

Review on the manuscript "Advancing Nearshore and Onshore Tsunami Hazard Approximation with Machine Learning Surrogates"

This manuscript introduces an approach in tsunami hazard and risk assessment through the use of machine learning (ML) surrogates. While the manuscript has been clearly improved since earlier versions, further revisions are required before it can be considered for publication.

Major comments:

While the manuscript is well-organized overall, some sections, particularly "Discussion and Conclusions", are overly detailed and repetitive. For instance, the discussion often reiterates technical results already covered in previous sections.

L437 "The general challenge with the use of such neural network surrogates for use in PTHA is that they have not been comprehensively validated if they generalise well beyond the specific type of events on which they are trained, such as events with different source mechanisms or when considering tsunami generated from multiple source regions for PTHA."

This sentence was unclear to me. I am not sure what you are trying to convey, and you may my suggestion if this looks fine:

"A challenge of using neural network surrogates in PTHA is their limited validation for generalizing beyond the specific event types used during training. For instance, their performance on events with different regions or multiple sources lacks sufficient validation."

L441 "The emphasis is on accurate predictions for larger magnitudes, which is crucial for early warning purposes. However, for the surrogate to serve as an effective hazard approximation in PTHA, it must accurately represent both large and small magnitude events despite limited training data."

This part also needs to be rewritten. The authors are trying to discuss the imbalance of the training data due to the significance of large events. However the meaning of the following sentence "it must accurately ..." is unclear.

L444 Remove "extend upon previous work and"

L448 "in a general setting and the influence ..." This is redundancy.

L485 "events, recognizing" -> "events. Recognizing"

L506 "compute time" -> "computation time" ?

The application and methodology appear to be similar to prior works, and it is unclear what is the significant expansion in this study. It is necessary to clearly state in the conclusion how this study represents an advancement over previous studies.

Comments:

Consider using a more specific description. For example, in abstract, the authors use "at large regional scale" and "large portions of the coast". This expression may be unclear to the readers.

L28 “select”->”limited”

Each subfigure in Figure 16 uses a different y-axis range, which makes it difficult to compare results. Consider using the same y-axis range or stating why the ranges differ.