Referee Report #2

General comments

This study provides the application of the statistical surrogate models to predict tsunami hazards from earthquakes at CSZ. The study employs the emulation approach to quantify the tsunami hazard, particularly the maximum onshore tsunami height at Victoria, British Columbia, considering varied conditions of seabed displacement over CSZ. Overall, the formats and writing are acceptable, but I have major comments on the novelty of this work and descriptions, including the model validation, fitting process, and conclusions. Therefore, I recommend revising the manuscript and resubmit for publication.

We would like to thank the reviewer for their thorough comments and suggestions to improve the manuscript. We hope that our point-by-point answers and the modifications on the manuscript will clarify and improve the aspects of our work that need further addressing.

Specific comments

1. For the novelty of this work.

The application of the emulator for tsunami hazard modeling in CSZ tsunami, particularly in Victoria, British Columbia, is already published by the third author (Guillas et al., 2018). It is indeed the first time to check the maximum onshore tsunami height using the surrogate model, as noted by the authors. Still, the model domain seems quite similar to the previous work. The fundamentals are already provided from other references except showing the surrogate model results at a specific area in British Columbia.

I think, the authors could earn more novelty from chapter 4, which is the model validation and fitting. Still, the process is relatively limited and needed to improve to get a novelty. Here are the related comments below.

The novelty of this work expands on the methodological aspects. The use of the adaptive sequential design algorithm by Beck and Guillas (2016) is implemented for the first time to construct the GP emulators towards a realistic tsunami hazard. This was not done in Guillas et. al. (2018). This allows the use of high resolution (30 m) over a coarser one (in Guillas et al.,2018) also through the use of multi-output GPs. To improve the novelty of the manuscript, we also associate our predictions with an occurrence exceedance probability. To our best knowledge, this link is implemented for the first time. Please also see our reply to reviewer #1 (also copied below) regarding this comment:

The novelty here is the use of the sequential design MICE by Beck and Guillas for the construction of the GP emulator of the tsunami model. This is done for the very first time towards a realistic case using High Performance Computing. Gopinathan et al (2021) and Giles et al. (2021) (now also cited in the revised manuscript) used a one shot random sampling for the training which lacks the information gain achieved by sequential design. Concretely, sequential design can reduce by 50% the computational cost, as demonstrated in (Beck, Guillas, 2016) for a set of toy problems, so applying the approach towards a realistic case is showcasing real benefits in the case of high resolution with a complex parametrization of the source and is new. Note, that Guillas et al. (2018) was neither using high resolution nor sequential design and the source was much simpler. The relationship of this work with existing literature is given below in the answers to specific comments.

2. For model “initial validation” (Line 195).

In general, it is hard to validate numerical tsunami model results, including the generation and propagation process, due to the lack of real observed data of CSZ events. The authors conclude that the maximum water elevation at a specific point shows good agreement with Fine et al. (2018) and other references and justify the current approach. However, the current comparisons are not clear and need to be improved. To specific, the authors pick scenario 24 for the validation, which shows the best agreement(?) to others works. If scenario 24 is chosen, it is better to show more detailed comparison results using figures or tables. The current comparison results describe the match at a particular point, but the authors need to show a spatial (map) comparison to justify the model to others. Also, it is better to provide detailed tsunami generation conditions of each reference and justify why the author chose scenario 24.

We agree. We clarify better in the manuscript:
“Scenario 24 is selected as it is the first scenario in our list of scenarios that has a maximum deformation of ca. 4m, similar to the maximum uplift in numerical studies of the event.”
“The maximum uplift was used as a guideline for this comparison due to its significant contribution to the tsunami excitation. As the experimental setting is controlled by MICE, the rest of the parameter values of scenario 24 do not necessarily match with the values of other numerical studies. For example, the maximum subsidence of scenario 24 is selected to be around 1.27 m as opposed to 2 m in the buried rupture model by Fine et al. 2018. This causes some discrepancies in the wave signal, the degree of which is not calculated since the scope of this comparison is to do an initial validation of our modelling as opposed to a reproduction of the currently existing work”.
We provide snapshots and spatial maps of scenario 24 but the exact locations of other works vary between them and are not always given. We added the following table of the results at the mouth of Victoria Harbour in the manuscript:

Study	Uplift (m)	Subsidence (m)	Approximate arrival time of Wave Trough (minutes)	Approximate arrival time of Wave Crest (minutes)	Approximate wave trough (m)	Approximate wave crest (m)
scenario 24	4.09	1.27	50	100	0.2	1.7-1.8
AECOM 2013	6.2	2.3	Not found	96	1-1.05	2.4-2.5
Cherniwasky et al., 2007	Not found, Mw 9	Not found, Mw 9	50	90	0.5-0.6	1.9-2.2
Fine et al 2018	4	2-2.5	52	88	0.96	1.6-1.7

3. For model “Fitting” (line 225)
a) As I understand, the emulator was trained from the single point in Fig. 8 (star). Need some explains why chose this point and any sensitivities on training and results.
The point in Fig. 8 (now Fig. 4 of the document) was selected to drive the design algorithm for the selection of the experiments. We add in the text: “This design location was selected as it provides variance in the response driven by each scenario and it is close to the centre of the region of impact. As it directs the sequential design there is some sensitivity of the design to this point, but in our opinion not that large as another point in the region would yield similar results since the influence of the parameters on impact points varies but not much. Furthermore, small variations in the design of experiments obtained have little influence on the construction of the emulator, but an agnostic one-shot design would greatly

differ from any of the sequential designs obtained by our approach and be much less efficient as it would ignore completely the concrete influence of the inputs on outputs to efficiently design the computer experiment. ”

b) Figure 9 shows the performance of the prediction from the emulator at three different points as the authors noted that the surrogate model couldn’t capture the pattern well at location 2 but show good agreement at location 3. Can authors explain why they are so different?

We add:
“As the waves propagate on land, the prediction becomes more challenging due to even the slightest variations caused by the surrounding topography. The sensitivity of the locations to the variance in the scenarios also plays a significant role. Location 2, for example, does not show large sensitivity to the variation of the parameters, the maximum elevation is close to zero in all of the cases. Location 3 is closer to the source and is experiencing the highest wave run-up and it is, therefore, less affected by these slight variations in the topography but more sensitive to the varying scenarios.”

c) Due to the small number of comparisons and fluctuation, those three points are not enough to represent the overall performance of the emulator. It is questionable that location 001 (I think it is location 01) could represent the general pattern of RMSE error of the emulator index (Fig. 9d). It is hard to conclude that 50 and 60 are good enough for your hazard results in Fig. 12. It is recommended to show more comparison results (somewhat similar to Fig. 9a,b,c but more efficiently) at different locations. We can observe a spatial variation of wave height at the shoreline in Fig. 12. The authors may consider checking the variance of RMSE for the emulator index at different locations.

Indeed, Location 1 is not in a general sense representing the RMSE error at all locations. The 3 locations are used for illustrative purposes, in order to demonstrate the emulators’ behavior at some points. Showing more locations might not necessarily add to the manuscript’s content as they may exhibit similar behavior, however the relative error might increase at locations further inland. This is related to the previous comment, and we also add:

“The RMSE is computed at these three locations for illustrating the behavior of the emulator’s predictions at certain points, the relative error might increase further inland at inundated locations.”

and also: “GP emulation is well suited for approximating nonlinear simulation behaviors, even when considering continuous outputs of low regularity and when restricted to small-sized experimental designs with space-filling properties. As shown by Owen et al. (2017), when considering two cases with computationally intensive simulators, more specifically, a land-surface simulator and a launch vehicle controller, GP emulation demonstrates good approximation properties even for small design sizes. By small design sizes, we refer to designs with the number of samples being about ten times the number of input parameters, a widely used rule of thumb for effective computer experiment design (Loeppky et al., 2009).”

Owen et al. (2017) “Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators. SIAM/ASA Journal on Uncertainty Quantification, 5(1): 403-435.

Loeppky et al. (2009) “Choosing the sample size of a computer experiment: A practical guide”, Technometrics, 51: 366-376.

4. About Probabilistic tsunami hazard,
One of the conclusions (at line 326) is that the emulator allows probabilistic hazard maps. Fig. 11 and 12 are not a typical probabilistic hazard map, which is not commonly accepted. The probabilistic tsunami hazard map provides the map (spatial distributions) of tsunami intensity measures (e.g., maximum tsunami flow depth) at a specific recurrence interval such as 500 yr, 1,000 yr, and 2,500 yr through a probabilistic tsunami hazard analysis (PTHA). The authors need to clarify that the current work from the typical probabilistic hazard map. Also, it is hard to agree that the current work is a kind of alternative format of hazard map as described in line 328. It is misleading to readers.
We agree that this is an important issue that needs to be addressed. Please see our response to the reviewer #1 below in an attempt to address this issue.

We agree and we refrained from using the explicit term PTHA for our work. The emulators can give the probability of a tsunami intensity measure to exceed a certain threshold given predefined metrics at any horizon. We now provide a correspondence between predictions and time horizons. We make a first attempt here, which is simplistic but we believe adds to the novelty of the manuscript as it is, to our knowledge, the first to incorporate emulators for probabilistic tsunami hazard (as the term is currently accepted). The following section will be added and will be discussed in more detail in the revised manuscript:

"Further, we associate the scenarios and predictions of the H2 distribution with probability of exceedance. We study the pattern of 1/1500 year exceedance rate wave heights over the region drawing elements from the process followed in Park et al. (2017). The probability of a full-rupture generated tsunami is considered, the earthquake magnitudes associated with such an event are in the ranges of Mw=8.7-9.3 (Goldfinger et al., 2012). We link the moment magnitude for each event with the maximum seabed uplift caused by the linearly decaying slip on the rupture surface similar to the Cascadia subduction interface, a simple linear solution, first introduced by Okada (1985) is used. Following this approach, the magnitudes of the H2 scenarios range between 8.77-9.28. To associate frequency of events with earthquake magnitudes, a Tapered Gutenberg–Richter (TGR) distribution is utilized which has been proved to give adequate predictions for the Cascadia subduction zone (Rong et al., 2014). The TGR Complementary Cumulative Distribution function for a given earthquake magnitude m can be computed as:

$$F(m) = \left[10^{1.5(m_t - m)}\right]^{\beta} exp\left[10^{1.5(m_t - m_c)} - 10^{1.5(m - m_c)}\right],$$

where m_t is a threshold magnitude above which the catalogue is assumed to be complete (here $m_t = 6.0$), m_c is the corner magnitude and β is the index parameter of the distribution. Considering the 10,000-year paleoseismic record, as reconstructed by Goldfinger et al. (2012) from turbidite data, m_c and β take values of 9.02 and 0.59 respectively. The discrete number of m_j magnitudes can also be computed by:

$$P\left[M = m_j\right] = G\left(m_j + 0.5\Delta m\right) - G\left(m_j - 0.5\Delta m\right),$$

where $G(m) = 1 - F(m)$ denotes the cumulative density function and Δm is the discretization interval. Following the above, the mean annual rate of exceedance and the number of earthquakes in 200K years can be estimated as shown in Figure 13a. These relationships can then be used to associate the probability of exceedance for each event of the event set H2 by computing the relative frequency of each event within the data (Fig.13b). The occurrence exceedance probability for the hazard values $h_1 > h_2 > \cdots > h_n$, with $n = 2000$ becomes $P_{ex}(h_1) = 0$ for h_1 (corresponding magnitude is equal to 9.3), $P_{ex}(h_2) = f_1$, and for the other hazard values can be computed as:

$$P_{ex}(h_{i+1}) = 1 - (1 - P_{ex}(h_i))(1 - f_i) \text{ for } i = 2,3 \ldots, n,$$

where f_i is the event frequency assigned according to the magnitude bin to which the event was linked by its maximum uplift. The above relation comes from basic probability theory and is often used e.g. in Monte Carlo probabilistic seismic hazard assessment (Musson, 2000). Accordingly, the mean annual exceedance rate can be computed for the hazard values at each location (Fig. 13c). Considering then a 1/1500 exceedance rate, the hazard curves of each location can be used to construct the hazard maps of Figure 13d which represent the maximum wave heights from the H2 events.

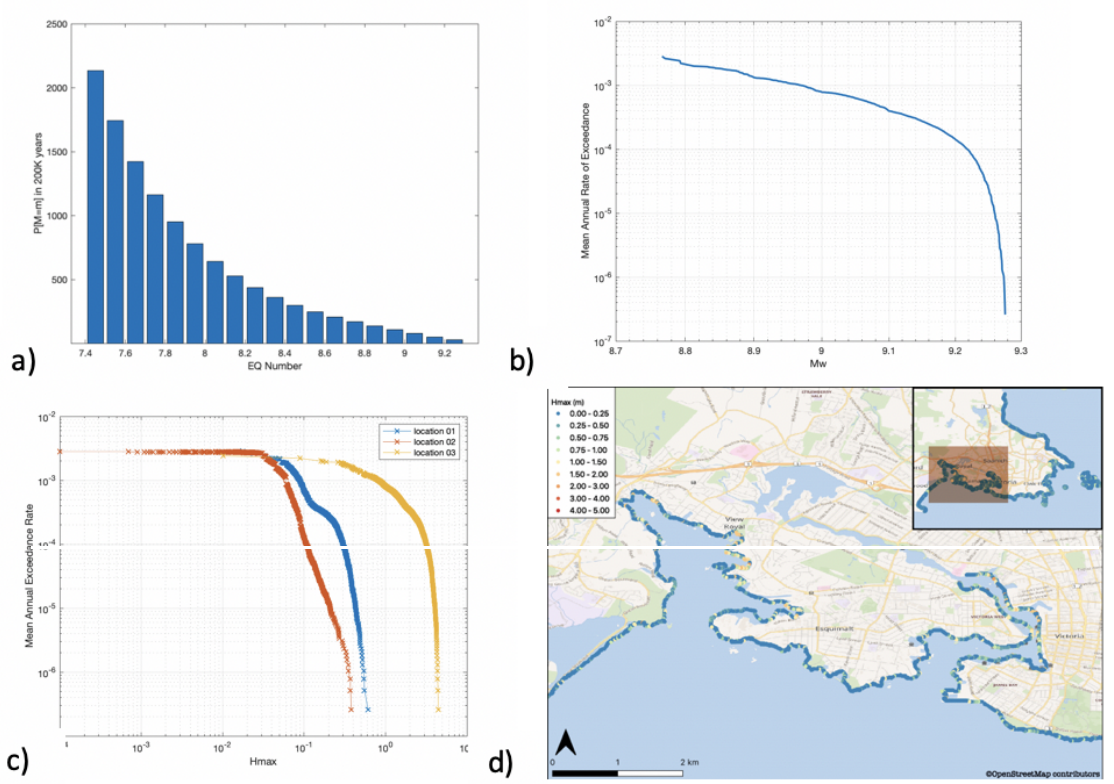

Figure 13

We note that as a first demonstration of how the emulators' predictions can be linked with the probability of exceeding a tsunami intensity measure over time, this is a simplified case. To better capture the probabilistic tsunami hazard in the region the seabed deformation parameter distributions used for generation of the predictions need to be precisely associated with Cascadia rupture characteristics, other more realistic magnitude-frequency relationships than the selected Tapered Gutenberg–Richter distribution can be considered in future research."

No technical comments.