4.2 LATEST DEVELOPMENT IN THE NWS' PROBABILISTIC EXTRA-TROPICAL STORM SURGE MODEL

Huiqing Liu[1*] and Arthur Taylor[2]
1. Ace Info Solutions, Reston, VA
2. NOAA / NWS / STI / Meteorological Development Laboratory, Silver Spring, MD
(*) Presented at the 100th AMS Annual Meeting, Boston MA, January 14, 2020

1. INTRODUCTION

The National Weather Service's (NWS) Meteorological Development Laboratory (MDL) implemented the Extra-Tropical Storm Surge (ETSS) model in 1995 (Kim et al. 1996) and Probabilistic Extra Tropical Storm Surge (P-ETSS) model in 2017 (Liu and Taylor 2018). In 2000, MDL also implemented a station bias adjustment post-processing methodology for ETSS to statistically account for components such as sea level rise, waves, river flooding, and model error. More recently, MDL enhanced P-ETSS by using the 42-member North American Ensemble Forecast System (NAEFS) instead of the 21-member Global Ensemble Forecast System (GEFS). Additionally, a station bias post-processing methodology, similar to the one in ETSS, was applied to P-ETSS (Liu et al. 2019). These enhancements are scheduled to be implemented in 2020.

However, a gap in skill still exists between the NWS probabilistic guidance for tropical and extra tropical storms. This is because Probabilistic tropical cyclone storm Surge (P-Surge) (Taylor and Glahn 2008) uses approximately 630 ensemble wind members, which is fifteen times more than the 42 ensemble members within the NAEFS-based P-ETSS model. To increase the spread of the P-ETSS results, MDL has worked to incorporate the European Centre for Medium-Range Weather Forecasts 51-member Ensemble Prediction System (ECMWF-EPS) into the P-ETSS model.

Additionally, P-ETSS does not account for model bias, nor water level components such as sea level rise, waves, and river flooding. As mentioned, MDL addressed this in 2019 by importing ETSS' station based post-processing methodology into P-ETSS. This enables recent observations to statistically account for those extra water level components. However, while this post-processing improves the station guidance, it doesn't improve the inundation calculation done by the model. In order to improve guidance over the whole grid, MDL has worked to statistically account for those components in a pre-processing stage through an initial water condition. This was done by averaging the anomaly at all station observations within the model domain to estimate an initial water level.

This paper describes the details of these two efforts and provides validation using historic events. Section 2 describes the improvements to P-ETSS. The motivations of doing these improvements are discussed in Section 3. Section 4 lists the historic storms along with observations used to validate the P-ETSS results. Section 5 presents the results. The paper concludes with a summary and discussion in Section 6.

2. IMPROVEMENTS TO P-ETSS

P-ETSS is a coastal inundation ensemble model forced by ensemble wind forcing systems. The P-ETSS system provides storm surge and overland inundation guidance four times a day based on surge and tide for all continental U.S and Alaska coastal areas. It provides finer resolution

---

* Corresponding Author: Huiqing Liu, 1325 East-West Hwy, Silver Spring, MD 20910-3280; e-mail: huiqing.liu@noaa.gov

guidance along the East Coast and the Gulf of Mexico by nesting high resolution basins. The current operational version (P-ETSS 1.0) is forced by the 3-hourly, 0.5-degree (55 km) GEFS, which has 21 ensemble members.

The next version, P-ETSS 1.1, scheduled to be implemented in October 2020, increases the number of ensemble members via the 42 member NAEFS. Additionally, the East Coast basin was expanded to cover Puerto Rico and the Virgin Islands in anticipation of future wave coupling within the model. The Gulf of Mexico basin was also expanded to cover the entire Gulf of Mexico and parts of the Yucatan Peninsula, which should allow P-ETSS to better model phenomena that come from outside the coastal area of interest (e.g. forerunner surge or reflection of waves off the Mexican coastline). P-ETSS 1.1 also incorporated a station based bias correction post-processing stage.

The latest efforts to improve P-ETSS are listed as green colored text in Table 1. The first improvement listed in the table is to increase the number of ensemble members from 42 to 93 by including the 51 ensemble members from the ECMWF-EPS. This is planned to be part of P-ETSS 1.2 which is scheduled to be implemented in 2022. The second improvement in Table 1 is to add a pre-processing stage to estimate an initial water condition. This is planned to be part of P-ETSS 1.1.

3. MOTIVATIONS

3.1 *Include ECMWF-EPS*

To reduce the threat to life and property, it is important that NWS forecasters and emergency managers receive highly reliable wind and storm surge guidance. However, wind forecasts now and in the near future have considerable uncertainty. A recent example is Hurricane Florence, which made landfall as a Category 1 hurricane near Wrightsville Beach, NC early on 14 September 2018.

Figure 1 shows the Hurricane Florence Advisory 47 ensemble forecast tracks from the GEFS (blue) and ECMWF-EPS (green) roughly 3 days before landfall. The GEFS indicated that Hurricane Florence would make landfall somewhere in the middle of North Carolina's outer banks, whereas the ECMWF-EPS indicated it would make landfall on the northern part of South Carolina's coastal shore. While this particular

Table 1. The current status and plans for P-ETSS. Green text indicates a feature that is discussed in this paper.

	P-ETSS 1.0 – Nov 2017	P-ETSS 1.1 – Oct 2020	P-ETSS 1.2 - 2022
Phenomena	Overland inundation based on surge + tide		
Area - No Nesting	Gulf of Alaska (Apr 2008) Bering Beaufort Chukchi Seas (Nov 2015) West Coast (Feb 2017)		
Area - Fine Resolution Nesting	East Coast (Feb 2009) Gulf of Mexico (Jan 2011)	East Coast (Feb 2018) Gulf of Mexico (Jun 2018)	
Forcing Resolution	3-hourly (GEFS)	3-hourly (NAEFS)	3-hourly (NAEFS) 6-hourly (ECMWF-EPS)
	0.5 degree (55-km)	0.5 degree (55-km)	0.5 degree (55-km)
Forcing Frequency	4x a day – 21 member GEFS	4x a day – 42 member NAEFS	4x a day – 93 member 42 NAEFS + 51 ECMWF-EPS
Post-Processing	No	Station based bias adjustment	
Initial water condition	No	Yes	

ECMWF-EPS forecast is better than the GEFS forecast in terms of where landfall occurred, neither ensemble system alone provided a wide enough spread to capture the uncertainty. Instead, to best capture the uncertainty, both sets of tracks should be combined together.

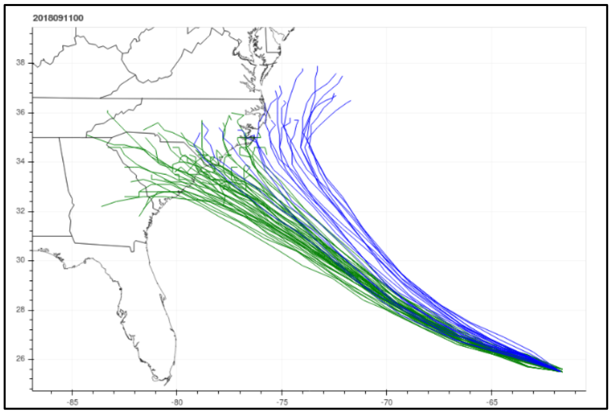

Figure 1. Hurricane Florence track forecast from Advisory 47. Blue represents the GEFS forecast tracks and green represents the ECMWF-EPS forecast tracks.

To see the impact on P-ETSS of this type of disparity in forecasts, three runs were made with Advisory 47: 1) used the 42-member NAEFS for forcing; 2) used the 51-member ECMWF-EPS for forcing, and 3) used both the NAEFS and ECMWF-EPS (93 members) for forcing. The results at Duck-Pier, NC (NOS CO-OPS id 8651370) are shown in Fig. 2. Duck-Pier, NC is located at the northern part of North Carolina's outer banks. The results from using the NAEFS as forcing (top panel) show over forecasting of the ensemble mean but a reasonable ensemble spread. The results from using the ECWMF-EPS as forcing (middle panel) show that the ensemble mean has a good agreement with observations but with a small ensemble spread. The results from using both NAEFS and ECMWF-EPS as forcing (bottom panel) show both a good ensemble mean and a reasonable ensemble spread.

The results from this particular Advisory for Hurricane Florence clearly show that including ECMWF-EPS is beneficial. To show that this conclusion is robust, a quantitative analysis on impacted stations using retrospective model runs from four recent historical hurricane cases is done in Section 4.

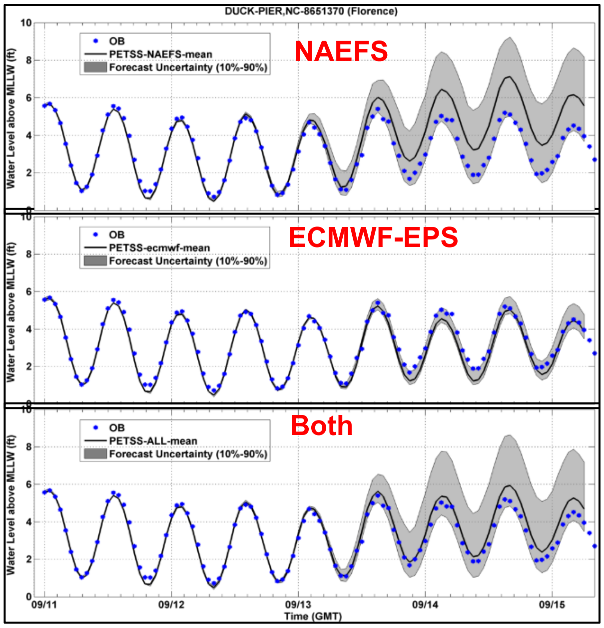

Figure 2. P-ETSS results from Advisory 47 at Duck Pier, NC (NOS CO-OPS id 8727520). The top panel is NAEFS-based P-ETSS, the middle panel is ECMWF-EPS-based, and the bottom panel is based on both NAEFS and ECMWF-EPS.

3.2 *Initial Water Condition Addition*

As P-ETSS does not account for model bias, nor water level components such as sea level rise, waves, and river flooding; an "anomaly" in the model guidance is unavoidable. The "anomaly" is defined as: Anomaly = Observation – Tide – Surge. A good example of such an anomaly is given in Fig. 3, which shows P-ETSS guidance without an initial water condition for 0600 UTC on 10 December 2019 at Cedar Key, FL (NOS CO-OPS id 8727520). It shows there is an approximately 1-foot water level anomaly (green line), which is the average anomaly from the previous five days of model runs. To resolve this, a station based post-processing methodology was implemented to statistically account for those components based on recent observations.

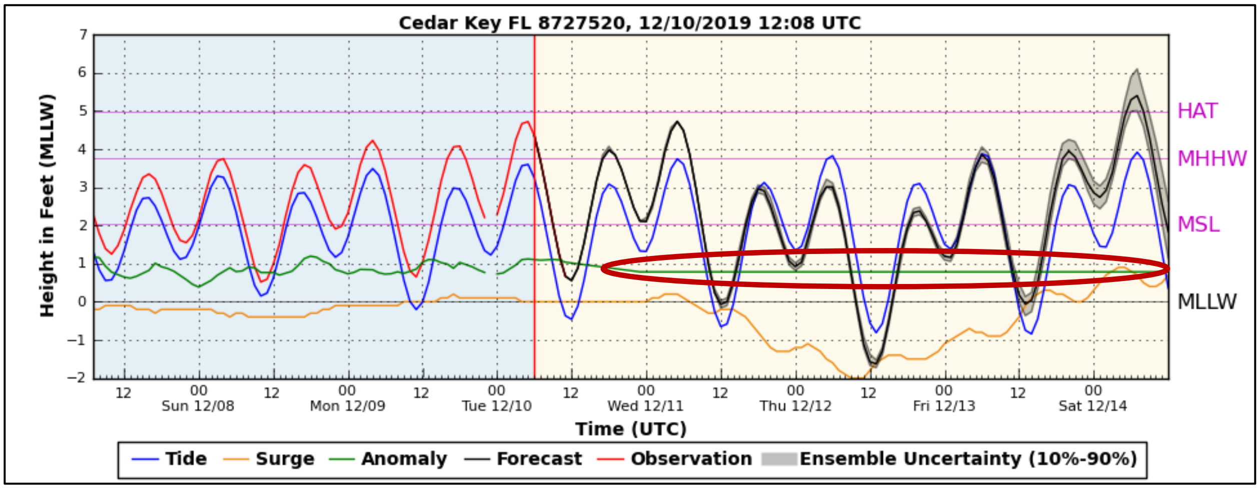

Figure 3. P-ETSS without initial water condition guidance at Cedar Key, FL for 0600 UTC on 10 December 2019.

Unfortunately, this doesn't help the model guidance at grid cells that are not coincident with station observations, nor does it help the inundation guidance since the statistical anomaly calculation is done after the inundation calculations. To improve the model results over the whole grid and the inundation guidance, a pre-processing stage based on station anomalies needs to be introduced as an initial water condition before the model run.

4. HISTORICAL STORMS

To evaluate the impact on P-ETSS of including ECMWF-EPS and adding an initial water condition, retrospective model runs were made for four storms that occurred during the past two years. Table 2 lists the four storms chosen to evaluate the P-ETSS model, which includes: Hurricanes Alberto and Florence in 2018, and Hurricanes Barry and Doran in 2019. Figure 4 shows NHC's hindcast tracks for these four storms along with the stations used to evaluate the model performance.

Table 2. The storms used to validate P-ETSS improvements.

Storm-name	Year	Forecast Start Time
Alberto	2018	May 23 18Z – 31 06Z
Florence	2018	Sep 09 18Z – 17 06Z
Barry	2019	July 06 18Z – 14 06Z
Dorian	2019	Aug 31 18Z – Sep 08 06Z

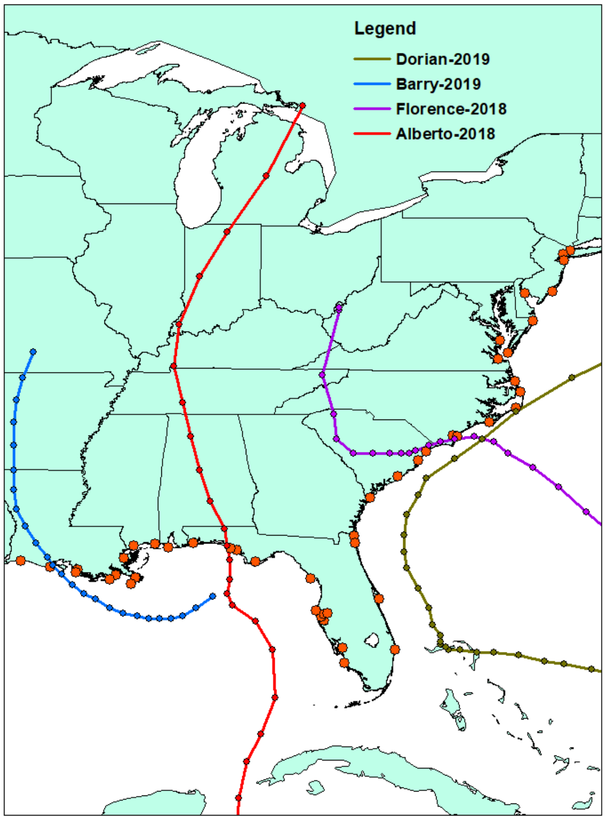

Figure 4. NHC's hindcast tracks for Hurricane Alberto-2018, Florence-2018, Barry-2019, and Dorian-2019, along with station locations.

5. RESULTS

5.1 *ECMWF-EPS*

P-ETSS ensemble mean skill scores for 12-, 24-, 36-, 48-, 60-, 72-, 84-, and 96-h projections were evaluated against the observations from tide gauges during a specific 96-h time window. The reason a 96-h time window was selected was to focus on when the water levels were most significantly impacted. The 96-h time series was created by splicing together 6-h slices from consecutive model runs. For example, the 24-h projection window spliced hours 19 to 24 from one model run to hours 19 to 24 from the next consecutive model run. This results in a relatively constant projection thereby reducing the impact of errors within different projections on the assessment. So the 24-h projection requires 16 consecutive model runs, and the 12-h projection requires 2 additional earlier model runs. The result is all 8 projections require a total of 30 consecutive model runs.

Model performance was then assessed based on the average of the following scores over the various tide gauge observation time frames:

1) Root Mean Squared Error (RMSE),

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}}$$

2) Peak Absolute Error (PAE),

$$PAE = abs(\max(X_{obs,i}) - \max(X_{model,i}))$$

To assess how well it captures the uncertainty of the forecast, we calculated another skill score:

3) Coverage of Observations by forecast area of Uncertainty (COU),
$COU = \frac{n}{96} * 100$

where n was the total number of hourly observations that fall inside the area of P-ETSS model uncertainty during the 96-h time frame.

The average RMSE and PAE for the tide gauge observation time frames for the 12-, 24-, 36-, 48-, 60-, 72-, 84-, and 96-h projections are shown in Fig. 5. The results show that for three of the four cases, the average RMSE scores for ECMWF-EPS-based P-ETSS were better than the NAEFS-based P-ETSS for most projections. The exception case was Hurricane Dorian, where the average RMSE scores for the NAEFS-based P-ETSS were better for the 12-, 24-, 36-, 48-, and 60-h projections. The ECMWF-EPS-based P-ETSS had a better average PAE in all four cases for almost all projection hours. The combined ECMWF-EPS- and NAEFS-based P-ETSS always performed well in terms of RMSE and PAE.

The average COU score versus projection hours for the four storms are shown in Fig. 6. It indicates that the COU score from ECMWF-EPS-based P-ETSS is the smallest for almost all projections for the four storm events. The COU score for the NAEFS-based P-ETSS was better than the combined NAEFS- and ECMWF-EPS-based P-ETSS for Hurricanes Alberto and Barry. The opposite was true for Hurricanes Florence and Dorian. Overall the COU score was still far from 80%, which indicates there were still a significant number of times where the observations were not within the estimated area of uncertainty. This implies that more ensemble members, or a more diverse set of ensemble members, are needed to capture the uncertainty. Overall, the combined ECMWF-EPS- and NAEFS-based P-ETSS is the best choice.

5.2 *Initial Water Condition*

The initial water condition methodology developed for P-ETSS and the deterministic ETSS model sets the initial water value to the mean anomaly of all stations within a given basin. Each cycle, an updated mean anomaly is computed with any new observation data. This is then added to the initial water level from the previous cycle and the model is run. Through this iterative process, the model adjusts to the best solution based on the updated observations and the resulting mean anomaly.

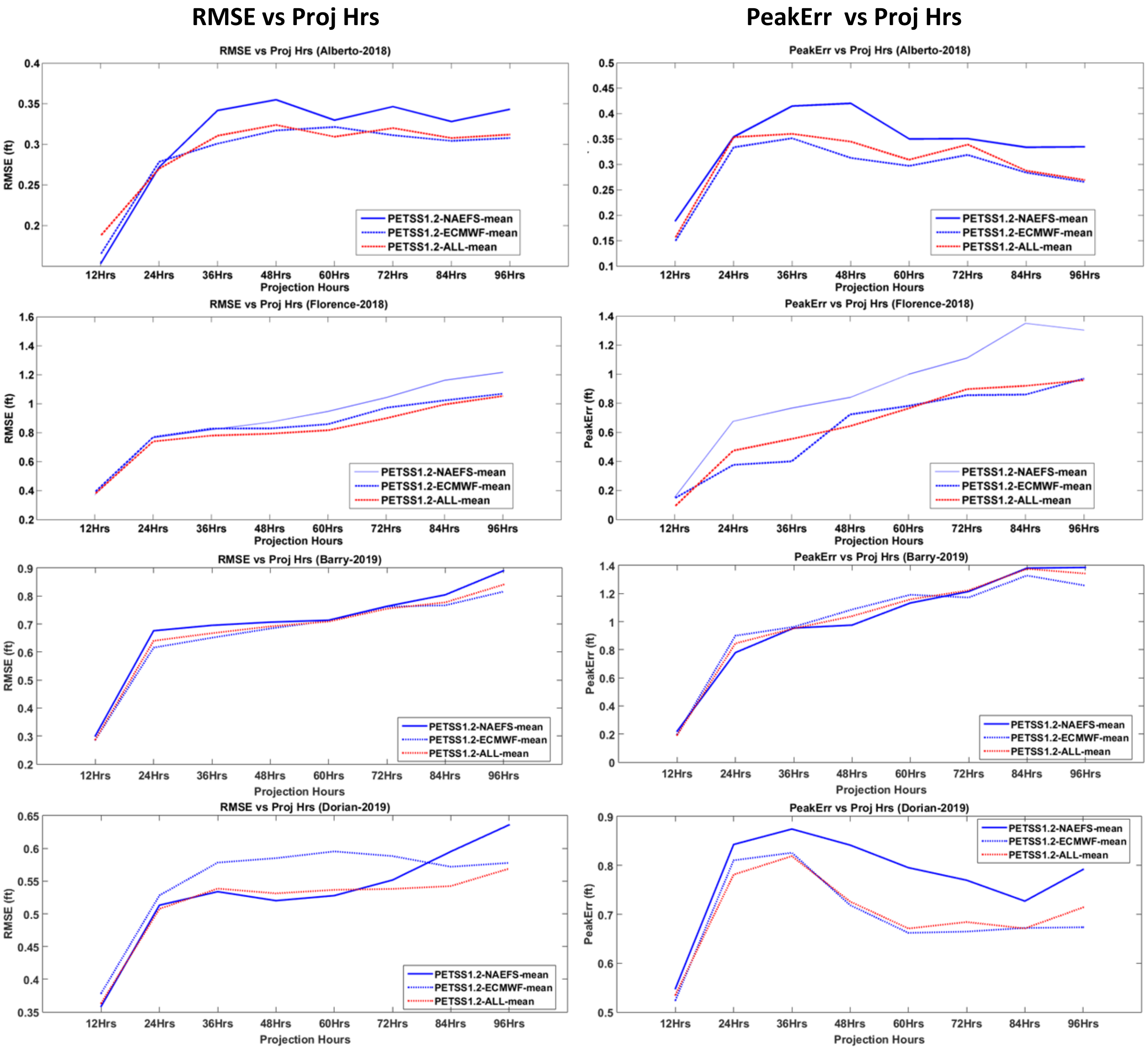

Figure 5. The left column is average RMSE versus projection hour. The right column is average Peak Error versus projection hour. From top to bottom the rows are: Hurricane Alberto, Florence, Barry and Dorian. The solid blue line represents the P-ETSS results from the NAEFS-based run, the dotted blue line represents the results from the ECMWF-EPS-based run, and the dotted red line represents the result from using both NAEFS and ECMWF-EPS.

A significant challenge with this methodology is determining the time periods over which to calculate the mean anomaly. Since the last cycle's initial water level is part of the initial water condition for the current run, in theory, the updated mean anomaly should only have to cover the most recent model cycle (6 hours). That way it would avoid double counting the anomaly. In practice, calculating the anomaly over a 6-h period is too short as it doesn't cover a full tide cycle, and doing so would result in a chaotic oscillation between cycle results at some stations. The likely reason for this is that there is a tidal phase shift at the station, due to incorrect tidal constituents or a non-linear interaction between the tide and surge. The phase shift results in a "fake" positive adjustment in the first 6 hours followed by a compensating negative adjustment in the second 6 hours.

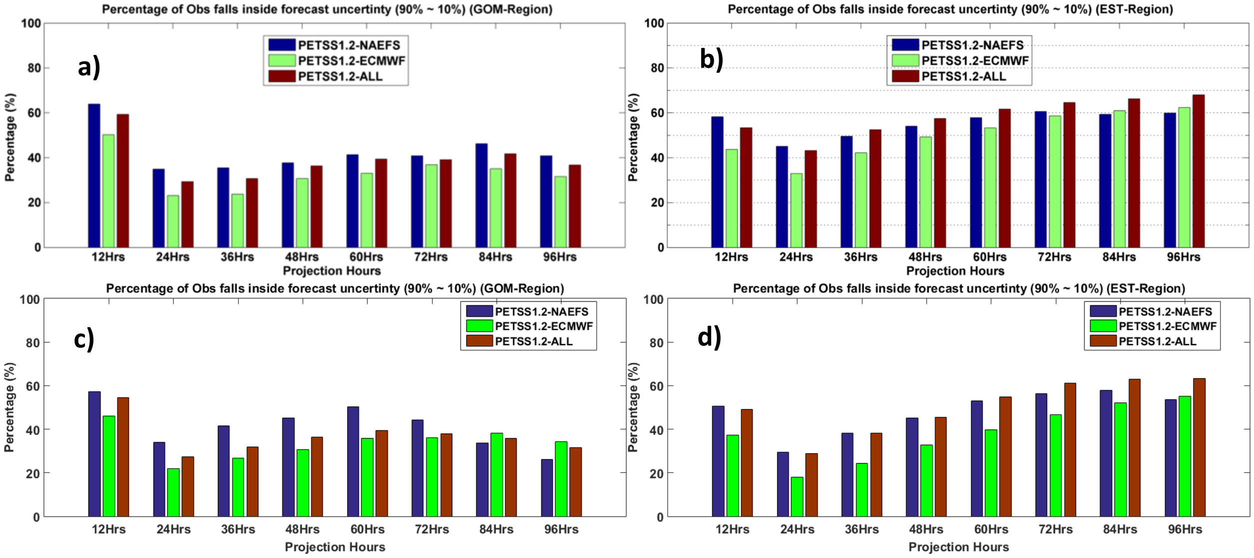

Figure 6. COU for P-ETSS per projection hour for runs based on NAEFS (blue), ECMWF-EPS (green), and both (red) for a) Hurricane Alberto, b) Hurricane Florence, c) Hurricane Barry, and d)

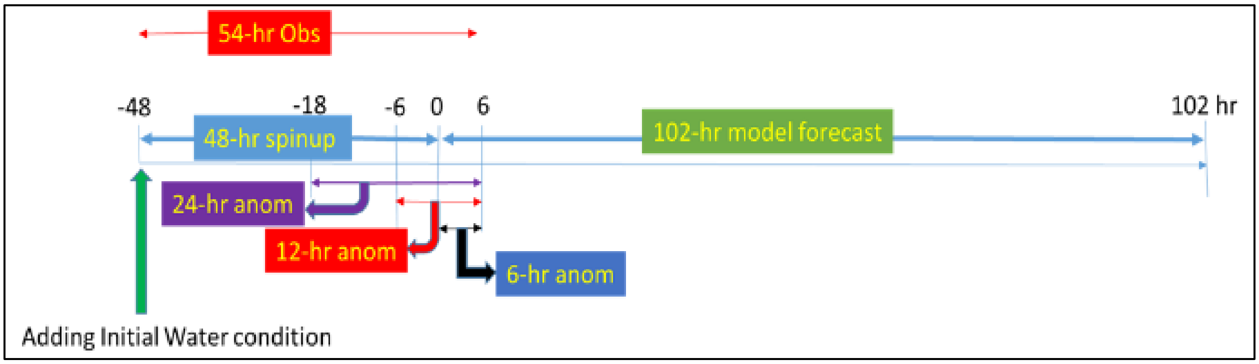

Figure 7. The process to create the initial water condition.

To determine the best time range for the anomaly calculation, four P-ETSS model runs were made with Hurricane Barry from 2019 (Advisory 10). These runs compared the case of no initial water condition with initial water conditions based on the mean anomaly calculated over 12, 18, and 24 hours. This is depicted in Fig. 7. The RMSE and PAE scores were used to evaluate the performance of P-ETSS. The difference in this section (versus the last section) being a focus on individual stations versus averages over all stations.

The RMSE and PAE for each of the 9 impacted stations are shown in Tables 3a and 3b. The RMSE and PAE show significant improvements in P-ETSS with initial water condition for almost all stations. The performance of P-ETSS was comparable when the mean anomaly was calculated over 12, 18, or 24 hours.

The same experiment was done with Hurricane Dorian from 2019 (Advisory 38), except we dropped the mean anomaly over 18 hours option. The RMSE and PAE for each of the 14 impacted stations are shown in Tables 3c and 3d. The results support a similar conclusion as with Hurricane Barry from 2019, with the notable exception of Ocean City Inlet, MD (NOS CO-OPS id 8570283) where both the RMSE and PAE preformed best with no initial water condition.

Table 3a) RMSE at stations for Hurricane Barry with no anomaly, or calculated over 12-, 18-, or 24-h. The green color indicates the run with lowest score.

stn	LA1	LA2	LA3	LA4	LA5	LA6	MS1	MS2	AL1
No	0.50	0.78	1.20	1.63	0.70	1.04	**0.55**	0.72	1.00
12-h	**0.50**	**0.35**	0.42	0.83	**0.32**	0.50	0.59	0.25	**0.16**
18-h	0.56	0.43	**0.31**	**0.72**	0.38	**0.48**	0.69	0.27	0.22
24-h	0.54	0.40	0.35	0.75	0.36	0.55	0.66	**0.22**	0.29

Table 3b) Same as Table 3a but for PAE.

Stn	LA1	LA2	LA3	LA4	LA5	LA6	MS1	MS2	AL1
No	0.83	1.59	1.63	1.91	1.13	1.87	0.51	0.79	1.12
12-h	**0.23**	0.69	0.73	1.01	0.43	1.17	**0.39**	0.11	**0.22**
18-h	**0.23**	**0.59**	**0.63**	**0.91**	**0.23**	**1.07**	0.49	0.11	0.32
24-h	**0.23**	**0.59**	0.73	1.10	0.33	1.27	0.49	**0.01**	0.42

Table 3c) RMSE at stations for Hurricane Dorian with no anomaly, or calculated over 12-, or 24-h. The green color indicates the run with lowest score.

stn	SC1	GA1	NC1	NC2	NC3	VA1	VA2	VA3	MD1	NJ1	NJ2	N3J	NY1	NY2
No	0.92	0.91	0.88	1.14	0.64	0.62	0.76	0.72	0.81	0.63	0.71	0.67	0.60	0.59
12-h	0.91	1.04	0.49	0.94	0.38	0.46	0.23	0.51	0.92	0.51	0.56	0.49	0.48	0.49
24-h	0.90	1.03	0.49	0.94	0.41	0.50	0.26	0.54	0.95	0.58	0.64	0.54	0.55	0.56

Table 3d) Same as Table 3c but for PAE.

stn	SC1	GA1	NC1	NC2	NC3	VA1	VA2	VA3	MD1	NJ1	NJ2	NJ3	NY1	NY2
No	1.77	1.87	1.33	3.62	0.83	1.02	1.04	0.76	1.28	0.51	0.50	0.31	0.12	0.54
12-h	0.67	0.87	0.63	2.82	0.03	0.02	0.14	0.04	2.18	0.39	1.30	1.11	0.92	0.26
24-h	0.67	0.97	0.53	2.72	0.08	0.02	0.14	0.04	2.28	0.49	1.50	1.21	1.02	0.36

After reviewing the station locations, Ocean City Inlet (station MD1 in the above tables) was the only station located along the open ocean. All the other stations are located inside bays or inlets. This meant that the model behaved differently at Ocean City Inlet, which in turn meant that there was a higher chance that the model would over forecast at Ocean City Inlet station and under forecast elsewhere (or vice versa). In that case, the initial water condition created by the mean anomaly from all stations would increase the water level at Ocean City Inlet station, thereby creating an over-forecast. To avoid this, we will need to develop a methodology to generate spatially varying initial water conditions for the ETSS and P-ETSS models.

The comparison of the time periods showed that the 12-h period had a slight improvement. It is also the closest time period to the idealized case (6-h period) that avoids the chaotic oscillation. So, the 12-h period was selected as the time period over which the mean anomaly is calculated.

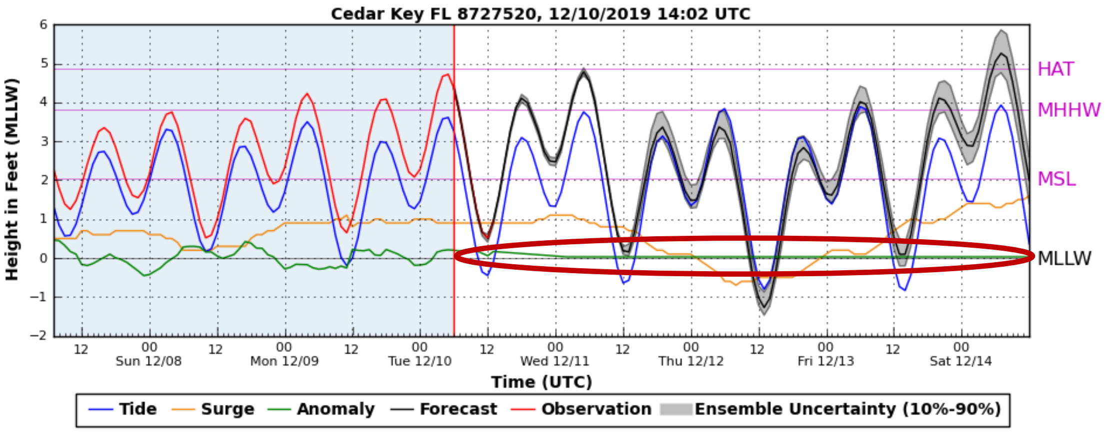

Figure 9. Same as Figure 3 but with an initial water condition.

Having chosen the time period of the initial water condition calculation, P-ETSS was run for the same case as in Fig. 3, but with an initial water condition. The guidance at Cedar Key, FL is shown in Fig. 9. Note that the 1-foot water level anomaly from Fig. 3 has disappeared. This means the mean model anomaly over the previous 5 days is zero. In other words, the model can now see the initial water condition and it can compute a more realistic field of inundation.

6. SUMMARY AND DISCUSSION

For all of the storm events, incorporating ECMWF-EPS provided better guidance for the 12- to 96-h forecast projections. Specifically, the ensemble mean of P-ETSS when forced by both NAEFS and ECMWF-EPS showed significant improvements over the version forced by NAEFS only. In terms of coverage of the uncertainty, the version forced by both NAEFS and ECMWF-EPS showed improvements in a large number of the storm events. However, the specific COU score per projection for the 4 storm events is still far from the goal of 80% for certain projection hours. More members will need to be added to expand its estimate of uncertainty, which will not only improve the COU, but also improve the performance of the ensemble mean. This improvement is planned to be implemented in P-ETSS 1.2.

Additionally, this paper described adding an initial water condition pre-processing stage to the P-ETSS model, which will be implemented in P-ETSS1.1. This pre-processing stage provides an efficient way to account for various biases such as mean sea level increase, omitted physical terms, and model errors. For most of the stations, adding the initial water condition to the pre-processing stage showed significant improvements over not doing so. Adding the pre-processing stage will also improve the inundation calculation as the model will be able to react to the extra water. Post-processing will still be done at stations, to adjust to local water level observations, but the adjustment will be smaller.

In the long term, MDL plans to enable ETSS and P-ETSS to use the parallel SLOSH version (Taylor and Liu 2020). That will enable ETSS and P-ETSS to utilize the updated (larger and finer) basins for South Florida, New Orleans, Texas, New York, etc. Finally, MDL plans to create finer basins to nest within Alaska's Bering, Beaufort, and Chukchi seas basin.

7. REFERENCES

Kim, S. C., J. Chen, and W. A. Shaffer, 1996: An Operational Forecast Model for Extratropical Storm Surges along the U.S. East Coast. Preprints, *Conference on Coastal Oceanic and Atmospheric Prediction,* Atlanta, GA, Amer. Meteor. Soc., 281-286.

Liu, H., and A. Taylor, 2018: Development of the NWS' Probabilistic Extra Tropical Storm Surge Model and Post-Processing Methodology. *16th Symposium on the Coastal Environment,* Austin, TX, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/98Annual/webprogram/Paper329410.html

______, A. Taylor, and K. Kang, 2019: Latest Development in the NWS' Extra Tropical Storm Surge, and Probabilistic Extra Tropical Storm Surge Model. *17th Symposium on the Coastal Environment,* Phoenix, AZ, Amer. Meteor. Soc., 3.8, https://ams.confex.com/ams/2019Annual/webprogram/Paper355246.html

Taylor, A., and Glahn, B., 2008: Probabilistic Guidance for Hurricane Storm Surge. *19th Conference on Probability and Statistics,* New Orleans, LA, Amer. Meteor. Soc., 7.4, https://ams.confex.com/ams/88Annual/techprogram/paper_132793.htm

______, and H. Liu, 2020: Latest Developments in the NWS' Sea Lake and Overland Surges from Hurricanes Model. *18th Symposium on the Coastal Environment,* Boston, MA, Amer. Meteor. Soc., 4.1, https://ams.confex.com/ams/2020Annual/webprogram/Paper370583.html