Responses to Reviewers' comments on:

"Quantifying cascading uncertainty in compound flood modeling with linked process-based and machine learning models"

David F. Muñoz, Hamed Moftakhari, and Hamid Moradkhani

Ms. Ref. No.: hess-2024-9

Submitted to: Hydrology and Earth System Sciences

We really appreciate the thoughtful comments and suggestions from the reviewer to improve the quality of the manuscript. For clarity, we have included the original reviewer's comments in blue text and our point-by-point response in black text.

Reviewer #1 (RC1)

A few places, the analysis requires attention in terms of Methods implemented, otherwise it is written well. Therefore, I suggest major revisions for this manuscript.

We thank the reviewer's feedback and further comments to improve this study. Please find below a detailed explanation of how changes are made according to your comments.

A. General Comments:

1. Line 42: CF events in low-lying areas are typically associated with tropical or extra tropical cyclones for which rainfall-runoff, wind-driven storm surges, total coastal water level including wave set-up and tidal variations, or all of the events concurrently or in a close sequence contribute to the severity compound events (Ganguli & Merz, 2019a; Ganguli & Merz, 2019b).

Indeed, we agree that extra-tropical storms such as the Nor'easters in the U.S. Atlantic Coast have the potential to drive CF events. Also, we thank for the important references related to CF in northwestern Europe. Please find below is the revised text.

"CF events in low-lying areas are typically associated with tropical or extra-tropical cyclones for which rainfall-runoff, wind-driven storm surge, or both can be classified as dominant flood hazard drivers (Bevacqua et al., 2020; Eilander et al., 2020; Ganguli and Merz, 2019a). In addition, the role of waves, tides, and nonlinear interactions on extreme water levels (WLs) can be crucial for the accurate simulation and/or prediction of CF events as reported in several studies (Ganguli and Merz, 2019b; Hsu et al., 2023; Nasr et al., 2021; Serafin et al., 2017)."

2. Line 117: multi-model ensemble methods: Kodra et al. (2020) proposed empirical Bayesian model that incorporates skill and consensus based weighing framework to narrow down uncertainty associated with large ensemble of earth system models in the projected climate.

Thanks for the sharing the contribution of Kodra et al., 2020 regarding multi-model ensemble methods. We have included this reference in the revised text as indicated below.

“Those methods include linear associations and first-order second moment approximations (Taylor et al., 2015; Thompson et al., 2008), generalized likelihood estimations (Aronica et al., 2002; Domeneghetti et al., 2013), sensitivity analyses (Alipour et al., 2022; Hall et al., 2005; Savage et al., 2016), multi-model ensemble methods (Duan et al., 2007; Kodra et al., 2020; Madadgar and Moradkhani, 2014; Najafi and Moradkhani, 2016), and data assimilation (Abbaszadeh et al., 2019; Moradkhani et al., 2018; Pathiraja et al., 2018).”

B. Typing errors:

3. Line 221: unstructured finite volume grid that consists of triangular 'elements' and not the 'cells'?

We respectfully disagree with the reviewer since triangular cells is the correct term when referring to unstructured grids. Below is a schematic of triangular cells extracted from the Delf3D-FM Technical Reference Manual for your consideration.

We now proceed by considering a cell attached to a node i in coordinate frame (ξ', η'), see Figure 3.6, and define an *optimal* angle Φ^{opt} between two subsequent edges that are connected to node i.

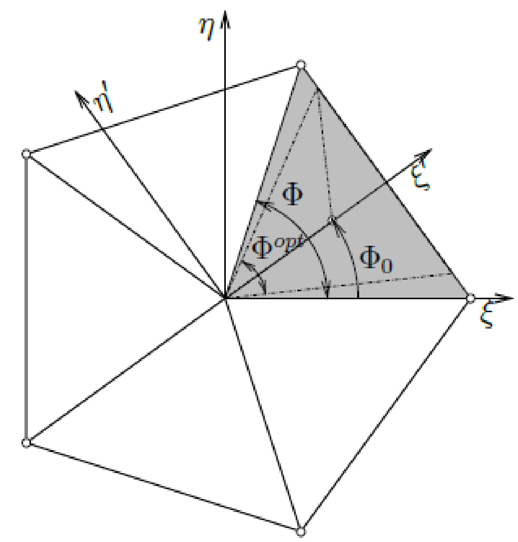

Figure 3.6: *non-rectangular triangular cell; the dashed cell is an optimal equiangular polygon, while the shaded cell is the resulting cell after scaling in η' direction; Φ_0 is the angle of the ξ'-axis in the (ξ, η)-frame*

4. Line 228: The word, 'in' appeared twice.

Thanks for catching this. We have revised the text as shown below.

“Similarly, the second hydrodynamic model is developed in 2D HEC-RAS using an unstructured finite volume grid.”

5. Figure 6 caption: Effect of individual sources of uncertainty. (a,b) initial condition, (g,h) model structure.

Thanks, we have corrected this typo as follows.

"Figure 6. Effect of isolated uncertainty on compound flood hazard assessment in Galveston Bay. Maximum water level residuals represent model scenarios with uncertainty stemming from (a, b) initial condition and (c, d) model structure. Water level residuals are calculated with respect to the best hydrodynamic model calibrated for Hurricane Harvey. Positive and negative residuals indicate overestimation and underestimation across the model domain, respectively. Right panel shows a zoom-in window over block census groups in Harris County at the northwest side of Galveston Bay."

5. Line 578: both agrees well with slope of the regression estimate. A linear regression yields two: slope and the intercept terms.

We agree with the reviewer and have clarified this in the text, accordingly.

"Similarly, the rank of permutation importance agrees well with both the ranks of regression weights (or slope terms) and those derived from feature importance (Figure 7i)."

C. Technical Comments:

7. Line 245-247: The discharge from the lake upstream and river gauge downstream are estimated simply the sum of two random variables. However, since both random variables are independent, the derived distribution can't be a simple sum – here convolution methods needs to be implemented to quantify sum of two continuous random variables:
https://dlsun.github.io/probability/sumscontinuous.html

We thank the reviewer for raising this important comment. Nevertheless, as explained in the link that you kindly provided, the sum of continuous random variables is necessary for estimating probability density functions (PDFs). This will also help estimate cumulative distribution functions and thereby return design periods. Here, we are simply estimating river discharge for Hurricane Harvey by leveraging available upstream time-series data as discussed in a previous peer-reviewed work in G-Bay (Muñoz et al., 2022). Another feasible alternatively consists in leveraging modeled river discharge from the National Water Model as discussed by Huang et al., (2021). In fact, our method led to comparable results with respect to the cited study in G-Bay. While we acknowledge limitations of our method, such as correctly capturing the timing of the peak flow, the sum of river discharge data is a reasonable proxy for the upstream boundary condition in the San Jacinto River. Moreover, this proxy leads to satisfactory water level simulations as shown in Figure 3 (e to h) and S2.

8. Line 91: Chezy's formula that is dependent on surface roughness, Reynold's number of fluid in contact and the mean hydraulic depth.

We appreciate the reviewer's recommendation. While we do not work directly with the Chezy's formula as per the model configuration in Delft3D-FM (e.g., Manning's equation), we account for the Reynold's number in the eddy viscosity concept for both laminar and turbulent flows, i.e., v_H is the horizontal viscosity parameter in Eq. (2) and (3). The viscosity concept expresses the Reynolds stress component as the product between flow and grid-dependent v_H as well as the corresponding components of the mean rate-of-deformation tensor. We believe this technical explanation is discussed in detail in any "Open Channel Flow" lectures. Therefore, for the sake of brevity, we limit our analysis to the continuity and momentum equations already included in the manuscript.

9. Line 336: 1:1 fit line to be fit of the linear regression

Here, we use a 1:1 line to evaluate the accuracy of simulated water levels using composites of the maximum values. We apologize for the confusion when referring to a 1:1 fit line from a linear regression model. The revise text reads as follows:

"We evaluate the accuracy of the composite maps by comparing observed and simulated maximum WLs (Figure 4b). Data points that fall along the 1:1 (diagonal) line represent a perfect match between those maximum WLs."

10. Figure 5: In flowchart: Also shows assessment wrt other machine learning methods.

We thank for the suggestion. However, we would like to keep the flowchart as simple as possible by focusing on the random forest (RF) regressor as it outperforms other widely used algorithms including artificial neural networks (ANN) and support vector regressor (SVR).

11. Lines 425-430: The comparative assessment with other machine learning methods should also be presented in supplementary.

We believe that such a discussion is unnecessary based on multiple peer-review studies and underlying evidence suggesting that ensemble learning methods (e.g., RF regressor) outperform other machine learning algorithms like SVM and ANN (Mosavi et al., 2018; Chen et al., 2020; Schoppa et al., 2020). For the reviewer's convenience, we include the requested assessment in this response as follows:

We compare the results obtained from the RF regressor with those of ANN (e.g., multilayer perceptron) and SVR (Figure RC1). ANN's and SVR's model parameters are calibrated by following an identical hyperparameter grid approach as described in the RF regressor (Section 2.4.2). The calibrated parameters for ANN are: hidden_layer_sizes=(100), activation='relu', solver='adam', learning_rate='adaptive', learning_rate_init=0.001, and tol=0.001. On the other hand, the calibrated parameters for SVR are: kernel="linear", C=10, gamma="auto", max_iter=500, and tol=0.001. ANN achieves satisfactory results with respect to those of RF regressor in terms of RMSE and both Pearson's and Kendall's correlation coefficients; the latter reported in parentheses. In contrast, SVR did not perform well even with nonlinear kernels such as the radial basis function (RBF).

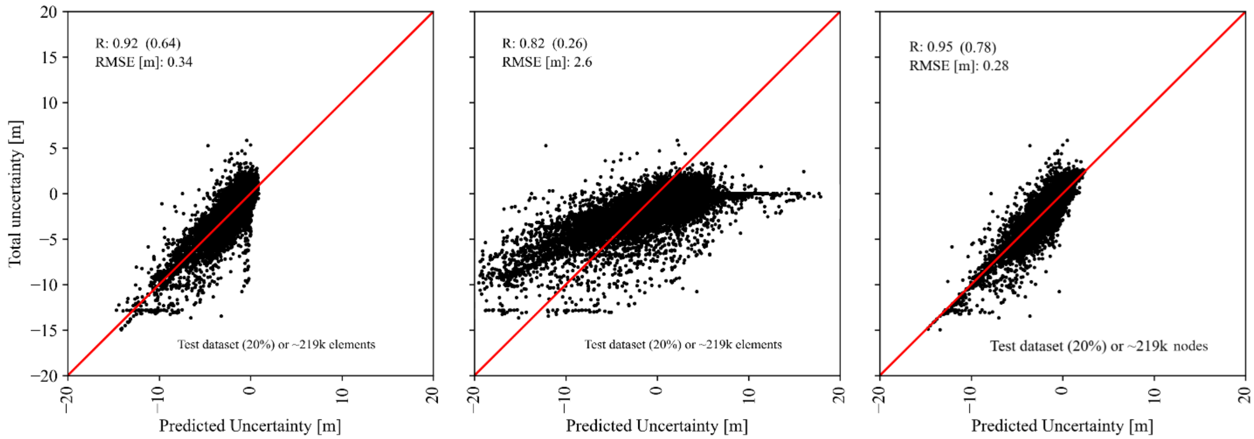

Figure RC1. Comparison of machine learning models for predicting total uncertainty. Results of RMSE, Pearson's, and Kendall's correlation coefficients are show for ANN (left panel), SVR (middle panel), and RF regressor (right panel). RF regressor outperforms both ANN and SVR.

12. Line 437: How outliers are identified?

We agree that an explanation of outlier removal is needed in the manuscript. The revised text reads as follows:

"In the context of hydrodynamic modeling, outliers are unrealistic WLs emerging around upstream and downstream BC lines as well as the edges of the model domain. Such values are extreme values, either positive or negative, that do not reflect WL dynamics within the model domain. Therefore, we masked out such values using a buffer polygon in ArcGIS and proceed with the training and validation dataset using realistic WLs (e.g., 1'093,501 data points)."

13. Sub-section heading 3.1: Effects of Individual and Aggregated Uncertainty

We thank for the suggestion. The heading has been modified accordingly. We prefer the term isolated and total uncertainty as suggested in your comment # 14.

"3.1 Effects of isolated and total uncertainty"

14. Line 513: Why the results of scenario S5 is shown in the Supplementary? It should be presented in the main text. Instead of cascading effects of the sources of uncertainty, the correct term would be total uncertainty considering all four sources that propagate in the system

We accept your thoughtful suggestions. Scenario S5 is now included in figure 6 as shown below. The figure caption is also modified to include the term "total uncertainty".

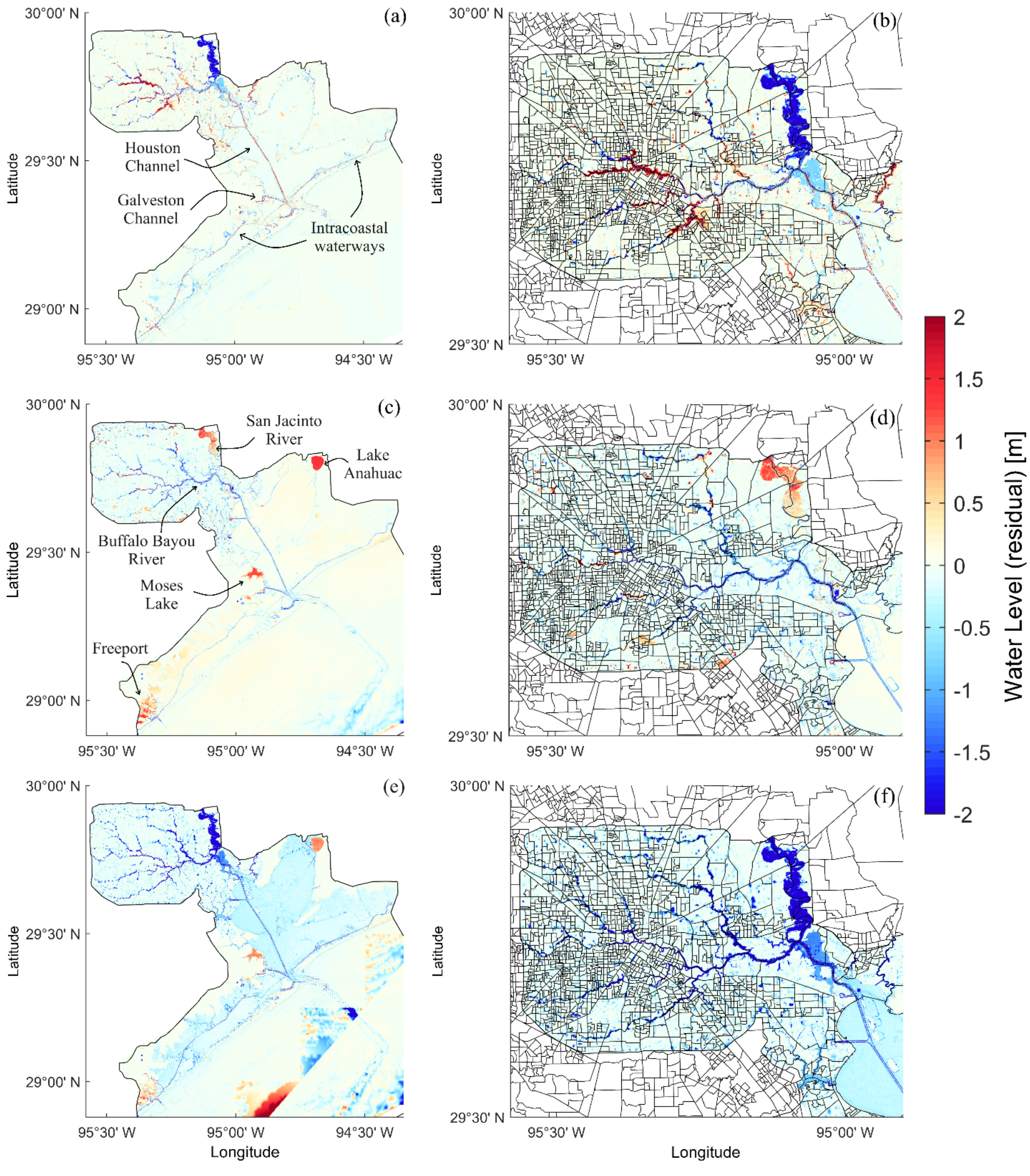

Figure 6. Effect of isolated and total uncertainty on compound flood hazard assessment in Galveston Bay. Maximum water level residuals represent model scenarios with uncertainty stemming from (a, b) initial condition, (c, d) model structure, and (e, f) total uncertainty. Water level residuals are calculated with respect to the best hydrodynamic model calibrated for Hurricane Harvey. Positive and negative residuals indicate overestimation and underestimation across the model domain, respectively. Right panel shows a zoom-in window over block census groups in Harris County at the northwest side of Galveston Bay.

15. Figure 7: Simply Pearson's r would not be suffice given highly nonlinear relation between individual and total uncertainty, please consider Kendall's tau instead.

We have included the Kendall's tau correlation coefficient in parentheses to address the reviewer's comment as shown below. Nevertheless, we see that both correlation coefficients lead to an identical conclusion regarding the benefit of machine learning approaches.

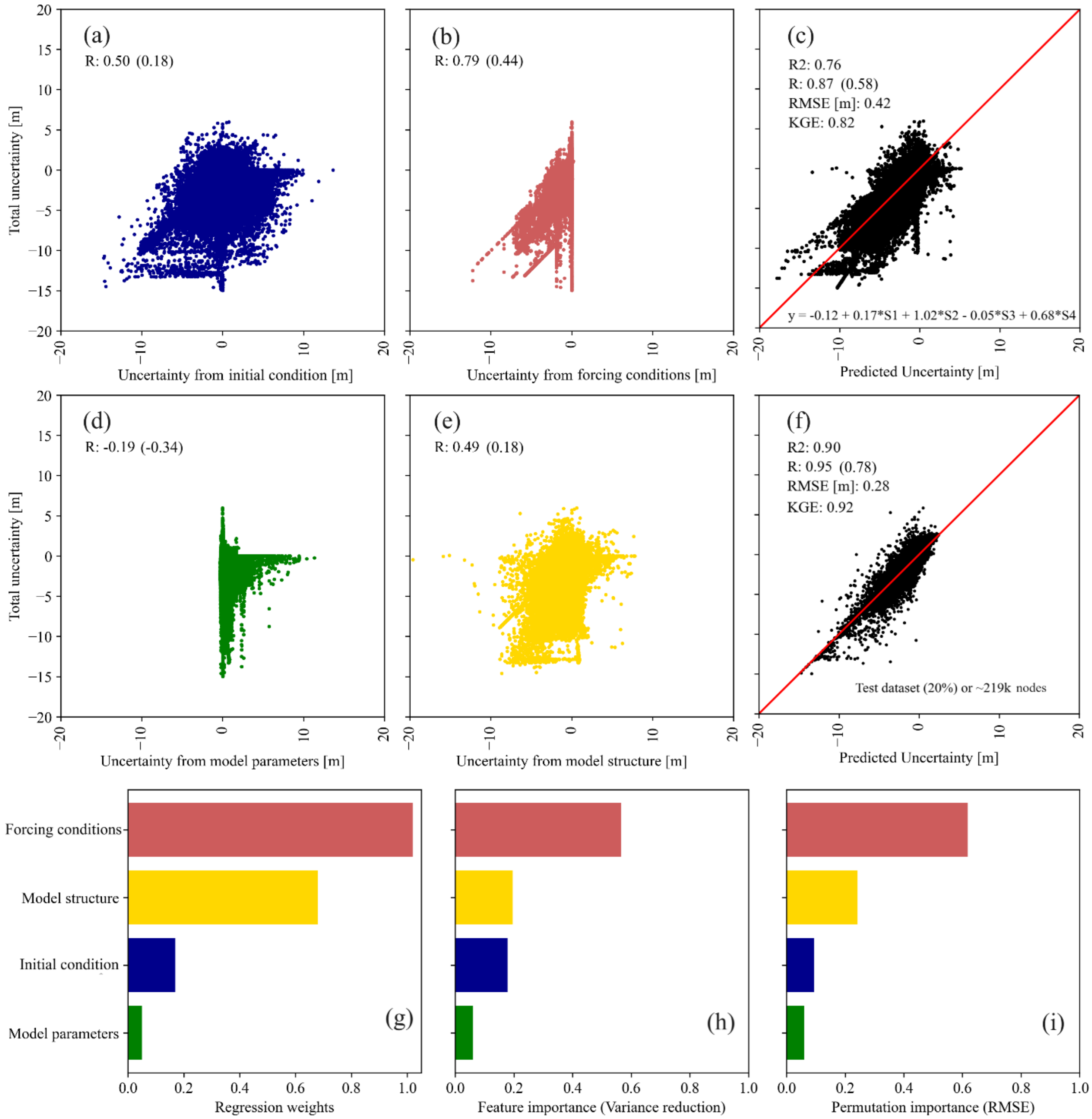

Figure 7. Isolated and total uncertainty reported in terms of water level residuals [m]. (a, b, d, e) Linear associations with the corresponding Pearson's and Kendall's correlation coefficients; the latter in parentheses. (c, f) Total and predicted uncertainty obtained from multiple-linear regression and RF regressor models. (g, h, i) Relative contribution of initial condition, forcing

conditions, model parameters, and model structure to total uncertainty in terms of regression weights, feature and permutation importance.

16. Table 4: No results shown for scenario S5. How the 95% confidence bounds are obtained in Table 4-please explain in Table footnote.

We realize that the reviewer is misinterpreting Table 4. The multiple linear regression model involves scenarios 1 to 4 as independent variables (e.g., sources of uncertainty). Scenario 5 is the dependent variable (e.g., total uncertainty). The 95% confidence intervals are obtained from the 'statsmodel' package available in both Python and R studio. We have included this information in the main text and Table 4, accordingly.

"We use the 'statsmodel API' package in Python to conduct a robust fitting of input features (https://www.statsmodels.org/stable/api.html) and report regression coefficients with the underlying statistical significance and confidence intervals (see Section 3.2)."

Table 4. *Multiple-linear regression fitting on maximum water level residuals.*

Scenario (Water level residuals)	***Input features***	***Regression weight***	***Confidence interval [5%, 95%]***
	Intercept	*-0.115*	*[-0.116, -0.114]*
S1	*Initial condition*	*0.175*	*[0.173, 0.176]*
S2	*Forcing conditions*	*1.017*	*[1.015, 1.019]*
S3	*Model parameters*	*-0.050*	*[-0.054, -0.046]*
S4	*Model structure*	*0.681*	*[0.679, 0.683]*

Confidence intervals are obtained from the 'statsmodel' package available in Python.

17. Line 544: Pearson's r doesn't give you rank. Only non-parametric methods are based on rank order transformation. For the former case, r is parameter.

We apologize for the confusion. We know that Kendall's tau and Spearman's rho are nonparametric methods that measure the association between two variables. In contrast to the Pearson's rho, they both rank data as correctly pointed out by the reviewer. Here, we refer to the rank resulting from the computed coefficients (either Pearson or Kendall) and its agreement with the absolute magnitude of regression weights. We do not refer to the rank of the data themselves. We have clarified the text as follows.

"Overall, the absolute magnitude of regression weights agrees well with the rank resulting from either Pearson's or Kendall's correlation coefficients."

18. One of the crucial steps in uncertainty quantification is narrowing down of uncertainty envelop & the identification of such method that can credibly narrow down the uncertainty. However, no such analyses were presented.

We fully agree with the reviewer's comment. However, the main objective of this study is to characterize the sources of uncertainty using process-based and machine learning methods as described in the abstract. Reducing uncertainty using residual learning techniques is the next step (our ongoing work) once the proposed methodology has been validated. To account for the reviewer's comment, we have edited the text in the conclusion section as follows:

"Following these results, we conclude that PB-ML models are a feasible alternative to conventional statistical methods for characterizing cascading uncertainty in compound coastal flood modeling and CF hazard assessment. The relative importance of the sources of uncertainty may also vary depending on catchment properties, storm characteristics, and dominant flood drivers, i.e., coastal to inland transition zones. Ongoing work is being conducted to effectively reduce uncertainty using residual learning techniques. Also, future work should focus on quantifying and reducing cascading and total uncertainty at large-scale and analyzing the effects of the four sources of uncertainty in flood risk assessment (e.g., damage cost)."

19. Line 613: A PB-ML to outperform ordinary linear MLR is pretty obvious. The assessment wrt other machine learning models should also be discussed.

As discussed in the comment # 11, we have included the suggested assessment in this response. We believe that a comparison is unnecessary based on multiple peer-review studies and underlying evidence suggesting that RFR outperforms ANN and SVM. Furthermore, our results presented in Figure RC1 support this claim.

Author's References:

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., et al. (2020). Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment* 701, 134979. doi: 10.1016/j.scitotenv.2019.134979

Huang, W., Ye, F., Zhang, Y. J., Park, K., Du, J., Moghimi, S., et al. (2021). Compounding factors for extreme flooding around Galveston Bay during Hurricane Harvey. *Ocean Modelling* 158, 101735. doi: 10.1016/j.ocemod.2020.101735

Mosavi, A., Ozturk, P., and Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water* 10, 1536. doi: 10.3390/w10111536

Muñoz, D. F., Abbaszadeh, P., Moftakhari, H., and Moradkhani, H. (2022). Accounting for uncertainties in compound flood hazard assessment: The value of data assimilation. *Coastal Engineering* 171, 104057. doi: 10.1016/j.coastaleng.2021.104057

Schoppa, L., Disse, M., and Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology* 590, 125531. doi: 10.1016/j.jhydrol.2020.125531