Summary

This study compares the flood hazard estimates generated using an event-based approach to those generated using a full suite of synthetic events, referred to herein as a response-based approach. Using Glouchester, NJ as a case study, the researchers estimate the compound hazard resulting from the joint probability distribution of 18-hr total precipitation and non-tidal residuals. From this distribution, they generate 5,000 synthetic storm events, each of which has a statistically derived spatiotemporally varying rainfall field and storm tide hydrograph (non-tidal residual + tide). From these, the authors sample 25 synthetic storm events with an Annual Expected Probability (AEP) of 1% (i.e., 'event-based') sampling from the isoline (representing maximum likelihood) and compare the estimated 1% AEP inundation map generated using the entire suite of synthetic storm events (i.e., 'response-based'). They find that the response-based approach generally generates higher hazard estimates, suggesting that selecting the 1% events based on boundary conditions alone does not generate the 1% hazard.

Overall, I found this study very interesting to read and many of my comments and questions come from a place of genuine interest in the modelers' choices. My major review comments are as follows:

First, while the experimental design itself is not particularly novel, the combination of statistically-derived synthetic storm events that can represent compound drivers (both non-tidal residual and precipitation) as inputs to a process-based model takes many of the concepts that have been discussed within the compound hazard modeling community a step further than previous studies, and thus I find that it is worthy of publication.

Second, while this work is new in the context of compound flooding – a topic of considerable interest – the authors rely heavily on the compound flood literature. In doing so, they overlook a larger body of research regarding stochastic flood hazard simulation for rainfall-induced floods. For example, there have been several other studies that draw similar conclusions about the use of design storms vs stochastically generated (see, e.g., Perez et al. 2024, several papers by Daniel B. Wright). The paper could be further improved by placing itself in this body of work, perhaps as part of both the introduction and discussion.

Third, I commend the authors on their work to validate the hydrodynamic model in a data-scarce environment, however, I would like to see more discussion of the potential limitations of the synthetically generated boundary conditions. The proposed approach also introduces considerable uncertainties into the hazard estimates that are difficult to disentangle, including questions about how the authors are accounting for different flood types (e.g., those driven by high river events vs coastal storms) and whether the statistically-derived framework can adequately account for these.

The comments below primarily ask for additional clarification or suggest that the authors contextualize some of their findings.

Comments

Line 34 the authors differentiate between event-based and response-based approaches. Is 'response-based' a commonly used term in flood hazard literature? Is there a citation that can be used to support this terminology? I have often heard them differentiated as 'design-storm' vs 'probabilistic' or 'stochastic' methods. Is there a citation that can be included to help support this choice of terminology?

In this same paragraph there is some discussion of the FEMA regulatory maps and process used for flood hazard delineation in the U.S. It is perhaps necessary here to differentiate from the approaches used for inland and coastal hazard. While they both rely on event-based approaches, the inland hazards are derived from design storms and assume a one-to-one relationship between the AEP of the precipitation event (spatially uniform, but varied in time according to a characteristic distribution) and the resulting flood flow which is then translated into hazard. In line 43, you state that temporal variability for inland flooding is neglected, but this is not true in the case of rainfall. It is true, however, in the case of flood mapping, where the peak flow is used to estimate the extent of the floodplain. There is some nuance here that gets lost in the way it is written now, and it might be worth revisiting this section.

With respect to coastal flood hazard estimation, I'm not sure I agree with what is written in lines 43-49. Perhaps you are referring to how boundary conditions are applied to the downstream end of an inland model(?), but with respect to the FEMA SFHA (V-zone), it is my understanding that the most recent version is derived from ADCIRC simulations (where events are created using JP-OMS) in which case the spatial and temporal variability of the drivers and also the resulting water levels are considered when estimating the resultant hazard.

Line 57 remove 'up to a century'

Line 74-76 I think you are correct that this has not been done for compound flooding; however, I would point out two relevant studies that undertake a similar analysis applied to rainfall flooding (one that uses SFINCS) and reach the same conclusion. I think you may want to point to these both in your introduction (and draw differences between your work from theirs) and again in the conclusion.

Perez, G., Coon, E. T., Rathore, S. S., & Le, P. V. (2024). Advancing process-based flood frequency analysis for assessing flood hazard and population flood exposure. *Journal of Hydrology*, *639*, 131620.

Baer, J. A. (2025). *Design Storms Underestimate Flood Hazard and Risk Derived From Stochastic Storm Transposition* (Master's thesis, The University of North Carolina at Chapel Hill).

Line 88-90 What is the total size of the model domain in km^2?

Line 93 Provide date range for the start of the federal disaster declarations, e.g., 'Between X and 2016, there were five federal disaster declarations…'

Line 94 Please provide more information about the building stock for context. You state that there are only 118 properties with coverage. Is this reflective of active policies based on OpenFEMA or where does this number come from? Do you also have the ability to provide a denominator for the total number of properties in Glouchester City and the number located within and outside of the FEMA SFHA? These would be useful numbers for context, given that insurance is only required at properties with a federally-backed mortgage within the SFHA.

Line 116-120 The downstream boundary condition at the Delaware River is assumed to be represented as a coastal water level? How do you consider the possibility of high river flow events (antecedent conditions that are not coastally driven) in your compound events framework? Have the previous works that are cited already established the relative frequency of high river flood events vs those that are driven by high coastal water levels and local rainfall? Do these events come from a different distribution than those that are considered 'compound coastal' events?

Section 3.3 You invested considerable effort in validating the model, which I think was excellent given how limited the data for this cite is, however, the process used to generate the synthetic events for the model also has many layers of uncertainty in it (e.g., statistically derived joint probabilities of precipitation and non-tidal residual, statistically derived spatio-temporal rainfall structures, scaled event magnitudes, lag). To what extent could your model framework allow you to explore some of these uncertainties? Is there any mechanism that you could use to test the compound events framework and whether you believe your hazards estimates are true (assume the process-based model results are valid)? Given that you are comparing the model against itself, this probably doesn't impact the main takeaways from your paper, but I think it is important to consider and an important question for the field: how do we validate the hazard estimates?

Figure 2 It would be helpful to see the univariate distributions of the two variables (rainfall and NTR). Here, you only show combinations that exceed either ~40 mm in 18-hr or ~0.6 m NTR. Why were these values selected to threshold the data? Given your findings that events with combined boundary conditions much, much smaller than the 1% AEP event can still generate flooding in excess of 1%. Do you have any concerns that there many be small events that should have been included in your stochastic event set?

Figure 2 What do your rainfall return period values look like compared to those estimated by NOAA Atlas 14? Is this distribution derived from the entire precipitation record (and over what scale)? I recognize that this information is likely provided in Maduwantha et al., but I think it is also important to reference that information in this paper as not every reader will have previously read Maduwantha.

Figure 2 How often is the NTR observed the Delaware River driven by coastal vs riverine flood events?

Figure 2 I would recommend splitting Figure 2 into two separate figures. One which shows panel a and a separate figure (and figure caption) for panels b-d. In addition, I would add a second column that shows the spatial distribution of the rainfall plotted next to each time series graph, similar to those shown in Supplementary Figures S5 and S6. This information is really valuable for understanding the resulting differences in flood inundation generated from these events that are described later in the paper.

Line 208-210 What is the size of the domain (total number of cells) in the model? Is 10 m the resolution of your model and 1 m the resolution of the subgrid?

Line 208 You state that the DEM was retrieved from CoNED. Does CoNED provide an accurate estimate for the channel bathymetry for the creeks in your model domain? Channel bathymetry is a big unknown in many hydrodynamic simulations and has been shown to be critical to correctly estimating flood inundation (e.g., they are too shallow, more water will be routed over the floodplain whereas if they are too deep, not enough water will be routed onto the floodplain). If they are not well-represented in CoNED, did you make any effort to 'burn in' or manually adjust the depth and width of the channels in your model? If not, to what extent do you think poor channel bathymetry may impact your validation results, particularly when you have to assume no infiltration in order to match a flood event in 2009 (lines 239-251)?

Line 237 You cite Flores et al. 2023 but there are other, more authoritative sources on this topic. Perhaps this study by the National Academies would be worth citing here in addition to Flores et al.

Lines 239-251 You do an excellent job of using what information you could find to try and validate the model. However, your primary focus is on infiltration and there is no discussion of whether local stormwater management (i.e., subsurface drainage) exists and whether that might contribute to error in the model performance since it can't be represented by SFINCS. There is a late reference to backwater (surcharges), but it does not come up in the validation. I would also suggest to include a description of the watershed as fully urbanized to help justify the decision to neglect infiltration.

Line 244-246 I would expect that neglecting infiltration has bigger implications for the smaller return period events and for those with low-intensity over the duration of the storm. Given that you are later comparing the design storms which only contain a few (or one?) event(s) with these characteristics against the probabilistic set which may contain many storms with these characteristics, what impact might this have on your final results. How could antecedent conditions in the 2009 events contributed to your choice to neglect infiltration? Could the land use (urban?) in this area also provide a justification for neglecting infiltration?

Line 275 Given that SFINCS is very computationally inexpensive, another option for validation would have been to increase the size of the model domain to include locations with gages. For example, if you increase the model to the scale of the HUC10, it appears that the Cooper River watershed, adjacent to Newton Creek has two active USGS gages. Knowing how the model performs at these locations would give you confidence in the parameterization of your model (e.g., infiltration, roughness, elevation), even if your later simulations are focused on a subarea of the entire model domain.

Line 285 I find it interesting that there was little variability in the pluvial flood extent between the different 1% AEP events. How different were the simulated rainfall intensities for these storms or did they all have similar spatial-temporal distributions? Were their centers of mass all located in the same place? Perhaps a supplemental figure that provides the boundary conditions associated with the 25 events would be useful (similar to Figure 2b-d) but with the addition of a panel showing the spatial distribution of the rain that fell.

Line 377 Replace 'Lots' with 'Much'

Line 410 You state that ignoring the spatio-temporal variability of extreme events by relying on a design event can lead to significant uncertainties in flood exposure, but it does not appear that you prove this point. In both your cases (event- and response-based) you are selecting events that vary in space and time, you just sample them differently. This point sounds more like its drawing a comparison with the standard FEMA approach to selecting design storms that are uniform in space and vary in time according to a characteristic distribution. One way to address this would be to cite the existing literature when you make this point before discussing how your approach (both event- and response-based) provide more information than the standard design storm approach.

Line 455-462 You again write here that 'not accounting for the spatial variability of rainfall…' Perhaps I misread, but per the previous comment, I understood that for both your event and response-based approaches, the precipitation was spatially and temporally varied across the model domain. This raises an interesting question… to what extent are your findings driven by the spatial variations in the rainfall in the rainfall field vs the temporal variations in the rainfall distribution? Could one take a design storm approach in which you select the accumulated rainfall from the curve (Figure 2) and apply a characteristic distribution to distribute it over time, compare that to the situation in which you allow it to be temporally varying, and then against the full Monte Carlo approach to provide insight into whether it's the spatial variability or the temporal variability of the rainfall that really matters for the flood hazard? At a minimum, I would be more intentionally about what conclusions can be drawn from the findings in this analysis vs those that are made more broadly in the literature.