# KALMAN FILTERING AND APPLICATION TO STORM SURGES

A Thesis

by

JESSE SLATEN

BS, Texas A&M University-Corpus Christi, 2017

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

MATHEMATICS

Texas A&M University-Corpus Christi
Corpus Christi, Texas

August 2019

KALMAN FILTERING AND APPLICATION TO STORM SURGES

A Thesis

by

JESSE SLATEN

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Alexey Sadovski, PhD
Chair

Baohua Chen, PhD
Co-Chair

Lei Jin, PhD
Committee Member

August 2019

ABSTRACT

Over the last decades with the advancement of computational power and access to data, the ability to create advanced forecasts and hind-casts too has grown. Recently, with an increase of the global population moving closer towards coastal areas. There is a much stronger presence to understand severe weather systems and there impact on the local population as well as the economy. With that there is still much work to be done within the field of weather forecasting specifically in tandem with real time decision making.This project will consider forecasting in the event of extreme weather systems. Precisely we will focus on the role of storm surge and investigating novel techniques in trying to increase accuracy in sea-level prediction models and decreases error associated with tidal gauge systems along the coast of Texas. Under the thought experiment that this will be used for some decision making process. Along with including an optimal warning time derived from the prediction methods.Tidal prediction methods have either been of two differing methods, statistical or deterministic. For typical usage most tidal predictions are given by deterministic methods, i.e. used by solving hydrodynamic equations in tandem with their astronomical constituents. Statistical methods have since been developed with the added benefit that we can include live measurements to improve tidal predictions based off a time series of observations.For high-impact weather systems we do not have the ability to solve the same hydrodynamic equations so quickly and readily as to aid emergency services and workers. Accurate models are desirable not only from a human life standpoint but from an economic standpoint as well. It may be the case that community is not endanger but it may affect local businesses that rely on precise forecasts. In these cases we must insist on the exact time closures must be necessary in order to minimize any economic impact or loss.In these cases we choose to employ a combination of both statistical and deterministic methods. The Kalman filter is one such method of combining these two methods in order to increase the accuracy in model.

ACKNOWLEDGEMENTS

I'd like to acknowledge every single professor I've ever had at TAMUCC, I'm more than grateful for the opportunities I've been given. Special thanks to my adviser Dr. Sadovski , Dr. Pal, and Dr. Zimmer for putting up with my handwriting for two years.

## DEDICATION

This is dedicated to Paul Walker.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES PAGE

# CHAPTER I: INTRODUCTION

Over the last decades with the advancement of computational power and access to data, the ability to create advanced forecasts and hind-casts too has grown. Recently, with an increase of the global population moving closer towards coastal area [10]. There is a much stronger presence to understand severe weather systems and there impact on the local population as well as the economy. With that there is still much work to be done within the field of weather forecasting specifically in tandem with real time decision making.

This project will consider forecasting in the event of extreme weather systems. Precisely we will focus on the role of storm surge and investigating novel techniques in trying to increase accuracy in sea-level prediction models and decreases error associated with tidal gauge systems along the coast of Texas. Under the thought experiment that this will be used for some decision making process. Along with including an optimal warning time derived from the prediction methods.

Tidal prediction methods have either been of two differing methods, statistical or deterministic. For typical usage most tidal predictions are given by deterministic methods, i.e. used by solving hydrodynamic equations in tandem with their astronomical constituents. Statistical methods have since been developed with the added benefit that we can include live measurements to improve tidal predictions based off a time series of observations.

For high-impact weather systems we do not have the ability to solve the same hydrodynamic equations so quickly and readily as to aid emergency services and workers. Accurate models are desirable not only from a human life standpoint but from an economic standpoint as well. It may be the case that community is not endanger but it may affect local businesses that rely on precise forecasts. In these cases we must insist on the exact time closures must be necessary in order to minimize any economic impact or loss.

In these cases we choose to employ a combination of both statistical and deterministic methods.

The Kalman filter is one such method of combining these two methods in order to increase the accuracy in models[6].

## 1.1 Prediction Methods

The most basic method in order to predict tidal height is to check the astronomical affects of the sun and moon. It is well known that the moon and sun both play a role in the how and low tides in the ocean. The gravitational forces of theses objects create a diurnal cycle produces the "buldge" of the Earth and its' oceans. This method is referred to as the harmonic method and is often associated within a single point or area on Earth. Because of the simplicity of this model it is usually calculated separately from the other model and superimposed in post-processing.

The other method for tidal prediction is calculated by the impact of meteorological affects on tides. These are mostly estimated by solving the linearized shallow water equations. The idea then is to combine the harmonic method with the operational method, in the most accurate way. For most purposes this method works fine, but for high impact weather systems the computational cost of these methods is too great so we must come up with ways of estimating the tidal height more quickly without losing accuracy.

More current prediction methods readily use live observations from tidal measuring stations placed along the coast. Not only do these stations contain the most up to date tidal height, but they also contain measurements of pressure , temperature, and wind speed and direction. This project will focus exclusively on a 1-Dimensional model. Working institutions will use two-dimensional models for there predictions, there exists some three-dimensional models but they are still too numerically expensive and require labor until they are deemed satisfactory for our current practices.

## 1.2 Kalman Fitlering

Filtering is one kind of a data assimilation technique. It's usage is in an effort to combine theory with observations. These techniques appear widely in meteorological predictions and forecasting.

With the abundance of field and forecasting data and the complexity of the equations involved, it is easy to see why this has great practical application.

With this we can combine the best characteristics of the deterministic and the statistical models. In estimating and modeling the water level we have made various assumptions about which parameters reflect the most influence on the outcome. Things like bottom friction and wind friction are usually determined prior to any simulations and are treated as constants. In the case of storm surges these constants and change greatly. We can use the Kalman filter to correct our predicted water-level as well as include our observations. With the added nicety that we are always increasing the amount of data and observational stations, the filter lends itself to this research.

The Kalman filter in particular has wide spread use from satellite navigation to autopilot features in airplanes and automobiles. This investigation will approximate the tidal movement by a one-dimensional model. Water-levels and velocities are to be estimated by the Kalman filter. With the added benefit that the model is being updated with the changing conditions to give accurate forecasts at lesser computational price.

It will be noted that with the shear amount of stations it could be with the variability of the bathymmetry and geography at these locations, there may be an “optimal” location of the measurement stations. Future work would be focused on investigating further if it is possible to design a system to designate an “optimal” location in order to more accurately predict water-levels in certain localized areas rather than have multiple stations within the same small area. We wish to employ the use of the filter not only to improve upon the state-estimate but to also provide more insight into the underlying mechanics of the model during storm surge times.

The paper will go as follows , Chapter II is a review of the relevant literature of the subject and corner cases. Chapter 3 is the methodology being employed. ChapterIV and V are then the results , discussions and conclusions.

# CHAPTER II: REVIEW OF THE LITERATURE

The actual use of Kalman filtering in storm surge tides is limited mostly to specific regions for the investigations. On the other hand the filter and other more complex filters appear widespread in general for water-level predictions and for stream flow predictions. The first related paper by Budgell [1] , in which he developed a filter from the one-dimensional shallow water equations used to predict water-levels at six minute intervals. These are useful but only to the extent that the period in which he is predicting is very small and not particularly viable in terms of emergency services decisions.For our investigation purposes we will choose to extend the the forecast lead time while sacrificing model accuracy, more importantly this filter only has application during little to no meteorological impact.

The next selection of papers comes from the area of hydrology and dam operations [14]. In the paper Zhu et al designed a neural network to simulate real-time flooding for a system of dams, the main focus of this procedure was to apply a optimization scheme based around minimizing the maximal uncertainty at each flow control station.

Other methods of data assimilation were also considered such as the adjoint method from optimal control theory. An unknown control function is used as a placeholder for some numerical model. With known data and information and used to minimize a cost function, usually just some function that describes the difference between the model results and data. This method in general applies to more problems but is very time consuming to decide on a function and then have to minimize it , usually by gradient descent. One section of key interests is in calculating the variances associated with ocean tide. Already estimating variances is a complex subject, but it's role in sea-level observation/prediction is not well understood[7]. From the standpoint of storm surges there is no literature to be found on the subject.

# CHAPTER III: METHODOLOGY

## 3.1 Filtering Problem

At it's most basic the filtering problem is creating a 'best' estimate for a real value in some given system. When given any equation we are simply modeling the state and never explicitly the exact state. In his text Øksendal writes clear enough, that based off of 'observations' what is the best estimate of the state? The linearized one-dimensional tidal surge equation is given by [6], as

$$\frac{\partial u}{\partial t} + g\frac{h}{\partial x} - fv + \lambda\frac{u}{D} - \gamma\frac{V^2\cos\Psi}{D} + \frac{1}{\rho_w}\frac{\partial p_n}{\partial x} = 0$$
$$\frac{\partial h}{\partial t} + \frac{\partial(Du)}{\partial x} = 0$$

with,

$u =$ u component of wind

$v =$ v component of wind

$h =$ water level

$D =$ depth of water

$f =$ Coriolis parameter

$\lambda =$ linear bottom friction coefficient

$\gamma =$ wind friction coefficient

$\Psi =$ wind direction

$\rho_w =$ density of water

$g =$ gravity

$p =$ atmospheric pressure

The numerical solution to this system is given by [3].

## 3.2 Kalman Filter

For these cases we will use a traditional Kalman filter, in order to create a better behaved vector of the state $m_k$ estimate at the time $k$, to estimate the true state value $x_k$.

$$x_{k+1} = \Phi x_k + w_k$$
$$m_k = H x_k + v_k$$

Here, $\Phi$ is the state transfer function which takes the previous heights at time $x_{k-1}$ into the present case, $w_k$ represents the noise assumed to be Gaussian and with mean 0 and covariance $Q_k$. $H_k$ is the observation transfer function which maps the true state into the observed space. Lastly, $v_k$ is the observation noise which is again assumed to have mean zero and be Gaussian white noise. From here on we will assume all states are mutually independent.

## 3.3 Model Validation

For this investigation we will primarily be using the Root Mean Square Error (RMSE) in areas that are further away from the shoreline to validate the reliability of the model. In cases closer to the shoreline we will refer to the Absolute Error (AE) for the purposes of forecasting and evacuation planning. There are other methods of cross-validation, for instance we could instead only take the largest difference in water level height as a maximum, since we are only interested in how badly a model approximates reality.

CHAPTER IV: FINDINGS/RESULTS

Each of these cases will be defined more easily by location, as each one of the NOAA Tidal Stations has different monitoring equipment. Along with each location has differing topographical nuisances whether it be attenuation from any buildings or structures nearby, or some natural disturbances.

## 4.1 Seadrift, Texas

Initially we begin our investigation with a situation where the linear filtering fails. At this station located in Seadrift Texas there is little to no tidal presence, and the water level is largely dominated by the surge at around the same time frame of August 25 - 31st 2017 of Hurricane Harvey. There are a few lost data points that we will choose to interpolate over for our analysis purposes See Figure 4.1.

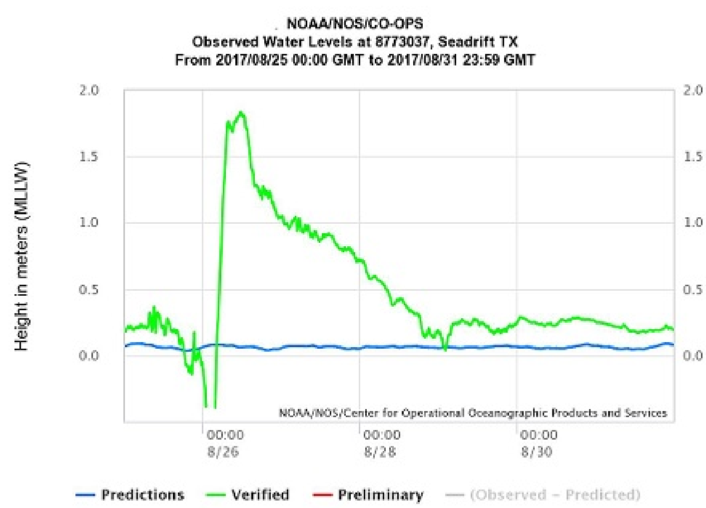



Figure 4.1
NOAA Seaddrift Tidal Station chart

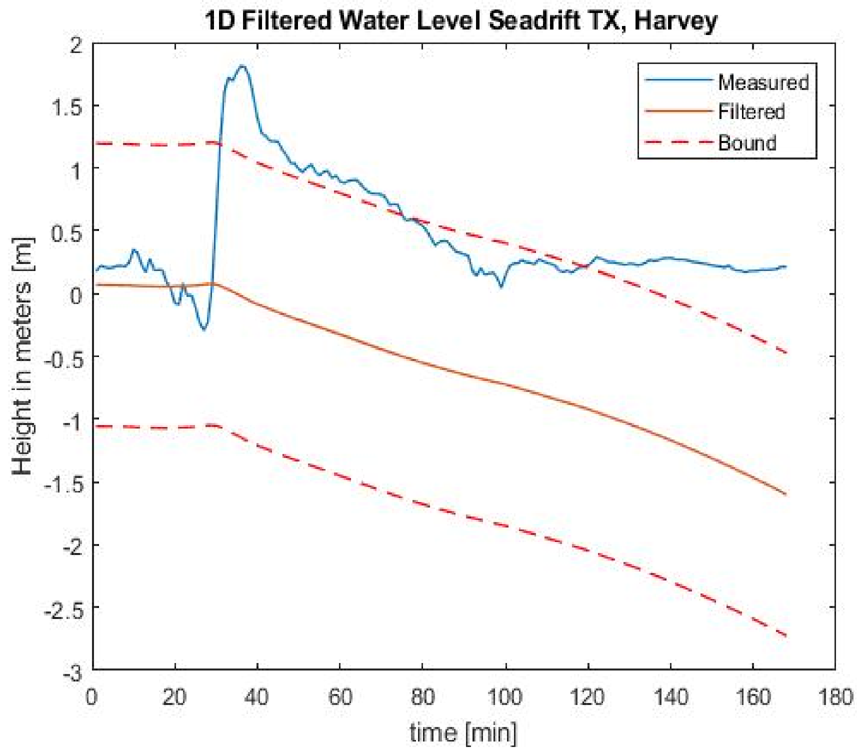



Figure 4.2
One Dimensional Filter , Seadrift TX tidal chart

In this case, we have a sharp rise brought upon from storm resulting in a sudden wave and slower decay as the storm leaves the area. Along with that the time series is given in intervals of one hour time steps rather than six minutes. These two things make it very difficult for the linear Kalman filter to work. So instead we bring our attention to a problem that has a smaller interval between time steps and also has has less of these nonlinear characteristics. See Figure 4.2, the one-dimensional filter fails to approximate the surge at all and only captures the descent.

## 4.2 High Island

It is clear that the non-linearity of the problem is preventing any kind of useful filtering of the data set. From here we can either, approximate our data as a linear problem, or try a different filter and see how it responds.

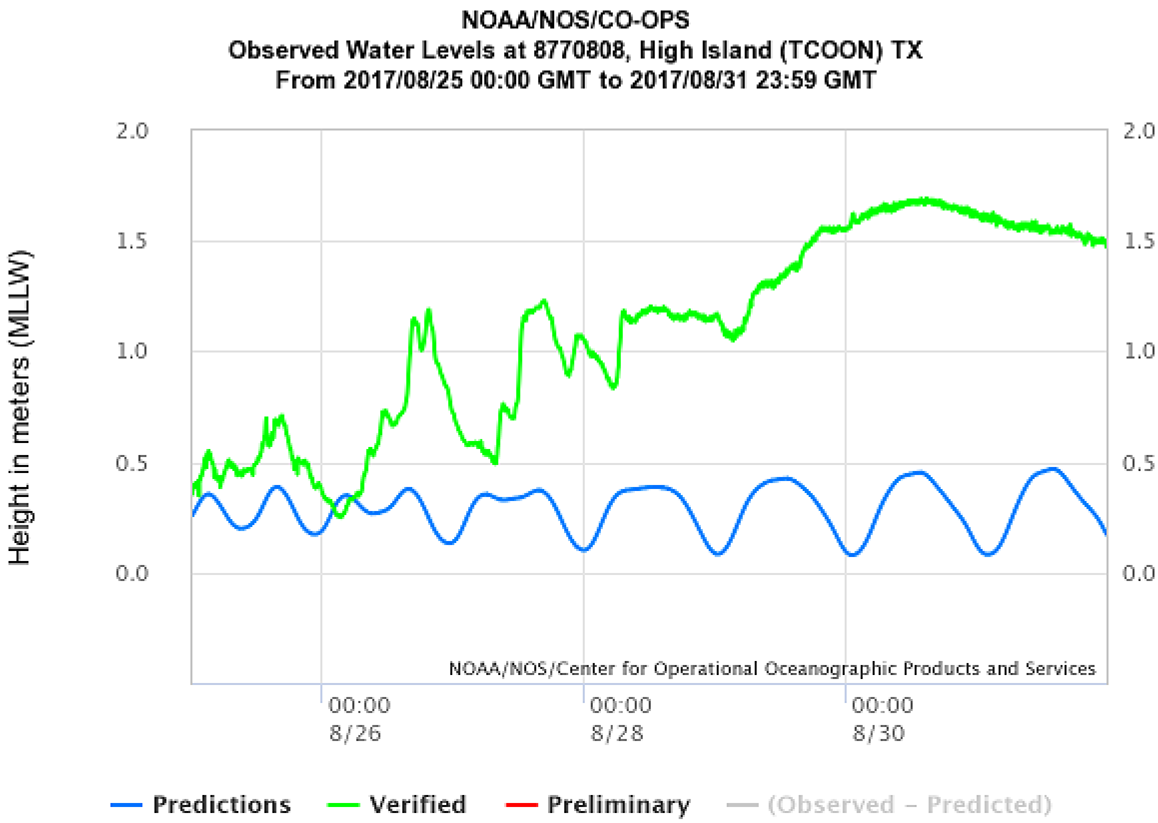



Figure 4.3
NOAA High Island Tidal Station Chart

Our next location is at High Island TX. Typical attempts at filtering these events fail as the surge lasts only for a few hours. With the surge being strongly associated with a severe weather system we can ignore tidal affects in order to better pinpoint the precise variables that are linked to storm flooding.

In the above Figure 4.4, we have chose to "detide" the water levels as to better hone in the distinct variables associated with the surge. It is clear that the filtered time series misses some of the important peaks and fails to have any sharp descents. While important for real time forecasting purposes these peaks and valleys are mostly related to tidal forcing. With that loss we gain the ability to analyze the time series to understand how and which variables are related as this series

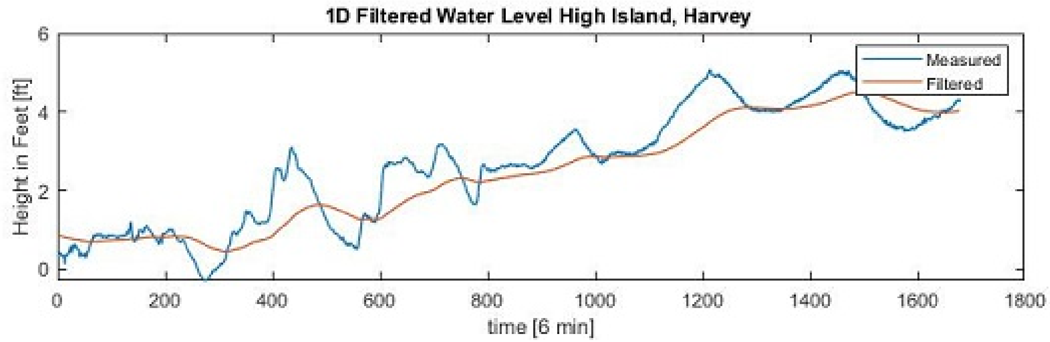



Figure 4.4
One Dimensional Filter, Detided High Island

has a general slower more linear progression. The first test is to see the correlation coefficient,$r$, amongst several variables.

The variables given from the station are wind speed and direction$[ms^{-1}]$, gusting $[ms^{-1}]$, and pressure $[mb]$. In the unfiltered case there is no clear sign of correlation between water level height and any of the variables. We know this is not the case as each of these variables is used in the processes of storm surge inundation.

Instead we opt to compare the filtered data sets and then examine there correlation. Wind speed in the horizontal direction is shown to have a correlation coefficient of $r = 0.54$. Which goes to show that there is a some positive correlation but not as important as in the case of the pressure. The pressure also needing to be filtered for noise in the measurements has with the filtered data set a correlation coefficient of $r = -0.955$.

In this case the only variables of important are $u_n$ normal component of wind speed with respect to an angle along the coast, $p$ barometric pressure at a station, and lastly $h_t$ height of the water at time t.

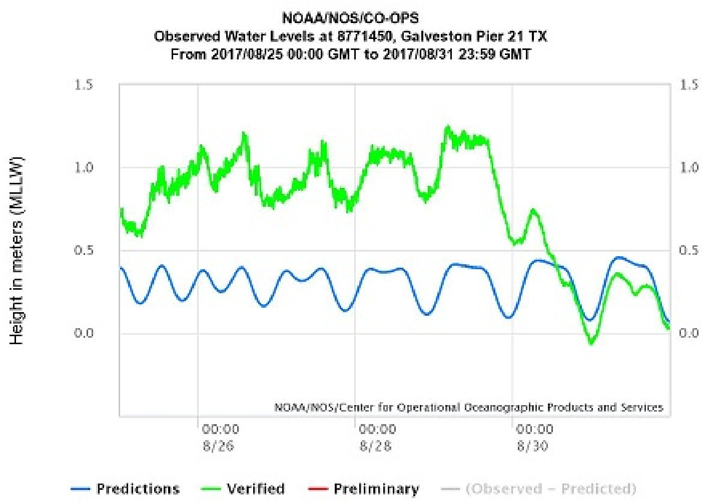



Figure 4.5
NOAA Galveston Pier 21 Tidal Chart

### 4.3 Galveston Pier 21

Next, is Galveston Pier 21, which requires some explanation. As Hurricane Harvey made landfall on August 24th it quickly diminished in strength and so was a slow moving system. Which leads to an extended period of it high water levels in this area. Again we can apply our filter in this situation for even better results as this time series is mostly linear despite the oscillations.

Unfortunately this station did not have an instrument to measure wind speed or direction, so we are only left with pressure. This time the correlation coefficient is $r = -0.7653$. Which again agrees with the general meteorological theory.

This time however we can better approximate and forecast the relevant nonlinear features with out filter by means of a first order Taylor approximation of the previous observations. As seen in Figure 4.6 the filter misses some peaks and valleys. Were we to include the previously obtained observations, $o_k$, and include them into the relevant formula

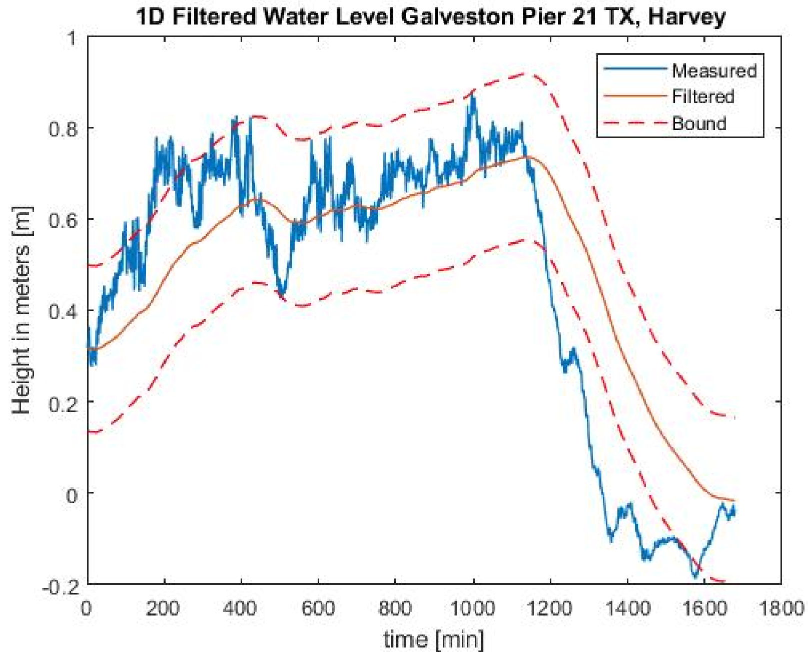



Figure 4.6
One Dimensional Filter Galveston Pier 21

$$x_{k+1} = \Omega x_k + w_k + \alpha \frac{\Delta x(k1)}{\Delta t} \tag{4.1}$$

$$m_{k+1} = H x_k + v_k. \tag{4.2}$$

Where now $\alpha$ is a weight factor, that we can choose to adjust when necessary. We could extend this notion to any dimensions with however many higher order terms that are needed. For this study we will be using the 1D Shallow Water equations given by Crowhurst [3] will be solved by finite difference method. This would then be another linear extension for the Kalman filter, aptly titled the extended Kalman filter. In Figure 4.7. With this novelty we are able to decrease the RMSE , but still run into the grievance of overestimating the lower end of of water levels.

By now we can see that the filter displays poor performance after the storm has passed and the water levels are allowed to recede to normal levels for the season. With this should we choose to adjust the filter and cut off for the longer period, and exclusively for the time of surge we can

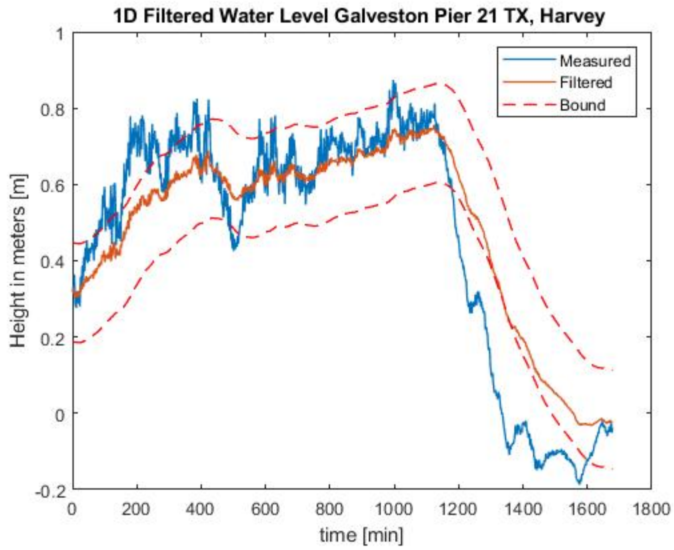



Figure 4.7
1D Taylor Approximated Filter Water Level Galveston Pier 21

improve our results. We can can further expand our forecast time to hour long intervals while losing accuracy we are able to improve precision in developing the slow moving head of our wave front. To improve the calculation of the first derivative of the function, we choose to apply a standard filter to the calculated first derivative of the observation data. Then applying this newly filtered first derivative we receive figure 4.8. We can observe one more that towards the end of the period the model fails to capture the fast movement that takes place once the system has passed through the station. We opt to cut off that portion, for the next approximation. In 4.9 we have lost the accuracy of the previous models but gain a longer forecast time.

## 4.4 Extending Forecast Lead Time

Should we choose extend the lead time from the hour to two or three we gain the ability to give better warnings but lose on accuracy greatly. In the case of Sargent TX. We then choose a smaller time frame to forecast closer to the landfall of the system. In figure 4.10 the worst error is approximately 0.48 meters. Which for evacuation purposes may be the decision between issuance of a

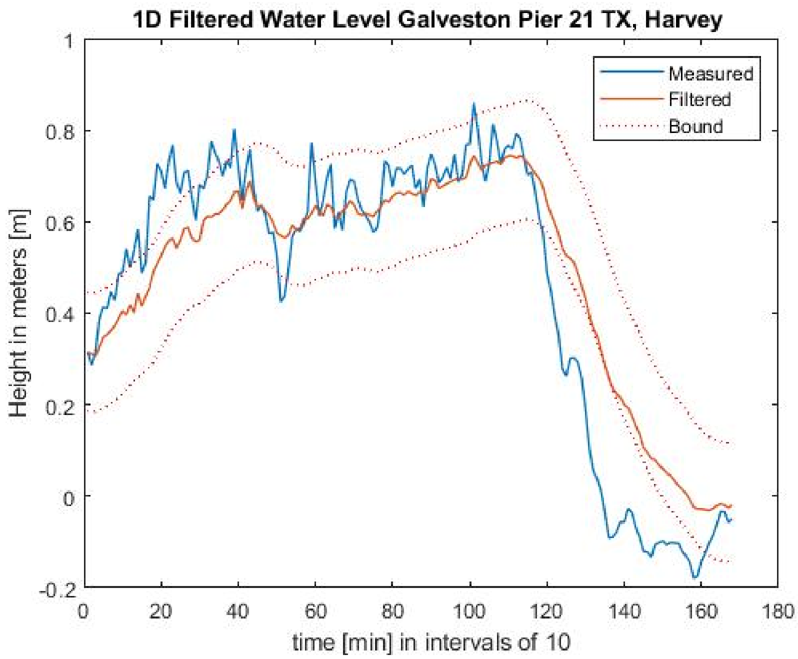



Figure 4.8
Filtered Sea-Level Height Over Time

mandatory evacuation or a flood-watch.

If we again extend the forecasting length by an additional hour we enter the ideal range to issue a warning within ample time for a safe evacuation. Unfortunate the 1-Dimensional problem overestimates the peaks and underestimates the lows. In Figure 4.11 we see at its worst there is an absolute error of approximately 3 hours. This forecast time is too unreliable to accurately give a warning.

## 4.5 Optimal Warning Time

In this section based off of our previous estimations and forecasting of sea level height we will create an optimal warning time to either safely evacuate a community or to warn local businesses of high surge. Along with issuing a warning we also consider the lag time needed before the detection before the evacuation procedures can start. These factors can vary from mechanism of

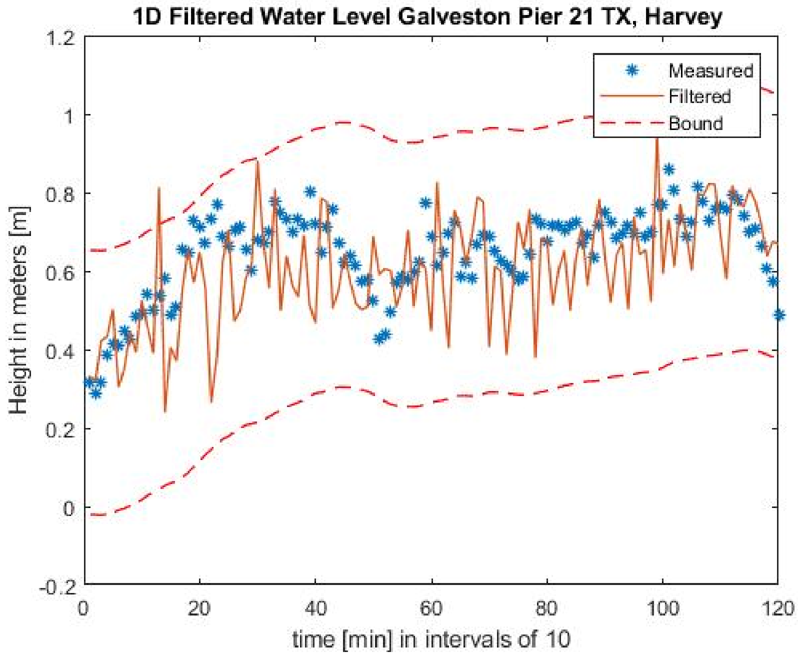



Figure 4.9
2nd Derivative approximation

warning, distance from impact, and structures inhabited [11]. Mobilization curves give an estimate for the total demand that departs within a given time frame. The US Army Corps of Engineers has given three 'speeds' for how quickly individuals will respond to a given evacuation warning 4.14.

An alternative model proposed by Tweedie modeling traffic as

$$F_t = 1 - \exp\{-\frac{t^2}{T}\}$$

where $F(t)$ is defined as the cumulative percentage of evacuees at time $t$. $T$ is the response rate to the disaster. $T$ should be adjusted depending on the environment that is trying to be modeled. $T(0)$ is the number of evacuees on the network that leave before a warning is issued. This number is difficult to estimate as there are many factors that lead to it such as, distance from disaster area, time of from last disaster for examples people who are familiar with hurricanes may be more likely to leave in the event of one already forming in the Gulf of Mexico prior to it landing. Another

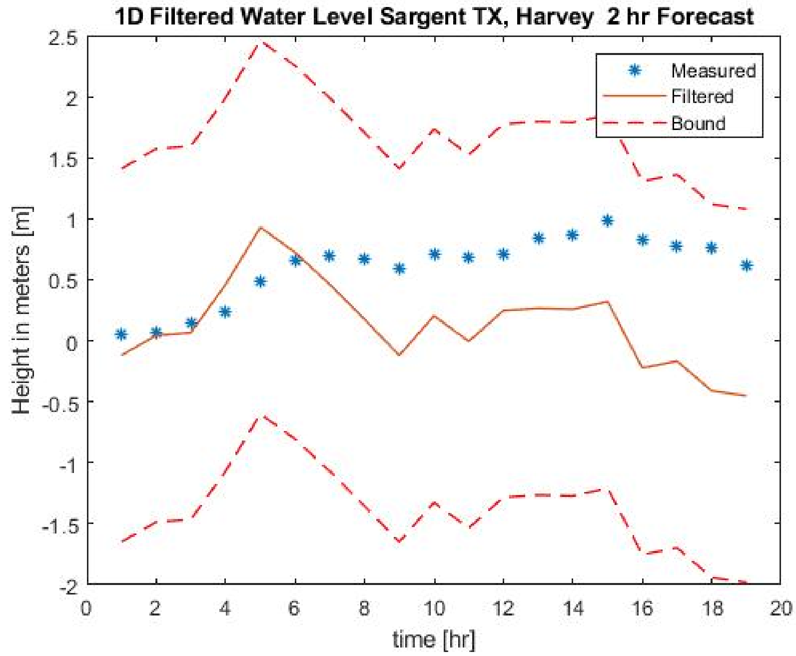



Figure 4.10
Forecast Time 2 Hr Sargent TX.

issue worth noting in this region is very common that during the period of time hurricanes are most prevalent there are more likely to be tourist and visitors present which will add another method of congestion to the roads.

All of the numbers organized were obtained from [2]. In the gulf coast region of Texas and Louisiana approximately 40,000 people were evacuated due to flooding. To account for tourist and unprepared locals we will allow for a large time for evacuation due to flooding being a slow process and to give evacuees enough time to locate a shelter or alternative safe place, collect pets and animals, and gasoline.

The methodology will be as follows, if the forested time series has a first derivative that changes within a threshold for a given region, we would then issue an alert. For example in the case of Sargent TX. If we were to obtain a given forecast of a change in height of over 1 meter in the time of a few hours, we would then have to issue an evacuation for low lying areas see Figure 4.15.

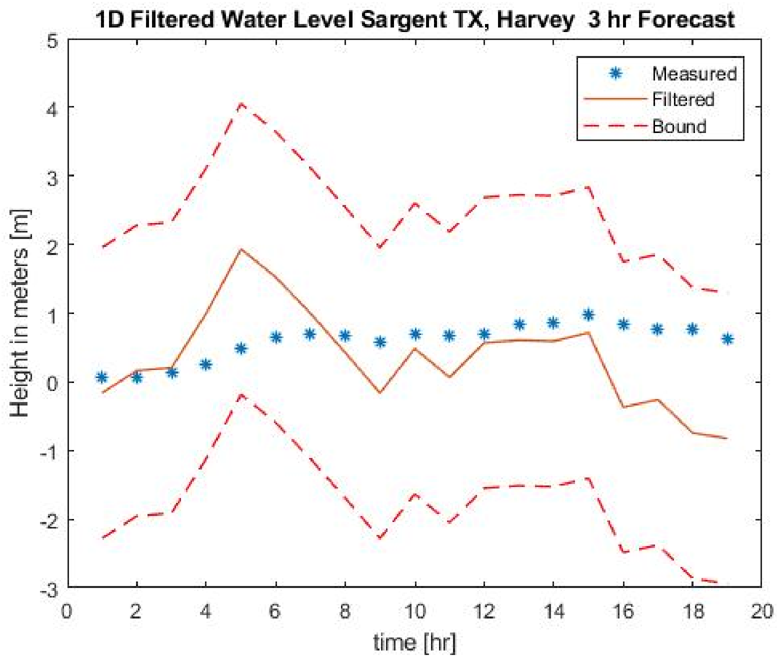



Figure 4.11
Forecast Time 3 Hr Sargent TX.

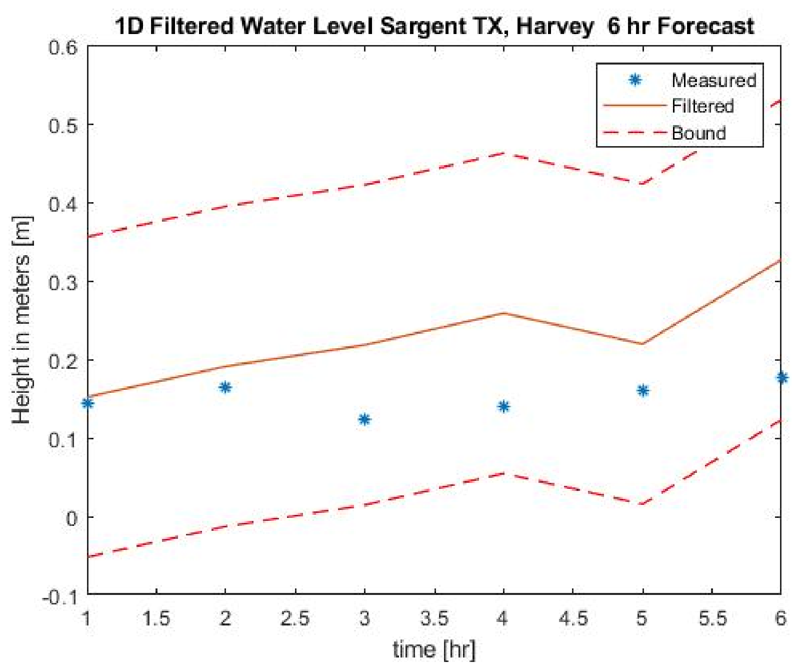



Figure 4.12
6 Hour Forecast

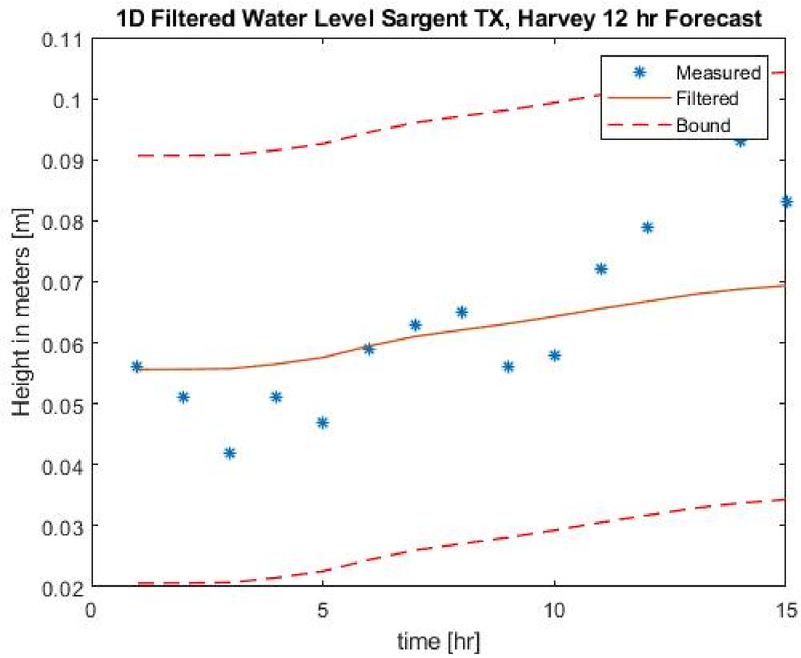



Figure 4.13
12 Hour Forecast

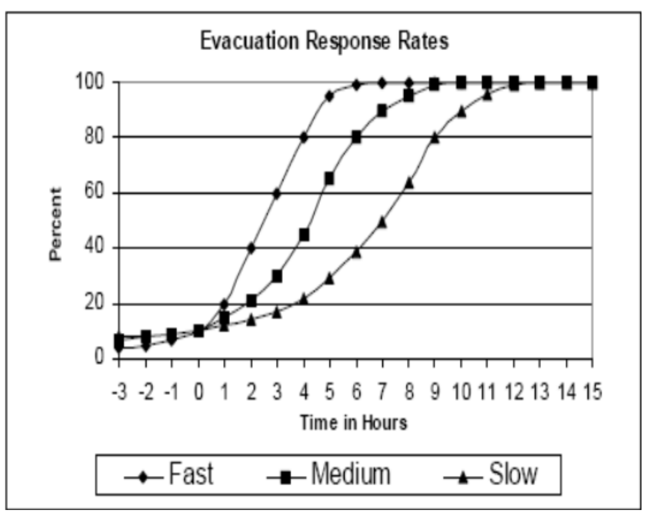



Figure 4.14
Hurricane Evacuation Mobilization Curve, Army Corps of Engineers (2000)

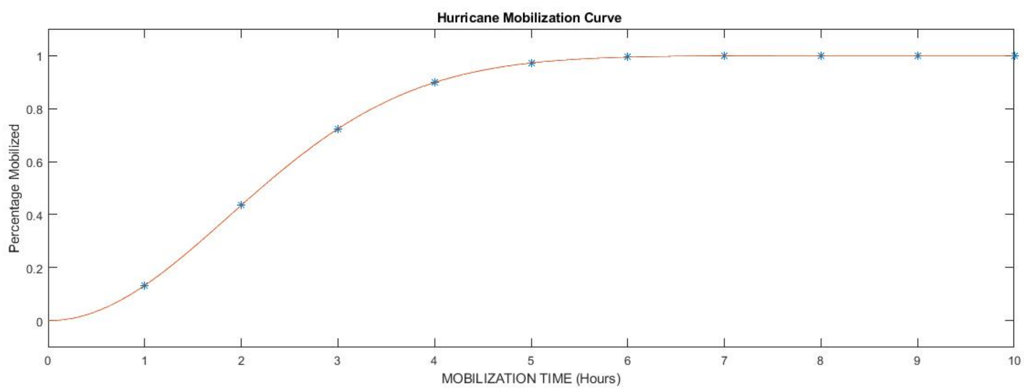



Figure 4.15
Example Evacuation Mobilization Chart

# CHAPTER V: DISCUSSION

From the previous material field data was used to attempt a short term forecast. It is clear from the above that pressure plays a significant role in the propagation of these storm surges. In estimating model uncertainty we must be sure to askew filter divergence by using an appropriate time step and measurement. In some of the cases the filter largely was dependent on the non-linear characteristics of the flood, there are ways of appropriately handling that. Along with that the one-dimensional approach has only a small feasible application for a very small region along the coast. It can however be useful in estimating certain model parameters that could be used for further simulations.

With the last demonstration it proved useful to apply the finite differences to the filter unfortunately this solution is no longer guaranteed to be optimal in this case. The computational effort needed is also an added benefit as it is far less intense than running and simulating the shallow water equations in the event of on ongoing time-sensitive situation.

# CHAPTER VI: SUMMARY AND CONCLUSIONS

In this study one kind of linear Kalman filter was developed in an effort to predict water-levels along the coast of Texas . The second was a finite differences extension that proved to be useful. Both of these models can be continually updated in real-time and adapt to fast changing circumstances. Unfortunately the forecast time greatly suffers from the linear portions, it is only applicable to small localized regions along the coast.

It would clearly be of interest to see the two-dimensional product of the filter along with actually producing the simulation estimates rather than obtaining more data from the NOAA. It may also prove to be useful in incorporating some spatial variance in the filter as within a portion of the coast there is not likely to be another Tidal gauge station for several miles/kilometers. In these cases it would be best to find a location that minimizes the uncertainty for a given region, or if it is even possible to suggest such a location exists.

Further investigations of the underlying mechanics of the shallow water equations are still desirable. In practice the condition of white noise being uncorrelated with time is likely not true as was assumed here. One of the larger drawbacks is all of the noise components all have to previously obtained or specified directly. Currently during these storm conditions the parameters and variables change so rapidly more investigation is needed on the effects these terms add. One very interesting idea would be to compare the benefits or losses of the filter to a satellite measurements to obtain water levels.

Lastly, major drawbacks of this filter is in application the noise and uncertainty have to be specified prior to any predictions. This would not work so readily in real scenario. This filter is not sensitive to any statistically significant changes in the input data.

In summary, the novel techniques provided and used in this text are useful and could be expanded further and incorporated into some real time forecasting and model predictions in the future.

## REFERENCES


[1] W. P. Budgell, *Stochastic filtering of linear shallow water wave processes*, Journal on Scientific and Statistical Computing, 8 (1987), pp. 152–170.

[2] A. O. Campbell, *Hurricane Emergency Evacuation Study for Coastal Regions of Rhode Island*, PhD thesis, University of Rhode Island, 2016.

[3] P. Crowhurst, *Numerical solutions of one-dimensional shallow water equations*.

[4] J. J. Dronkers, *Tidal computations in rivers and coastal waters*, North Holland Publishing Company.

[5] D. C. Heath and P. L. Jackson, *Modeling the evolution of demand forecasts ith application to safety stock analysis in production/distribution systems*, IIE Transactions, 26 (1994), pp. 17–30.

[6] I. Heemink, K. Bolding, and M. Verlaan, *Storm Surge Prediction Using Filtering*, PhD thesis, Twente University of Technology, 1986.

[7] D. L. Kriebel, J. D. Geiman, and G. R. Henderson, *Future flood frequency under sea-level rise scenarios*, Journal of Coastal Research, 31 (2015), pp. 1078–1083.

[8] O. ksendal B., *The filtering problem. in: Stochastic differential equations.*, (2003).

[9] B. Pei, *An error quantification methodology for hurricane storm surge simulations*, (2012).

[10] E. N. Rappaport, *Fatalities in the united states from atlantic tropical cyclones: New data and interpretation*, Bulletin of the American Meteorological Society, 95 (2014), pp. 341–346.

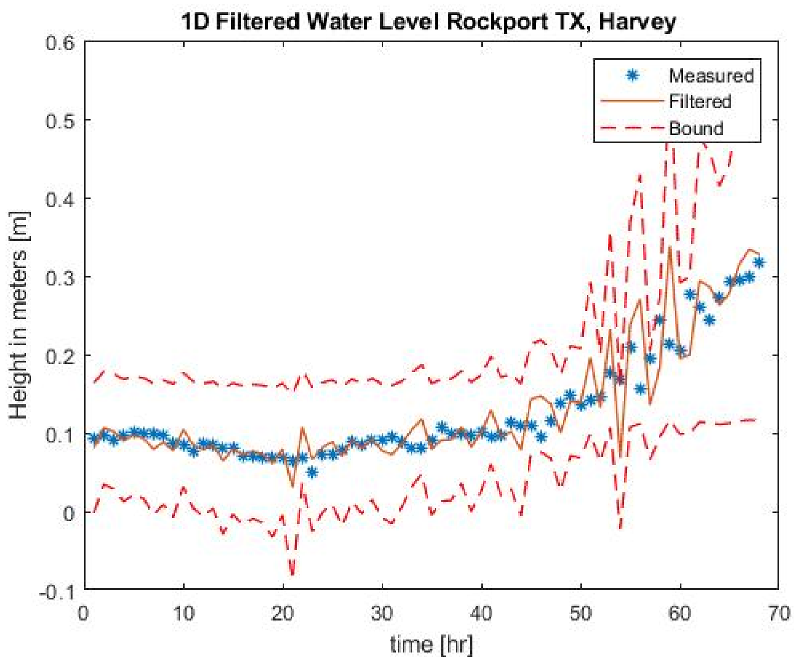
[11] J. H. Sorensen, *When shall we leave? factors affecting the timing of evacuation departures*, International Journal of Mass Emergencies and Disasters, 19 (1991), pp. 153–165.

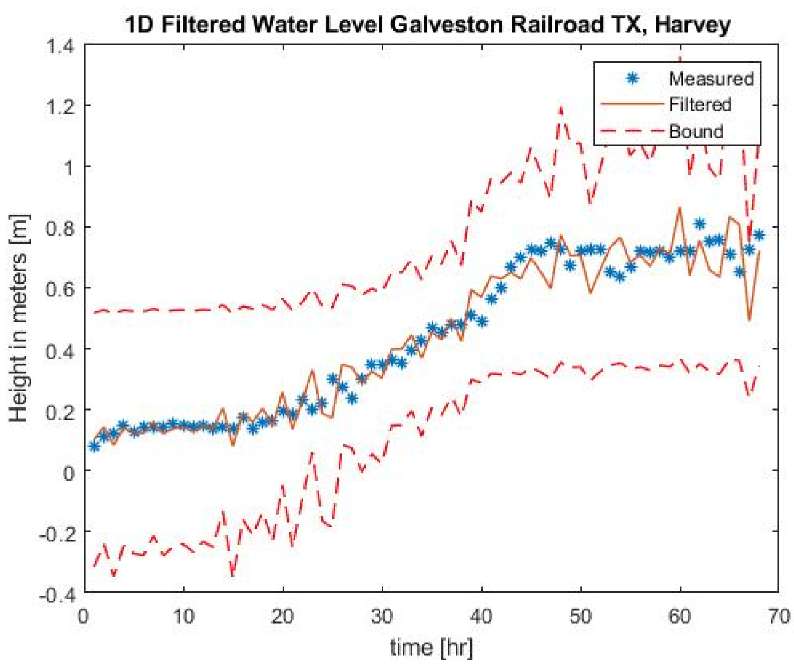
[12] T. Zhao, X. Cai, and D. Yang, *Effect of streamflow forecast uncertainty on real-time reservoir operation*, Advances in Water Resources, 43 (2011), pp. 495–504.

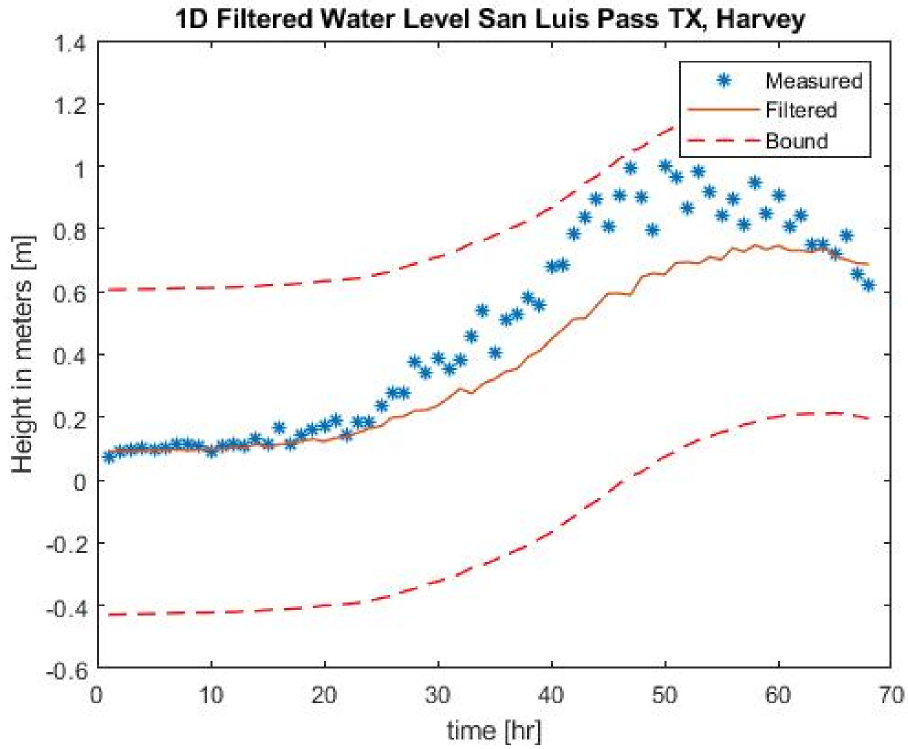
[13] T. Zhao, J. Zhao, D. Yang, and H. Wang, *Generalized martingale model of the uncertainty evolution of streamflow forecasts*, Advances in Water Resources, 57 (2013), pp. 41–51.

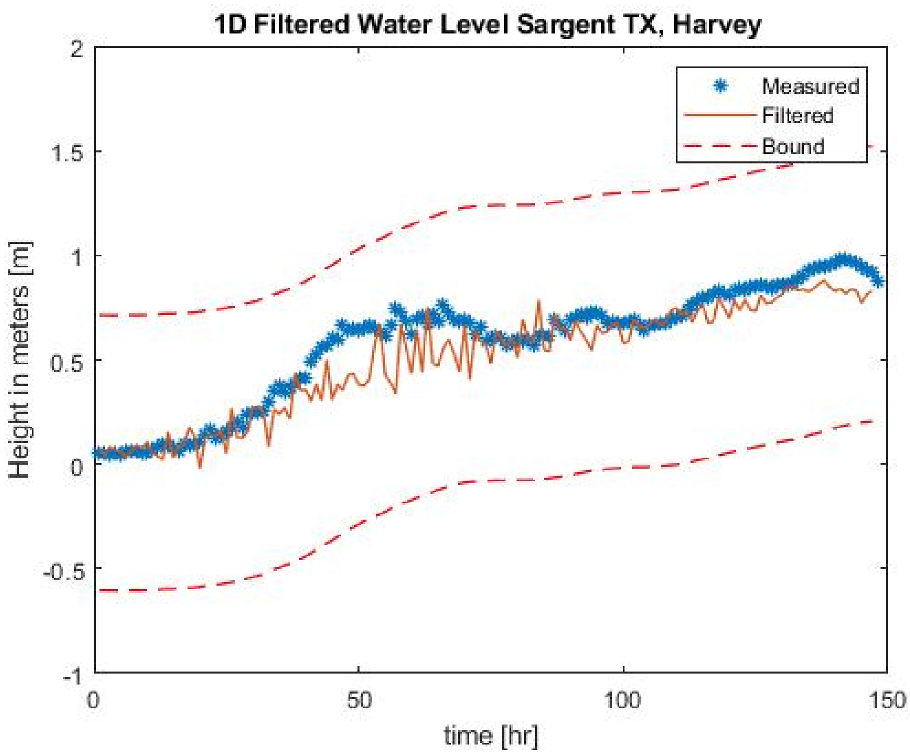
[14] F. Zhu, P.-A. Zhong, Y. Sun, and W. W.-G. Yeh, *Real-time optimal flood control decision making and risk propagation under multiple uncertainties*, Water Resources Research, 53 (2017), pp. 635–654.

APPENDIX A: 2nd Derivative Filter Approximations

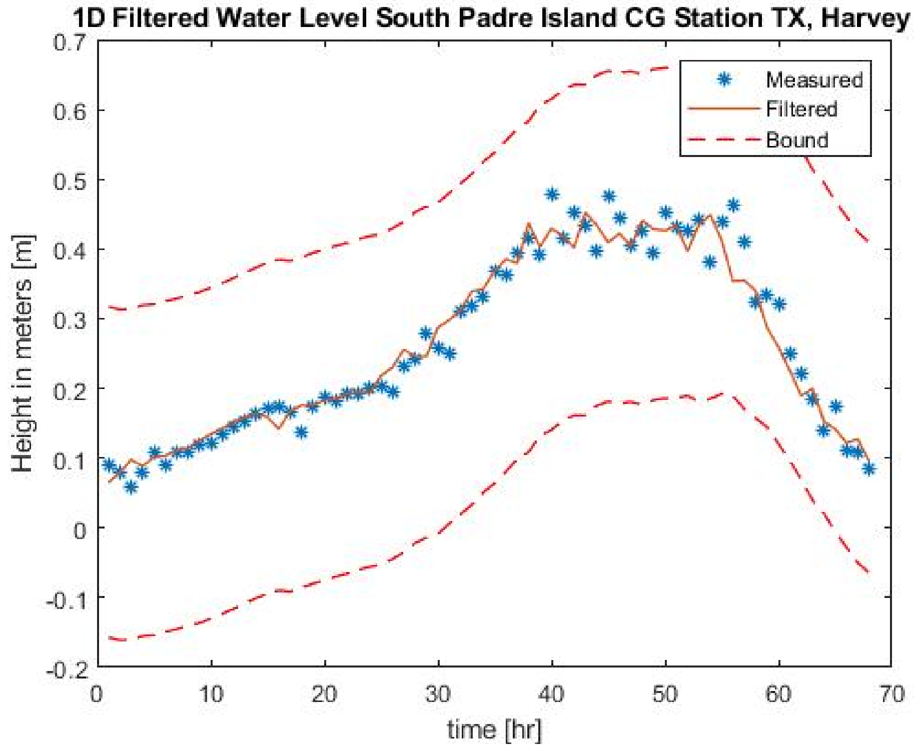


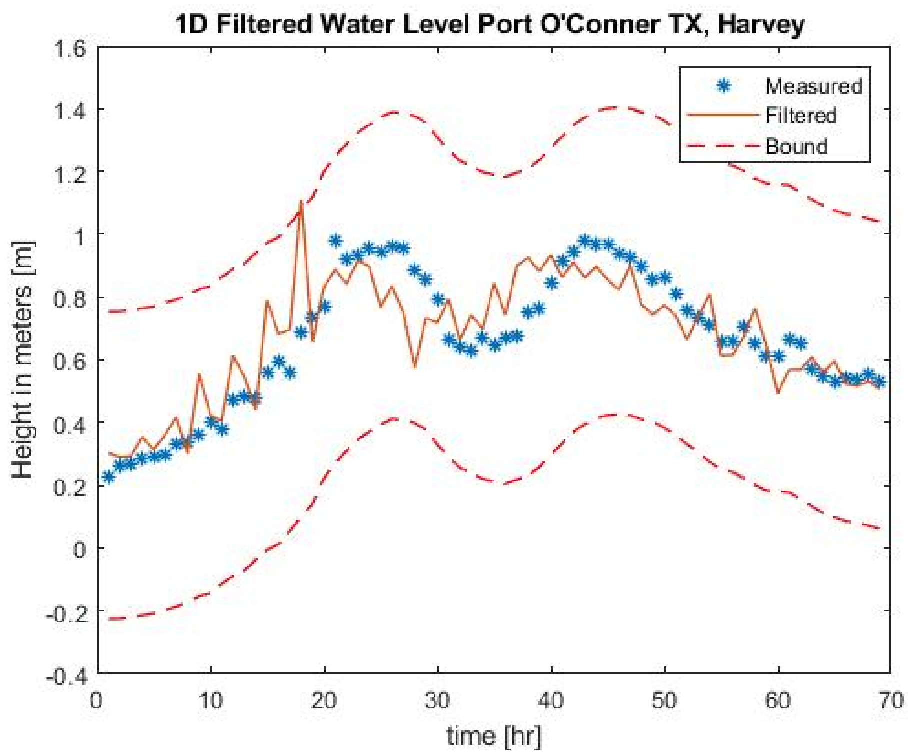




1D Filtered Water Level San Luis Pass TX, Harvey
Measured
Filtered
Bound
Height in meters [m]
time [hr]
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
10
20
30
40
50
60
70


1D Filtered Water Level Sargent TX, Harvey
Measured
Filtered
Bound
Height in meters [m]
time [hr]
2
1.5
1
0.5
0
-0.5
-1
0
50
100
150

}


1D Filtered Water Level South Padre Island CG Station TX, Harvey
Measured
Filtered
Bound
Height in meters [m]
time [hr]


1D Filtered Water Level Port O'Conner TX, Harvey
Measured
Filtered
Bound
Height in meters [m]
time [hr]