Comment on nhess-2021-63

Anonymous Referee #1

Referee comment on "Probabilistic, high-resolution tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation" by Dimitra M. Salmanidou et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-63-RC1, 2021

Referee Report for

Title: Probabilistic, high-resolution tsunami predictions in North Cascadia by exploiting sequential design for efficient emulation
Author(s): Dimitra M. Salmanidou et al.
MS No.: nhess-2021-63
MS type: Research article

GENERAL COMMENTS

Dear authors,

This study implements and describes the results of the application of an emulator for the modelling of tsunamis generated by co-seismic seafloor displacement, which produces a very satisfying trade-off between accuracy and computational cost. This approach can be advantageous for tsunami hazard analysis typically requiring computationally intensive exploration: 1) of the expected natural variability (the aleatory uncertainty), and 2) of different yet still credible modelling choices consistent with - limited - observations (the epistemic uncertainty).

I consider the technical topic dealt with here of great relevance. Saving computational resources allows in principle their more efficient usage to quantify and even reduce uncertainties through a smarter exploration of the parameter space characterising the natural phenomena under scrutiny.

However, I see at least three issues that need to be addressed:

- the limited originality or the need for a better framing with respect to the state of the art;
- the weakness of some elements of the modelling approach;
- the overstatement of the focus/message.

For this reason, I suggest that the paper undergoes major revisions before being published.

In what follows, I explain the three issues listed above.

1) Originality.

It is not completely clear to me how much this paper presents a definite advancement over previous work, mostly by some of the authors of this study.

The authors state at lines 9-10 in the Abstract: "This approach allows for a first emulation-accelerated computation of probabilistic tsunami hazard in the region of the city of Victoria, British Columbia."

Actually, the one presented is far from being a probabilistic tsunami hazard assessment, as detailed in point 3) below.

Nevertheless, this might still be a methodological paper regarding some aspects of the tsunami hazard assessment workflow.

However, the authors - or some of them I'm not sure - already beautifully clarified the potential impact of their emulator in combination with the VOLNA simulator.

In fact, they say (lines 44-45): "Statistical emulators (also known as a statistical surrogate models) can be called to address these issues (Behrens and Dias, 2015)."

And also: (lines 46-48): "Such approaches have been implemented for uncertainty quantification of tsunami hazard at various settings (Sraj et al., 2014; Salmanidou et al., 2017, 2019; Guillas et al., 2018; Denamiel et al., 2019; Snelling et al., 2020; Gopinathan et al., 2021).

Then, I'm struggling to understand the real advancement here. I see this work mostly as a kind of reshuffling of previous methods, for example, the one used in Gopinathan et al. (2021), which relies on Beck and Guillas (2016), or the one applied by Giles et al. (2021, https://doi.org/10.3389/feart.2020.597865), not cited here, which uses the Gaussian Process emulator code but not, if I get it right, the adaptive sequential design.

But maybe there are technical details that really represent a leap ahead and just deserve to be better highlighted.

So, let me assume for a moment that this is not the case and the novelty brought by this study can be shown beyond any reasonable doubt: even so,

- the description of the relationship of the present work with the existing literature should be improved (see specific comments);
- the advantage over the previously applied techniques should be clearly demonstrated by means of the use-case, it cannot be only declared based on the author's experience.

Conversely, if a novel significant methodological development was not achieved, and the study is only the application of an existing yet pretty sound technique to a slightly different use-case, the paper may lack enough scientific significance.

Please, address this first issue with great care.

2) Modelling.

Even if, overall, the approach is certainly up to the best international standards, and probably even pushing the limit a bit beyond, it remains very weak if not insufficient in certain parts and aspects of the proposed workflow.

The approach to the modelling of the seafloor co-seismic displacement is too simplistic and completely subjective. There's a bunch of methods out there, some of which have been used for decades now, ranging from very simple modelling techniques to much more

sophisticated ones.

For example, Gopinathan et al. (2021), with one of the coauthors of this study, apply a much more realistic approach.

In general, the different degree of complexity used by these methods may be chosen to match the "amount" of underlying physics, geology, and available constraints, for example deriving from the understanding of the regional seismotectonics and/or seismic and tsunami history.

As a term of comparison, among others, a couple of quite popular reviews focussing on seismic source modelling for tsunami application are

Geist, E. L., and Oglesby, D. D. (2014). Earthquake mechanism and seafloor deformation for tsunami generation, in Encyclopedia of earthquake engineering. Editors M. Beer, I. A. Kougioumtzoglou, E. Patelli, and I. S.-K. Au (Berlin, Heidelberg: Springer), 1–17. doi:10.1007/978-3-642-36197-5_296-1

Geist, E. L., Oglesby, D. D., and Ryan, K. J. (2019). "Tsunamis: stochastic models of occurrence and generation mechanisms," in Encyclopedia of complexity and systems science," Editor R. A. Meyers (Berlin, Heidelberg: Springer), 1–30. doi:10.1007/978-3-642-27737-5_595-2

There are two options that I foresee to improve the current situation. Either you make your choice and revise the modelling approach, or at least state clearly up front that you focus on different aspects than the earthquake modelling in this paper, and as a consequence prefer using a very simplistic approach for illustrative purposes only. It should also be stated very clearly that this approach should be replaced by a more realistic one for any real application.

Yet, how much the (over-)simplifications introduced here affect the characteristics of the resulting tsunamis remains to be addressed. Wave features in fact interact with bathymetry and topography features during tsunami evolution and they can strongly influence the inundation characteristics and extent. Different features of the waves may also result in a different performance of the emulator. These aspects should be discussed upfront in the paper as caveats for the readers to help them understand the real limitations of this approach.

3) Focus/Message: this is not a probabilistic hazard assessment, rather it is an illustration of a method to deal with a component potentially useful for a hazard assessment.

The title and some descriptions in this paper may be confusing since one may think it deals with probabilistic tsunami hazard analysis, which is not true in my opinion. This paper, as stated at the very beginning of this report, deals with a technique for reducing computational cost. This technique is potentially applicable for hazard analysis. So, this ambiguity and some related overstatements should be fixed.

Let me try and clarify this concept. Can the emulation (or simulation) of an arbitrarily chosen set of scenarios be called "probabilistic tsunami hazard" (as for example in the abstract, at line 9)? What is a "probabilistic tsunami hazard"?

In its most common acceptance, a probabilistic tsunami hazard analysis provides the probability of exceedance in a given time interval of different thresholds of the chosen hazard intensity at a specific location. In general, this is made in the following way. Even if limited only to tsunamis generated by earthquakes, an attempt is typically made to take into account the full range of earthquake (parameter) variability in the hazard assessment; then, a model of their temporal occurrence, combined with the effect of each modelled tsunami scenario, allows the tsunami probability to be estimated. Otherwise, modelling a subjectively chosen range of scenarios, maybe similar to some historical events, is a what-if experiment, addressing the consequences of the hypothesised set.

On the contrary, in this study, a set of scenarios is quite arbitrarily chosen; then a beta distribution is assigned from which the scenarios to be modelled are sampled (Fig. 11). It is correctly stated that "The shape parameters of the distributions can be utilised to express the scientific knowledge on the source" (lines 254-255), but no effort is made to link the parameters of this distribution to reality, except for a generic correspondence of the maximum values of the uplift-subsidence to those estimated for historical events or previous studies. This is a scenario analysis, based on a quite generic parameter setting, not a full probabilistic hazard assessment. This is perfectly fine for illustrative purposes of the potential usage of the emulation technique, provided that things are correctly framed.

So, please, improve the description throughout the entire manuscript to resolve this ambiguity, by eliminating any attempt to call it probabilistic hazard analysis.

Many additional comments related to the 3 points above are described in the specific comments in the following of this report.

SPECIFIC COMMENTS

Line 2:
"Traditional"? what do you mean by traditional? That the PTHAs for Cascadia have either

low-resolution or use few scenarios? or that this is true all over the world?
I'm aware of several studies combining many scenarios at high resolution. Please, make a better survey of the relevant scientific literature.

Line 2:
Hazard curves, perhaps, not "hazard maps"; the PTHA produces hazard curves, from which the hazard maps can be extracted.

Line 3:
By "cost", do you mean: "computational cost"?
Actually, as stated above, there are several recent studies using many scenarios at high resolution. Mostly limited to one specific site indeed.
So, there is at least a third variable to consider, further than the range of scenarios, and resolution: the extent of the target coastal stretch.
Having said that, it is always a good idea to save resources, to be able then to expand in the right direction if needed.
However, it should be also mentioned that there are several recent approaches to deal with an optimal trade-off between cost and quality of the analysis.
So, this whole sentence should be reconsidered. This comment is connected to another one below on the need for citing relevant literature in the Introduction.

Line 9:
As already explained above, this cannot be considered a probabilistic tsunami hazard analysis.

Lines 30-35:
I suggest mentioning the global hazard assessment of Davies et al. and the Seaside hazard assessment of Gonzalez et al., both addressing PTHA at different resolution for the same target zone.

Lines 39-41:
It would be appropriate to mention, discuss, and cite the different existing techniques aimed at computational resource optimisation such as the class of methods based on the similarity between scenarios (e.g. offshore wave matching, cluster analysis), on importance sampling, and most recently on deep learning.

Line 42:
I believe that it is not appropriate to state that high-resolution studies are necessarily needed for coastal planning. There are several meaningful - in my opinion - strategies based on lower resolution assessments. They are either approximations or stochastic treatments, or both, of the inundation probability, which are being used for evacuation or long-term coastal planning in different countries. Please, rephrase.

Line 44:
About how "large" and "unaffordable", is it possible to provide some numbers? Otherwise, this would remain quite vague. What about using some references here?

Line 45:
Probably, also Sarri et al. (2012) can be cited in this context.

Lines 47-48:
Also, Giles et al. (2021) can be probably cited here. However, as said above, the progress with respect to this list of papers should be clarified.

Line 49:
For the sake of self-consistency, a short intro on what emulators are and how they work, and the role of training data would be useful.

Line 52:
This comment on interpolation/extrapolation is crucial in my opinion, can you quantitatively prove your statement or at least qualitatively describe the reason in some detail?

Lines 63-64:
Please add a reference for MOGP

Lines 90-125:
This Section is inadequate for a real PTHA, as already explained.
The natural variability of earthquakes as currently understood is not represented. Nor are their temporal occurrence features.
The ones used here are kind of toy scenarios for tsunami initial conditions, and this makes the present study a "what-if" hazard assessment, which can have its own importance depending on the context, but it is not a full PTHA at all, for which among the most important aleatory variables are those describing the earthquake kinematics and/or mechanics, and the characteristics of the seismic cycle, along with the analysis of the epistemic uncertainty related to our understanding and to the (lack of) data available regarding the seismicity.
Your approach in my understanding is as follows: define an arbitrary shape related to a vague notion of how a co-seismic displacement profile would look like, parameterise it and vary these parameters within the ranges inferred by the characteristics of a few past earthquakes.
Please, use modern geology, physics and seismology, plus all available observations to constrain and validate your model, making it hopefully more realistic, while quantitively dealing with the large uncertainty involved.
Data and deep knowledge of the phenomena may help to limit subjectivity.
Alternatively, you may say that despite you are not dealing with PTHA and you are not using a state-of-the-art probabilistic earthquake model, you are proposing a method to

deal with a specific segment of the analysis, aimed at reduction of the computational cost of the scenario simulation phase.

Line 94:
These by Stake and Wang are not two different methods for representing the seabed displacement.
The first one has to do with the reconstruction of a single event. The second one deals with the interseismic period, in turn indirectly linked to the slip, which causes the displacement. Unclear why both of them are indicated here as paradigmatic examples of the tsunamigenic displacement on this subduction zone.

Line 95:
The function used is not smooth, as I understand it.
Moreover, it is not completely clear how the full-length "rupture" is built and what's the temporal history of this "rupture".

Lines 125-128:
A vast scientific literature exists on empirical, numerical and theoretical earthquake scaling relations, that cannot be ignored. How were these parameterisation and ranges chosen?

Line 145:
Please explain what hyperparameters are.

Line 146:
Please, clearly explain, at least qualitatively, the training procedure at least once throughout the paper. I understand that it is about finding outputs for unexplored inputs because of the computational cost. It must be clearly specified that as such it is a mathematical interpolation(?) based on some assumptions regarding the nature of the uncertainty, but no data were used for calibration purposes? If this is the case, this must be clearly specified as a limitation of the method quite upfront in this paper. It seems to me that quite naive initial conditions and only results of simulation (not real data) are used for calibration. Regarding the data, this is understandable, due to tsunami data scarcity. BTW, has this emulator approach ever been trained/calibrated using tsunami observations? If yes, please state it clearly, because, under a kind of ergodic hypothesis, you might assume that training elsewhere would make the usage here more trustable. Moreover, why should the emulators be independent of each other? Aren't they correlated to a degree that depends on their mutual location in the parameter space? Could you please clarify and explain better in the text? Probably, I don't understand completely because I'm not an expert on this specific methodology but, please, make an effort to make your descriptions more accessible to the NHESS general reader.

Line 237:
It looks like the predicted response is not defined positive as the maximum elevation. Am

I wrong? Can't this be corrected?

Line 238:
Underestimation (or even a bias) is generally considered worrisome in the context of the hazard assessment, particularly when conservatism is desired at a later stage by decision-makers.
Can you comment on this aspect?

Lines 271-272:
It is not clear at all what is the physical meaning of this probability. It looks like this just the probability of the impact conditional to a subjectively defined source variability modulated by a beta distribution that "can be utilised to express the scientific knowledge on the source".
What is the relation of this "probabilistic" source model to the expected future seismicity of the region? Without clarifying this, it is very difficult to say what is the relation of this "probabilistic hazard map" with the estimate of future tsunami probability, which is the core objective of hazard assessment.
It is then hard to believe that these can be defined as "The probabilistic hazard maps in the South-East part of Vancouver island".
The statement at lines 271-272 is one of those motivating comment 3) at the beginning of this report.
Like it is, this is an interesting tool for reducing computational cost. Nothing else. And the one presented is just an illustrative example with purposely oversimplified assumptions.

Lines 303-304:
Here many of the concerns just expressed are even adopted by the authors. This sounds like these histograms can be translated into probabilities with some meaning.
Yet, it should be said that often also data (e.g. seismic catalogues) are used to constrain probabilities, not only expert judgement.

Line 321:
I doubt this can be judged a "real-case hazard prediction", for the reasons already expressed.

Line 327:
Same as above: "to produce probabilistic hazard maps that assess the tsunami potential in the area".

Lines 328-330
So, we are ready for starting coastal planning now? BTW, any comparison/validation with respect to other studies?
Or this sentences mean that a totally different approach would be needed for that? Please, clarify.
BTW, the probability of something happening in the future should be referred to a given

time interval.

Line 340-345:
It looks like we at least partly agree, eventually: "As so, there are some aspects that need to be considered in future work to further refine the probabilistic outputs. These span from the tsunami generation to the inundation.". This is not enough in my opinion, for the reasons that were already explained several times.
Yet, no reference to the temporal behaviour of earthquakes is made, which is necessary to address the hazard.

One last comment regards the (lack of) Discussion and the Conclusion.
The limitations of the approach are not deeply investigated, or at least their analysis not completely disclosed. First of all, it should be better clarified how much the approach may suffer from the lack of experimental data for calibration. However, this is a common issue for all the hazard assessment methods heavily based on simulations. Nevertheless, I'd like to see what would happen when trying to predict inundation and whether and to what extent nonlinearity would challenge the emulator. Moreover, no attention is paid to addressing the uncertainty related to the simulator itself - simulations are unfortunately always presented without uncertainty, we all do that - but the issue should be at least mentioned in this context. Several methods exist relying on the sensitivity of the simulators for example to the input variations. Last, how much the uncertainty on the bathymetry would affect the results?

Technical Corrections

Figure 1:
Figure 1 is never invoked. Please check also the ordering of the other Figures, for example, Figure 8 is called out before previous ones. Check all of them.

Line 93:
Varying —> Various?

Line 153:
Add a comma "," after 2008 or remove the next one after "criterion".

Line 167:
Please add a link and/or reference for Gmsh.

Figure 5:
Where is the 0 of the profiles? Show it in the Figure or use a classic AA' BB' CC' notation

for the 3 profiles and adjust the axes of the Figures to the right accordingly.

Line 205:
burried —> buried

Line 330:
"would particularly" —> "would be particularly important"? Please check this sentence.