Dear Reviewer,

Thanks very much for taking the time to help us improve the manuscript. We have copied your questions/comments below for your reference and respond to each. 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 general comment>

To me the novel aspect of this work is having very much detailed representation of hydrology (via NWM based on WRF), river morphology and coastal ocean dynamics (via a 3D baroclinic model). The manuscript is well written and can be of interest of community, thus I suggest publication in NHESS, however after a major revision. The major issue to me is that, despite the efforts of authors to calibrate/validate their model, on the hydrologic fluvial side the model does not seem well calibrated, and the significant error in NWM prediction (both timing and magnitude of peak) can pose a significant error to the overall estimates. Also, not clear to me how the pluvial flooding processes (from direct rainfall) is treated here. More detailed comments below.

<Response to general comment>

We need to clarify a possible misunderstanding on our model. NWM (based on WRFHydro) is only used as a boundary condition like HYCOM and ERA, which all have uncertainties but are the state-of-the-art. NWM segments inside the model domain are only used to guide mesh generation, because these segments correspond to thalwegs in the DEM.
We are not the developers of NWM, so the quality of NWM prediction is out of our control. It is the best hydrologic products available to us for this regional-scale study; we are fully open to other hydrologic models.

The merit of this work is being an effort toward a fully-coupled compound inland flood and ocean related surge model. Here "fully-coupled" means the model solves its included processes with the same set of governing equations (Santiago-Collazo et al., 2019). Specifically, we include enough watershed region in our model domain so that the interaction between pluvial/fluvial processes and estuarine/oceanic processes can be simulated directly, without the need for "coupling" or "link" among different types of models. In the watershed, the movement of water (flood routing) is purely controlled by physics (governing equation), just like how the estuarine and the ocean circulation are simulated. This poses great challenge on the model's robustness and efficiency.

We probably did not make this clear enough in the first paragraph of "3.2 Coupling with NWM". In the revised manuscript, we will add more details (see responses to Reviewer's Major Comment #1 and Reviewer's Major Comment #2 below) and convey a clearer message on NWM's role in our model.

<Reviewer's Major Comment #1>

Hydrologic processes are not well represented here. River flow hydrograph estimates from NWM are significantly different than the observed flow at USGS gauges (see Figure 6). About 2-3 days of lag in peak flow estimation and up to 100% error in peak flow magnitude can pose a significant error in overall inundation and coastal water level dynamics if propagated through the system. Timing (and magnitude) of peak runoff plays a significant role in extreme water level dynamics in freshwater-influenced coastal systems and getting these hydrologic characteristics right plays a major role in accurate estimation of extent/intensity of compound flooding. This might partially explain the wide range of difference (up to 4 m) between simulated and estimated high water marks (Figure 11), with a considerable number of points having estimation error greater than 1m (Figure 11e).

<Response>

We agree with the reviewer that part of our model errors can be attributed to the errors from NWM. We add the following text after Line 190 of the original manuscript to explain our choice:

"The observation indicates the peak streamflow occurs about 7 days after the landfall, which is the time it takes for the rainfall induced flood to reach the coastal rivers. Note that there is typically a time lag of 1-2 days between the peak flow in NWM and the gaged flow. 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 Major Comment #2>

Could not find any information on how the pluvial processes are handled in the model. Rain-on grid modeling (L174) is a good contribution to the field which I acknowledge, but how the underlying processes and variables that contribute to the runoff generation are accounted for is vague. How infiltration capacity is accounted for? How drainage infrastructure contribution is accounted for (in urban settings)? How the generated runoff is routed between pixels? This is even more important when the results suggest that flooding in a significant portion of coastal land in dominated by "precipitation only"(Figure 15c).

<Response>

We removed the original text on Line 171-173. We state that the neglection of infiltration and drainage is a limitation of the model that warrants further work. The relevant texts in the revised manuscript read:

[Revised Line 106 of the original manuscript] This model solves the physical processes from the watershed to the ocean with the same set of governing equations, qualifying for Santiago-Collazo et al. (2019)'s definition of a fully-coupled compound surge and flood model.

...

[Revised Line 171-174 of the original manuscript] Over the model domain, streamflow injected at the land boundary and precipitation are directly handled by the hydrodynamic core of the ocean model (SCHISM). This fully coupled configuration is rare in the existing compound flooding simulations (Santiago-Collazo et al., 2019). To ensure the accuracy and robustness of SCHISM in simulating hydrological and hydraulic processes, we examined the model's performance in both lab-scale and field-scale tests in a previous study

(Section 2.2 and 2.3 in Zhang et al. (2020)) and applied the model in the Delaware Bay watershed including part of the Delaware River up to 40 meters above the NAVD88 datum with a hydraulic jump (Fig. 14 in Zhang et al. (2020)).”

…

[Line 184] As a model limitation, infiltration is neglected in this work. In the case of Hurricane Florence induced flooding, we expect the effect of infiltration to be minor. According to the NOAA’s weather map (https://www.wpc.ncep.noaa.gov/dailywxmap/), there was continuous rainfall along the US east coast from Sep 11, 2018 to the date of Florence’s landfall (Sep 14, 2018), so the infiltration capacity of the soil was already reduced. Moreover, “wet” storms like Hurricane Florence (2018) and Hurricane Harvey (2017) tend to dump large amount of rain fall at a location for days because of the slow movement of the storm, so most of the rainfall is on saturated soil. 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 likely responsible for most of the large errors in Fig. 11c) for operational use.

…

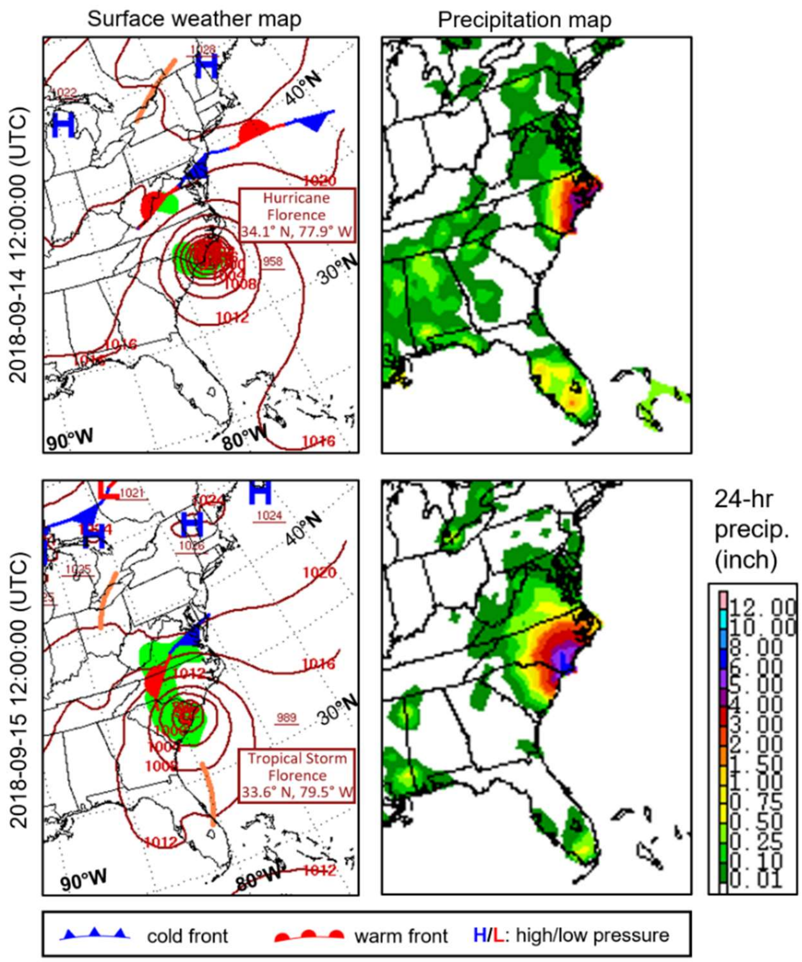
[Line 300-302] The only remaining large error in the “baseline” occurs in an urban area away from the river, likely due to the building or drainage effects that have not been incorporated in the model (Fig. 13d).

< Reviewer’s Minor Comment #1>

Figure 1: red text (in the left panels) is not readable in print. I’d also add some space between rows in the legend.

<Response>

Revised as suggested:

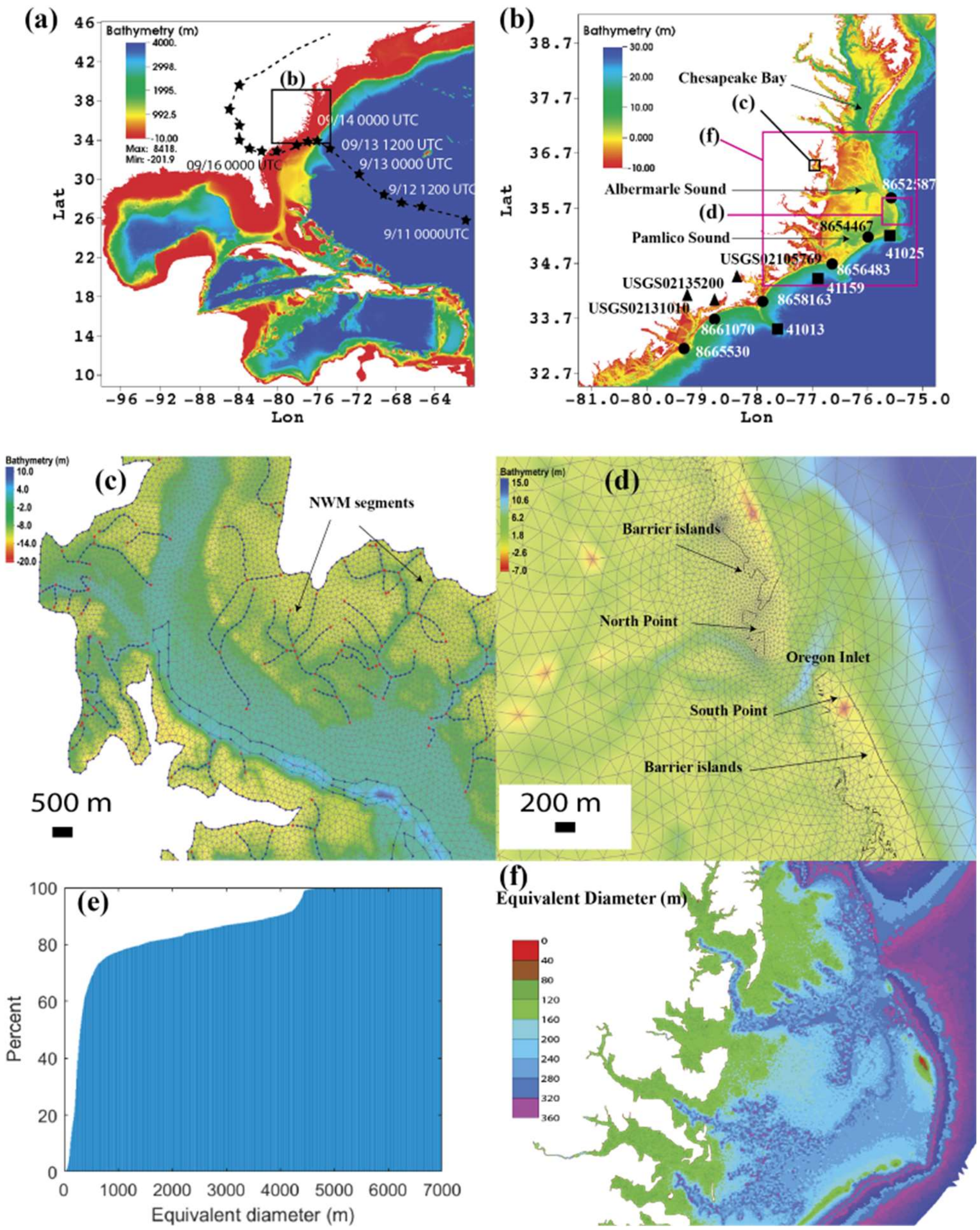


<Reviewer's Minor Comment #2>

Figure 3: Black text is really hard to read on the dark blue background.

<Response>

Revised as suggested:

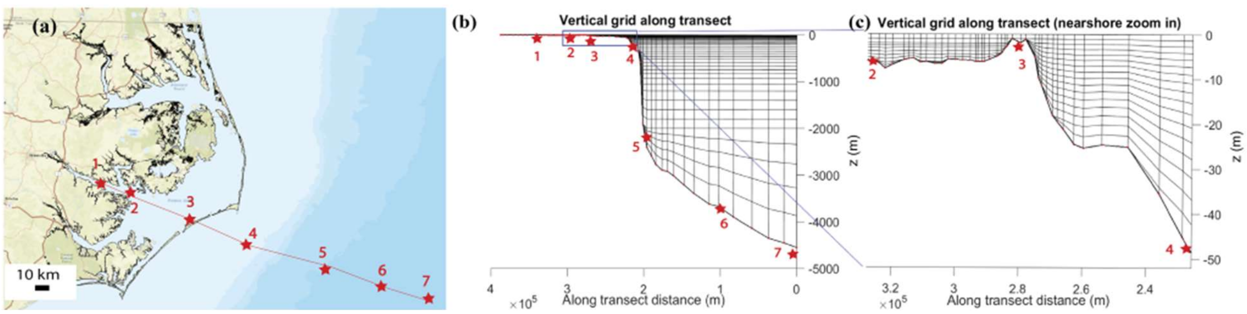


<Reviewer's Minor Comment #3>

Figure 4 is confusing. Hurricane track is numbered right-to-left (Panel a) and vertical grid points are numbered Left-to-Right (Panels b and c)

<Response>

Revised as suggested:



<Reviewer’s Minor Comment #4>

L110: what "LSC" stands for?

<Response>

We change this sentence to:

“… that combines a hybrid triangular-quadrangular unstructured grid in the horizontal dimension and localized sigma coordinates with shaved cells (dubbed as LSC2; Zhang et al., 2015) in the vertical dimension.”

<Reviewer’s Minor Comment #5>

L120: "10 m above sea level", you mean mean sea level? if yes, is it regional or global. Please be more specific about this.

<Response>

Changed to “10 m above the NAVD88 datum”.

<Reviewer’s Minor Comment #6>

L133: what "CFL" stands for?

<Response>

Changed to “Courant–Friedrichs–Lewy (CFL) condition”

<Reviewer’s Minor Comment #7>

L143: Explain what is HYCOM first time use it in the manuscript.

<Response>

In the revised manuscript, we use the term “HYbrid Coordinate Ocean Model (HYCOM; https://www.hycom.org/)” upon the first occurrence.

<Reviewer’s Minor Comment #8>

L174: Simply said rainfall is routed. Please, elaborate more on this, how exactly routed? How infiltration capacity incorporated?

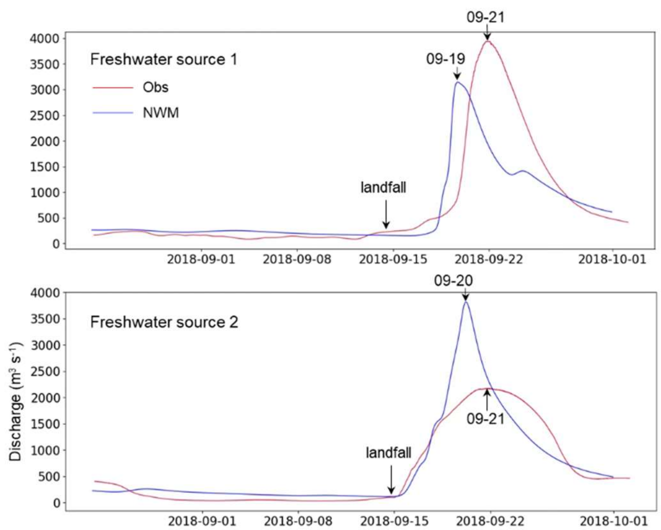
This is answered in the response to <Reviewer’s Major Comment #2>.

<Reviewer’s Minor Comment #9>

L189: "4-6 days" in observed or modeled events? The modeled and simulated peaks seem to be few days apart them-selves.

<Response>

We clearly labeled the date of the peaks in Fig. 6:



We change the sentence to:

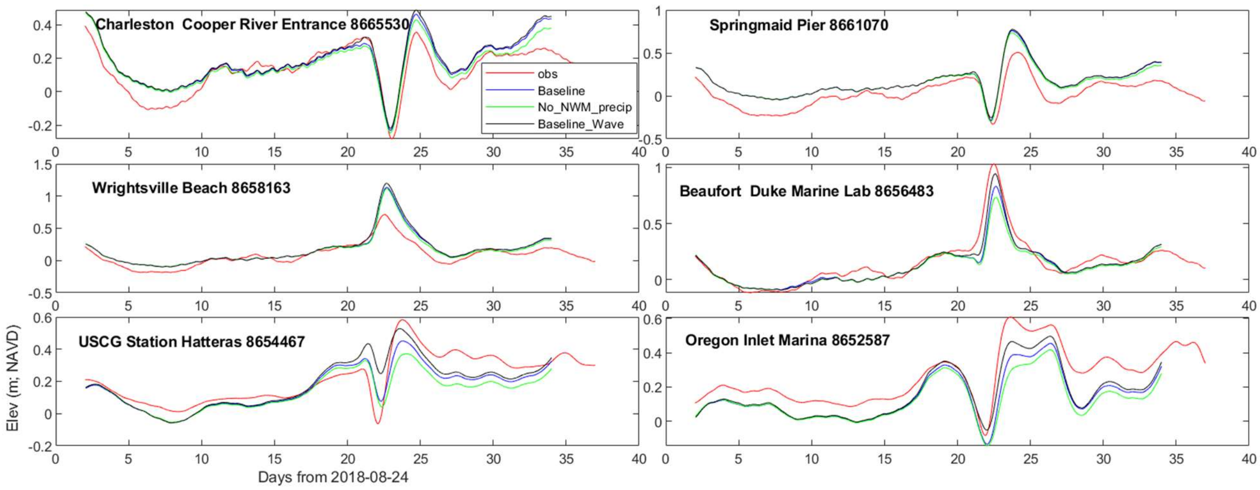
“The observation indicates the peak streamflow occurs about 7 days after the landfall, which is the time it takes for the rainfall induced flood to reach the coastal rivers. Note that there is typically a time lag of 1-2 days between the peak flow in NWM and the gaged flow. The forcing errors in the magnitude and timing of NWM’s peak flow should explain part of the model errors especially in the watershed.”

<Reviewer’s Minor Comment #10>

Figure 9: Better to compare detided signal here. These gauges are extremely tidal influenced and estimation skills of the model for stochastic processes is diluted by highly predictable astronomic tides.

<Response>

We agree that de-tided signal will give a clearer picture of stochastic processes. We add plots on sub-tidal signals (see the figure below) to Fig. 9.



We change the caption of Fig. 9 to: “Fig. 9: Comparison of elevation at 6 NOAA gauges: (a) total elevation; (b) subtidal elevation. …”

We also add more discussion in the text:

After Line 242: “…, as seen from the total elevation (Fig. 9a) and the sub-tidal signals (Fig. 9b). The latter applies a low-pass Butterworth filter (Butterworth, 1930) only preserving longer-period (longer than 2 days) components.”

After Line 251: “The mechanism causing the water level set-downs at the two South Carolina stations is similar to that causing the set-downs behind the barrier islands in North Carolina. The two South Carolina stations (Charleston and Springmaid) are located to the south of the landfall site and the wind direction is from the land to the ocean, pushing water away from shore.”

Line 253: “The averaged MAE for the subtidal comparison is 8.6 cm and averaged correlation coefficient is 0.92.”

<Reviewer’s Minor Comment #11>

L275 & L284: describing errors "quite satisfactory" and "reasonable agreement", when mean absolute error in estimated high water levels is 0.73 m and with nearly 20% of points having error >1m, is a subjective description. I’d avoid such statements and simply let the readers decide whether if such estimation errors are reasonable according to their required accuracy and project objectives.

<Response>

We remove all subjective claims such as "quite satisfactory" and "reasonable agreement" in the revised manuscript.

<References>

Butterworth, S. (1930), On the theory of filter amplifiers, Wireless Eng., 7, 536–541.

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).

Santiago-Collazo, F.L., Bilskie, M.V., Hagen, S.C.: A comprehensive review of compound inundation models in low-gradient coastal watersheds. Environ. Model. Software 119:166–181. https://doi.org/10.1016/j.envsoft.2019.06.002, 2019.

Zhang, Y.J., Ye, F., Stanev, E.V. and Grashorn, S., 2016. Seamless cross-scale modeling with SCHISM. Ocean Modelling, 102, pp.64-81.

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.