Summary

This article presents a compound flood assessment of Hurricane Irene in the Delaware River Basin. The authors used a new two-way/tightly-coupled approach to quantify the uncertainty of different aspects without the modeling approach, such as the environmental forcing and coupling technique.

General Comments

First, the authors need to highlight better the novelty of the paper since it is current version is difficult to identify clearly. For example, they say that their main goal is to assess the uncertainties in the compound flood model, but this has been done before by Muñoz et al. (2024). However, the authors mention this gap (L69-72) but fail to mention the work by Muñoz despite citing them for another reason in the discussion. Another example is their claim that the compound flood needs to be redefined to include AMC and not just TC impacts. However, Bilskie et al. (2021) and Santiago-Collazo et al. (2023) simulated CF from a TC but were preceded by an antecedent rainfall event, which affects the AMC. My point here is that the manuscript needs to go into an intensive literature review process so it can better define the knowledge gaps and cite the appropriate references. In my judgment, presenting the manuscript novelty as an uncertainty assessment of this specific model they used is not enough to impact the science community, as we expect with any manuscript in top-tier journals like this one.

Second, the authors need to improve their introduction section so it can be more comprehensive than its current state. For example, one of the main points of discussion in the manuscript is the coupling technique that their model uses, such as the two-way/tightly coupling approach, but fails to explain to the reader (at least briefly) how this technique works and compares it with others. At a minimum, the authors should briefly define and explain this and cite a reference such as Santiago-Collazo et al. (2019). Similarly, the paragraph (L49-58) that describes the E3Sm model should be moved to methods and include more details of the model itself. Without this, the authors are expecting the reader to go into another reference to read basic details and model configuration, which is not appropriate. At a minimum, they should talk/show their numerical modeling domain, validation results and simulation set, and model assumptions. The authors only cite a paper from Feng et al. (2024) to point the reader to all the important details about the model for this study, which it should not. Furthermore, the author makes various unsupported claims that are not true. For example, (L90) said that model coupling uncertainties have not been studied before (which is not true; see Muñoz et al., 2024) since this coupling technique has been recently developed. However, the authors fail to comment on the many available models that can do this, such as SFICNS, SCHISIM, and the work done by George XU from LSU.

Third, the study limitations should be shown before the results section so the reader is aware of all of them before "believing" the results. For example, I was wondering about the lack of the coastal flood model (deep ocean, for example) since it was not explained in the methods, but then in the limitation section, I learned that the model configuration used in this study does not have one. It is not clear how they model the coastal flooding using their proposed framework since they do not have the coastal flood model (L443) nor use boundary conditions. Thus, I am not even sure if they model coastal processes. I also understand that the authors had a different purpose than to provide a high-resolution model, but the coarse resolution inland is producing an overestimation of the fluvial flood inland. However, the authors said that it is prudent to overestimate the potential flooding from a flood hazard risk assessment perspective (L244-245). As a civil engineer myself, I am in complete disagreement with such a statement. It is not the same to overestimate the flood depth by a couple of inches, but to estimate areas being severely flooded, where there should not be any flood, is just a "bad model." Thus, the author's use of this type of statement represents a lousy justification for their poor validation/calibration processes. I strongly recommend going back to the MOSART model component and calibrating it to a better fit before moving forward with the manuscript. This has a big implication in their study findings since if the simulated storm was a pluvial/fluvial-dominated CF event, then the results would be overestimated greatly, including the uncertainty, since Irene was a coastal flood-dominated event.

Specific Comments

- Figure 1: remove the adjectives (e.g., large, heavy, high) for the flood drivers labels. This will make the figure more general and applicable to a broader audience.
- Figure 2: why are the MOSART network lines so different from the real stream network? The authors say that is a limitation but do not offer details on why.
- L227: The word "effectively" should not be used since the authors do not show at least a time series hydrograph, which they are multiple along the river, of its results and the observed to verify how effective the simulation was.
- L 231: does the stream network used in MOSART include the real bathymetry? If so, the data was available?
- L283: there was no discussion/mention of the temporal scale of any model in the manuscript.