Replies to Reviewer #1's comments on "Partitioning the uncertainty contributions of dependent offshore forcing conditions in the probabilistic assessment of future coastal flooding at a macrotidal site ". (nhess-2021-271)

We would like to thank Reviewer #1 for the constructive comments. We agree with most of the suggestions and, therefore, we have modified the manuscript to take on board their comments. We recall here the reviews and we reply to each of the comments in turn.

Reviewer #1:

General comment

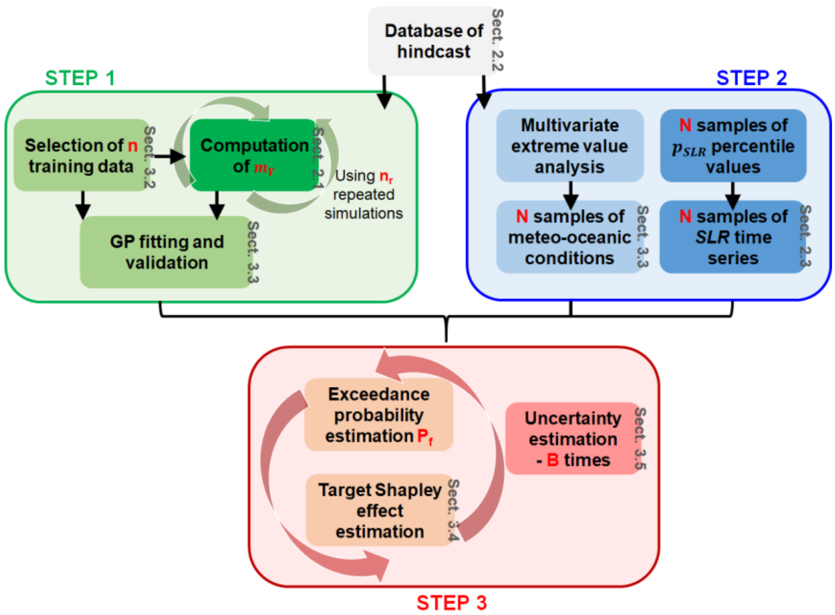
The authors present an interesting methodological framework to assess the uncertainty contributions of forcing conditions to the probability of future flooding. Then, the framework is applied to a macrotidal study site at the French Atlantic coast, and its sensitivity to different framework assumptions is also tested.

The paper is quite interesting both in terms of proposed methods and yielded results, and will be of interest for NHESS readers.

We thank Reviewer #1 for his/her positive analysis.

The methodological framework is complex, and because of the way the paper is structured, some parts of the method are under the Section 4 (Application) instead of Section 3 (Methods). As a general comment, I would recommend a slight review of the current structure of the manuscript, trying to wrap all methodological information in section 3, leaving sections 4 and 5 to purely present results. A methodological chart might also help in such a complex study.

Thank you for this suggestion. We have rewritten the methodological section in this sense (by moving any methodological description of Sect. 4 in Sect. 3). In this view, we have also rewritten the data description in Sect 2. Finally, Sect. 3.1 has been rewritten to better describe the different steps as well as the links between the sections, and a flowchart (new Figure 5) has been added to clarify the different steps.



New Figure 5: Flowchart of the procedure. The sections describing the methods/data are indicated in grey next to the boxes.

In line with the previous comment, current section 6 presents a summary of results and future works. I would also recommend splitting this in Discussion and Conclusions. The results summary is repetitive and the manuscript could be improved with a broader discussion comparing the used methods (e.g. the Heffernan and Tawn 2004 approach, or the GPs) with other methods available in literature such as hierarchical copulas or RBFs (e.g. in Goulby et al., 2014). Some discussion on limitations is done in current chapter 6 although it seems short for such a complex study such as the one performed here, with many methodological steps.
We agree with this suggestion. Sect. 5 is now dedicated to the discussion by assessing the impact of the modelling choices in Sect. 5.1 (as previously done) and the remaining limitations in Sect. 5.2 (regarding the modelling assumptions, the drivers of the flood processes and the *SLR* effect). In particular, we have shortened Sect. 5.1 by focusing on the description of the results of new Figure 9 (old Fig. 14) and by placing the details in Supplementary Materials E.

Regarding the use of alternative methods for extreme modelling, we have added this aspect in Sect. 5.2 by highlighting the interest of comparing to copula-based approaches; in particular by referring to the recent comparison exercise of Jane et al. (2020).

Regarding the use of alternative metamodelling techniques, we acknowledge that other methods could have been used. Though of interest, given the high predictive capability of the fitted GP (Q^2 >99%, see new Figure 6) in our case, we believe that this comparison would bring little added value. We preferably focus on the uncertainty related to the approximation of the true numerical model by a metamodel, i.e. the GP error. Contrary to other methods, GP can easily account for this type of error using the sampling-based approach described in Sect. 3.5. This is now better emphasized in Sect. 3.1. We have also underlined this aspect in the concluding remarks as well as in the abstract.

Added reference:

Jane, R.; Cadavid, L.; Obeysekera, J.; and Wahl, T.: Multivariate statistical modelling of the drivers of compound flood events in south Florida. Natural Hazards and Earth System Sciences, 20(10), 2681-2699, 2020.

Specific comments

Line 33: "…flood severity is significantly increased..
This is now corrected.

Line 112: Maybe the sentence is lacking the verb: "..were built on…".
This is now corrected.

Line 266-270: This is general methodology. Consider defining the base case scenario and the alternative scenarios to test sensitivity on different assumptions under the methodological sections.
Thank you for this suggestion. Given the comments of the two other reviewers, we preferably keep the original structure as such, because we believe that the readability is improved when the test sensitivity on different assumptions (originally described in Sect. 4) is discussed directly in the new Sect. 5 "discussion".

Line 304: What does aggregating mean in this context? From what I understand, results of Q2 in Table 1 should be given as a mean +- std as they are the result of a 10-fold cross validation procedure. How is the single value given in table 1 computed from the 10 folds?
We estimate a global performance indicator (here defined as the coefficient of determination Q^2), and to do so we use the prediction errors calculated at all iterations of the cross-validation procedure. That is why there is only one value. Please refer to Hastie et al. (2009): Sect. 7.10 for further details. This is now clarified in Sect. 3 as follows.

"To validate the assumption of replacing the true numerical simulator by the kriging mean (Eq. 2a), we measure whether the GP model is capable of predicting "yet-unseen" input configurations, i.e. samples that have not been used for training. This can be examined by using a K-fold cross-validation approach (e.g. Hastie et al., 2009: Sect. 7.10). To do so, the training data is first randomly split into K roughly equal-sized parts. For the k^{th} part, we fit the GP model to the other K−1 parts of the data, and calculate the prediction error of the fitted model when predicting the k^{th} part of the data. We do this for k = 1,2,...,K and combine the K estimates of prediction error as follows.

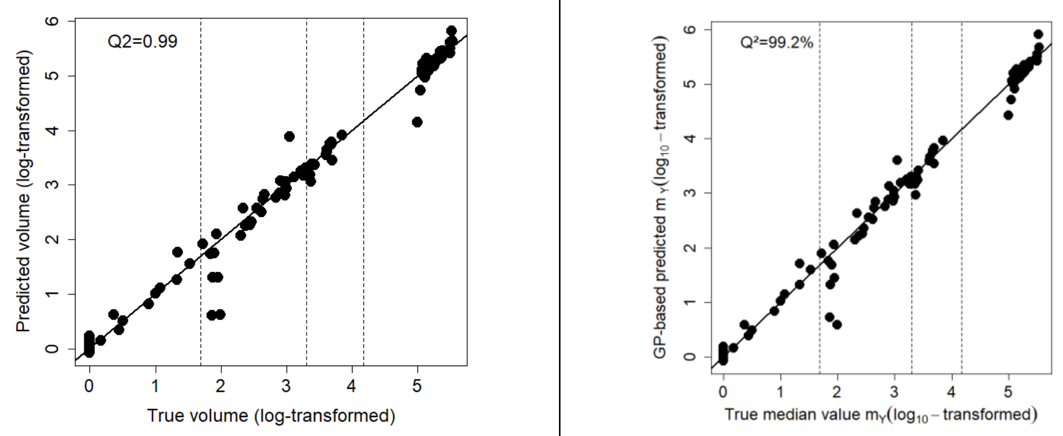
Let us consider $\Lambda:\{1,\dots,n\} \to \{1,\dots,K\}$ an indexing function that indicates the partition's index to which each data point (of the training dataset) is allocated by the randomization, and denote by $\widehat{m}_Y^{-k}(\boldsymbol{x})$ the prediction at $\boldsymbol{x}$ using the GP model fitted using the k^{th} part of the data removed. Then, the cross-validation estimate of the coefficient of determination denoted Q^2 holds as follows:

$$Q^2 = 1 - \frac{\sum_{i=1}^{i=n}\left(m_Y^i - \widehat{m}_Y^{-\Lambda(i)}(\boldsymbol{x}_i)\right)^2}{\sum_{i=1}^{i=n}\left(m_Y^i - \bar{m}\right)^2}, \tag{3}$$

where m_Y^i is the i^{th} median value of Y computed using the modelling procedure of Sect. 2, and $\bar{m}$ is the average value of the numerically computed median values. A coefficient Q^2 close to 1.0 indicates that the GP model is successful in matching the new observations that have not been used for the training".

Figure 5: is it showing 1 fold or the 10 folds? A word on the effect that the poorer performance of the GP approach around Yc = 50 m3 might have on the results might be interesting in the discussion.

We confirm that new Figure 6 (old Figure 5) is showing 10 folds. We thank Reviewer #1 for noticing a possible problem around Yc = 50 m3. To check for any problem in our procedure, we have repeated the 10-fold cross validation procedure. New results are shown in new Figure 6: the same behavior can be noticed (though with some differences because the split of the dataset is done randomly as afore-described). Both figures clearly show a possible lack of predictability around Yc = 50 m3, and we now have clearly indicated in Sect. 4.1 that this potential problem is a motivation for accounting for the uncertainty in our results thanks to the procedure described in Sect. 3.5. The width of the error-bars in the Shapley effects' estimations confirm that the impact of this GP error is here only minor.



(a) Previous cross validation | (b) New cross validation

Line 332-333: the ZCA-cor procedure and number of neighbors are both parameters of the R function, namely n.knn and rescale. This part and associated reference can be moved to the methods section, as it is confusing here.

Thank you for this suggestion. This description is now placed in Sect. 3.5.

Line 335. It is not standard to present Figure 9 before Figures 7 and 8. Consider moving this part later in the manuscript.

We agree with this comment. The initial objective was to highlight the low uncertainty in the estimates of the Shapley effects, but this is only visible in Figure 9. To avoid referring to this figure, we propose to provide a new Table (Table 1 in Sect. 4.3) with the error estimates for the base case.

Table 1. Shapley effects relative to the occurrence of the event $\{m_Y > Y_C = 2,000\text{m}^3\}$ given RCP4.5 scenario, estimated by computing the median value from B=50 replicates of the estimation procedure (Sect. 3.5) accounting for GP and sampling error. The numbers in brackets correspond to the minimum and maximum value computed from the estimation procedure.

Year	SLR	SWL	Hs	Tp	Dp	U	Du
2050	0.143 [0.126, 0.160]	0.319 [0.293, 0.344]	0.110 [0.095, 0.125]	0.102 [0.086, 0.119]	0.104 [0.095, 0.115]	0.119 [0.104,0.130]	0.104 [0.091, 0.112]
2100	0.410 [0.398, 0.420]	0.279 [0.270,0.291]	0.082 [0.078, 0.085]	0.045 [0.043, 0.047]	0.047 [0.044, 0.051]	0.091 [0.086, 0.096]	0.047 [0.044, 0.050]
2200	0.542 [0.536, 0.550]	0.282 [0.277, 0.286]	0.050 [0.048, 0.053]	0.020 [0.019, 0.022]	0.025 [0.022, 0.027]	0.056 [0.053, 0.059]	0.024 [0.022, 0.025]

Line 360-366: As in Line 266-270, this is general methodology. Consider defining the base case scenario and the alternative scenarios to test sensitivity on different assumptions under the methodological sections.
Thank you for this suggestion. Given the comments of the two other reviewers, we preferably keep the original structure as such, because we believe that the readability is improved when the test sensitivity on different assumptions (originally described in Sect. 4) is discussed directly in the new Sect. 5 "discussion".

Line 428. Do you men Figure 8a,b?
Thank you for noticing this problem. This is now corrected.

Orleans,
March 28th, 2022
J. Rohmer[1] on behalf of the co-authors

[1] BRGM, 3 av. C. Guillemin - 45060 Orléans Cedex 2 – France