Dear Reviewer,

Thank you for your time and effort in helping us improve our manuscript. Here, we respond to all your comments, with your original comments repeated for your reference. Some of the responses have texts/figure copied from "AC1: Short comments in reply to Review #1", which are labeled as "repeating AC1". Following the journal's publishing guideline, the revised manuscript is not uploaded at this time, so all line numbers and figure numbers are based on the original manuscript.

<Reviewer's Comment #1>

The model is not calibrated. The authors selected a few constant values as the friction values at various locations in the models based on previous studies. Although the authors tried to justify this by saying "in favour of simplicity", I am not convinced that a model without calibration will be of great value.

<Response> (repeating AC1)

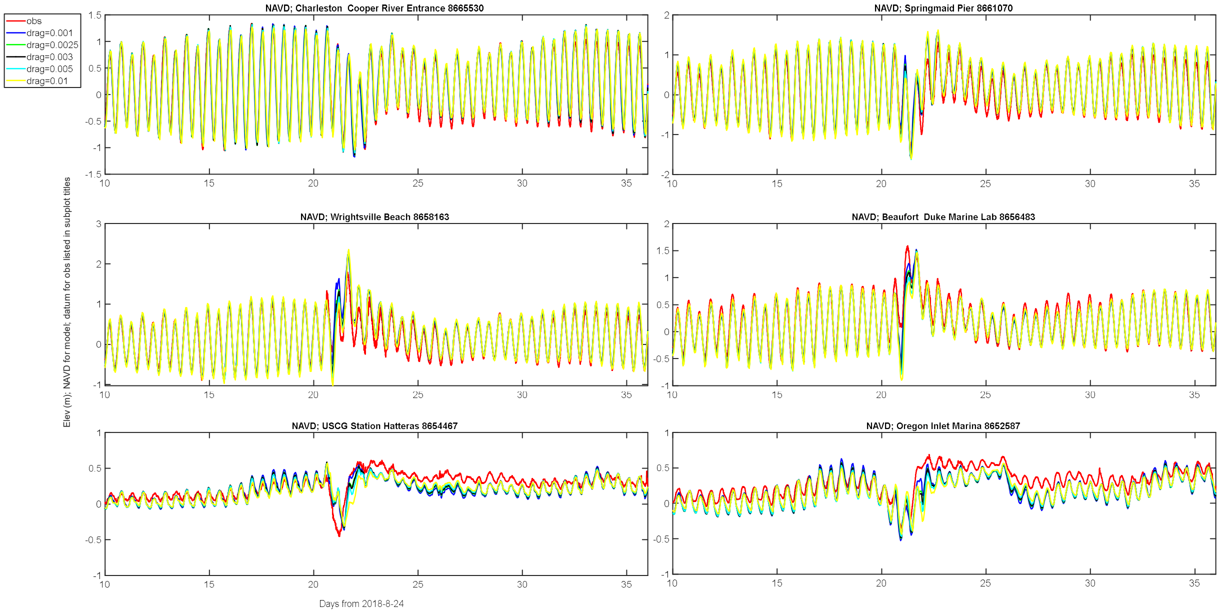
We agree with the reviewer that "The tuning of friction values is intentionally kept to minimum in favour of simplicity" is not a good statement. Calibration was indeed an important step during our model setup.

We remove the sentence on Line 156-159 and add more details on calibration:

"The friction of the baseline model was tuned in the wet area (river, estuary, ocean; lower than 1 m above MSL) and on higher grounds (higher than 3 m above MSL) separately. In the wet area, drag coefficients were tested within a range of 0.001-0.01. Commonly accepted default value of 0.0025 gave good error statistics near the landfall site; but the sensitivity is small in the range of 0.001-0.005."

"In the watershed, drag coefficients were tested within a range of 0.01-0.5. The optimal value was chosen based on the High Water Marks (HWMs) comparisons at over two hundred locations, collected by USGS. A small friction value within this range tended to under-predict the elevation at HWMs, and a large value led to over-prediction. Values with the range of 0.02-0.05 gave good error statistics. We chose 0.025 because it gave slightly better results in the Cape Fear River watershed near the landfall."

"Note that this is the parameterization based on the region influenced by Hurricane Florence. Spatially varying parameterization of bottom friction for different systems is an on-going effort as we study more recent hurricanes and operationalize the model along the East Coast and Gulf Coast. However, as we presented in Ye et al. (2020), Zhang et al. (2020) and Huang et al. (2021), the choices described above seem to work fine in general for other systems as well."



(The figure is for the reviewer's reference)

<Reviewer's Comment #2>

Is the NWM model calibrated?

<Response>

Yes, according to Gochis et al. (2018), "all USGS GAGES-II reference basins" and additional gages/basins are included in the calibration and validation; multiple evaluation criteria are applied with emphasis on bias reduction; then parameter sets are regionalized for the whole domain and re-validation.

We are not the developer of the National Water Model (NWM). In our model study, NWM is an external forcing like HYCOM and ERA, which all have uncertainties but are the state-of-the art and the best products available to us. We are open to use any other hydrologic sources to drive our model.

To clarify this, we add the following text after Line 190 of the original manuscript:

"The forcing errors in the magnitude and timing of NWM's peak flow should explain part of the model errors especially in the watershed. For example, we found that replacing the NWM streamflow with gaged flow at USGS Station 02109500 WACCAMAW RIVER AT FREELAND, NC improves the model skill locally. However, this is not cost-effective for our goal of operationalizing this compound flood model along the US East Coast and Gulf Coast. The developers of NWM (Gochis et al., 2018) showed that NWM's model skill was improved by each version update, with 44% of the gauges having bias < +/- 20% in the latest version (NWM v2.0). We will adopt the newest and best NWM version in our ongoing study and operational forecast as soon as it is available. And we are open to using any other hydrologic sources to drive our model."

<Reviewer’s Comment #3>

With such detailed 3D modelling of such a large area, what the efficiency of the model is like? E.g. what is the computational time for a flood event lasting for a specific period of time (e.g.3 days)? What is the specs of the computing facility used?

<Response>

We add the information on computational cost after Line 164 in “Section 3. Model description” of the original manuscript:

“Although the model covers a large domain, most of the elements are quasi-2D, making it efficient enough for operational forecast. For the baseline run, the real time to simulation time ratio is 80 with 1440 cores on TACC’s Stampede2 and 30 with 480 cores on W&M’s Sciclone. Intel Skylake cores with a nominal clock speed of 2.1 GHz were used on both clusters. A 3-day simulation will take about 0.9 hours using 1440 cores or 2.4 hours using 480 cores.”

<Reviewer’s Comment #4>

The authors reported the average MAE. But often it is the peak error that is important. What is the peak error, when and where did it occur? What is the potential impact of this peak error?

<Response>

Revised as suggested. The revised manuscript reads:

(Line 253) “The peak errors at different stations occur with the storm surge, with a maximum over-prediction of 0.64 m at Spring maid Pier, SC. The overpredicted peak surge can lead to overpredictions in elevation on high water marks (HWMs).”

…

(Line 274) “There is a slight positive bias on near-shore HWMs, corresponding to the over-prediction of peak elevation on coastal stations (Fig. 11).”

…

(Line 371) “Note that the ocean dominance near Wrightsville Beach, NC (NOAA Station 8658163) and Springmaid Pier, SC (NOAA Station 8661070) may be exaggerated (see locations in Fig. 3), considering the overestimated peak elevation there (Fig. 11)

<Reviewer’s Comment #5>

Figure 11: The figure shows modelling errors of up to +/-4 meters at various locations. Do the authors have an explanation on the large errors at these locations (apart from just saying calibration can improve model performance)? - The authors did a good job explaining model performance in relation to grid resolution. A similar explanation here will be good.

<Response>

In short, the few large errors are likely the result of similarly **under-resolved bathymetric/topographic features** or the **neglection of drainage system in the urban area**, with minor contributions from other factors such as the uncertainties in the NWM prediction and DEM. We add more explanations into the revised manuscript:

Line 184: "The drainage in urban settings is not included in our model. This may lead to some big errors in the prediction of elevations on high water marks, for example the one large error in the urban area of Fig. 13d. We do have a plan of explicitly accounting for infiltration as volume sinks based on NWM (or other hydrologic models). However, considering the additional uncertainty this would bring, for now we choose to continue improving more important aspects of the model (especially the quality of model grid, which is very likely responsible for the few large errors in Fig. 11c) for operational use."

…

Line 293: "Our tests show that the simulated elevation on the High Water Marks (HWMs) in the watershed is sensitive to: grid resolution, precipitation, river inputs through the land boundary, and bottom friction. Grid resolution and quality is the most important factor, as explained in the last paragraph of Section 4.3. Misrepresentation of flood routing can easily lead to errors of a few meters near some very localized features like ditches, highways etc. How to resolve all important small features efficiently is an on-going research area. The defects in grid quality can lead to large errors that are not likely to be rectified by tuning other parameters. To fix the remaining few large errors away from the landfall site, grid quality should be examined first. The continuous improvement on this model grid is part of an ongoing effort of operationalizing the model along the US East Coast and Gulf Coast, and we will report this in future studies. Afterwards, the inclusion of urban drainage should reduce the occasional large errors there (Fig. 13d). Other factors such as uncertainties in DEM, precipitation and the river flow through land boundary also play minor roles."

…

[Line 300-302] The only remaining large error in the "baseline" occurs in an urban area away from the river, likely due to the building or drainage effects that have not been incorporated in the model (Fig. 13d).

<Reviewer's Comment #6>

What is the return period of flooding at various locations during this event? A comparison on the return period for floods caused with or without compounding effect will give readers a clearer picture of the impact of the compound effect.

We agree with reviewer that a clearer picture of the impact of the compound effect is needed here. US Geological Survey (USGS) provides return period (or annual exceedance probability) of streamflow at selected gages (Feaster et al. 2018). However, many of these gages are at upstream locations of the modeled watershed. On one hand, they are not much affected by the ocean; on the other hand, these streams will be nearly dry without the NWM streamflow injected at the land boundary, making the interpretation of compound effects difficult.

So, we would like to propose an alternative with a more direct measure on the elevation, serving the same purpose of clarifying the compound effect. We propose to calculate the "percent inundated area" or "average inundation depth" inside the North Carolina and South Carolina watersheds and compare this index between the baseline and each sensitivity test.

<Reviewer’s Comment #7>

Please see comments on Figure 11.

<Response>

Fig. 14’s main purpose is to show the adverse effect of excluding NWM and precipitation. Fig.14 shows more large errors (under-prediction, blue) than in Fig. 11 (baseline), because Fig.14’s has no stream flow from NWM or direct precipitation.

We add more explanations in the 2nd paragraph of Section 5 to convey the idea more clearly:

“Not surprisingly, without rivers and precipitation, watershed is mostly dry as the storm surge cannot propagate over steep terrains (Fig. 12). As a result, the predicted HWMs are biased too low (Fig. 14) as the steep topography quickly damped out any surges brought in by the ocean. *This leads to systematically underpredictions in the watershed and a 64% increase in MAE compared to the baseline (Fig. 11)*. … *Note that there is no apparent deterioration of model skill at the near-shore HWMs, because those locations are predominately affected by the oceanic processes.*”

<Reviewer’s Minor Comment #1>

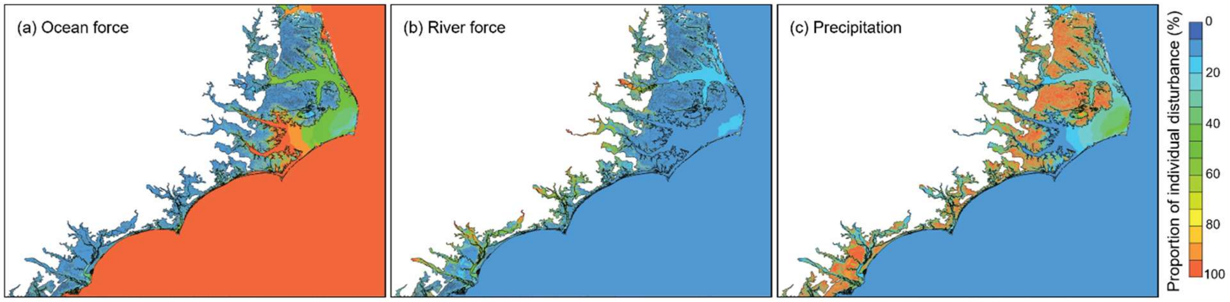
Figure 15. I understand this figure is used to show the impact from different compound flood driver. It is difficult to interpret the results. The caption can include some accompany text on how the figure can be interpreted (what the proportion values imply).

<Response>

Following the reviewer’s suggestion, we will change Fig 15’s caption to:

Fig. 15. Regional map showing the spatially-varying importance of each forcing factor: (a) ocean force (“Ocean” in Table 1); (b) river force (“NWM” in Table 1); and (c) precipitation (“Rain” in Table 1).” The value is the proportion of a factor’s individual “disturbance” (see definition in Section 5) to the sum of the disturbance from all factors. The colors from blue to red represent increasing importance of a factor at a specific location.

We also add texts to the labels inside the figure:



<References>

Feaster, T.D., Weaver, J.C., Gotvald, A.J. and Kolb, K.R., 2018. Preliminary peak stage and streamflow data for selected US Geological Survey streamgaging stations in North and South Carolina for flooding following Hurricane Florence, September 2018 (No. 2018-1172). US Geological Survey.

Huang, W., Ye, F., Zhang, Y.J., Park, K., Du, J., Moghimi, S., Myers, E., Pe'eri, S., Calzada, J.R., Yu, H.C. and Nunez, K., 2021. Compounding factors for extreme flooding around Galveston Bay during Hurricane Harvey. *Ocean Modelling*, *158*, p.101735.

Gochis, D.J., Cosgrove, B., Dugger, A.L., Karsten, L., Sampson, K.M., McCreight, J.L., Flowers, T., Clark, E.P., Vukicevic, T., Salas, F. and FitzGerald, K., 2018, December. Multi-variate evaluation of the NOAA National Water Model. In *AGU Fall Meeting Abstracts* (Vol. 2018, pp. H33G-01).

Ye, F., Zhang, Y.J., Yu, H., Sun, W., Moghimi, S., Myers, E., Nunez, K., Zhang, R., Wang, H.V., Roland, A. and Martins, K., 2020. Simulating storm surge and compound flooding events with a creek-to-ocean model: Importance of baroclinic effects. *Ocean Modelling*, *145*, p.101526.

Zhang, Y.J., Ye, F., Yu, H., Sun, W., Moghimi, S., Myers, E., Nunez, K., Zhang, R., Wang, H., Roland, A. and Du, J., 2020. Simulating compound flooding events in a hurricane. *Ocean Dynamics*, *70*(5), pp.621-640.