Nat. Hazards Earth Syst. Sci. Discuss., community comment CC1
https://doi.org/10.5194/nhess-2021-77-CC1, 2021

Comment on nhess-2021-77

Randall LeVeque

Community comment on "Rapid tsunami force prediction by mode-decomposition-based surrogate modeling" by Kenta Tozato et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-77-CC1, 2021

This paper contains a very interesting analysis and you have obtained some impressive results. Here are a few comments / suggestions that I hope might be useful.

There were a few things I was confused about regarding the mathematical description of the algorithm that perhaps you could clarify:

In forming the covariance matrix (9), I guess you are assuming that the columns of X have already had their means subtracted?

In line 121, j=1,...,n should be j=1,...,N I think? and also in the summation in the denominator of the equation on line 126 it should be N not n.

In line 138 where you say C and C' have common eigenvalues, I think in general they have a different number of eigenvalues, but any excess ones are all zero. You never make it clear how many rows the matrix X has in this section, i.e. the number of observations, but I guess this is greater than N
in general so that C is a larger matrix than C'?

In line 156 do you mean only the coefficients α_{ij} for j=1,..,r not for j=1,...,p ? I guess so from (16), but worth clarifying. So f(\beta) in general is a vector of length r.

In this paragraph why do you use subscript j rather than k as used in (15) and (16)?

In (17), I think the \exp(...) after the second = sign should be
= \sum_{i=1}^N w_i \exp(...)

At first glance (18) seems to be a square N by N linear system so it seems no regression is needed, so maybe it's worth pointing out that it is really Nr by N since f(\beta) has length r.

Lines 188-190 weren't clear to me. Maybe say the slip was varied from 0.7 to 1.4 times the original slip in the model of Figure 2. When I first read this I also thought you were varying the rake between -20 and +25 degrees, which would be wrong for a subduction event, so maybe also make it even clearer that these are the range of perturbations in the rake angle from the ones given by Fujii-Satake?

In Figure 5 the cyan lines for the Obs. are very hard to see, maybe make these lines black or red?

Line 209, by "concrete connection method" I think you mean the method for coupling the 2D and 3D methods together, but this was confusing to me at first. Maybe say something like "To couple the 2D and 3D models together, the method used in the study of Takase et al (2016) was employed." (By "concrete" I think you mean the specific method employed here, and it might be better to use "specific" here and some other places. Since in English concrete is also a building material, and you are talking about forces on buildings, there could be some confusion.)

Line 231, discussing the 2D mesh used for evaluating the tsunami force: do you average (or sum?) the force over all vertical building faces that happen to lie in the 10m cell? It seems like this would vary a lot from cell to cell just based on the particular geometry of the buildings. In particular some cells might include no walls and hence have 0 force (?) while neighboring cells might have one wall or perhaps at the corner of a building a cell has two walls. So I'm surprised the plots of forces look as smooth as they do and perhaps you can say more about this?

It is great that you can get a surrogate model that reproduces the time evolution as well as it does, in addition to the spatial patterns. But I wonder if this is mainly because you are only considering perturbations to the Fujii-Satake model in which the basic spatial structure of the fault slip is always the same and so the time evolution shows similar sets of waves and arrival times, just somewhat varying magnitudes? The results are still impressive, but I wonder if you can comment on how this might be extended to developing a surrogate model that could be useful in real time for some earthquake that is not a small perturbation of 2011 (which the next big one almost certainly won't be).

To develop a surrogate model that would handle a greater variety of quakes, perhaps it would be necessary to give up on trying to model the full time-dependent solution and instead build a surrogate model that only attempts to predict the maximum inundation depth and force, which might be much easier to do and still very useful.

I don't understand some of the discussion in the paragraph just below Table 2 (lines 312-320). You say "the ratio of mean values was 434%...". I think a ratio should just be a number, not a percent. Do you mean the ratio was 4.34? And what mean values is this the ratio of? Is it the ratio of the error to the true result, i.e. the relative error? This isn't clear.

Maybe this would be clearer if you included also tables of the raw numbers you are comparing, it's not clear where these numbers come from or how they relate to Table 2.

In line 316-317 you give ratios like 0.78%. Again I'm not sure what you mean by a ratio as a percent, do you mean the ratio of error to true value is 0.0078?

In spite of my questions and possible confusion, in general I like the paper and believe it should be published after some clarifications.