Response to Reviewers “Beyond the 100-Year Flood: Probabilistic Flood Hazard Assessment for King and Pierce Counties under Future Climate Scenarios.”

Dear Editor, dear reviewers,

On October 28, 2025, we submitted our manuscript to Copernicus's *“Natural Hazards and Earth System Science (NHESS)”* titled: "Beyond the 100-Year Flood: Probabilistic Flood Hazard Assessment for King and Pierce Counties under Future Climate Scenarios”. On January 18, 2026, we were informed that a decision had been made. We received comments from three reviewers, who provided very constructive feedback on the work done and valid suggestions. We would like to acknowledge their time and efforts, which have led to an improvement in the quality of our manuscript. Below you find a point-by-point reply to all specific questions and suggestions. Attached, you also find the revised manuscript with the changes made to address the review comments tracked.

Kind regards,

C.M. (Kees) Nederhoff, Ph.D.

Anonymous Referee #1 (Nov 22, 2025)

This cell by cell extreme value analysis is presented as novel, but it has been carried out by other, similar studies as well. Therefore, the authors need to clarify what the novelty of their manuscript is, in light of what has been done before. It might be that the novelty is the specific model train used, or the location of application, or a combination of these. Either way, please clarify the novelty.

For example of other applications of the cell-by-cell EVA:

- Deb, M., Sun, N., Yang, Z., Wang, T., Judi, D., Cooper, M. G., & Wigmosta, M. S. (2025). Extreme flood return levels in a US mid-Atlantic estuary using 40-year fluvial-coastal model simulations. Scientific Data, 12(1), 1459.
- Son, S., Xu, C., Davlasheridze, M., Ross, A. D., & Bricker, J. D. (2025). Effectiveness of the Ike Dike in mitigating coastal flood risk under multiple climate and sea level rise projections. Risk Analysis.
- Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C., & Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. Nature communications, 7(1), 11969.

We thank the reviewer for this important point and for providing these relevant references. We agree that cell-by-cell extreme value analysis is an established technique, and we did not intend to claim it as novel. We have revised the introduction to explicitly acknowledge prior implementations, citing Muis et al. (2016), Deb et al. (2025), and Son et al. (2025).

Our novelty lies not in cell-by-cell EVA itself, but in the Continuous Flood Response Modeling (CFRM) framework and its specific contributions:

- **Simulation length and empirical EVA**: Unlike the 40-year simulation of Deb et al. (2025) or event-based approaches, CFRM applies 100 years of continuous boundary forcing, enabling direct empirical estimation of 100-year return levels without fitting theoretical extreme value distributions, thereby avoiding the uncertainty introduced by distribution selection and parameter estimation.
- **Moment-of-flooding functionality**: We introduce new SFINCS model functionality that records the timestamp of maximum water levels at each grid cell, enabling spatially-explicit attribution of flood hazards to specific historical events. This supports both validation (Section 4.1.4) and stakeholder communication by linking probabilistic results to recognizable events. While cell-by-cell extreme value analysis is not novel, examining the spatial distribution of extreme events on a cell-by-cell basis and from a continuous 100-year timeseries is, at least to our knowledge.
- **Quantified comparison with deterministic methods**: We demonstrate that deterministic single-event approaches differ from CFRM-derived estimates by up to 0.5 m in compound flood zones, with systematic spatial patterns reflecting the transition from coastal to fluvial dominance, a finding with direct implications for flood mapping practice.

We have revised the introduction to clearly articulate these contributions.

The introduction is somewhat long. Much space is spent describing the transition from event-based to probabilistic hazard analysis, but this is already quite common knowledge in the community, so this section can be shortened.

We agree with the Reviewer. The extended discussion of the transition from deterministic to probabilistic methods is well-established in the literature and does not require detailed explanation for this audience. We have largely rewritten the introduction so it is more condensed and now focuses on positioning CFRM relative to existing methods and clearly stating our specific contributions.

Fig 1. The label "bathymetry" is confusing, because this cover topography over land. Perhaps just call it "elevation". Also, the lable of Weirs should be explained more in the caption, as these are not apparent on the map.

We agree and have revised the labeling in Figure 1 and changed it to Topo-bathymetry (Elevation). Also, in order to make the weirs more clearly visible in the figure, we changed the color to Purple.

Line 216. You mention roughness values but don't specify which roughness parameterization. Darcy roughness length? Chezy? Hazen Williams? I assume this is Manning, but you need to specify this. We have used Manning's n friction values and added this to the revised manuscript (L156).

Line 230. There is a variation of assumed fraction of wave height for wave setup, typically from 5% to 20%. You choose 20%, which is OK, but should be tested via a sensitivity analysis. For example of the 5% assumption see

- Feng X, Yin B, Yang D, William P (2011) The effect of wave-induced radiation stress on storm surge during Typhoon Saomai (2006). Acta Oceanol Sin 30(3):20–26. https://doi.org/10.1007/s13131-011-0115-6
- Yamanaka Y, Shibata R, Tajima Y, Okami N (2020) Inundation Characteristics in Arida City Due to Overtopping Waves Induced by 2018 Typhoon Jebi. APAC 2019. Springer Singapore, Singapore., pp 199–206Return to ref 2020 in article

We agree with the Reviewer and performed some sensitivity testing in order to test the importance of this assumption. This information was added in the revised discussion (L625-640). The results for the largely estuarine systems (small, locally forced fetch-limited waves) were found not to be sensitive to this assumption.

Line 330. Why did you randomize tidal phase instead of applying the actual phase during the reanalysis simulation? Also, what reanalysis dataset did you use? We ran the reanalysis dataset from Parker et al. (2026). Since this dataset only had 82 years and we wanted to run 100 yr, we randomized the tidal phase for the 18 additional years. We clarified this in the revised manuscript (L274-280).

In your CMIP-driven simulations, how did you determine the upstream river discharges? Did you couple these with a hydrological model for each river watershed? We coupled our flood computations with a downstream coastal model (Parker et al. 2026) and an upstream hydrological model (Buettink et al. 2026). For more information, please see L274-276 in the revised manuscript.

How does the PGW approach deal with shifts in frequency of events due to climate change, if it is based on multiplying the historical time series by a factor? Correct, the PGW approach computes a change with the Global Climate Model (which might be biased) and then uses this computed change on the hindcast period (which is more robust). We apply this change to the individual water level and discharge inflows, see also L285-309. A strength of this approach is that it limits biases from CMIP6 climate models by using the real observed reanalysis timeseries. As the reviewer points out, a limitation is that it does not capture shifts in the frequency of events. With observed biases (and disagreement across the ensemble) in the CMIP6 models, we did not feel confident that directly using the CMIP6 time series was robust, and so a PGW approach was chosen.

Section 3.6 model skill. What are the data being compared here? Historical data vs. present climate simulated data?

We have compared water levels (or gauge height for river points) with metrics such as RMSE and MAE. We have compared flood extents with the binary classification metrics introduced by Wing et al. (2017). We only validated model skill for the present climate, see Section 3.6.

Fig 5 is confusing. It needs a larger-scale locator map, and also needs more visible borders between each subplot.

We agree with the reviewer and recreated Figure 5 to include a larger locator map and made the visible borders between each subplot more clearly visible.

Fig 6 also needs a locator map and better boundaries. Also, what is the source for the extent of each reported flood event? How were these areas determined?

We also recreated Figure 6 to include a larger locator map and made the visible borders between each subplot more clearly visible. The areas flooded were determined by numerical simulations. We clarified this in the revised manuscript.

Since you used 100 years of synthetic analysis instead of statistical distributions, do you have PDFs or CDFs of the resulting 100 years of water level or depth data for some of the cells you analyzed? It would be informative to compare these to what would come from fitting standard statistical distributions, to help determine which would be a better predictor in the practice, moving forward.

We thank the reviewer for this constructive suggestion. We indeed have this information and have added a dedicated discussion in the revised manuscript (L640-660) comparing the empirical distributions against standard parametric fits, focusing on the Duwamish Waterway and the South Park neighborhood in Seattle.

The comparison reveals an important distinction between channel and floodplain flood responses. At the Duwamish Waterway, water levels follow a smoothly shaped empirical distribution that is amenable to fitting with standard GEV or GPD models (Figure 9). However, the floodplain response at South Park exhibits threshold-dependent behavior: flooding only initiates when return periods exceed approximately 10 years. This non-stationarity — driven by the changing elevation difference between channel water levels and floodplain elevation, combined with the fundamentally different shape characteristics of the distribution — violates a core assumption of EVA, namely that block maxima are drawn from a stationary distribution. As a result, parametric distributions fitted to channel gauge data cannot reliably be extrapolated to predict overland flood hazard.

This ambiguity in EVA application across the field — where different practitioners apply different theoretical distributions and arrive at substantially different extreme value solutions — was in fact one of the primary motivations for adopting an empirical approach in this study. The CFRM framework preserves physical discontinuities such as levee overtopping thresholds and backwater effects that parametric distributions would smooth out. We believe Figure 9 now makes this case more concrete and directly addresses the reviewer's question about which approach is a more reliable predictor in practice.

And please clarify. IS SFINCS used in a wave-phase-resolving way, so as to quantify wave runup and/or overtopping? Or is it only still water level that is being assessed?

We ran SFINCS as a compound water level modeled force with downstream still water levels and upstream discharges. Unfortunately, the computational expense of dynamically including wave runup and overtopping was prohibitive. We clarified this in the revised manuscript (L625-640).

This is a very nice use of a variety of methods for GCM application, hydrodynamic modeling, and risk analysis.
We thank the reviewer for their kind words and time reviewing our manuscript!

Anonymous Referee #2 (Dec 7, 2025)

This study applies a compound flood modeling framework to risk estimation in King and Pierce Counties in the US. While the study itself is well carried out, I am concerned that the model performance to reproduce water level is poor as shown in Figure 3-5. Future projections of compound flood risk are also illustrated later in the article, but I am not sure the reliability of the overall framework of the study.

We thank the reviewer for their careful evaluation. We would like to clarify the distinction between coastal and inland model performance, as well as explain why the framework remains reliable for its intended purpose, flood hazard assessment.

First, coastal water levels are reproduced with an RMSE of 14–17 cm (Table 1), which is comparable to or better than similar regional modeling studies. This study is primarily focused on coastal hazards, and so this metric is a better quantification of the fidelity of the modelling approach. The larger errors (uRMSE of 49–116 cm) occur at inland riverine stations where bathymetric data quality is inherently limited, a challenge recognized globally for coastal–riverine interfaces. Regional-scale hydrologic models often avoid validating against stage (rather than using discharge), because of this challenge. This limitation is further compounded by ongoing bed aggradation in the study area rivers. Czuba et al. (2010) documented aggradation of up to 7.5 feet on the lower Puyallup River between 1984 and 2009 based on surveyed channel cross-sections, and estimated aggradation rates of approximately 5 inches per year at streamflow gaging stations on the nearby Nisqually River. Such dynamic channel bed evolution means that the stage-discharge relationship itself is non-stationary, which introduces fundamental uncertainty into any discharge-based flood forecasting approach — independent of model skill.

Second, the primary objective of our framework is to characterize flood extent and return period statistics, rather than to perfectly reproduce instantaneous water levels at all times. The validation against FEMA Special Flood Hazard Areas demonstrates high spatial agreement (hit rates of 0.75–0.83), indicating the model captures flood-prone areas reliably.

Third, we explicitly account for model uncertainty through our "low" and "high" boundary condition simulations (Section 3.4), which bracket the plausible range of flood response. The uncertainty ranges presented in Table 5 reflect this propagated uncertainty in our hazard estimates.

We have revised the discussion (Section 5.1) to more explicitly acknowledge these error magnitudes and contextualize them relative to the study's objectives (L560-585).

Anonymous Referee # 3(Dec 22, 2025)

The manuscript introduces a framework for mapping compound coastal–riverine flood hazards in the Salish Sea via dynamic modeling. The approach, based on cell-by-cell EVA across historical and future scenarios, is relevant for regional hazard assessments. While the work is timely, the methodology needs better justification and more concise justification, specifically regarding uncertainty handling, driver dependence, and validation of results.

We thank the reviewer for this summary assessment. We have addressed these concerns through several revisions:

1. **Novelty and justification**: We have revised the introduction to clearly distinguish our CFRM framework from existing cell-by-cell EVA implementations, emphasizing the 100-year continuous simulation approach, new timestamp functionality, and quantified comparison with deterministic methods (see Response 1 above).
2. **Uncertainty handling**: We have clarified that the "low" and "high" simulations represent deterministic sensitivity tests inspired by the 95% confidence interval concept, rather than formal statistical confidence bounds (L377-392).
3. **Driver dependence**: We have clarified that for 82 years, the historical record preserves natural correlations between coastal water levels, waves, and discharge, while for the 18 synthetic years, we assume a correlation between NTR and discharge/waves. We acknowledge this limitation in the revised discussion (L418-420).

Validation: We have reframed the FEMA comparison as a qualitative benchmark rather than formal validation, given the unknown methodologies underlying FEMA maps (revised L215-227).

The Introduction is repetitive, particularly between lines 45–60. It should be tightened to focus on the paper's specific methodological novelty. Similarly, the Study Site section contains unnecessary geographic and ecological information (e.g., lines 118–120, 138–145). Focus strictly on the physical processes and boundary conditions relevant to the flooding model.

We agree with both observations and have made the following revisions:

- Introduction (L28–70): We have removed the repetitive content explaining why deterministic methods fail to capture compound events, consolidating this into a single sentence. The revised text now proceeds directly from the problem statement to existing approaches and then to our specific contributions.
- Study Site section: We have removed content not directly relevant to flood modeling boundary conditions. The Study Site section has been reduced by approximately 600 words (from 1400 to 785 words).

Figure 1. The map is currently hard to read. The weir locations are nearly invisible, and the "yellow" color scheme lacks sufficient contrast. Please change the legend from "bathymetry" to "topobathymetry" and add labels for major cities mentioned in the text (Seattle, Tacoma). Move the basemap metadata to the acknowledgments or remove (same for Figure 5). We agree with the reviewer and have revised Figure 1 to make it easier to read. Changes include a change of the colormap, a change in labels, inclusion of major cities, and making weirs easier to see.

Lines 221–228: You cite Parker et al. (in prep) for water levels and waves but provide no detail on the spatial coverage or specific extraction points. These boundary locations must be indicated in Figure 1.

We clarified details of the modeling, spatial coverage, and output extraction at ~20 m depth from Parker et al., which is based on a regional Delft3D Flexible Mesh (Delft3D FM) model and computes tides and surges across the Salish Sea, see (L160-175) in the revised manuscript. We also added the boundary location in Figure 1 (white circles for downstream water level boundaries and blue circles for upstream discharge inflows).

Line 228: The description of "sum of locally generated wind waves and the linear transformation of Pacific Ocean swell" is vague. Are you using spectral superposition or a simple bulk-wave approximation? Where exactly is this summation performed?
We are using a simple bulk wave approximation for wave height and wave period is determined with a weighted average. We have clarified this in the revised manuscript (L169-175)

Using the 0.2 Hs for wave setup is a significant simplification for a mixed-energy environment like the Salish Sea. Please justify why this uniform approach is valid despite the region's high spatial heterogeneity.
We agree that this is a significant simplification, but one that was deemed necessary. This modelling choice was made because the computation cost was prohibitive for a continuous simulation compound flood modeling approach for a full 100-year timeseries. A version of this modelling framework was developed to resolve wave setup, but was found to be too computationally heavy. Furthermore, while the Salish Sea in general is mixed-energy, King and Pierce counties are all local seas, and wave setup was found to be a relatively small contribution to total water levels. To test our simplification, we included some sensitivity testing in the revised discussion (L625-640).

The comparison with FEMA SFHAs should be framed as a qualitative benchmark rather than a formal validation.
We agree with the reviewer and changed the wording around the comparison in the revised manuscript.

Line 317: The “low” and “high” simulations are described as representing a 95% CI but these simulations should be treated as deterministic sensitivity tests rather than formal statistical confidence intervals.
We agree with the reviewer and have revised the manuscript. The low and high estimates are not formal statistical confidence intervals but instead deterministic sensitivity tests inspired by the 95% CI (L255-265).

The model shows water-level errors over 1 m and 1-day delays in peak timing. These are significant discrepancies. Would a sensitivity analysis on riverine bathymetry help in this validation?
We thank the reviewer for this observation. The larger inland water level errors (uRMSE 49–116 cm) are indeed driven primarily by uncertainties in riverine bathymetry, particularly the overestimation of tidal propagation upstream due to hydro-flattened DEMs that lack accurate subaqueous channel geometry. We acknowledge this limitation and have expanded the discussion in Section 5.1 to contextualize these errors.

A comprehensive sensitivity analysis of riverine bathymetry would help quantify its contribution to inland water level errors. However, such analysis was not performed in this study due to (1) the computational expense of re-running thousands of annual simulations for multiple bathymetric configurations, and (2) the lack of high-quality riverine bathymetric survey data against which to calibrate alternative configurations. We did perform manual deepening of channel profiles along the Green and Puyallup Rivers (Section 3.2.1), which improved hydraulic connectivity.

Importantly, despite these water level discrepancies at individual stations, the primary objective of this framework is flood hazard characterization rather than instantaneous water level reproduction. The high agreement with FEMA flood extents (hit rates 0.75–0.83) demonstrates that the model reliably identifies flood-prone areas. Furthermore, our uncertainty quantification through low and high boundary condition simulations (Section 3.4) propagates these errors into the final hazard estimates (Table 5), providing stakeholders with realistic bounds rather than single deterministic values.

Why use static SLR scenarios when you have a dynamic framework capable of handling temporally evolving boundaries? This feels like a missed opportunity.
This is, in fact, a very deliberate choice. First, time horizons in sea level rise projections are strongly influenced by carbon emission trajectories across our planet. Secondly, specific projections vary rapidly with new science coming out. Therefore, depending on the users' risk tolerance, one can select a certain emission trajectory and find the latest guideline for SLR to relate to timing. This ensures flexibility for our stakeholders and allows our output products to be used for a longer time.

Averaging deltas across CMIP6 members inherently smooths the extremes. This seems to contradict your goal of capturing rare, high-impact events. Comparisons between static SLR scenarios and averaged CMIP6-based scenarios should be more carefully justified
We agree with the reviewer: averaging CMIP6 members smooths the extremes. However, we are less focused on end members within the ensemble and more focused on finding a robust change in the extremes. We take the view that the CMIP6 ensemble is a statistical sample with uncertainty/randomness, in which we use averaging to improve robustness and limit individual model biases and incorporate uncertainty. This may smooth extremes, but with the added gain of confidence in the utilized signal. Beyond this, the focus is more on moving practitioners away from stationary 100-yr univariate assessments and towards an acceptance of the instationarity of the climate across return periods for compound flooding.

Line 338: Assuming discharge and waves are fully correlated with non-tidal residuals is a very strong assumption that likely masks the true variability of compound events. What is the impact of this simplification on your results?
For 82 years, we have let the historical record of waves, tides, and non-tidal residuals determine the compound flood patterns, but the reviewer is correct that for 18 years, we have assumed that waves and non-tidal residuals are correlated. We have acknowledged this limitation in the revised manuscript, including the possible impacts on the results.

Figure 5: The SFINCS-FEMA comparison currently includes permanent river channels. To get a real sense of the model's skill in predicting floodplain inundation, you should exclude the river cells from the wet/dry validation as you did with coastal waters.
River channels were intentionally retained in the validation because, unlike coastal waters, which represent prescribed boundary conditions, channel water levels within the estuary are model outputs that emerge from the interaction of fluvial and tidal forcing. Excluding channels would remove the transition zone where compound flood dynamics are most relevant to this study. Additionally, FEMA flood zones include river channels, so a consistent comparison requires their inclusion in both datasets.

Line 303: Fix typo "precdeded".
Line 644: Fix typo "eproduces".

We fixed both typos and thank the reviewer for pointing these out.

Define EAFA and SLR once at first mention and then use the acronym consistently.
In the revised manuscript, we introduce acronyms once and use them consistently. We do repeat the acronyms in figure captions, abstract, discussion & conclusion since these are often read independently from the manuscript.