Responses by authors in blue:

We thank the referee and editor for taking their time to provide valuable comments and suggestions to further revise the manuscript. Here, we provide responses to their questions and comments. Blue text shows our response to the referee comments, updates which will be incorporated in the revised manuscript are highlighted as track change and black text shows the referee comments.

Review by Referee

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.

We find that summarizing the main point of each test set is very useful for framing the discussion; however, we appreciate the point that we include too many technical details in the "discussion and conclusions" section. We tried to reduce the technical details to broad statements in order to make the discussion more easily readable in the revised manuscript.

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."

This has been updated in the revised first paragraph of "Discussion and Conclusions" and reads as below.

A key challenge in using neural network surrogates for probabilistic tsunami hazard analysis (PTHA) is their limited validation in generalising beyond the specific event types used during training. For instance, their performance on events from different regions, involving multiple sources and mechanisms, has not been sufficiently validated in previous studies. By testing ML surrogates on a diverse set of events, this research broadens their applicability and makes a significant contribution to validating the generalisation capabilities of ML surrogates for PTHA. Furthermore, training datasets often require thousands of events from each source region to fully capture the inherent variability. In many cases, previous studies using ML have prioritised accurate predictions for larger magnitude events, which are crucial for early warning systems. However, for a comprehensive PTHA, it is equally important that surrogates can predict both large and small magnitude events with high accuracy. Recognising this, we designed two specialised ML surrogates to address the distinct challenges of approximating nearshore and onshore tsunami hazards that help offset the related computational costs, while overcoming the limitations of imbalanced and limited training datasets. The nearshore surrogate predicts the time history of tsunami wave at the shore, while the onshore surrogate predicts the inundation depths across vast locations overland. The hybrid ensemble approach introduced here leverages model and parameter sensitivity to enhance prediction accuracy, marking a step forward in integrating uncertainty quantification into ML-based PTHA. This expanded capability, which has been under explored in prior research, is essential for extending tsunami hazard and risk assessment.

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.
This has been updated in the revised paragraph above.

L444 Remove "extend upon previous work and"
Removed.

L448 "in a general setting and the influence …" This is redundancy. L485 "events, recognizing" ->"events. Recognizing" L506 "compute time" -> "computation time"?
Corrected.

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.

The updated paragraph above summaries the main advancements in the study, all addressing the applicability of ML surrogates for PTHA and PTRA

- Validation and generalisation testing – for broader applicability needed in PTHA.
- Handling imbalanced and limited training information – overcoming the challenges of sparse training data.
- Modelling two tsunami hazard measure – wave time history and max inundation.
- Uncertainty quantification from the surrogates.

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.

We have updated the abstract to clearly mention it now reads as below:

Probabilistic tsunami hazard and risk assessment (PTHA and PTRA) are vital methodologies for computing tsunami risk and prompt measures to mitigate impacts. However, their application across extensive coastlines, spanning hundreds to thousands of kilometres, is limited by the computational costs of numerically intensive simulations. These simulations often require advanced computational resources like high-performance computing (HPC), and may yet necessitate reductions in resolution, fewer modelled scenarios, or use of simpler approximation schemes. To address these challenges, it is crucial to develop concepts and algorithms for reducing the number of events simulated and more efficiently approximate the needed simulation results. The case study presented herein, for a coastal region of Tohoku, Japan, utilises a limited number of tsunami simulations from submarine earthquakes along the subduction interface to build a wave propagation and inundation database. These simulation results are fit using a machine learning (ML) based variational encoder-decoder model. The ML model serves as a surrogate, predicting the tsunami waveform on the coast and the maximum inundation depths onshore at the different test sites. The performance of the surrogate models was assessed using a five-fold cross-validation assessment across the simulation events. Further, to understand their real world performance and generalisability, we benchmarked the ML surrogates against five distinct tsunami source models from the literature for historic events. Our results found the ML surrogate capable of approximating tsunami hazard on the coast and overland, using limited inputs at deep offshore locations and showcase their potential in efficient PTHA and PTRA.

L28 "select"->"limited"

Thanks, corrected.

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.

Thanks for suggesting this, updated the figures as below: