Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/nhess-2022-59-RC2, 2022

Comment on nhess-2022-59

Anonymous Referee #2

Referee comment on "Time-dependent Probabilistic Tsunami Hazard Analysis for Western Sumatra, Indonesia, Using Space-Time Earthquake Rupture Modelling and Stochastic Source Scenarios" by Ario Muhammad et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-59-RC2, 2022

The paper "Time-dependent Probabilistic Tsunami Hazard Analysis for Western Sumatra, Indonesia, Using Space-Time Earthquake Rupture Modelling and Stochastic Source Scenarios" by Muhammad et al develops PTHA for western Sumatra using two alternative magnitude-frequency models; one that is time-dependent (and relatively novel), and another that is time-independent (and more conventional). The paper details the methodologies used to fit each model, and compares their results.

The subject of this paper is of interest for tsunami hazard science and the readership of NHESS. The presentation is mostly clear although could be improved in places (mentioned in details below). But as far as I can tell, there are a number of weakness of this paper that will require major revisions to address, before it is suitable for publication. These involve problems with the statistical methods, including: using very different data to fit each model; not treating issues of time-varying completeness in the long-term historical data; use of questionable methods to set Bayesian priors. The paper also neglects key parameter uncertainties controlling the frequencies of earthquakes, that will probably have a large impact on the results (i.e. greater than the current differences between the time-dependent and time-independent models). The authors also need to adjust the introduction to better reflect controversy regarding the performance of time-dependent models (in the literature)

I hope the authors can consider the comments below and either fix their analyses, or (where I am mistaken) edit the manuscript to make their approaches clearer and more obviously defensible. While this will take significant work, I think it is certainly doable, and will make for a useful contribution to NHESS.

HIGH-LEVEL COMMENTS

- Generally the paper argues that time-dependent modelling is more accurate, but doesn't provide strong justification for this. To my knowledge there are contrasting views on this in the literature, which should be represented in this paper. The language should be softened, and uncertainties better discussed (see detailed comments).

- The tsunami hazard results do not seem account for uncertainties in the scenario-frequency model parameters (e.g. b, rate of earthquakes, maximum-magnitude, and other parameters). This is true for both the time-dependent and time-independent models, although details of their parameters are different. Variation of the model parameters within the statistical uncertainties will likely have a substantial impact on the results, especially given only 10 events have been used to constrain the time-dependent model (which has many parameters). When these uncertainties are accounted for, I expect they will be larger than the current difference between the time-dependent and time-independent results. Given the 'many synthetic catalogues' approach used in this paper, the uncertainties could be accounted for by randomly drawing different model parameters for each synthetic catalogue.

- The maximum-magnitude is set to Mw9. In reality maximum magnitudes are quite uncertain (justified further below) and yet very impactful for the results. Again I expect they are likely more important than the effect of time-dependence in the current modelling. Many other PTHA studies treat this as an uncertain parameter (details below), and I suggest that issue is also addressed in this paper.

- In so far as I can tell, there are a number of technical weaknesses in the scenario frequency modelling that should be addressed or clarified.
  + The long-term data (10 events over the last 450 years) likely has a time-varying magnitude of completeness; the earliest 8 events all have Mw>=8.3, while only the most recent 2 events (2007+) have Mw<=8. This is not surprising - a-priori we certainly expect that it would be harder to detect smaller earthquakes in the paleo data. But the statistical methods seem to ignore this issue. This could have a large impact on the fit of the time-dependent model.
  + The time-independent model is fit to different data than the time-dependent model, and the time-independent fit is dominated by small earthquakes (mostly having magnitudes well below the Mw 7.65+ that are of interest in this study). Even if there were no differences between the models, the use of such different data would lead to differences in their results. This makes it hard to determine the significance of the time-dependent model structure for the PTHA results. To remedy this, the long-term data should be used to constrain the time-independent model for larger magnitudes. This may require accounting for time-variations in the catalogue completeness (citations below), and placing less weight on low magnitude earthquakes (so they don't dominate the fit at higher magnitudes, which currently doesn't agree especially well with the data). This should help reduce the under-estimation of the earthquake frequencies with Mw>=8.3 (currently about three times less common in the time-independent model, vs the long term data).
  + There appear to be some anomalies in the Bayesian fit of the time-dependent model. The priors for some parameters seem to be set using the same data used for fitting, which should not be done with Bayesian statistics. Also, the figures show differences between the priors and posteriors that suggest a poor specification of the priors (details in comments below).
  + The fit of the time-dependent model seems to ignore the change in completeness

magnitude of the long-term data.

DETAILED COMMENTS

Near L25: 'Over the next decades, major tsunamigenic events are anticipated in .... '. It sounds like "we expect large tsunamis in each of these subduction zones within a few decades". I don't think this is well justified. Do the references really backup the 'major events in the next-few-decades' claim? Historically, time-dependent predictions over these kinds of timescales have not performed well for subduction zones (e.g. Rong et al., 2003).

Near L40: "...assuming a lack of memory between major earthquake occurrences is often viewed as a first approximation .." -- I think there are contrasting views on this in the literature, that should be represented in this part of the paper. For example Rong et al. (2003) are quite critical of assumed quasi-periodic earthquake recurrence (on empirical grounds). OTOH there is empirical evidence that large earthquakes tend to be weakly periodic, but without correlation between successive inter-event times (Griffin et al., 2020).

L53-54: "Recent work has also used high-resolution spatial grids .... to produce more accurate tsunami hazard results (e.g. < 90m, ...)". I don't think we should describe "< 90m" as high-resolution for onshore work, that is quite coarse. I might describe resolutions of 10m or less as high-resolution (e.g. Gibbons et al., 2020).

Near L67: "Since time-dependent hazard estimation leads to more realistic short-term results" -- this really needs justification, or removal. To my knowledge this point has not been demonstrated in general, and it may-or-may-not be true. I suppose for aftershock modelling there would be lots of evidence, but this study is using quasi-periodic modelling for large events, and I believe there is less evidence on this matter. In the Paleo record, some sites look more time-dependent than others, e.g. Griffin et al. 2020.

Near L74: "A uniform-slip was used, which may underestimate the hazard..." -- I believe Horspool et al. (2014) used a log-normal distribution to predict the (uncertain) heights at the coast from the uniform-slip scenarios, as a way of accounting for uncertainties due to the slip model and uncertain geometry. In principle this is supposed to compensate for the lack of slip heterogeneity. In practice it could either underestimate, or overestimate, the variability of natural earthquake-tsunamis. If their sigma were sufficiently large, it may even predict greater hazard than your model (I haven't checked whether it does, just clarifying the principle).

L86: Please add a statement about why you use segments to define the rupture extents (I think it is related to the space-time modelling?).

L115: "magnitude-frequency distribution" -- Should this be "probability density function"? I think the MFD would include the factor lambda_i.

L117: "frequency-magnitude distribution" -- It think this should be "Cumulative Distribution Function"? Furthermore I think you need to say that f_i(M) is the derivative of F(M) (and consider whether you need a subscript _i for F).

L125: In Equation 4, the subscript '_i' might be confused with the same subscript used to denote the source in Equation 2. Also, I think Eq 4 should use 'j' for consistency with notation in the paragraph just before Equation 4?

L127: Here I am concerned that you are not using the long-term paleo data to fit the GR model. Why not? The longer term data suggests a high rate of Mw >= 8.3 (8 events in 450 years, rate around 0.018), quite a bit more frequent than suggested by your time-independent model (visually seems ~ 0.006 in Figure 1C, or one-third the frequency -- noting this fit is dominated by low-magnitude earthquakes, below magnitudes of practical interest for this study). As well as taking the opportunity to improve the model accuracy, this would be good because the long-term data is used for fitting the time-dependent model. The use of very different data to fit the two models allows for a substantial 'arbitrary' difference between their results, which is not related to their structure (temporal/non-temporal). I am concerned that this may dominate the differences in your results. I would suggest you fit the time-independent GR model using both the long-term and catalogue data (there are various approaches to treating the varying completeness magnitude, e.g. Weichert, 1980), while removing the instrumental events from the long-term data. Also, you might want to use fewer low-magnitude earthquakes to constrain the fit (to reduce the influence of low-magnitude earthquakes on the fit, and better represent the data at magnitudes that matter for this PTHA).

L135: Around here, could you please explicitly state that the time-dependent model does not have Mw-frequency curves that follow the GR distribution, over any time-scale. I didn't realise this initially, and it is obviously a very important point for the subsequent analysis. Perhaps a sentence highlighting that instead the Mw-frequency distribution will reflect correlations between rupture on different segments, which is parameterised by the model itself.

L150: It looks like the magnitudes only go up to 9? I think this is neglecting the large uncertainties in Mw-max. Neglect of those uncertainties may have a strong impact on the results. A few relevant points: Berryman et al. (2015) suggested uncertain Mw-max values in this region ranging from 9.0 - 9.6 based on scaling relations and the historical record. Such highly uncertain Mw-max values have been represented in PTHAs (e.g. Davies et al., 2017; Davies and Griffin, 2020). Horspool et al. (2014) allowed Mw-max on Sumatra to vary in 9.3 - 9.7. We know the nearby 2004 event had a magnitude exceeding 9 (around 9.2). From Tohoku we also know that Mw 9.1 can occur in relatively compact regions, smaller than the extents of your study. On this basis I don't think we can exclude the possibility of higher magnitude earthquakes.

L153: "..for each of those 21 rupture scenarios" -- suggest to add "geometrical" before "rupture scenarios", to be consistent with previous sentences. Here there are a few interacting concepts: "geometrical rupture scenarios (seems to be a magnitude plus a set of segments?)", "scenarios", "events" (is this the same as "scenarios"). I suggest you pick one term for each concept, and then use it consistently throughout the paper.

L 164: "(one height for one simulation catalog)" -- does the height vary with space, or are we looking at the 'maximum height anywhere in the model'?

L 172: "The results confirm that N_{sim} = 100,000 catalogues are sufficient to produce a stable result" -- stable in terms of what? The mean over all catalogues? Please make this clear, as I suppose individual catalogues must vary greatly.

L175-179: This section is confusing me. Above I understood that you used N_{sim} = 100,000 to get a stable result. But now it is suggested that many more catalogues were required for 1-50 years. Please edit to make this clearer. [NOTE: Some sentences from the 'Results' section may help in this regard, mentioned below.].

L197-199: "This number is consistent with the GR model". In my judgement they are "not very consistent", with the model under-predicting the frequency of large events (as discussed above, the GR model has a substantially lower frequency of Mw>=8.3). Note the 450 year record contains 10 events (Mw 7.8-8.9), but the first 8 events have Mw>=8.3, and the last two events are from the recent instrumental period. This suggests changes in the magnitude of completeness of the 450 year catalogue over time. A-prior we expect this would happen because Paleo records find it more difficult to detect small events. This issue should be accounted for when comparing the GR model with the long term data (and above I suggest that the long-term data should also be used to fit the time-independent GR model -- doing that will probably lead to significant increases in the modelled frequency of large earthquakes).

L205: "see Figure and Figure 5" -- missing Figure number.

L217: Suggest you use a word other than "scenarios" to denote the 21 "magnitude + set-of-segments" combinations.

L250-ish: Above I argued that the long-term data (10 events, 450 years) is likely subject to a varying completeness magnitude, noting the only two events with Mw<8 events are recent instrumental events, and all others have Mw>=8.3. From what I can see, this 'changing completeness magnitude' is not accounted for in the statistical fit of the time-varying model (Section 2.2.2). I expect this would have a large effect on time-varying model fit - for example, overestimating the conditional probability of multi-segment rupture (which also effects the frequency of high Mw events), and affecting the BPT model parameters.

L250: "The prior median of mu for each segment is different, namely .... These values represent the median interarrival time of earthquake rupture on each segment over the last 450 years". It sounds like you are using the same data both to specify the priors, and then to fit the model (?). In Bayesian statistics, the priors should be specified in a way that doesn't use the fitting data, or at least doesn't use it in important ways. Another potential problem with the methodology is suggested by Figure 9, where we see the prior and posterior for 'mu' are very different on some segments -- the posterior is more diffuse and often has a very different average (e.g. Panels A, I, K). This suggests the priors have been overly constrained in the analysis. Typically priors would be set either using data different used for fitting, or given weakly informative values.

L310: This source zone has some history of "tsunami-earthquakes", with waves much larger than might be expected from the magnitude (e.g. Mentawai 2010). Can the current model produce similar large waves for scenarios with magnitude below 8, using the rigidity of 40GPA? I would be surprised if it can, although that will also depend on how concentrated the slip is allowed to be. Please add a comment on the capacity for the model to make 'tsunami-earthquake' type scenarios.

L317 "... 300 stochastic models are sufficient to simulate stable and consistent tsunami heights and depths" -- I think this must depend on the model region, and what you are interested in. For instance it would not give an accurate representation of the 99.5th percentile. Also for a model where only a very small part of the source-zone could affect the site of interest, one might need to generate many scenarios to get enough relevant scenarios. In summary, I don't think you can refer to stability tests from another study to provide justification for using 300 models in this study. Instead, can you report on a test that is specific to this case?

L347: "... the final parameter estimates are taken from the maximum a posterior". It would be better to account for the model uncertainty (also in Mw-max, b, etc), which should be substantial given the limited data available to fit the model, and will probably have a substantial impact on the hazard. One way to do this would be to draw a different parameter set for each of the large number of synthetic catalogues that are simulated.

Section 3.1: As discussed earlier, please comment on why the 'mu-priors' for some segments are so different to the posteriors (little overlap for Fig 9 panels A, I, K). This is surprising given especially considering that the priors were apparently constructed using the same data used for fitting. To me it suggests weaknesses in how the priors were constructed, or some other problem.

L356-360: This is a very clear description of how the catalogue duration was defined. I suggest you move this to the earlier methods section (where I expressed confusion about the method).

L360 and Figure 10: Regarding the validation of the annual seismic moment release: Considering that the data was used to fit the model, I don't think the observations/model are particularly consistent on segments 3 and 4. In both cases the observed data exceeds the 90th percentile of the model. Again this seems to suggest some under-estimation in the model, as discussed repeatedly above. Please check that this is all correct following revisions, and if it is, add a comment explaining why this is nonetheless reasonably consistent.

L379: Figure 11C is not a strong basis for making a point about which segments rupture more or less, because it is only 1 catalogue. Can you please make a figure that better justifies the points made in this paragraph?

Line 447 and Figure 12: The conditional probability of Mw9.0 (if an earthquake occurs) is larger in the time independent case. But I doubt that these results will be robust to parameter uncertainties in the time-dependent model, considering that limited data (10 events, that likely has time-varying completeness) was available to fit its many parameters. This further suggests the importance of considering model parameter uncertainties in the PTHA.

Line 451: One factor neglected in this discussion is the effect of using different datasets to fit the 2 models, which could cause differences in the results even if there was no other difference between the two kinds of models. I think the calculations in this paper should be revised so that the time-independent model is informed by the long-term data, and that parameter uncertainties in both the time-independent and time-dependent models are accounted for. In my judgment it is likely that the parameter uncertainties will could lead to differences in the results that are substantially larger than differences between the current time-dependent and time-independent models.

TECHNICAL CORRECTIONS

None for now.

REFERENCES

Berryman, K.; Wallace, L.; Hayes, G.; Bird, P.; Wang, K.; Basili, R.; Lay, T.; Pagani, M.; Stein, R.; Sagiya, T.; Rubin, C.; Barreintos, S.; Kreemer, C.; Litchfield, N.; Stirling, M.; Gledhill, K.; Haller, K. & Costa, C. The GEM Faulted Earth Subduction Interface Characterisation Project: Version 2.0 – April 2015 GEM, GEM, 2015

Davies, G.; Griffin, J.; Løvholt, F.; Glimsdal, S.; Harbitz, C.; Thio, H. K.; Lorito, S.; Basili, R.; Selva, J.; Geist, E. & Baptista, M. A. A global probabilistic tsunami hazard assessment from earthquake sources Geological Society, London, Special Publications, Geological Society of London, 2017

Davies, G. & Griffin, J. Sensitivity of Probabilistic Tsunami Hazard Assessment to Far-Field Earthquake Slip Complexity and Rigidity Depth-Dependence: Case Study of Australia Pure and Applied Geophysics, 2020, 177, 1521–1548

Gibbons, S. J.; Lorito, S.; Macías, J.; Løvholt, F.; Selva, J.; Volpe, M.; Sánchez-Linares, C.; Babeyko, A.; Brizuela, B.; Cirella, A.; Castro, M. J.; de la Asunción, M.; Lanucara, P.; Glimsdal, S.; Lorenzino, M. C.; Nazaria, M.; Pizzimenti, L.; Romano, F.; Scala, A.; Tonini, R.; Manuel González Vida, J. & Vöge, M. Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations. Frontiers in Earth Science, 2020, 8, 623

Griffin, J. D.; Stirling, M. W. & Wang, T. Periodicity and Clustering in the Long-Term Earthquake Record Geophysical Research Letters, American Geophysical Union (AGU), 2020, 47

Horspool, N.; Pranantyo, I.; Griffin, J.; Latief, H.; Natawidjaja, D. H.; Kongko, W.; Cipta, A.; Bustamam, B.; Anugrah, S. D. & Thio, H. K. A probabilistic tsunami hazard assessment for Indonesia Natural Hazards and Earth System Sciences, 2014, 14, 3105-3122

Rong, Y.; Jackson, D. D. & Kagan, Y. Y. Seismic gaps and earthquakes Journal of Geophysical Research: Solid Earth, Wiley-Blackwell, 2003, 108

Weichert, D. H. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes Bulletin of the Seismological Society of America, 1980, 70, 1337-1346