Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/nhess-2022-74-RC2, 2022

Comment on nhess-2022-74

Anonymous Referee #2

Referee comment on "Multilevel multifidelity Monte Carlo methods for assessing uncertainty in coastal flooding" by Mariana C. A. Clare et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-74-RC2, 2022

In all, the methodology, as presented, seems to be more adequate for conducting uncertainty quantification analysis and not for coastal flood risk assessment. Not much evidence is presented in the paper supporting the usefulness of the methodology in risk assessment. The estimation of cumulative distribution functions (CDFs) is just one piece of a proper risk assessment. In areas like Myrtle Beach, SC, and throughout most of the U.S. Atlantic and Gulf coasts, risk assessment must be concerned with understanding and characterizing coastal hazards (e.g., storm surge, waves, tides) due to different storm populations (e.g., tropical cyclones, extratropical storms), and must rely on some form of extreme value analysis. None of these elements are present in this paper.

A methodology developed for uncertainty analysis is not necessarily transferable to coastal hazard analysis or risk assessment, therefore, I would consider revising the title of the paper as it might be inadvertently misleading.

Although my background and experience encompass both probabilistic hazard analysis and hydrodynamic modeling, I found this paper to be quite difficult to follow. The abstract states: "Here, we apply the multilevel multi-fidelity Monte Carlo method (MLMF) to quantify uncertainty by computing statistical estimators of key output variables with respect to uncertain inputs,…"; but there is no discussion about how are these uncertain inputs identified or prioritized. The three cases presented in the manuscript, including 2D real-world case, are highly idealized and seem to consider only one uncertain input per case; this is, Manning's coefficient, beach slope, and offshore water level, respectively. In real-world applications, rarely there is just one uncertain input parameter.

The paper discusses how to consider multiple output locations, but is the methodology applicable when there are multiple uncertain input parameters?

The initial case (presumably Level 0) does not seem to be well defined. Equation #14 is used to determine the optimal number of samples, but how can the initial number of samples be estimated without prior knowledge of key output variance and the input-to-output relationship?

Also, there is no mention of other approaches that are used for similar purposes, this is, to determine the optimal number of events to be subsequently simulated using high fidelity models. Such approaches could leverage methods like Latin hypercube sampling, genetic algorithms, joint probability methods, and even recent machine learning techniques that directly account for input-output relationships. It’s difficult to judge the benefits of the proposed methodology without a discussion, at least conceptually, of some of these other approaches.

Specific comments:

Several numerical models are introduced in the abstract and the manuscript introduction without defining the name (or acronym); e.g., SFINCS is not defined (Super-Fast INundation of CoastS) until the third time that the model is mentioned.

Figure 1 mentions “Level 1”, but the concept of what a level constitutes in this methodology has not been established at this stage of the manuscript. Also, in Figure 1, is the Euro symbol meant to describe typical computational costs or time associated with the different levels of fidelity and number of samples? It might help to provide additional context. In terms of order of magnitude, is each “Euro symbol” representing hours, weeks, or months?