Improving the Skill of National Weather Service's Extratropical Total Water Level Forecast System

Rosimar Ríos-Berríos[1] and A. A. Taylor[2]

[1]2009 NOAA Educational Partnership Program Scholar, University of Puerto Rico at Mayaguez

[2]National Weather Service's Meteorological Development Laboratory

Background

Extratropical cyclones are storms that form at middle and high latitudes outside of the tropics. The atmospheric conditions of these storms include; low atmospheric pressure, large amounts of precipitation and high speed winds. One effect of these cyclones is extratropical storm surge, which is the water moved to shore by the high speed winds of the cyclone. This surge, in conjunction with tidal forces, can cause severe flooding, putting all who live along the coast at risk. To warn people about this hazard, the NWS' Meteorological Development Laboratory (MDL) created the Extratropical Storm Surge (ET Surge) model to forecast storm surge. The model extracts the pressure and surface wind speed and direction from the NWS's Global Forecast System (GFS) atmospheric model and uses these as forcing for its numerical storm surge model. Currently, it is run at 00Z, 06Z, 12Z, and 18Z to generate 96 hourly forecasts. After the model is run, the NWS uses it to provide a total water level forecast on its website (http://www.weather.gov/mdl/etsurge), to inform people about their current risk of flooding.

Purpose of this study

The total water level forecast system is the principal source of guidance for forecasting floods caused by extratropical storms. The forecasts generated by the system are currently used operationally by emergency managers, weather forecast offices, and the public. The question is, how accurate is the current NWS's total water level forecast system and how can we improve it?

Methodology

✓ This project required the following general procedure:

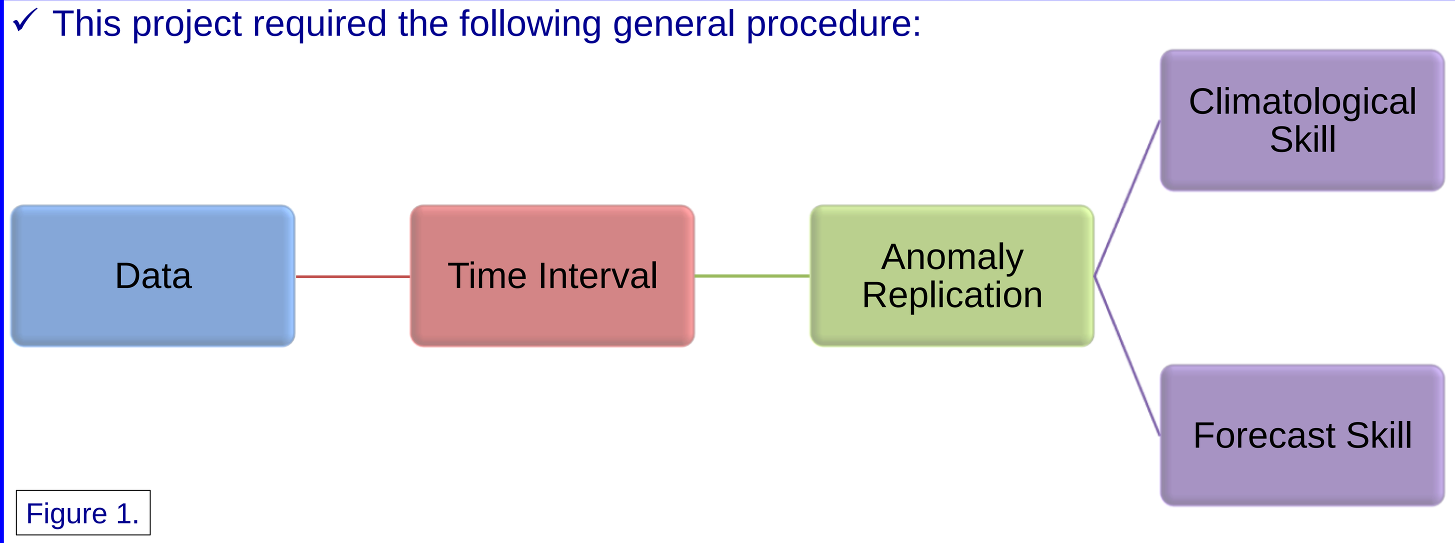

Figure 1.

✓ To determine the climatological skill of the total water level system we plotted its Mean Absolute Error (MAE) and bias as a function of time **(see Fig. 2)**. We then compared this to a baseline error, which did not compute storm surge.

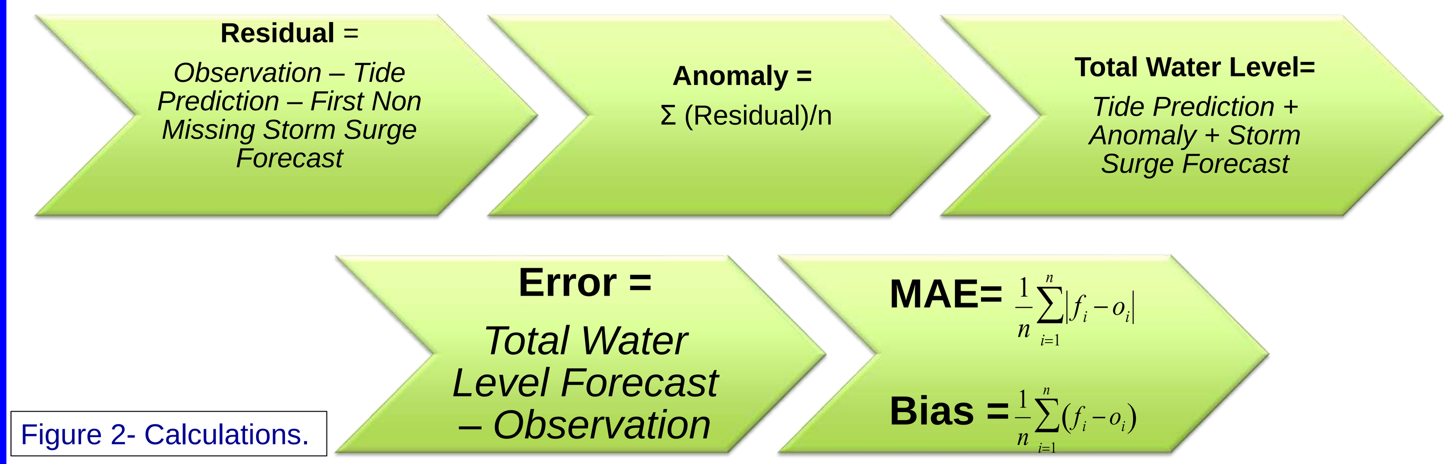

Figure 2- Calculations.

✓ To determine the skill of the total water level system to detect high water level and warn people, we said that a high water event occurred if the Highest Astronomical Tide (HAT) was exceeded in a 12-hour period. Given this definition we could generate a 2x2 contingency table and calculate the Threat Score (TS) and Probability of Detection (POD) **(see Fig. 3)**.

FORECAST		OBSERVATIONS	
		YES	NO
	YES	A	B
	NO	C	D

TS= A/(A+B+C) POD= A/(A+C)

Figure 3- Contingency Table.

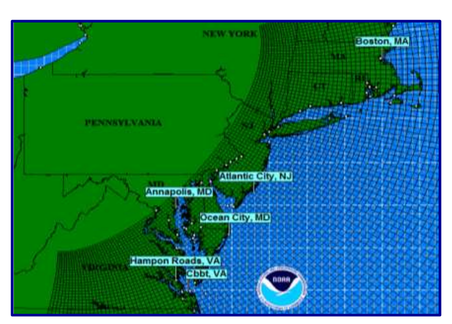

Figure 4- Stations studied of the East Coast if the US.

Figure 5- Stations studied of the Gulf of Mexico.

Verification

✓ In order to detect possible mistakes in our methodology and identify climatological changes, such as sea level rise, we initially worked with individual years for Boston, MA; Chesapeake Bay Bridge Tunnel, VA; and Hampton Roads, VA **(see Fig. 4)**.

✓ After we were confident in our methods we merged multiple years together and added more stations of the East Coast of the U.S. and of the Gulf of Mexico **(see Fig. 5)**.

Experiments

✓ Once we completed the verification efforts we explored the possibility of making some changes to the total water level system. In particular we considered what would happen if we changed the way the anomaly was calculated from taking a 5-day average (M5) of the excess water (also known as residual) to a 3-, 7-, or 30-day average (M3, M7 or M30).

Results (Climatological Skill)

Verification

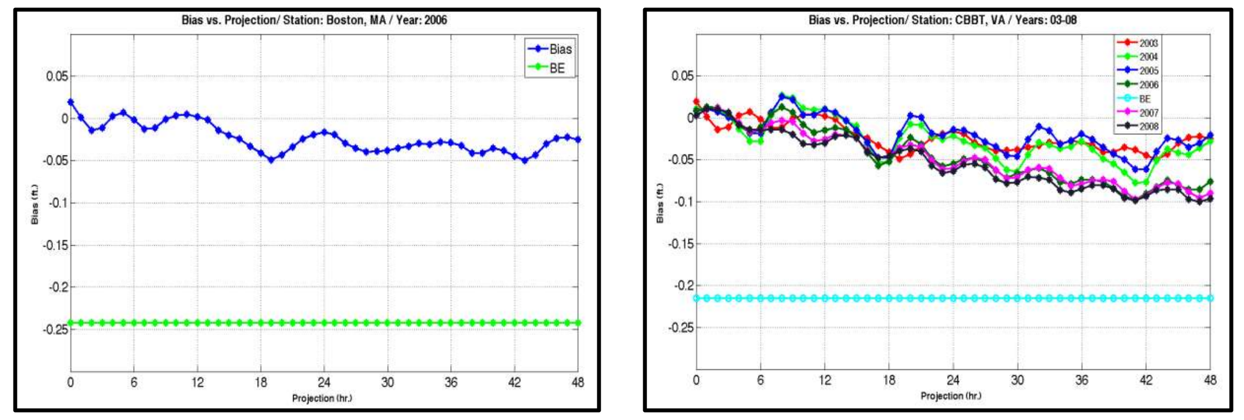

Figure 6. (a) (b)

Figure 6a shows that the values of the bias were relatively low in comparison to the baseline error. One interesting fact that we found was that for 2006, the bias had a repeating "W" pattern which indicated a diurnal pattern to the bias.

In the results plotted in **Figure 6b**, we found that each year has a very similar pattern in the bias, but we noticed that the lowest values were in the last 3 years.

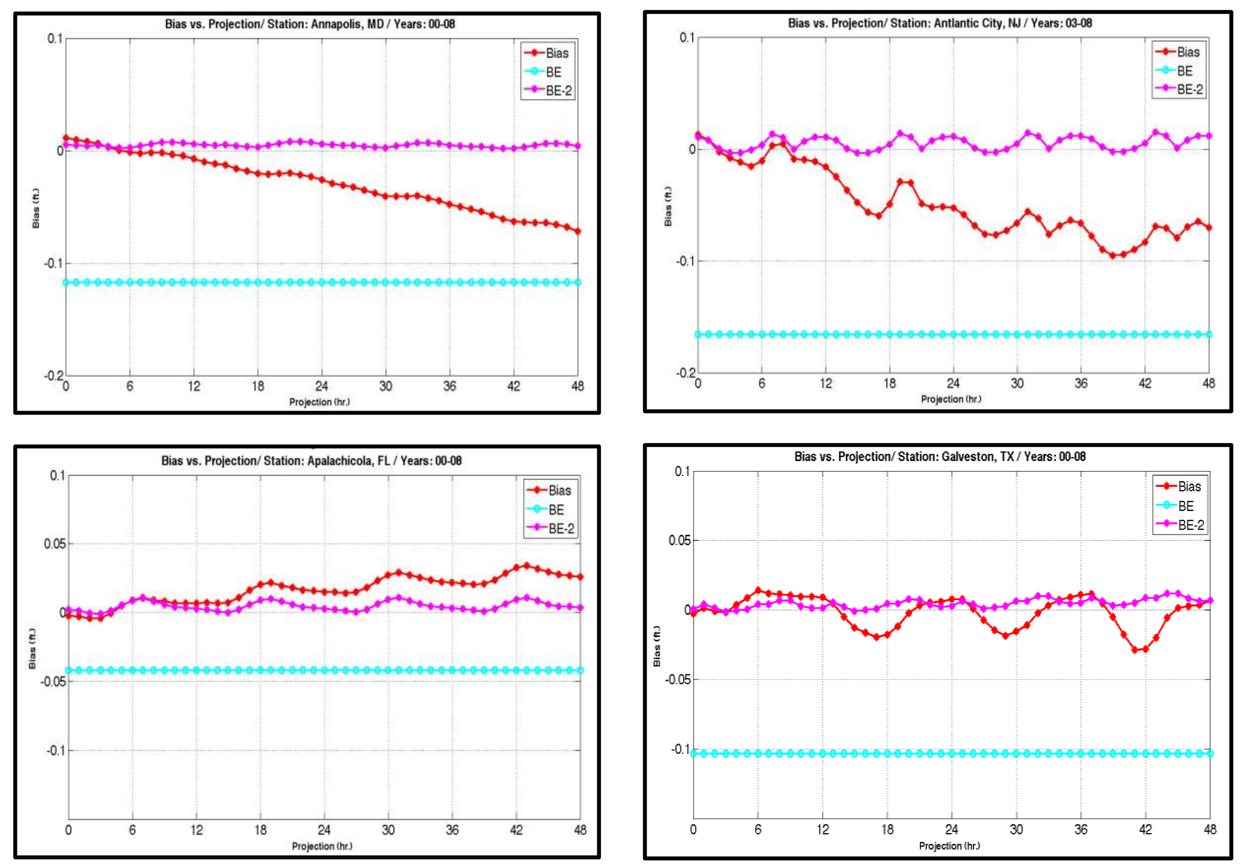

Figure 7. (a) (b) (c) (d)

This figure shows the results found by studying multiple years together for the stations of the East Coast of the US **(Figs. 7a and 7b)** and the Gulf of Mexico **(Figs. 7c and 7d)**. Notice that we included a second type of baseline error (BE-2) which adjusted the original baseline error with a 5-day average of the residual, without a storm surge forecast. This removes the bias from the baseline error and indicates that the storm surge forecast is introducing a negative **(Figs. 7a, 7b, and 7d)** or a positive **(Fig. 7c)** bias in the late periods.

Experiments

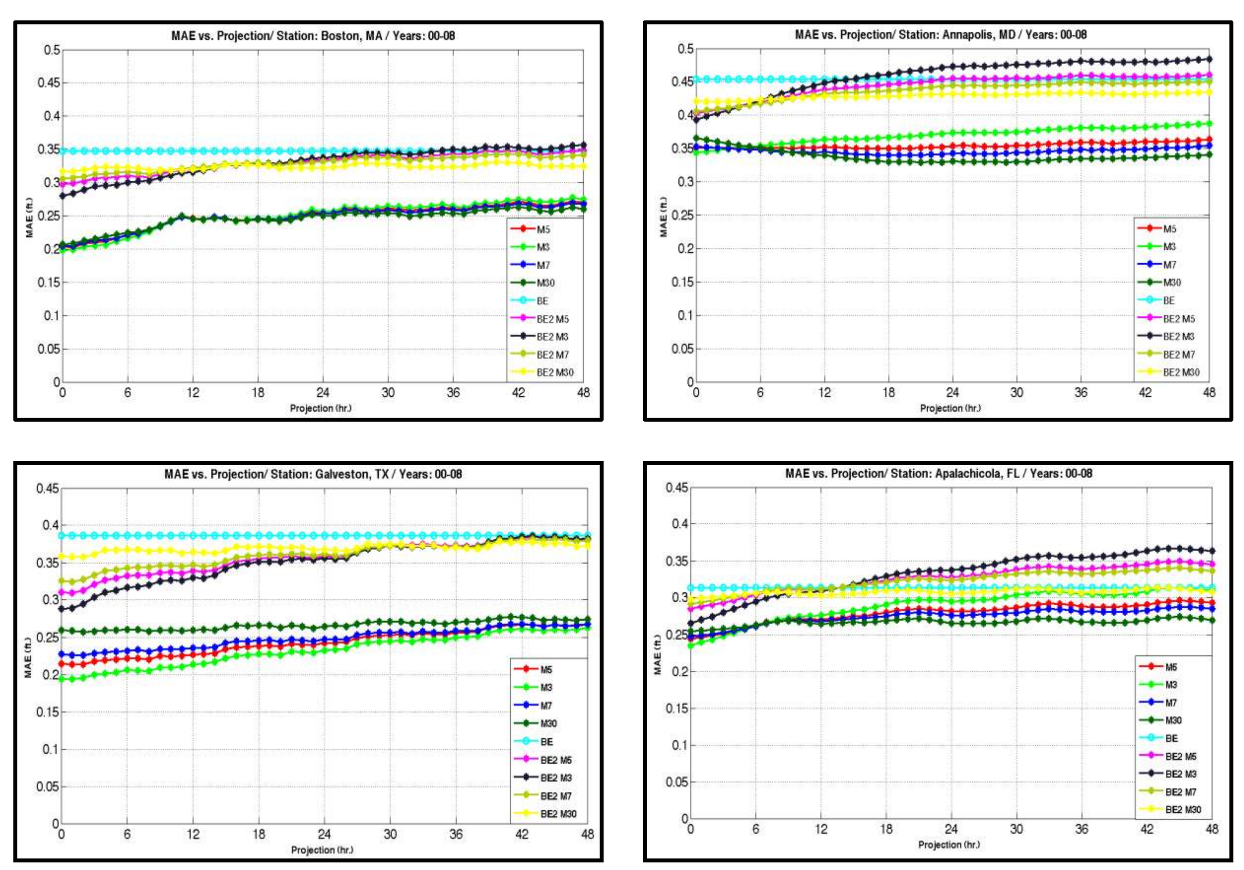

Figure 8. (a) (b) (c) (d)

Figures 8a and 8b show that if the system uses a 3- or 7-day anomaly, or a combination of both, it would generate total water level forecasts with a smaller MAE for the stations of the East Coast of the US. The same happens for the stations located the Gulf of Mexico **(Figs. 8c and 8d)** for which the difference is more noticeable. However, note that the improvement is on the order of 0.025 ft. which may not be operationally significant.

Results (Forecast Skill)

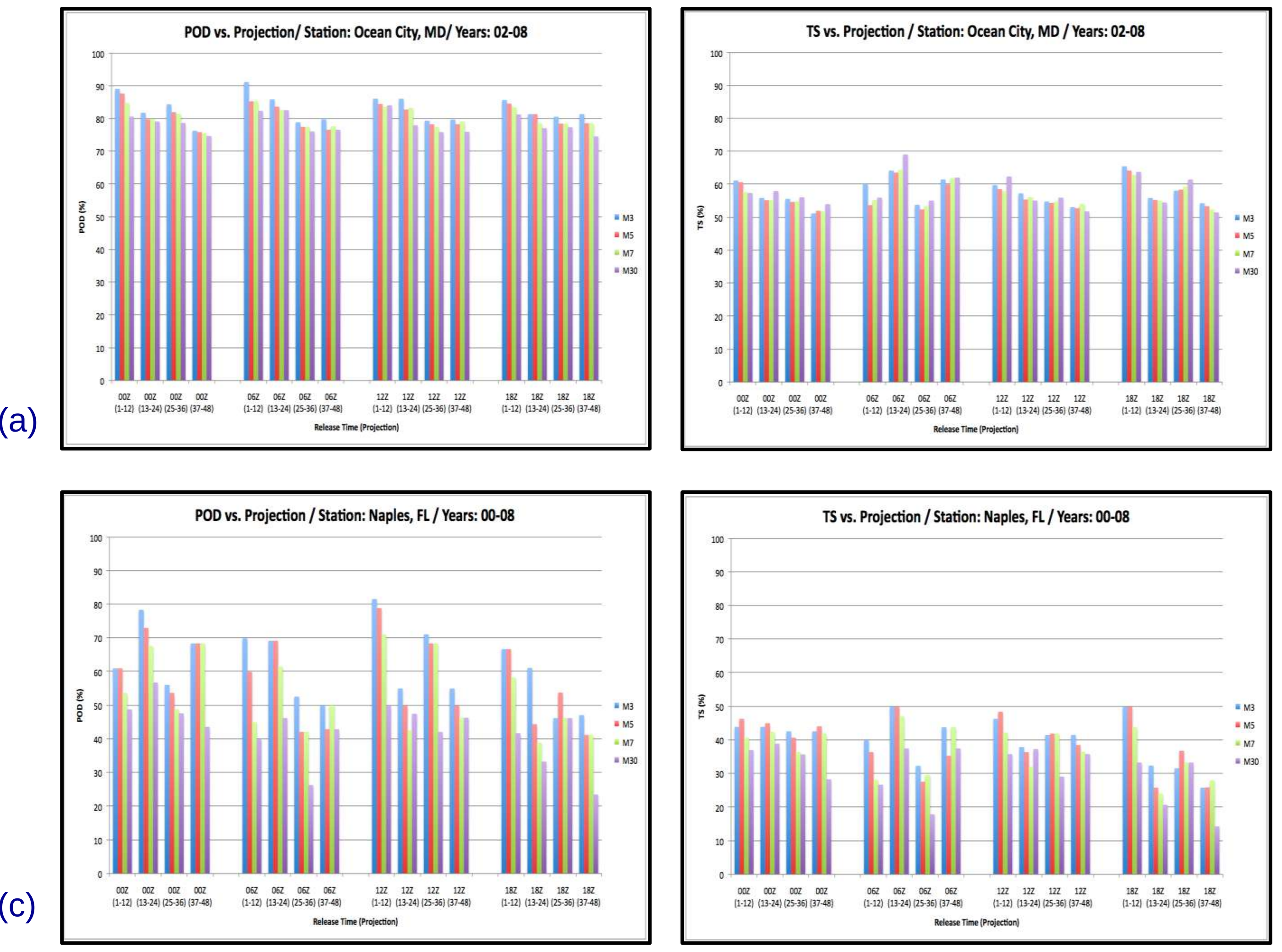

Figure 9. (a) (b) (c) (d)

This figure displays the charts of Probability of Detection (POD) and Threat Score (TS) that were generated using the 2x2 contingency tables. The perfect mark for both scores is 1, which means that most of the time the best results for the regions studied are obtained using the M3 anomaly **(see Figs 9a, 9b, 9c and 9d)**. However, one can notice that both scores are higher in the station of Ocean City **(Fig. 9a and 9b)** which belongs to the East Coast of the US. This fact applies for all the anomalies studied, including the 5-day average that the system actually uses.

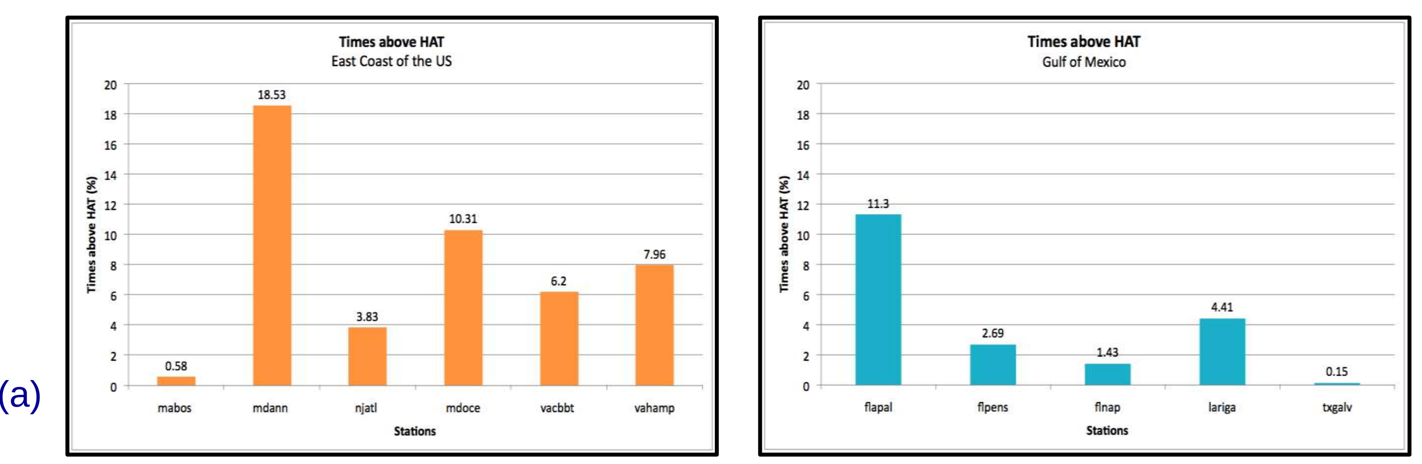

Figure 10. (a) (b)

As shown in **Figures 10a and 10b**, there is a large disparity in the percent of times the observed total water level exceeded the HAT among the stations.

Conclusions

•The results show that the current system is accurate and has skill.

•The MAE and bias charts reflect that using a different anomaly could improve the forecasts, but the difference is not very large.

•The system works better for the East Coast of the US than for the Gulf of Mexico, possibly due to the difference in the amount of extratropical cyclones that affect each region.

•There is a large disparity in the percent of times the observed total water level exceeded the HAT among the stations.

Recommendations

•Further investigation is needed to determine the best solutions, but using a 3- or 30-day anomaly would improve the system. In doing so, one should consider whether the improvement is significant enough to justify the additional complexity.

•We recommend investigating the use of a different threshold chosen so that the chance of exceeding it is consistent.

Acknowledgements

I would like to thank my mentor, Mr. Arthur Taylor, Physical Scientist of MDL, Dr. Wilson Shaffer, Chief of the Evaluation Branch of MDL; Dr. Bob Glahn, Director of MDL; and Ms. Valery Dagostaro, Meteorologist of the MDL, for their contributions and support for this project. Also I would like to thank the NOAA Educational Partnership Program for funding this scholarship.