year $T$ corresponding to the annual maximum water level $x$ can be obtained from the following equation (Chow et al., 1988) in terms of CDF of the distribution $F(x)$:

$$T = 1/(1 - F(x))$$  \hspace{1cm} (4.17)

For each water level value $x_i$, the corresponding return period $T_i$ can be obtained from Equation (4.17) for plotting $T_i$ - $x_i$ graphs. Before plotting, we need datum shift back before plotting, $x = x + x_0$. For Weibull and Lognormal distributions, $x_0$ is the mean high water level from NOAA observations: $x_0 = MHHW$. Then, data are converted to NGVD datum. After estimating the parameters of each distribution using FEMA recommended Maximum-Likelihood Estimation method, we can get $F(x)$ for each model. Plots of $T_i$ verse $x_i$ for each method are shown in Fig. 2-4 and Fig. 2-5 for Pensacola and Fernandina, respectively.

As shown in Fig. 2-4 and Fig. 2-5, predictions from traditional probability models generally follow the data trend for return years less than 20 years. Table 2-3 shows that the correlation coefficients are more than 0.85 for all models. In Fernandina site, the correlation coefficients even higher than 0.96 for Gumbel and Lognormal models and the RMSE are less than 0.1 meter for these two models in Fernandina.

The major deficiency for all these methods is that, as the return years increase, predictions deviate from the data trend at the high end. The maximum error occurs in corresponding to the maximum return year available in the data. For the Pensacola station, the observed annual maximum water level with an 83-year of return period is 3.11 meter NGVD, which was recorded during hurricane Ivan in 2004. Predictions of annual maximum water level from Gumbel, Weibull, Lognormal and GEV methods are much low 1.990 m, 1.786 m, 1.673 m, and 2.146m, respectively. The errors are above 31 % for all four methods. For the Fernandina station, the observed annual maximum water level occurred in 1898 with return period of 95 years is 3.30 m in comparison to the predictions of 2.694 m, 2.585 m, 2.566 m, 2.850 m from Gumbel, Weibull Lognormal and GEV methods, respectively. The errors range from 13.6% to 22.2%.
Comparing long-term observed data with return period of 83 and 95 years at two stations on Florida coast shows the high frequency data and low frequency data are not consistent on the single fitted line by using Gumbel, Weibull, lognormal and GEV models. In analysis, the high frequency data are mostly caused by astronomical water levels such as tides. However, the low frequency data are mainly caused by storm surges, especially top high water levels which are induced by major hurricanes. During hurricanes, the storm surge is affected by astronomical tides, moving lower pressure, winds, and heavy runoff from rainfall. As a result, the water level during a hurricane year is usually much higher than the maximum water level during the years without hurricanes. This phenomenon is quite obviously in Pensacola, where the two highest annual maximum water levels are 3.11 m and 2.56 m. However, more than 90% of annual maximum data are in the range of 0.3 m to 1.52 m. Similarly, the highest water elevation in Fernandina is 3.33 m, but the rest of data are higher than 1.52 m and lower than 2.13 m. These show that characteristics of annual maximum water level distributions in hurricane affected Florida coastal area are quite different from those in Pacific coast as given in Sobey’s study (2005), in which most common extreme value distributions have been widely tested.

6. A Proposed Frequency Analysis Method

As discussed above, the physical characteristics of the annual maximum water level in coastal areas vulnerable to hurricanes may be different from those in Pacific coast. Although the study by Sobey (2005) indicates that Lognormal model may perform better in extreme water level prediction for the Pacific coast of USA, results given above show that traditional methods are unable provide satisfactory predictions of 1% annual maximum flood elevations in Pensacola and Fernandina Stations (Fig 2-4 and Fig 2-5). By analyzing the data distributions, a new probability distribution algorithm was proposed as given below and tested with observations. Assume \( x \) is the observed annual maximum coastal water level, and \( T \) is the return year; the following the quadratic polynomial is proposed to describe the relationship between \( x \) and \( T \):

\[
x = aT_i^2 + c \tag{5.1a}
\]

Where, \( T_i = \ln T \) \tag{5.1b}
Parameters $[a, c]$ can be obtained from Least-Square fit of the quadratic polynomial for a given data set of $[x, T]$. There are 82 annual maximum water level data for Pensacola Station (Fig 2-2) and 94 annual maximum water level data for Fernandina Station (Fig 2-3). For each annual maximum water level, the corresponding return year was determined from the plotting position equation (Eq. 4.16). Then each data point $[x, T]$ represents an annual maximum water level and the corresponding return year. With more than 82 data points of $[x, T]$ for Pensacola Station (Table 2-2) and 94 data points of $[x, T]$ for Fernandina Station, it is possible to find 2 parameters of $[a, c]$ in the 2nd order polynomial equation given in Equation 5.1a.

The probability density function (PDF) and cumulative distribution function (CDF) corresponding Equation 5.1a can be derived by the following step by step procedures:

1. Based on Equation 5.1a, 5.1b and Equation 4.16, we can define $T_i \geq 0$ and $x \geq c$.

2. Eq 5.1a can be written as:

$$\frac{(x-c)}{a} = T_i^2 \quad (T_i \geq 0, \ x \geq c) \quad (5.2)$$

   Where $[T_i=0, x=c]$ is the extreme point of the parabola.

3. Find the solution of $T_i$ in term of $x$ for $T_i \geq 0, \ x \geq c$ by solving Equation 5.2.

   $$T_i = \left(\frac{x-c}{a}\right)^{1/2} \quad (5.3)$$

4. Substitute Eq 5.1b into Eq 5.3, the following explicit function is obtained

   $$T = e^{\left(\frac{x-c}{a}\right)^{1/2}} \quad (5.4)$$

5. Based on the definition (Chow et al., 1988), Cumulative Distribution Function $F(X \leq x)$ and Probability Density Function $f(x)$ can be defined as

   $$F(X \leq x) = 1 - \frac{1}{T} \quad (5.5a)$$
   $$f(x) = \frac{dF(X \leq x)}{dx} \quad (5.5b)$$

6. Substitute Equation 5.4 into Equation 5.5a, the cumulative distribution function $F(X \leq x)$ for annual maximum water levels is derived as shown in Equation 5.6a. Substitute Equation 5.6a into 5.5b, the Probability Density Function $f(x)$ is obtained as shown in Equation 5.6b.
Equations 5.6a and 5.6b are similar to Weibull distribution as shown in Table 2-1. In the case of \( c=0 \), the probability density function and cumulative distribution function given in Equations 5.6a and 5.6b are reduced to the special case of the 2-parameter Weibull distribution with \( k=1/2 \) (as shown in Table 2-1). However, parameters in the probability distribution given in Equation 5.6a and 5.6b have different geometric and physical meanings. Instead of solving the parameters in the probability density function given in Eq 5.6 using the maximum-likehood estimation method, the estimation of the parameters \([a, c]\) of the 2\(^{nd}\) order polynomial of Equation 5.1a now be conveniently obtained by the common least square fit method using the observation data of annual maxima water levels \((x)\), and log values of return year \((\ln T)\) or \((T_i)\).

There are 82 data points of \([x, T_i]\) for Pensacola Station, and 94 data points \([x, T_i]\) for Fernandina Station. Two parameters of \([a, c]\) in Equation 5.1a can be conveniently determined by least square fit. Once parameters \([a, c]\) are determined, the PDF and CDF functions given in Equation 5.6a and Equation 5.6b can be described. For Pensacola, applying least square fit method with 82 data points of \([x, T_i]\) to Equation 5.1a, parameters of \(a = 0.121\), \(c = 0.570\) were obtained. For Fernandina station, applying least square fit method with 94 data points of \([x, T_i]\) to Equation 5.1a, parameters of \(a = 0.0737\), \(c =1.739\) were derived.

After obtaining the coefficients of \([a,c]\), the probability density function (Equation 5.6b) and cumulative distribution function (Equation 5.6a) are plotted against observations in Fig 2-6 and Fig 2-7 for Pensacola station and Fernandina stations, respectively. Results show that the PDF and CDF proposed in this study as given in Equations 5.6b and 5.6a reasonably match with
observations. The special features probability distributions of coastal flood are adequately characterized. For Pensacola station, predictions from the proposed probability equations follow majority of data falling below 1 m water level. It is also extends to capture the maximum recorded annual maximum water level at 3.11 m with 83 return year. For Fernandina station, the PDF and CDF also reasonably reproduced the probability distributions of the observed water levels.

Annual maximum water levels are plotted against return years in Fig 2-8 and Fig 2-9. For both Pensacola and Fernandina stations, predictions from probability distribution equations presented in this study are much more accurate than those from traditional models. The predictions match well with the general trend of data sets in both low end and high end of the data sets. Correlation coefficients as given in Table 2-3 are 0.98 and 0.92 for Pensacola and Fernandina, respectively. Compared to those obtaining from traditional methods, the present method results in the best correlation coefficient and the lowest root-mean-square error in Pensacola Station. Although the correlation coefficient is slightly lower than other method in Fernandina station, the most important feature from the present method is that it is capable of reproducing maximum observed water levels with the longest return periods of 83 and 94 years in Pensacola and Fernandina, respectively. In other words, the method presented in this study has been tested with observations for a long-term period close to 100-year return period, which is used as a criteria for coastal flooding mitigation (FEMA, 2005).

Comparison of predictions with the highest recorded annual maximum water levels is given in Table 2-4. For Pensacola Station, model presented in this study provides the closest prediction among all models for the maximum recorded extreme water level of 83-year of return year during Hurricane Ivan in Pensacola in 2004. The prediction of 2.981 m from Xu-Huang’s model given in this paper is much close to the observation of 3.11 than 1.990 m, 1.786 m, 1.673 m, 2.146 m obtained from Gumbel, Weibull, Lognormal, and GEV methods, respectively. While all other methods produce above 30% error, the error from the presented method in this paper is 4.1% which is generally acceptable in common engineering design and planning. For Fernandina Station, Xu-Huang’s method provides an even better prediction with 0.92% error for the recorded maximum annual maximum water level of 95-year of return period that occurred during
1898. Compared with the observed 3.33 m for 95-year of return period, the prediction of 3.267 m at Fernandina station by Xu-Huang’s model is much better than 2.694m, 2.585 m, 2.566 m, 2.850 m predicted by the traditional models Gumbel, Weibull, Lognormal, and GEV methods respectively.

7. Comparison between Frequency Analysis and Simulation Method

Results of model predictions of extreme annual maximum water levels are compared to those obtained from more costly simulation method. Dean and Zhu (1986) conducted numerical hydrodynamic model modeling study in some coastal areas in Florida. In Dean and Zhu’s study, Monte Carlo method was used to characterize a storm based on probability distributions established for the parameters needed. The important parameters are: \( p_0 \) (central pressure), \( R \) (radius to maximum winds), \( v_F \) (forward speed), \( \theta \) (hurricane translation direction) and landfall or alongshore characteristics. All of the parameters were considered to be independent. Storms were constructed by randomly choosing a value for each parameter by generating a random value uniformly distributed between 0 and 1, and then entering the cumulative distribution at this value and selecting the corresponding parameter value. Each storm selected by this Monte Carlo procedure was simulated with the hydrodynamic model, and shoreline elevations are recorded. Simulating a large number of storms in this way is equivalent to simulating a long period of history, with the frequency connection established through the rate of storm occurrence estimated from a local storm sample. With the available statistics of hurricanes, a set of numerical models were applied on the side of interest with a random astronomical tide from the hurricane season as boundary condition. After the 500 storm events had been simulated, the peak water levels are ranked and the return year \( T \) is \( T=500/M \), where \( M \) is the rank of the combined total tide level (for example, if the simulation is carried for a 500 year period).

Predictions of annual maximum water levels for 100 year and 200 year return period from the simulation method by Dean and Zhu (1986) and the proposed frequency analysis method in this study are given in Table 2-5. For Pensacola Station, results from simulation method and frequency analysis are very close with about 0.2 m difference. For Fernandina station, results of extreme water level from simulation method is 0.88 m higher for 100-year event and 1.04 m
higher for the 200-year return period in comparison to those obtained from the proposed frequency analysis method. Because the frequency analysis method presented in this study has been satisfactorily validated for the annual maximum water levels for the 95-return year (close to 100 year) for Fernandina station, predictions of 100-year annual extreme water level from the frequency analysis would be more reliable than that from simulation method.

8. Conclusion

Extreme coastal flood elevation is an important factor in coastal engineering design and coastal flood mitigation. When long-term observation data of water levels are available, frequency analysis method has been recommended by FEMA (2005) in deriving extreme water levels for 100 year. However, because most data set are limited and less than 50 years in most coastal observation stations, it is usually difficult to have sufficient observed data to evaluate the accuracy of the 100 year extreme water levels predicted by traditional probability distribution methods. In this study, data from two long-term observation stations were applied to evaluate the performance of the traditional frequency analysis methods. Annual maximum water levels of 94-year data from Fernandina stations and 82-year data from Pensacola were obtained to support this study. Both stations are located in Florida, USA. Results show that traditional extreme value methods such as Gumbel, Weibull, Lognormal and GEV are able to reasonably predict the short-term return period such as less than 30-year return period, but inaccurate in predicting 100-year extreme water levels. For the maximum recorded extreme water level, predictions from Gumbel, Weibull, Lognormal, and GEV methods show of 30%-46% error in Pensacola station for the longest recorded return period of 83 years, and 11%-30% error in Fernandina station for the recorded return period of 95 years. In addition, the deviation from the data trend in the maximum recorded return years from the traditional methods raise concerns regarding the accuracy of the predictions of coastal flood with longer return periods such as 100 years or 200 years for coastal hazard mitigation.

In this study, probability distribution function and cumulative density function have been proposed, and have satisfactorily validated using long-term extreme water level data in Pensacola and Fernandina of Florida. Two parameters in the probability equation can be directly obtained from least square fit of a 2nd order polynomial of the observed water level in terms of
the logarithm of the return year. Comparison with available data indicates that the method presented in this study produces much more accurate predictions of maximum recorded extreme water levels, with 0.92% error for 95 return years in Fernandina, and 5.5% error for 83 return years in Pensacola. With good correlation coefficients of 0.98 in Pensacola station and 0.92 in Fernandina station, predictions of extreme water levels from the presented method follow well with the data trend that produces more reliable predictions of extreme water levels beyond the maximum recorded data period, such as flooding elevation in 100 or 200 return years. The method presented in this study provides a valuable reference to support coastal and water resources management community in coastal flood mitigation.

9. Acknowledgements

The authors would like to thank Maria Little in NOAA, for providing the data for this study.
Fig. 2-1. Study sites: Pensacola and Fernandina Stations, Florida, USA
Pensacola annual maximum water level during 1923-2004 (by 82 years data)

Fig. 2-2. Time series of annual maximum water levels in Pensacola, USA

Fernandina annual maximum water level during 1898-2004 (by 94 years data)

Fig. 2-3. Time series of annual maximum water levels in Fernandina, USA
Fig. 2-4. Extreme water levels predicted by traditional frequency analysis methods in Pensacola, Florida, USA.

Fig. 2-5. Extreme water levels predicted by traditional frequency analysis methods in Fernandina, Florida, USA.
Fig. 2-6. PDF and CDF in Pensacola station

Fig. 2-7. PDF and CDF of Fernandina station
Fig. 2-8. Extreme water levels predicted by Xu-Huang’s in Pensacola, Florida, USA.

Fig. 2-9. Extreme water levels predicted by Xu-Huang’s in Fernandina, Florida, USA.
Table 2-1 Traditional Probability Distributions for Extreme Values

<table>
<thead>
<tr>
<th>Probability Model</th>
<th>Range</th>
<th>PDF- $f(x)$</th>
<th>CDF- $F(x)$</th>
<th>Parameter Estimation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel (Extreme Value I)</td>
<td>$-\infty \leq x \leq +\infty$</td>
<td>$f(x) = \frac{1}{\alpha} \exp \left[ -\frac{x-\xi}{\alpha} - \exp \left( -\frac{x-\xi}{\alpha} \right) \right]$</td>
<td>$F(x) = \exp \left[ -\exp \left( -\frac{x-\xi}{\alpha} \right) \right]$</td>
<td>$\alpha = 0.78 \cdot \sigma$; $\xi = \mu - 0.5772 \alpha$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x = \xi - \alpha \ln \left[ -\ln F \right]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull (Extreme Value III)</td>
<td>$x \geq 0; \alpha, k \geq 0$</td>
<td>$f(x) = \left( \frac{k}{\alpha} \right) x^{k-1} \exp \left[ -\left( \frac{x}{\alpha} \right)^k \right]$</td>
<td>$F(x) = 1 - \exp \left[ -\left( \frac{x}{\alpha} \right)^k \right]$</td>
<td>Estimated by Maximum-Likelihood Estimation for $\alpha, k$</td>
</tr>
<tr>
<td>Log Normal</td>
<td>$-\infty \leq y \leq \infty$; $(0 \leq x \leq \infty)$</td>
<td>$f(y) = \frac{1}{x\sigma_y \sqrt{2\pi}} \exp \left[ -\frac{(y - \mu_y)^2}{2\sigma_y^2} \right]$</td>
<td></td>
<td>Estimated by Maximum-Likelihood Estimation $\mu_y$, $\sigma_y^2$</td>
</tr>
<tr>
<td>Generalized Extreme Value (GEV)</td>
<td>$-\infty \leq x \leq +\infty$; $\sigma &gt; 0$</td>
<td>$F(x; \mu, \sigma, \xi) = \exp \left{ \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right}$</td>
<td>$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma \left{ \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi}} \exp \left[ -\left( \frac{x - \mu}{\sigma} \right) \right]$</td>
<td>Estimated by Maximum-Likelihood Estimation for $\xi$, $\sigma$, $\mu$</td>
</tr>
</tbody>
</table>
Table 2-2 Plotting positions of return years $T$ and corresponding annual maximum water levels $x$

<table>
<thead>
<tr>
<th>$T$ (years)</th>
<th>$x$ (m)</th>
<th>$T$ (years)</th>
<th>$x$ (m)</th>
<th>$T$ (years)</th>
<th>$x$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>83</td>
<td>3.109</td>
<td>2.86</td>
<td>0.75</td>
<td>1.46</td>
<td>0.613</td>
</tr>
<tr>
<td>41.5</td>
<td>2.564</td>
<td>2.77</td>
<td>0.741</td>
<td>1.43</td>
<td>0.613</td>
</tr>
<tr>
<td>27.67</td>
<td>1.744</td>
<td>2.68</td>
<td>0.735</td>
<td>1.41</td>
<td>0.613</td>
</tr>
<tr>
<td>20.75</td>
<td>1.601</td>
<td>2.59</td>
<td>0.735</td>
<td>1.38</td>
<td>0.613</td>
</tr>
<tr>
<td>16.6</td>
<td>1.452</td>
<td>2.52</td>
<td>0.726</td>
<td>1.36</td>
<td>0.61</td>
</tr>
<tr>
<td>13.83</td>
<td>1.251</td>
<td>2.44</td>
<td>0.724</td>
<td>1.34</td>
<td>0.598</td>
</tr>
<tr>
<td>11.86</td>
<td>1.132</td>
<td>2.37</td>
<td>0.723</td>
<td>1.32</td>
<td>0.595</td>
</tr>
<tr>
<td>10.38</td>
<td>1.122</td>
<td>2.31</td>
<td>0.72</td>
<td>1.3</td>
<td>0.583</td>
</tr>
<tr>
<td>9.22</td>
<td>1.04</td>
<td>2.24</td>
<td>0.714</td>
<td>1.28</td>
<td>0.583</td>
</tr>
<tr>
<td>8.3</td>
<td>1.01</td>
<td>2.18</td>
<td>0.705</td>
<td>1.26</td>
<td>0.583</td>
</tr>
<tr>
<td>7.55</td>
<td>0.967</td>
<td>2.13</td>
<td>0.705</td>
<td>1.24</td>
<td>0.552</td>
</tr>
<tr>
<td>6.92</td>
<td>0.888</td>
<td>2.08</td>
<td>0.705</td>
<td>1.22</td>
<td>0.552</td>
</tr>
<tr>
<td>6.38</td>
<td>0.888</td>
<td>2.02</td>
<td>0.705</td>
<td>1.2</td>
<td>0.531</td>
</tr>
<tr>
<td>5.93</td>
<td>0.888</td>
<td>1.98</td>
<td>0.699</td>
<td>1.19</td>
<td>0.525</td>
</tr>
<tr>
<td>5.53</td>
<td>0.879</td>
<td>1.93</td>
<td>0.693</td>
<td>1.17</td>
<td>0.522</td>
</tr>
<tr>
<td>5.19</td>
<td>0.866</td>
<td>1.89</td>
<td>0.692</td>
<td>1.15</td>
<td>0.522</td>
</tr>
<tr>
<td>4.88</td>
<td>0.857</td>
<td>1.84</td>
<td>0.687</td>
<td>1.14</td>
<td>0.522</td>
</tr>
<tr>
<td>4.61</td>
<td>0.842</td>
<td>1.8</td>
<td>0.677</td>
<td>1.12</td>
<td>0.522</td>
</tr>
<tr>
<td>4.37</td>
<td>0.83</td>
<td>1.77</td>
<td>0.674</td>
<td>1.11</td>
<td>0.522</td>
</tr>
<tr>
<td>4.15</td>
<td>0.821</td>
<td>1.73</td>
<td>0.674</td>
<td>1.09</td>
<td>0.516</td>
</tr>
<tr>
<td>3.95</td>
<td>0.797</td>
<td>1.69</td>
<td>0.674</td>
<td>1.08</td>
<td>0.51</td>
</tr>
<tr>
<td>3.77</td>
<td>0.796</td>
<td>1.66</td>
<td>0.671</td>
<td>1.06</td>
<td>0.491</td>
</tr>
<tr>
<td>3.61</td>
<td>0.793</td>
<td>1.63</td>
<td>0.668</td>
<td>1.05</td>
<td>0.491</td>
</tr>
<tr>
<td>3.46</td>
<td>0.772</td>
<td>1.6</td>
<td>0.653</td>
<td>1.04</td>
<td>0.491</td>
</tr>
<tr>
<td>3.32</td>
<td>0.766</td>
<td>1.57</td>
<td>0.647</td>
<td>1.02</td>
<td>0.461</td>
</tr>
<tr>
<td>3.19</td>
<td>0.766</td>
<td>1.54</td>
<td>0.619</td>
<td>1.01</td>
<td>0.43</td>
</tr>
<tr>
<td>3.07</td>
<td>0.754</td>
<td>1.51</td>
<td>0.613</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.96</td>
<td>0.751</td>
<td>1.48</td>
<td>0.613</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2-3 RMSE and correlation coefficient $r$ for each models

<table>
<thead>
<tr>
<th>Stations</th>
<th>Method</th>
<th>Gumbel</th>
<th>Weibull</th>
<th>Lognormal</th>
<th>GEV</th>
<th>Xu-Huang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pensacola</td>
<td>$r$</td>
<td>0.8531</td>
<td>0.8519</td>
<td>0.8982</td>
<td>0.9556</td>
<td>0.9814</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.212</td>
<td>0.211</td>
<td>0.211</td>
<td>0.188</td>
<td>0.090</td>
</tr>
<tr>
<td>Fernandina</td>
<td>$r$</td>
<td>0.9617</td>
<td>0.9364</td>
<td>0.9624</td>
<td>0.9799</td>
<td>0.92304</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.078</td>
<td>0.094</td>
<td>0.084</td>
<td>0.058</td>
<td>0.109</td>
</tr>
</tbody>
</table>

Note: RMSE is the root-mean-square error in meter
$r$ is the correlation coefficient

Table 2-4 Errors In Predicting Maximum Recorded Water Levels (m)

<table>
<thead>
<tr>
<th>Maximum Recorded Elevation</th>
<th>Models</th>
<th>Gumbel</th>
<th>Weibull</th>
<th>Lognormal</th>
<th>GEV</th>
<th>Xu-Huang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pensacola 3.11 m Return year=83</td>
<td>Elevation</td>
<td>1.990</td>
<td>1.786</td>
<td>1.673</td>
<td>2.146</td>
<td>2.937</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-1.119</td>
<td>-1.323</td>
<td>-1.436</td>
<td>-0.963</td>
<td>-0.173</td>
</tr>
<tr>
<td></td>
<td>Error%</td>
<td>36%</td>
<td>42.6%</td>
<td>46.2%</td>
<td>31.0%</td>
<td>5.5%</td>
</tr>
<tr>
<td>Fernandina 3.30 m Return year=95</td>
<td>Elevation</td>
<td>2.694</td>
<td>2.585</td>
<td>2.566</td>
<td>2.850</td>
<td>3.267</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>-0.604</td>
<td>-0.713</td>
<td>-0.732</td>
<td>-0.448</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>Error%</td>
<td>18.3%</td>
<td>21.6%</td>
<td>22.2%</td>
<td>13.6%</td>
<td>0.92%</td>
</tr>
</tbody>
</table>

Note: Maximum recorded water level in Pensacola is 3.11 m in 2004.
Maximum recorded water level in Fernandina is 3.30 m in 1898
Table 2-5 Comparison of 100-year and 200-year annual maximum water levels predicted by the proposed method to those obtained from more costly and complex simulation method by Dean and Zhu in Pensacola and Fernandina (in NGVD datum)

<table>
<thead>
<tr>
<th>Stations</th>
<th>Return year</th>
<th>Simulation Method (Hydrodynamic modeling and Monte Carlo) By Dean and Zhu</th>
<th>Frequency Analysis by Xu-Huang(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pensacola</td>
<td>100</td>
<td>3.35</td>
<td>3.152</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>3.78</td>
<td>3.979</td>
</tr>
<tr>
<td>Fernandina</td>
<td>100</td>
<td>4.18</td>
<td>3.301</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>4.85</td>
<td>3.807</td>
</tr>
</tbody>
</table>
Neural Network and Harmonic Analysis for Recovering Missing Extreme Water Level Data during Hurricanes

1. Abstract

Predictions of extreme coastal water levels are important to coastal engineering design and hazard mitigations in Florida. Annual maximum water levels are often used in frequency analysis of 1% annual chance flood in coastal flood hazard mapping (FEMA, 2004). However, in many coastal areas, long history data of water levels are unavailable. In addition, some water level records may be missed due to the damage of measurement instruments during hurricanes. In this study, a method has been developed to employ artificial neural network and harmonic analysis for predicting extreme coastal water levels during hurricanes. The combined water levels are decomposed into tidal signals and storm surge. Tidal signal can be derived by harmonic analysis, while storm surge can be predicted by neural network modeling based on the observations of local wind speeds and atmospheric pressure. The neural network model employs three-layer feed-forward back propagation structure with advanced scaled conjugate training algorithm. The method presented in this study has been successfully tested in Panama City Beach and Apalachicola located in Florida coast for Hurricane Dennis and Hurricane Ivan. In both stations, model predicted peak elevations match well with observations in both hurricane events. The decomposed storm surge hydrograph also make it possible for the analysis potential extreme water levels if storm surge occurs during spring high tide.
2. Introduction

In coastal flood hazard analysis and hazard zone mapping, 1% annual chance flood is defined as the base by FEMA (2004). When observed data cover a period over 30 years including years with and without hurricanes, one of the methods recommended by FEMA (2004) for estimating 1% annual flood is the frequency analysis of observed annual maximum water levels by using Generalized Extreme Value (GEV) method. When 1% annual flood is statistically estimated by annual maximum water levels, it represents the combined effects of coastal forcing. Generally, annual maximum water level during a hurricane year occurs responding to the strongest hurricane events. Missing data of annual maximum water levels is one of the obstacles for providing reliable estimations of 1% annual flood for coastal flood hazard mitigations. During hurricane storm surges, some water level measurement station may be damaged. As a result, observations of water level data may be not be available. However, during a hurricane event, wind speeds and atmospheric pressure may be available.

In this paper, neural network and tidal harmonic analysis are presented for recovering missing extreme water levels. During hurricane events, water levels consist of mixed signals of astronomic tidal harmonic components, and storm surges induced by hurricane winds and atmospheric pressure. Tidal harmonic components can be predicted by tidal harmonic analysis of observed water levels for a period of observations before the hurricane event, while the hydrograph of the storm surge in response to the wind and atmospheric forcing can be predicted from local winds and atmospheric pressure. Decoupling tidal signal from storm surge signal provides an effective and simple approach for neural network modeling study of storm surge by identifying principal components of input forcing parameters. In addition, it allows coastal engineers and managers to examine extreme water level scenarios of mixed storm surge and spring high tide. In the following sections, a review of neural network’s applications and methodology are described. Two hurricane data sets are illustrated for model validations. Storm surge hydrograph are presented by subtracting tidal harmonic components from observed water level signals. Neural network training and verifications are presented by using two hurricane data sets in Panama City Beach and Apalachicola, Florida, USA.
3. **Review of artificial neural network applications**

Artificial neural networks (ANN) have proven their usefulness in a number of research areas such as electronics, aerospace, and manufacturing engineering (Hagan et al., 1995). An ANN can correlate multiple input variables with the output signal through nodes or neurons. It is capable of directly correlating input time series of forcing functions to the output variables through the interconnected nodes with trainable weights and bias. In contrast to traditional harmonic analysis (Ippen, 1966), which is used only in the predictions of periodic tidal component, the neural network model can be trained to recognize and predict both nonlinear and non-periodic signals.

Carl et al (2005) in their paper “Using an Artificial Neural Network to Improve Predictions of Water Levels Where Tide Charts Fail” In predicting water levels where tide charts fail, recognized wind forcing as the main variable not included in harmonic analysis. Their research has shown that Artificial Neural Network (ANN) models including input variables such as previous water levels, tidal forecasts, wind speed, wind direction, wind forecasts and barometric pressure can greatly improve water level predictions at several coastal locations including open coast and deep embayment stations. By using 24 hours of previous water levels at the studied site Rockport, 24 hours of previous water levels at nearby station Bob Hall Pier, and 12 hours of previous wind speeds and wind directions at Bob Hall Pier lead to the most accurate water level forecasts without using wind forecasts. However, In their study, the special hurricanes are not specially shown, so we can not see that this ANN model can predict water level well during hurricanes.

Phillipe Tissot, Daniel T. Cox, M.ASCE, and Patrick Michaud, in 2001, in their paper “Neural Network Forecasting of Storm Surges Along The Gulf of Mexico”. The ANN model they use to predict storm surge also selected water levels, wind stress, barometric pressure as well as tidal forecasts and wind forecasts as inputs. Models including wind forecasts outperform other models without wind forecasts input 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. 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. However, we can see the
water levels of verification are not quite high which show no storm surge induced by major hurricanes are included in verification, even so, the performance of the model are not quite good in peak water level, though the Neural Network Forecast has a good trend and much closer to measured water levels than tide tables. We can see the ANN model that include water level, wind speed, wind direction and atmosphere pressure can include more factors that may influence the water level. Thus, the ANN model used above can perform better than tide tables or could be a good compliment Where Tide Charts Fail. However there is not enough prove in their study that the ANN model mentioned before can predict high storm surge well during major hurricanes.

Huang and Murray (2003) applied the ANN in the coastal water level predictions. The RNN—WL model was developed to enable coastal engineers to predict long term water levels in a coastal inlet, based on the input of data in a remote NOAA station in the region. The RNN—WL model was tested in an application to Long Island South Shore. Located about 60–100 km away from the inlets there are two permanent long-term water level stations. The satisfactory results indicate that the RNN—WL model is able to supplement long-term historical water level data at the coastal inlets based on the available data at remote NOAA stations in the coastal region.

In applying a trained and validated ANN, output variables are directly calculated without iteration from the input variables and the vectors of weights and bias in the network nodes. This functioning is similar to directly find the output from a linear regression function. Therefore, applying an ANN model takes much less computational time than the traditional fluid mechanic models, as long as data is available to establish the ANN model. For this reason, some researchers have combined fluid mechanics modeling with neural networks to improve the efficiency of model applications (Bibike and Abbot, 1999). Greenman and Roth at the NASA Ames Research Center incorporated neural networks with finite element fluid mechanics models to optimize airfoil design (Greenman and Roth, 1999). A fluid mechanics model can be used to provide time series outputs of system responses under a few study scenarios for a period of time. The time series outputs from fluid mechanics model simulations and forcing functions can then be used as “data” for neural network model development. A validated neural network model can serve as a cost-effective tool in quickly assessing the system response to the input factors. The application of neural networks in oceanographic study is relatively new due mainly to highly nonlinear characteristics. Hsieh and Tang (1998) discussed several typical obstacles and
provided some suggestions to incorporate neural networks with other time series forecasting approaches. There are some successful ANN applications in coastal engineering. For example, Bibike et al. (1999) used ANN to encapsulate numerical hydrodynamic model simulations for cost-effective forecasting of water levels. Mase et al. (1995) adopted ANN to assess the stability of armor unit and rubble-mound breakwater and found satisfactory agreement between observations and model predictions. Huang and Foo (2000) employed neural networks to directly correlate time series of salinity to the forcing functions of winds, water levels, and freshwater inputs in a multiple-inlet estuary of Apalachicola Bay, Florida. Deo and Naidu (1999) applied ANN to perform real-time wave forecasting. Tsai and Lee (1999) conducted a study that applied ANN in tidal-level forecasting using historic data at the same station, which did not address the non-periodic sub-tidal sea levels and the correlation with tidal data at other stations.

4. **Neural network methodology**

4.1 **Multiple-layer ANN Model**

In practical applications, a neural network often consists of several neurons in several layers. A schematic diagram of a three-layer neural network is given in Fig 3-4, where $X_i$ ($i=1, ..., n$) represents the input variables (such as boundary forcing functions of wind and water levels); $Y_i$ ($i=1, ..., m$) represents the outputs of neurons in the hidden layer; and $Z_i$ ($i=1, ..., p$) represents the outputs of the neural network such as water levels and currents in and around coastal inlets. The layer that produces the network output is called the **output layer**, and all other layers are called **hidden layers**. The weight matrix connected to the inputs is called the **input weight** ($W_{ij}$) matrix, and the weight matrices coming from layer outputs are called **layer weights** ($W_{jk}$).

4.2 **Standard Network Training Using Gradient Descent Method**

Multiple-layer neural networks using backpropagation training algorithms are popular in neural network modeling (Hagan et al., 1995) because of their ability to recognize the patterns and relationships between nonlinear signals. The term backpropagation usually refers to the manner in which the gradients of weights are computed for non-linear multi-layer networks. A neural network must be trained to determine the values of the weights that will produce the
correct outputs. Mathematically, the training process is similar to approximating a multi-variable function, \( g(X) \), by another function of \( G(W,X) \), where \( X=[x_1,x_2,\ldots,x_n] \) is the input vector, and \( W=[w_1,w_2,\ldots,w_n] \) the coefficient or weight vector. The training task is to find the weight vector \( W \) that provides the best possible approximation of the function \( g(X) \) based on the training input \( [X] \).

The standard or basic training method is the **Gradient Descent Method**. In this method, weight changes move the weights in the direction where the error declines most quickly. Training is carried out by assigning random initial weights to each of the neurons (usually between 0.1 and 1.0) and then presenting sets of known input and target (output) values to the network. The network estimates the output value from the inputs, compares the model predicted output to the target value, and then adjusts the weights in order to reduce the mean squared difference between the network output and the target values. The complete input-output sets are often run through the network for several iterations (or epochs) until either the mean square error is reduced to a given level or reaches a minimum, or until the network has been trained for a given number of iterations.

If we let \( w_m \) represent the value of weight \( w \) after \( m-th \) iteration in a neuron, then

\[
w_m = w_{m-1} + \Delta w_m
\]

where \( \Delta w_m \) is the change in the weight \( w \) at the end of iteration \( m \). It is calculated by

\[
\Delta w_m = -\epsilon d_m
\]

where \( \epsilon \) is the user-specified parameter controlling the proportion by which the weights are modified. The term \( d_m \) is given by

\[
d_m = \sum_{n=1}^{n} \frac{\partial E}{\partial w_m}
\]

where \( N \) is the total number of examples and \( E \) is the simulation output error.

The basic backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient). This is the direction, in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. Therefore, the basic gradient descent training algorithm is inefficient due to its slow convergent speed and at times the poor accuracy in model predictions. Two types of algorithms that can be used to significantly improve training speed are given below.
4.3 Faster Training with Variable Learning Rate

With standard steepest descent, the learning rate is held constant throughout training. The performance of the algorithm is sensitive to the proper setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate is too small, the algorithm will take too long to converge. It is not practical to determine the optimal setting for the learning rate before training, and, in fact, the optimal learning rate changes during the training process, as the algorithm moves across the performance surface. The performance of the steepest descent algorithm can be improved if we allow the learning rate to change during the training process. An adaptive learning rate will attempt to keep the learning step size as large as possible while keeping learning stable. The learning rate is made responsive to the complexity of the local error surface. An adaptive learning rate requires some changes in the training procedure. First, the initial network output and error are calculated. At each epoch new weights and biases are calculated based on the current learning rate. New outputs and errors are then calculated. If the new error exceeds the old error by more than a predefined ratio, the new weights and biases are discarded. In addition, the learning rate is decreased. Otherwise the new weights, etc., are kept. If the new error is less than the old error, the learning rate is increased. This procedure increases the learning rate, but only to the extent that the network can learn without large error increases. Thus a near optimal learning rate is obtained for the local terrain. When a larger learning rate could result in stable learning, the learning rate is increased. When the learning rate is too high to guarantee a decrease in error, it gets decreased until stable learning resumes.

4.4 Faster Training by Network Optimization with Conjugated Gradient Algorithm

In practical ANN applications, optimized training algorithms are often used. There are several optimized training algorithms as described by Haykin (1999), such as scaled conjugated gradient backpropagation. One of the optimized methods is the scaled conjugate gradient algorithm (SCG) (Moller, 1993; Hagan, 1995). The SCG optimized training algorithm was designed to avoid the time consuming line search. In the conjugate gradient algorithm a search is performed along conjugate directions, which produces faster convergence than steepest descent directions. It differs from the previously mentioned error backpropagation in gradient calculations and subsequent corrections to weights and bias. Here, a search direction, $d_k$, is computed at each
training iteration, k, and the error function, \( f(w) \), is minimized along it using a line search. The gradient descent does not move down the error gradient as in the above backpropagation method, but along a direction that is conjugate to the previous step. The change in gradient is taken as orthogonal to the previous step, with the advantage that the function minimization, carried out in each step, is fully preserved due to lack of any interference from subsequent steps. The basic iteration process (Fitch, et al., 1991; Hagan 1995) is given as follows:

1. Initialize weight vector, \( \overrightarrow{w} \), by using uniform random numbers from the interval (-0.5, 0.5).
   Calculate error gradient, \( \overrightarrow{g}_0 \), at this point. Select initial search direction
   \[ \overrightarrow{d}_0 = -\overrightarrow{g}_0 \]  
   (4)

2. For each iteration \( k \), determine constant \( \alpha_k \), which minimizes the error function \( f(\overrightarrow{w} + \alpha_k \overrightarrow{d}_k) \) by line search where \( \overrightarrow{d}_k \) is the search direction at iteration \( k \). Update the weight vector \( \overrightarrow{w}_k \) to \( \overrightarrow{w}_{k+1} \) using:
   \[ \overrightarrow{w}_{k+1} = \overrightarrow{w}_k + \alpha_k \overrightarrow{d}_k \]  
   (5)

3. If error at this iteration, \( k+1 \), is acceptable, or if specified number of computations of the function and gradients is reached, terminate the algorithm.

4. Otherwise obtain new direction vector, \( \overrightarrow{d}_{k+1} \):
   \[ \overrightarrow{d}_{k+1} = -\overrightarrow{g}_{k+1} \]  
   (6)
   if \( k + 1 \) is an integral multiple of \( N \), where \( N \) is the dimension of \( \overrightarrow{w} \). Otherwise,
   \[ \overrightarrow{d}_{k+1} = -\overrightarrow{g}_{k+1} + \beta_k \overrightarrow{d}_k \]  
   (7)
   where
   \[ \beta_k = \frac{(\overrightarrow{g}_k \overrightarrow{g}_k^T)}{(\overrightarrow{g}_{k-1} \overrightarrow{g}_{k-1})} \]  
   (8)

If the algorithm is not converged, continue from step (2) for next iteration
In addition to above basic iteration algorithm, Hagan (1995) introduces function comparison and linear search methods to locate the minimum of a function in a specified direction. For quadratic functions the algorithm will converge to the minimum in at most $k$ iteration, where $k$ is the number of parameters to be optimized. The mean-squared-error performance index for multilayer networks is not quadratic, therefore the algorithm would not normally converge in $k$ iteration. The development of the conjugate gradient algorithm does not indicate what search direction to use once a cycle of $k$ iterations has been completed. There are many procedures suggested, Hagan (1995) recommend the simplest method to reset the search direction to steepest decent direction after convergence at $k$ iterations. More details regarding this complex algorithm can also be found in Fitch et al. (1991) and Moller (1993).

5. Data Analysis and Separating Tidal Signal from Storm Surges

5.1 Data Sets at Two Study Sites with Two Hurricane Events

Two study sites, Panama City Beach and Apalachicola (Fig 3-1) located in Florida coast waters, have been selected in this study because two hurricane data sets are available for each of the study site. Two hurricanes include Hurricane Ivan in 2004 and Hurricane Dennis in 2005. For each hurricane at each study site, hourly wind speeds and directions, atmospheric pressure, water levels were obtained from NOAA. Data set for one hurricane event was used for ANN model training, while another data set was used for model verification. Water levels obtained from NOAA measurement stations are total water levels combined with effects of tides, winds, and atmospheric pressure. In order to effectively predict storm surge induced by hurricane winds and atmospheric pressure, data processing was conducted to filtering tidal effects. To make ANN model work better, the input variables are scaled before applied in the model. ANN model always work well for the variables scaled between -1 and 1. Variables were scaled for ANN model simulations for wind $u$ and wind $v$, water level $x$, air pressure $p$ as described below:

$$u_{\text{scale}} = 0.8 \times u / \max(u)$$
$$v_{\text{scale}} = 0.8 \times v / \max(v)$$
$$p_{\text{scale}} = -0.8 \times (p - p_0) / \max(abs(p - p_0))$$
where, \( u, v, p \) and \( x \) are the variables before scaled, \( u_{scale}, v_{scale}, p_{scale} \) and \( x_{scale} \) are the input variables after scaled. \( \max(u) \) and \( \max(v) \) are the maximum velocity of wind \( u \) and \( v \) of both training and verification cases. \( p_0 \) is atmosphere pressure, \( \max(abs(p - p_0)) \) is the maximum difference of \( p \) and \( p_0 \) for both training and verification cases. \( \max(x) \) is the maximum water level of both training and verification cases.

### 5.2 Separating Storm Surge hydrographs and Tidal Signals from Observed Water Levels

Tidal signals can be predicted by harmonic analysis by least square method (Boon and Killey, 1978; Huang et al 2002). Tidal harmonic analysis is based on a decomposition of the observed tidal height into basic periodic tidal harmonic components.

\[
H(t) = H_0 + \sum_{n=1}^{N} f_n H_n \cos\left(a_n t + (V_0 + u)_n - k_n\right)
\] (9)

Where \( H(t) \) = height of the tide at any time \( t \); \( H_0 \) = mean water level above some defined datum such as mean sea level; \( H_n \) = mean amplitude of tidal constituent \( n \); \( f_n \) = factor for adjusting mean amplitude \( H_n \) values for specific times; \( a_n \) = speed of constituent \( n \) in degrees/unit time; \( t \) = time measured from some initial epoch or time, i.e., \( t = 0 \) at \( t_0 \), \( (V_0 + u) \) = value of the equilibrium argument for constituent \( n \) at some location and when \( t = 0 \); \( k_n \) = epoch of constituent \( n \), i.e., phase shift from tide-producing force to high tide from \( t_0 \) , the harmonic components of the studied period in Panama City Beach and Apalachicola are shown in Table 3-1, Table 3-2, Table 3-3 and Table 3-4.

A period of water levels at normal condition before the hurricane events as shown in Fig 3-3 can be used to derive tidal harmonic components. Once the tidal components have been obtained, the tidal water level can then be predicted to extend to the period covering hurricane events. By subtracting tidal water levels predicted from harmonic analysis from observed water levels during hurricane events, storm surge hydrograph induced by winds and atmospheric pressure can be derived. For Panama City Beach Station, storm surge hydrograph is given in Fig 3-4 for Hurricane Dennis and in Fig 3-5 for Hurricane Ivan, which was derived by filtering tidal
signals predicted from tidal harmonic analysis. For Apalachicola Station, storm surge hydrograph is given in Fig 3-6 for Hurricane Dennis and in Fig 3-7 for Hurricane Ivan, which was derived by filtering tidal signals predicted from tidal harmonic analysis.

6. ANN Modeling of Storm Surge Affected by Winds and Atmospheric Pressure

6.1. ANN Network Design for Modeling Storm Surge

A network model was designed to correlate the storm surge hydrograph to forces of winds and atmospheric pressure. Wind effects were accounted for by the north-south and east-west directions of wind vectors. A three-layer feed-forward backpropagation network (Haykin, 1999) with a nonlinear differentiable log-sigmoid transfer function in the hidden layer (Fig 3-2) was employed. To avoid network overfitting, Fletcher and Goss (1993)'s approximation was applied to approximately estimate the number of neurons in the hidden layer. The network was trained using the data set and then validated with another data set. Through sensitivity study, the optimal network size was selected as that size which resulted in the minimum error and maximum correlation in the validation data set. After a series of sensitivity tests, a network with 25 neurons in the hidden layer was adopted. Data analysis indicates that there are phase differences between peak storm surge hydrograph and the boundary forcing. In order to account for the phase difference between inputs and outputs, four consecutive hourly data points from the input time series of wind speed components \([u,v]\) and atmospheric pressure were used in the ANN model to predict currents at the given time step. Four consecutive hourly data were used in this study because the maximum phase difference between currents and water levels is less than four hours, which will allow the ANN network to have sufficient historic data at each time step to train and recognize the historic pattern and phase lag between inputs and outputs. This is similar to the autoregressive and moving average variables in stochastic modeling. In physical meaning, this means that previous conditions wind and atmospheric pressure have effects on storm surge elevations at current time step. The neural network for hurricane storm surge can be illustrated in the following equation to account for the effects of wind speed \([u,v]\) and atmospheric pressure \([p]\) on the storm surge elevation \(\eta (t_i)\) at time step \(t_i\).
\[ \eta(t_i) = ANN \begin{bmatrix} u(t_i), u(t_{i-1}), u(t_{i-2}), u(t_{i-3}) \\ v(t_i), v(t_{i-1}), u(t_{i-2}), v(t_{i-3}) \\ p(t_i), p(t_{i-1}), p(t_{i-2}), p(t_{i-3}) \end{bmatrix} \] (10)

Where \([t_i]\) is current hour, and \([t_{i-1}, t_{i-2}, t_{i-3}]\) are the corresponding 1, 2, 3 consecutive hours before the current time step; \(u\) is the east-west component of wind speed, and \(v\) is the north-south component of wind speed. The ANN network structure designed for this study is given in Fig 3-8. For two study sites, the measurement stations are located in the coastal where effects of river flow on water levels are not significant. Therefore, river flow was not used as an input factor.

Several model training method have been tested in this study. To compare different model training techniques, model simulations were conducted with a network consisting with 25 neurons and 0.01 training goal. Using the standard gradient decent training algorithm, model convergence speed is very slow. By applying a variable learning rate, model convergence speed increases. Scaled conjugate training algorithm has the advantage of requiring less epochs and CPU time, yet achieves the highest correlation value and lowest error. Therefore, the scaled conjugate training algorithm was selected for the neural network model in the following sections.

6.2 ANN Model Training and Verification in Panama City Beach

Data set from Hurricane Dennis was used in model training. Using wind speeds and atmospheric pressure, model simulations were performed with initially defined coefficients. By comparing model outputs of water levels to observations, the coefficient matrix of weight and bias for each layer of the network was adjusted so as to reduce the root-mean-square (RMS) error between model predictions and observations. The network training was completed when the RMS error was reduced to an allowed value. The training method and the number of neurons used in the model have effects on the model performance. An appropriate number of neurons, not too few or too many, will provide the best performance. A model with too few neurons may not be able to realistically describe the non-linear feature, while too many neurons in the network may cause overfitting of data and miss the generality of the physical characteristics. In this study, model sensitivity studies were conducted to determine optimal number of neurons that provides better predictions. Advanced conjugate descent back propagation method was used in model training. For each iteration, the weights and biases of the network were updated in the conjugate
directions, which produced faster convergence than the traditional steepest descent directions. During the training period, model coefficients were adjusted until model predictions of water levels fit well with the target observations. Fig 3-9 shows the time series of hourly wind speeds and atmospheric pressure as inputs to the ANN model, and the outputs of storm surge during hurricane Dennis in 2005. It indicates that the ANN model coefficients were adequately adjusted to reasonably reproduce the storm surge hydrograph during the training period.

In the model verification phase, the 2nd data set for Hurricane Ivan was used to verify the ANN model. Using the coefficient matrixes of weight and bias trained from the first data set for Hurricane Dennis, the model directly predicts the corresponding storm surge hydrograph from the forcing inputs of u and v components of wind speeds and atmospheric pressure p. Similar to applying an empirical regression equation, the neural network model simulation in verification phase required no iterative computation once the model was satisfactorily trained. As shown in neural network structure in Equation 10, four consecutive hourly data points from the input time series of wind speed components [u,v] and atmospheric pressure were used in the ANN model to predict currents at the given time step to account for the effects of previous forcing conditions. Comparison of the model predictions of storm surge hydrograph with observations is given in Fig 3-10. Model prediction of storm surge hydrograph shows good agreement with observations. The correlation between observed and predicted hydrograph of storm surge is 0.96 (shown in Table 3-6) for Hurricane Ivan in model verification. The difference of peak water levels during the storm surge is 0.01m (shown in Table 3-5), or 0.4%, which is generally considered as acceptable in coastal engineering design and analysis.

6.3 ANN Model Training and Verification in Apalachicola

The neural network model was further tested in another study site, Apalachicola. Using the same model training and verification techniques that were applied to Panama City Beach station, the neural network model was trained using the data set from Hurricane Dennis, and then verified using another data set from Hurricane Ivan. Fig 3-11 shows time series of wind speeds and atmospheric pressure used in the model training, and the resulting storm surge hydrograph from the model training period during Hurricane Dennis in 2005. Fig 3-12 presents the inputs of
wind speeds and atmospheric pressure and the resulting storm surge hydrograph for the verification period during the Hurricane Ivan in 2004. Because Apalachicola located in a longer distance than Panama City from the Center of Hurricane Ivan, atmospheric pressure changes in Apalachicola during Hurricane Ivan was less. There were some difference between model predicted water levels and observations before and after the storm surge. This may due to the stronger correlation between peak surge elevation with peak pressure deficit and wind speeds. For small fluctuations of water levels near the beginning and the end of the hurricane period, the correlation pattern may be weaker that may lead to the less accurate from ANN model predictions. However, the overall patterns of the storm surge hydrograph from model predictions and observations are very similar. The peak storm surge hydrograph from model predictions and observations are very close (shown in Table 3-5), which provide useful supplemental data for frequency analysis of coastal flood.

7. **Extreme Water Levels from Combined Storm Surge and Tides**

When missing data of historic storm surge hydrograph has been derived from observed data of wind speeds and atmospheric pressure, extreme water levels during a hurricane event can be recovered by superposition of storm surge and harmonic tides. For the examples given in this study, combined water levels for Hurricane Ivan are given in Fig 3-13 for Panama City Beach, and given in Fig 3-14 for Apalachicola. The maximum combined water levels obtained from neural network and harmonic analysis of tidal components match well with observations at both Panama City and Apalachicola during Hurricane Ivan.

Tidal amplitude in spring tide is usually much stronger than that in neap tide. When storm surge occurs during spring high tides, the extreme water level scenario during a hurricane event can be derived. As shown in Fig 3-15 and Fig 3-16, Hurricane Ivan occurred in neap tide in 2004. If it happened during spring tidal condition, the peak extreme water levels would be about 0.22 m higher in Panama City Beach and Apalachicola (shown in Table 3-7). The difference may be much higher in other coastal areas if tidal amplitudes are large.
8. Conclusion

The technique using neural network in conjunction with harmonic analysis for recovering extreme water levels during hurricane events has been presented in this paper. In coastal flood hazard analysis and mapping, annual maximum coastal water levels, which usually occur during hurricanes, are recommended by FEMA (2004) for usages in frequency analysis for predicting 100 year flood elevation. However, during hurricane events, some water measurement stations may be damaged and some annual maximum water level data may be missed. In the proposed method, extreme water levels during hurricane events are decomposed into tides which can be predicted by tidal harmonic components, and storm surges which can be predicted from neural network based on inputs of winds and atmospheric pressure. The method has been satisfactorily tested in two study sites, Panama City Beach and Apalachicola, located in Florida. For each study site, the neural network model for storm surge was trained using a data set obtained from Hurricane Dennis in 2005, and verified by another independent data set obtained from Hurricane Ivan in 2004.

In the application of the proposed method, harmonic analysis was firstly applied to extract tidal signals from observed coastal water levels, which combined astronomical and storm surge. Before the hurricane event, water levels for a period of normal water levels were used to obtain tidal harmonic components by harmonic analysis. Storm surge hydrograph was then derived by subtracting tidal predictions from observed water levels during hurricane events. By removing tidal signals, storm surge hydrographs induced by forces of winds and atmospheric pressure were obtained from the time series of observed water levels in the study sites of Panama City Beach and Apalachicola for Hurricane Dennis in 2005 and Hurricane Ivan in 2004. The three-layer feed-forward backpropagation neural network developed for storm surge predictions has been tested using two data sets in Panama City Beach and Apalachicola. For each station, the first data set for Hurricane Dennis was used in neural network model training, and the 2nd data set for Hurricane Ivan was used for model verification. The ANN model employs the inputs of hourly local atmospheric pressure and wind speeds to predict time series of storm surge hydrograph. Optimized method of scaled conjugate training algorithm was employed in the neural network model training. Results from model verifications using the 2nd data set indicate that model
predictions of storm surge hydrograph reasonably follow the general trend of the observations. The peak elevations of the storm surge were satisfactorily predicted by the ANN model using the atmosphere pressure and wind speed in both study sites.

Predictions of combined extreme water levels during hurricane events can be re-composed by adding tidal harmonic components to the storm surge hydrograph obtained from neural network predictions. In this way, missing extreme water level data during hurricane events can be recovered by using the supplemental data resulted from harmonic analysis and neural network modeling. In addition, risk assessment can be conducted for analyzing the extreme scenario if peak storm surge occurs during spring high tide.

9. Acknowledgement

This study is funded by US National Oceanic and Atmospheric Administration (NOAA) though Florida Hurricane Alliance
Fig. 3-1. Study Areas in the coastal of North Florida

Fig. 3-2. A three-layer feed-forward neural network for multivariate signal processing.
Fig. 3-3. Harmonic analysis for Panama City Beach during 05/01/05 to 07/31/05:

Fig. 3-4. Harmonic filtering of tidal signals from storm surge for Hurricane Dennis 2005 in Panama City Beach
Fig. 3-5. Harmonic filtering of tidal signals from storm surge for Hurricane Ivan 2004 in Panama City Beach

Fig. 3-6. Harmonic filtering of tidal signals from storm surge for Hurricane Dennis 2005 in Apalachicola
Fig. 3-7. Harmonic filtering of tidal signals from storm surge for Hurricane Ivan 2004 in Apalachicola

Fig. 3-8. Neural network structure for predicting storm surges using wind and atmospheric pressure after harmonic filtering tidal signal
Fig. 3-9. Panama City Beach for Hurricane Dennis: ANN model training using wind and atmospheric pressure to predict storm surges

Fig. 3-10. Panama City Beach for Hurricane Ivan: ANN model verification using wind and atmospheric pressure to predict storm surges
Fig. 3-11. Apalachicola for Hurricane Dennis: ANN model training using wind and atmospheric pressure to predict storm surges

Fig. 3-12. Apalachicola for Hurricane Ivan: ANN model verification using wind and atmospheric pressure to predict storm surges
Fig. 3-13. Compositions of water levels from storm surges and tidal variations in Panama City Beach during Hurricane Ivan

Fig. 3-14. Compositions of water levels from storm surges and tidal variations in Apalachicola during Hurricane Ivan
Fig. 3-15. Scenarios of extreme water levels composed by spring high tides and ANN predicted storm surges in Panama City Beach. a) harmonic tides, b) storm surge, c) combined tides and storm surge.

Fig. 3-16. Scenarios of extreme water levels composed by spring high tides and ANN predicted storm surges in Apalachicola. a) harmonic tides, b) storm surge, c) combined tides and storm surge.
Table 3-1 Tidal Harmonic components in Panama City during Hurricane Dennis starting from local time 0.00 hour on 05/01/07 to 23:00 hour on 07/31/07

<table>
<thead>
<tr>
<th>Component</th>
<th>Period (hour)</th>
<th>am (m)</th>
<th>Phase (radian)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>3</td>
<td>0.0001</td>
<td>1.7753</td>
</tr>
<tr>
<td>M4</td>
<td>6.21</td>
<td>0.0069</td>
<td>1.2465</td>
</tr>
<tr>
<td>M2</td>
<td>12.42</td>
<td>0.0290</td>
<td>-1.1741</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0267</td>
<td>2.2939</td>
</tr>
<tr>
<td>N2</td>
<td>12.66</td>
<td>0.0063</td>
<td>-0.6940</td>
</tr>
<tr>
<td>K2</td>
<td>11.97</td>
<td>0.0092</td>
<td>0.7045</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0072</td>
<td>-1.0352</td>
</tr>
<tr>
<td>L2</td>
<td>12.19</td>
<td>0.0026</td>
<td>0.9772</td>
</tr>
<tr>
<td>K1</td>
<td>23.93</td>
<td>0.1585</td>
<td>-1.6236</td>
</tr>
<tr>
<td>O1</td>
<td>25.82</td>
<td>0.1647</td>
<td>-0.6435</td>
</tr>
<tr>
<td>P1</td>
<td>24.07</td>
<td>0.0499</td>
<td>2.8902</td>
</tr>
<tr>
<td>Q1</td>
<td>26.87</td>
<td>0.0287</td>
<td>-0.6989</td>
</tr>
<tr>
<td>M1</td>
<td>24.86</td>
<td>0.0045</td>
<td>-2.4316</td>
</tr>
<tr>
<td>J1</td>
<td>23.10</td>
<td>0.0112</td>
<td>-1.8943</td>
</tr>
</tbody>
</table>

Table 3-2 Tidal Harmonic components in Panama City during Hurricane Ivan starting from local time at 0.00 hour on 07/01/04 to 23:00 hour on 09/30/04

<table>
<thead>
<tr>
<th>Component</th>
<th>Period</th>
<th>am</th>
<th>phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>3</td>
<td>0.0003</td>
<td>0.3221</td>
</tr>
<tr>
<td>M4</td>
<td>6.21</td>
<td>0.0041</td>
<td>0.8719</td>
</tr>
<tr>
<td>M2</td>
<td>12.42</td>
<td>0.0307</td>
<td>1.4386</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0234</td>
<td>2.0739</td>
</tr>
<tr>
<td>N2</td>
<td>12.66</td>
<td>0.0048</td>
<td>1.5942</td>
</tr>
<tr>
<td>K2</td>
<td>11.97</td>
<td>0.0056</td>
<td>-0.8391</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0046</td>
<td>-2.0474</td>
</tr>
<tr>
<td>L2</td>
<td>12.19</td>
<td>0.0018</td>
<td>-2.7600</td>
</tr>
<tr>
<td>K1</td>
<td>23.93</td>
<td>0.1582</td>
<td>-2.6050</td>
</tr>
<tr>
<td>O1</td>
<td>25.82</td>
<td>0.1513</td>
<td>2.9724</td>
</tr>
<tr>
<td>P1</td>
<td>24.07</td>
<td>0.0424</td>
<td>-2.4697</td>
</tr>
<tr>
<td>Q1</td>
<td>26.87</td>
<td>0.0313</td>
<td>2.2195</td>
</tr>
<tr>
<td>M1</td>
<td>24.86</td>
<td>0.0096</td>
<td>-2.6167</td>
</tr>
<tr>
<td>J1</td>
<td>23.10</td>
<td>0.0105</td>
<td>-2.7897</td>
</tr>
</tbody>
</table>
Table 3-3 Tidal Harmonic components in Apalachicola during Hurricane Dennis starting from local time at 0.00 hour on 05/01/07 to 23.00 hour on 07/31/07:

<table>
<thead>
<tr>
<th>Component</th>
<th>Period</th>
<th>am</th>
<th>phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>3</td>
<td>0.0001</td>
<td>2.4189</td>
</tr>
<tr>
<td>M4</td>
<td>6.21</td>
<td>0.0024</td>
<td>-1.8807</td>
</tr>
<tr>
<td>M2</td>
<td>12.42</td>
<td>0.1017</td>
<td>1.8415</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0413</td>
<td>-0.8472</td>
</tr>
<tr>
<td>N2</td>
<td>12.66</td>
<td>0.0228</td>
<td>1.6998</td>
</tr>
<tr>
<td>K2</td>
<td>11.97</td>
<td>0.0151</td>
<td>-2.0978</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0128</td>
<td>1.4367</td>
</tr>
<tr>
<td>L2</td>
<td>12.19</td>
<td>0.0064</td>
<td>-2.0169</td>
</tr>
<tr>
<td>K1</td>
<td>23.93</td>
<td>0.1342</td>
<td>-1.0079</td>
</tr>
<tr>
<td>O1</td>
<td>25.82</td>
<td>0.1266</td>
<td>-0.0603</td>
</tr>
<tr>
<td>P1</td>
<td>24.07</td>
<td>0.0444</td>
<td>-2.5518</td>
</tr>
<tr>
<td>Q1</td>
<td>26.87</td>
<td>0.0245</td>
<td>0.4575</td>
</tr>
<tr>
<td>M1</td>
<td>24.86</td>
<td>0.0090</td>
<td>2.3585</td>
</tr>
<tr>
<td>J1</td>
<td>23.10</td>
<td>0.0045</td>
<td>-2.2290</td>
</tr>
</tbody>
</table>

Table 3-4 Tidal Harmonic components in Panama City during Hurricane Ivan starting from local time at 0.00 hour on 07/01/04 to 0.00 hour on 09/30/04:

<table>
<thead>
<tr>
<th>Component</th>
<th>Period</th>
<th>am</th>
<th>phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>3</td>
<td>0.0001</td>
<td>-0.9519</td>
</tr>
<tr>
<td>M4</td>
<td>6.21</td>
<td>0.0023</td>
<td>3.1265</td>
</tr>
<tr>
<td>M2</td>
<td>12.42</td>
<td>0.1028</td>
<td>-1.9027</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0195</td>
<td>2.0010</td>
</tr>
<tr>
<td>N2</td>
<td>12.66</td>
<td>0.0218</td>
<td>-2.5275</td>
</tr>
<tr>
<td>K2</td>
<td>11.97</td>
<td>0.0337</td>
<td>0.9573</td>
</tr>
<tr>
<td>S2</td>
<td>12.00</td>
<td>0.0352</td>
<td>-1.6846</td>
</tr>
<tr>
<td>L2</td>
<td>12.19</td>
<td>0.0059</td>
<td>1.8901</td>
</tr>
<tr>
<td>K1</td>
<td>23.93</td>
<td>0.1367</td>
<td>-1.9979</td>
</tr>
<tr>
<td>O1</td>
<td>25.82</td>
<td>0.1358</td>
<td>-2.6492</td>
</tr>
<tr>
<td>P1</td>
<td>24.07</td>
<td>0.0477</td>
<td>-1.9534</td>
</tr>
<tr>
<td>Q1</td>
<td>26.87</td>
<td>0.0229</td>
<td>-2.9231</td>
</tr>
<tr>
<td>M1</td>
<td>24.86</td>
<td>0.0031</td>
<td>-1.9972</td>
</tr>
<tr>
<td>J1</td>
<td>23.10</td>
<td>0.0063</td>
<td>-2.3686</td>
</tr>
</tbody>
</table>
Table 3-5 The extreme water level predictions combining storm surge with tides

<table>
<thead>
<tr>
<th>Hurricane</th>
<th>Panama City Beach</th>
<th>Apalachicola</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observation</td>
<td>Model Prediction</td>
</tr>
<tr>
<td>Dennis (train)</td>
<td>1.873</td>
<td>1.851</td>
</tr>
<tr>
<td>Ivan (verify)</td>
<td>1.634</td>
<td>1.628</td>
</tr>
</tbody>
</table>

Note: Datum is mean sea level

Table 3-6 Correlation coefficients between models predicted and observed storm surge hydrograph.

<table>
<thead>
<tr>
<th>Station</th>
<th>Hurricane</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panama city Beach</td>
<td>Dennis</td>
<td>0.9811</td>
</tr>
<tr>
<td></td>
<td>Ivan</td>
<td>0.9618</td>
</tr>
<tr>
<td>Apalachicola</td>
<td>Dennis</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Ivan</td>
<td>0.9668</td>
</tr>
</tbody>
</table>

Table 3-7 Extreme water levels (m) combining storm surge in spring and neap tides

<table>
<thead>
<tr>
<th>Hurricanes</th>
<th>Panama City Beach</th>
<th>Apalachicola</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spring</td>
<td>Neap</td>
</tr>
<tr>
<td>Ivan</td>
<td>1.89</td>
<td>1.67</td>
</tr>
<tr>
<td>Dennis</td>
<td>2.09</td>
<td>1.77</td>
</tr>
</tbody>
</table>


Dean RG and Chiu TY. 1986. Combined total storm tide frequency analysis for Escambia County, Florida. Report of Beach and Shores Resources Center, Florida State University, Tallahassee, FL.


http://www.fema.gov/plan/prevent/fhm/frm Cfham.shtm


Website:
National Data Buoy Center, NOAA: http://www.ndbc.noaa.gov/hurricanes.shtml
BIOGRAPHICAL SKETCH

EDUCATION

2004–present  Ph.D. candidate, Depart of civil and environmental engineering, Florida State University, Tallahassee, FL.

2001–2004  M.S., Civil and Environmental Engineering, Hohai University, Nanjing, China.
Thesis: Study on waterway regulation of typical rocky rapids in a branching channel by 2-D hydrodynamic model.
Advisor: Dr. Wei Zhang.

1997–2001  B.S., Coastal Engineering, Hohai University, Nanjing, China.