Response to Reviewers

Title: Disentangling Atmospheric, Hydrological, and Coupling Uncertainties in Compound Flood Modeling within a Coupled Earth System Model

Author Response 2nd revision

Editor comments:

My decision is that the paper can be published after minor revisions. This is based on multiple rounds of review and reports by three reviewers. Currently, one reviewer is satisfied with the revisions, but the other reviewer (who did not review the first version of the manuscript) has a few major comments that need to be addressed. The authors should address all comments from Reviewer #2 before resubmitting the manuscript. Specifically, they should expand and improve the literature review related to uncertainty analysis methods. Second, they should improve the presentation of the results by comparing the relative uncertainties from different sources.

Author Response:

We would like to sincerely thank the editor and reviewers for their valuable comments and recommendations. We have carefully addressed the reviewer's suggestions as follows.

Reviewer 2
Reviewer Comments:

This manuscript applies an earth system model, E3SM, to study the coupling and meteorological uncertainties associated with compound flood modeling. The model is applied to simulate inundation in the Delaware River Basin and Estuary during Hurricane Irene.

Author Response:

We appreciate the reviewer for the critical assessment of our work. In the following we address your comments point by point. Our responses and all changes in the revised paper are marked in blue.

Major comments:

R2C1:

Given that the focus of this paper is on assessing sources of uncertainty in the coupled compound flood framework, the literature review should be expanded to discuss the approaches used for uncertainty analysis in previous studies and their pros/cons. This would help to justify the choice of SEM and ANN for this study.

Author Response:

Thank you for the insightful comment. We agree that a more comprehensive review of uncertainty quantification methods would benefit the manuscript and help clarify the rationale behind our approach.

We would also like to emphasize that our focus is on uncertainty quantification within a fully coupled Earth system modeling framework, which remains an underexplored area in compound flood research. To assess coupling uncertainty, we conducted controlled experiments (as shown in Table 1), comparing simulated discharge and inundation under different coupling configurations, all driven by the same atmospheric ensemble. For cascading meteorological uncertainty, we applied two complementary methods: (1) ensemble spread metrics (NMAD and CV), and (2) Structural Equation Modeling (SEM), which reveals interdependencies among atmospheric, hydrological, and oceanic variables. SEM is particularly well-suited for identifying cause-effect relationships in complex systems with multiple interacting drivers. Due to the high computational cost of fully coupled simulations with the global ocean model activated, we did not apply ANN-based methods to the meteorological uncertainty analysis. However, we were able to expand the ensemble size for hydrological uncertainty analysis, which enabled us to use ANN to examine the relative influence of hydrological drivers on discharge and inundation. This follows a similar approach to Muñoz et al. (2024), who used random forests to identify dominant drivers, although our ANN model allows for multivariate predictions (discharge and inundation).

In the revised manuscript, we have expanded the introduction (L118~L133) to include a brief review of existing methods for uncertainty analysis, such as ensemble modeling, statistical modeling, machine learning approaches and SEM, along with their advantages and limitations.

"The above-mentioned uncertainties are complicated but must be carefully evaluated for ESMs as they will be more frequently applied for CF simulations in the context of climate change. A variety of approaches have been adopted for understanding the uncertainties of CF modeling. These approaches offer trade-offs between computational cost, physical interpretability, and the ability to disentangle complex drivers. Ensemble-based methods remain a primary strategy for characterizing the cascading uncertainty from the forcing data (Hamill et al., 2011; Hou et al., 2017; Villarini et al., 2019). Multiple realizations with perturbed initial conditions and/or model physics represent a range of scenarios that evolve differently based on the dynamics of the models (Blanton et al., 2020; Nederhoff et al., 2024; Saleh et al., 2017; Wang et al., 2024). Probabilistic frameworks, such as Bayesian inference (Beven and Binley, 1992), provide more robust treatment of parameters and model uncertainties (Naseri & Hummel, 2022), but often rely on strong assumptions and intensive sampling. Machine learning techniques have been increasingly applied to flood modeling (Hu et al., 2019) and are effective at capturing nonlinear relationships of CF drivers (Muñoz et al., 2024), though they require large training datasets and may sacrifice physical interpretability (Shen et al., 2023). Structural equation modeling (SEM; Wright, 1921) has also been adopted to disentangle complex, interacting processes (Du et al., 2015; Santoro et al., 2023). SEM offers a balance between statistical rigor and interpretability in multi-driver systems without a significantly amount of data. Despite these advances, uncertainty quantification within fully coupled ESM frameworks remains relatively underexplored due to high computational demands and limited methodological integration across domains."

R2C2:

The manuscript could be improved by better highlighting and comparing the relative uncertainties from different sources. Different methods were used to quantify the coupling uncertainty, the meteorological uncertainty, and the uncertainty from hydrological driver propagation, and I don't see a definitive comparison between or synthesis of the results. Line 438-442 states that "uncertainty from the atmosphere simulations is comparable to that of two-way river-ocean coupling… but is considerably smaller than the uncertainty of two-way river-ocean coupling if the MPAS-O modeled inundation is excluded." However, it is unclear to me where the "uncertainty from the atmosphere simulations" is clearly reported and how the magnitudes were compared.

Author Response:

We appreciate the reviewer's suggestion and acknowledge that the comparison between different sources of uncertainty, particularly between atmospheric forcing and model coupling, was not clearly presented in the original manuscript. While the cascading meteorological uncertainty and the hydrological uncertainty target different aspects of the system and are therefore not directly comparable (the former concerns the propagation of uncertainty through interconnected system drivers, while the latter focuses on the influence from distinct hydrological drivers), we agree that a more explicit comparison between the atmospheric forcing uncertainty and coupling uncertainty is necessary.

To clarify, the standard deviation values reported in lines 438–442 of the previous manuscript (e.g., σ = 0.015 for HR, 0.014 for FR, etc., in Experiment 3) represent the variability in flood metrics across all ensemble members of atmospheric forcing, thereby quantifying the uncertainty introduced by atmospheric variability. In contrast, coupling uncertainty can be assessed by comparing the flood metrics across Experiments 1, 2, and 3 using the mean values of the ensemble simulation.

In response to this comment, we have added a new Supplementary Table S1 that directly compares these two sources of uncertainty. This table reports: (1) the magnitude of differences in flood metrics between Experiments 1, 2, and 3 (reflecting coupling uncertainty) and (2) the standard deviation of flood metrics across all ensemble members (reflecting atmospheric uncertainty). Additionally, we have revised the discussion section in the main text to elaborate on this comparison (L461-L469).

"Additionally, we directly compared the uncertainty introduced by model coupling with that from atmospheric forcing (Table S1). The atmospheric uncertainty, quantified as the spread of flood metrics across ensemble members in Experiment 3, is comparable to the uncertainty introduced by two-way land-river and river–ocean coupling when only riverine inundation is considered (Exp3 – Exp 1 and Exp 3 – Exp 1). However, when coastal inundation is included, the coupling-induced uncertainty becomes substantially larger across all metrics. The discharge also shows more variability in the river-ocean coupling experiment (Exp 3) than in the atmospheric ensemble, indicating the significant influence of the two-way river-ocean coupling configuration on flow dynamics. These results highlight the need to consider both meteorological variability and structural model uncertainty when evaluating flood risk in the coupled ESM framework."

Table S1 Comparison of uncertainty in flood and discharge metrics due to model coupling and atmospheric forcing. Coupling-induced uncertainty is represented by the difference in the metrics between Experiments 2 or 3 and Experiment 1, averaged across ensemble simulations. Atmospheric uncertainty is quantified as the standard deviation of metrics across ensemble members in Experiment 3.

<table>
<tr><th>Uncertainty source</th><th>HR</th><th>FR</th><th>SI</th><th>FA [$\times 10^7 m^2$]</th><th>Q [m^3/s]</th></tr>
<tr><td>Two-way land-river coupling (riverine flooding) (Exp 2 – Exp 1)</td><td>-0.076</td><td>0.007</td><td>-0.020</td><td>-6.186</td><td>-124.342</td></tr>
<tr><td>Two-way river-ocean coupling (riverine flooding) (Exp 3 – Exp 1)</td><td>0.015</td><td>-0.004</td><td>0.005</td><td>2.480</td><td rowspan="2">1641.695</td></tr>
<tr><td>Two-way river-ocean coupling (riverine&coastal flooding) (Exp 3 – Exp 1)</td><td>0.346</td><td>-0.098</td><td>0.129</td><td>54.185</td></tr>
<tr><td>Atmospheric forcing
(standard deviation in Exp 3)</td><td>0.015</td><td>0.014</td><td>0.01</td><td>8.500</td><td>326.094</td></tr>
</table>

Other Comments:

R2C3:

Line 179-180: Is the elevation adjustment of 1 meter spatially uniform? Is this justified by the data?

Author Response:

We apologize for the confusion. To clarify, the 1-meter threshold for determining inundation in MPAS-O cells is not a spatially uniform adjustment of elevation data itself, nor does it reflect a deficiency in the underlying GEBCO bathymetry data. Instead, it is a practical criterion introduced during post-processing to reduce minor inundation signals that can appear when aggregating from the 250-m MPAS-O mesh to the coarser MOSART grid. Therefore, this 1-meter threshold is justified not by the original elevation data, but rather by the resolution gap and the need to maintain consistency in identifying meaningful inundation extents when aggregating results across different spatial scales. We will add a brief clarification in the manuscript to reflect this point explicitly (L197~L200).

"Within each MOSART grid cell, the inundation fraction is determined by the percentage of MPAS-O cells with a simulated water depth over 1 m. This threshold does not imply a spatially uniform adjustment of the GEBCO bathymetry data used by MPAS-O. Instead, it serves as a practical criterion to mitigate biases arising from upscaling inundation extents from the higher-resolution MPAS-O mesh to the coarser MOSART grid."

R2C4:

Line 198-199: Please clarify here that Experiment 1 used one-way coupling while Experiment 2 used two-way coupling.

Author Response:

Thanks. This has been clarified in the revision (L216~L217):

"The first two experiments implement one-way and two-way coupled ELM and MOSART, respectively, while the third experiment interactively couples MPAS-O with MOSART."

R2C5:

Line 248-249: By "roughly even distribution", do the authors mean that the discharge and precipitation values are sampled at even intervals across the range of values modeled, or that the values are applied over the study domain in an event spatial pattern? Please clarify.

Author Response:

Sorry for the confusion. We confirm that by "roughly even distribution," we mean that the selected ensemble members were chosen such that their simulated discharge and precipitation span the full range of modeled values, with values approximately evenly spaced across that range. We have revised the manuscript to clarify this point (L266~L267).

"The original EAM ensemble is expanded by first selecting 5 ensemble members whose river discharge and precipitation values span the full range of the ensemble and are approximately evenly spaced across that range during Hurricane Irene"

R2C6:

Line 264-265: It was not clear what data was used to train/test the ANN. Are the 625 ensemble simulations from the coupled model used? What was the split for testing and training?

Author Response:

Thanks. The input and output variables shown in Figure 3 were derived from the full set of 625 ensemble simulations. These simulations provided the dataset used for both training and testing the two-stage ANN. As now clarified in Appendix B, the dataset was randomly split into 80% for training and 20% for testing.

Revised Main Text: "To quantify the relative importance of each hydrological driver of CF, we employed a two-stage Artificial Neural Network (ANN) approach (Fig. 3), trained and tested using data from all 625 ensemble simulations."

Revised Appendix B: "Before training, the data were randomly split into training and testing datasets, with 80% used for training and 20% for testing,"

R2C7:

Line 290-294 and Fig 4: In panel (a), which only uses MOSART, it is not clear to me why the cells immediately adjacent to the river and bay are not inundated but the adjacent inland cells are. It seems that if the flood is propagating from the river into the floodplain, the shoreline cells should also be inundated, with or without tides and surge. Or is the flood propagation occurring in a different way? Since flooding in these cells is the main source of the stated improvement in the model when MPAS-O is incorporated, it is important to clarify the flood propagation process in these areas.

Author Response:

We appreciate this comment. To clarify, the ELM–MOSART configuration in Figure 4a does simulate riverine inundation along the mainstem of the Delaware River and some upstream tributaries (as indicated by the red areas aligned with the river network). These inundated areas are generated by excessive precipitation and routed through the river network using MOSART's macroscale inundation scheme, which represents subgrid-scale flooding within individual grid cells. As noted in the Methods section (Section 2.1), this scheme does not explicitly simulate lateral flood propagation between cells, which can lead to some limitations.

However, the lack of inundation near the coastline in Figure 4a is not due to lateral propagation issues, but rather the absence of coastal processes (specifically tide and surge) that are necessary to raise water levels enough for those low-lying coastal cells to become inundated. When MPAS-O is included (Figure 4b), its dynamic two-dimensional wetting and drying scheme enables storm surge and high coastal water levels to intrude into these shoreline areas, triggering inundation that the ELM–MOSART configuration alone cannot represent. We have clarified this point in the revised manuscript (L310~L316).

"More importantly, although ELM–MOSART simulates extensive riverine inundation along the Delaware River mainstem and tributaries through precipitation-induced runoff (Fig. 4a), it does not capture inundation in low-lying shoreline areas near the coastline. This is because tide and storm surge that elevated local water levels sufficiently to exceed the inundation threshold in coastal cells are not included in this configuration. MOSART's macroscale inundation scheme does not simulate lateral water propagation across grid cells, and coastal inundation requires dynamic oceanic forcing. By integrating MPAS-O (Fig. 4b), which includes two-dimensional wetting and drying, the model captures these near-coastline inundations more accurately."

R2C8:

Line 330-332 and Fig 5: Is the discharge reported as an absolute value? Or is the graph showing the discharge after the coastal water levels have receded and the river begins to flow downstream again? It would be helpful to see the time series of streamflow to understand the temporal effects.

Author Response:

Thanks for the comment. We believe the comment refers to Figure 6, which shows the peak discharge values along the Delaware River mainstem during Hurricane Irene, as noted in the figure caption. The intention of this figure is not to illustrate temporal dynamics, but rather to compare the maximum discharge values among the three model configurations. The elevated peak discharge observed near the river outlet in Experiment 3 is primarily due to the backwater effect induced by high coastal water levels during the storm. While the time series of streamflow and water level are not shown in this manuscript, a full temporal analysis is available in Figure 5 and 7 of Feng et al. (2024). We have clarified this point in the revised text (L353~L355).

"The elevated water levels due to tide and storm surge force an upstream propagation of ocean water into the river channel, resulting in a local increase in peak river discharge and riverine inundation near the outlet, where the highest coastal water levels during Irene lead to elevated maximum discharge values along the lower Delaware River."

R2C9:

Line 442-444: How do the sigma values from Experiment 3 show "the critical need to account for the meteorological uncertainty and is cascading effects through the coupled system"? Please provide more explanation here.

Author Response:

Thanks. We have revised our discussion. Please see our response to R2C2.

R2C10:

Line 473: Exposure implies that there are assets in the flood zone, which I don't think was examined here. "Hazard" is a better word choice.

Author Response:

Thanks. We acknowledge that "risk" is broadly defined, encompassing flood hazard, exposure, and vulnerability (Kron, 2005). Specifically, flood hazard and exposure risks represent the frequency or intensity of flooding events and the extent of human exposure to these events, respectively, as was also discussed in Feng et al. (2023). While we did not explicitly assess the distribution of assets or population, our use of the term "exposure" was intended in a broader sense, referring to the spatial extent of inundation, which implies increased potential for exposure under more severe flood scenarios. To avoid confusion, we have revised the manuscript to replace "flood exposure" with "flood extent" here.

R2C11:

Line 478-479: Why is this the case if Irene was associated with the 75th percentile AMC scenario, as mentioned earlier (Line 252)?

Author Response:

Thanks for the comment. Here we intend to describe that although Hurricane Irene coincided with a relatively wet antecedent soil moisture condition (approximately the 75th percentile), the modeled inundation area still aligns more closely with the best-case scenario, as well as simulations with lower AMCs. This outcome is because reducing soil moisture below Irene's level has a limited effect on peak discharge and inundation, whereas increasing AMC beyond Irene's level (i.e., AMC100) leads to disproportionately larger flood impacts, as shown in Figure 11. This asymmetry is due to the fact that, despite Irene's wet initial conditions, the soil still retained some infiltration capacity at the onset of the storm. In contrast, scenarios with saturated soils (AMC100) overwhelm that capacity, resulting in significantly enhanced runoff and flood extent. We have revised the manuscript to clarify this important point, which we consider a key finding of the study (L503~L509).

"Interestingly, the modeled inundation area for Irene closely aligns with the best-case scenario (Fig. 11b and 11c), despite the fact that Irene occurred under a relatively wet AMC (i.e., 75th percentile AMC). This reflects an asymmetric hydrological response: while drier AMC scenarios show only modest reductions in flood extent, the scenario with saturated soils (AMC100) leads to a disproportionately large increase in peak discharge and inundation. This is likely because, despite the wet soils prior to Irene, there remained sufficient infiltration capacity at the storm's onset. In contrast, further increases in AMC rapidly exceed that capacity, exacerbating surface runoff and flood hazards. This nonlinear amplification highlights the critical role of AMC in modulating compound flood severity."

R2C12:

Line 489: Did the authors confirm that exposure increased? If not, "increasing exposure risks to coastal residents" should be changed to "increasing flood hazards."

Author Response:

We rephrased "increasing exposure risks to coastal residents" to "increasing the extent of flooding, raising potential risks to coastal residents."

R2C13:

Fig 11: In panel (c), what does the Irene scenario (shown in yellow) represent? I thought the purple outline was showing the observed flooding during Irene.

Author Response:

We apologize for the confusion. The purple outline in Figure 11c represents the observed flood extent derived from satellite data (Tellman et al., 2021), consistent to that in Figure 4. The "Irene" scenario shown in yellow corresponds to the best model simulation that represents the Hurricane Irene event at the AMC75 condition, which is also indicated by the dashed lines in panels (a) and (b). We have updated the figure caption to clarify this distinction.

"(c) Spatial map of plausible inundation extents in the DRB, showing the minimum (best-case scenario), Irene (AMC75), and maximum (worst-case scenario) simulated inundation. The purple outline represents the observed flood extent from satellite data, consistent with Figure 4."

R2C14:

Line 496-498: The runtime of the various model configurations was never mentioned, so this statement is unsupported.

Author Response:

We thank the reviewer for pointing this out and have added information on the runtime performance of the simulations to support this statement. Specifically, the global ELM–MOSART simulation required less than 10 minutes using 400 CPUs, while the fully coupled ELM–MOSART–MPAS-O simulation took approximately 5 hours. These runtimes are relatively efficient in ESMs given the global scope and high-resolution coastal refinement in E3SM, and they demonstrate the model's suitability for ensemble-based uncertainty quantification. The relevant sentence has also been revised for clarity.

In the revised Section 2.1, we added the runtime information: "The global ELM–MOSART simulations are computationally efficient, requiring less than 10 minutes using 400 CPUs, while the fully coupled ELM–MOSART–MPAS-O simulations take approximately 5 hours."

In the discussion (L527~L530), the original statement has been rephrased to "Given the global domain and component complexity, the relatively efficient runtime makes the framework suitable for disentangling interconnected drivers of complex physical processes and their cascading effects through ensemble-based analyses."

R2C15:

Line 542-544: “Significantly enhances the accuracy” compared to what baseline? The actual observed flooding was not well predicted by any of the models considered.

Author Response:

Our intention was to convey that the fully coupled E3SM simulation—including the MPAS-O ocean component—demonstrated improved performance in representing compound flood processes compared to the ELM–MOSART configuration alone. However, we agree that the purpose of this framework is not yet to achieve high-accuracy flood prediction relative to fine-resolution regional models. Rather, as discussed in the introduction, this integrated ESM-based framework serves as an intermediate step that enables the identification and analysis of cascading uncertainties across Earth system components. To avoid overstating model performance, we have revised the conclusion text accordingly (L574~L576).

“Our research demonstrates that an integrated atmosphere, land, river and ocean system improves the representation of multivariate flooding processes relative to partially coupled configurations, while enabling the analysis of cascading uncertainties through the multi-component Earth system modeling framework.”

Below is the list of newly added references.

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrological processes, 6, 279-298, 10.1002/hyp.3360060305, 1992.

Du, X., García-Berthou, E., Wang, Q., Liu, J., Zhang, T., and Li, Z.: Analyzing the importance of top-down and bottom-up controls in food webs of Chinese lakes through structural equation modeling, Aquatic Ecology, 49, 199-210, 10.1007/s10452-015-9518-3, 2015.

Hou, X., Hodges, B. R., Feng, D., and Liu, Q.: Uncertainty quantification and reliability assessment in operational oil spill forecast modeling system, Marine pollution bulletin, 116, 420-433, 10.1016/j.marpolbul.2017.01.038, 2017.

Hu, R., Fang, F., Pain, C., and Navon, I.: Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method, Journal of Hydrology, 575, 911-920, 10.1016/j.jhydrol.2019.05.087, 2019.

Naseri, K. and Hummel, M. A.: A Bayesian copula-based nonstationary framework for compound flood risk assessment along US coastlines, Journal of Hydrology, 610, 128005, 10.1016/j.jhydrol.2022.128005, 2022.

Nederhoff, K., van Ormondt, M., Veeramony, J., van Dongeren, A., Antolínez, J. A. Á., Leijnse, T., and Roelvink, D.: Accounting for uncertainties in forecasting tropical-cyclone-induced compound flooding, Geoscientific Model Development, 17, 1789-1811, 10.5194/gmd-17-1789-2024, 2024.

Saleh, F., Ramaswamy, V., Wang, Y., Georgas, N., Blumberg, A., and Pullen, J.: A multi-scale ensemble-based framework for forecasting compound coastal-riverine flooding: The Hackensack-Passaic watershed and Newark Bay, Advances in Water Resources, 110, 371-386, 10.1016/j.advwatres.2017.10.026, 2017.

Santoro, S., Lovreglio, R., Totaro, V., Camarda, D., Iacobellis, V., and Fratino, U.: Community risk perception for flood management: A structural equation modelling approach, International journal of disaster risk reduction, 97, 104012, 10.1016/j.ijdrr.2023.104012, 2023.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., and Li, L.: Differentiable modelling to unify machine learning and physical models for geosciences, Nature Reviews Earth & Environment, 4, 552-567, 10.1038/s43017-023-00450-9, 2023.

Wang, Z., Leung, M., Mukhopadhyay, S., Sunkara, S. V., Steinschneider, S., Herman, J., Abellera, M., Kucharski, J., Nederhoff, K., and Ruggiero, P.: A hybrid statistical–dynamical framework for compound coastal flooding analysis, Environmental Research Letters, 20, 014005, 10.1088/1748-9326/ad96ce, 2024.