Review of manuscript NHESS-2017-41 "Integrating faults and past earthquakes into a probabilistic seismic hazard model for peninsular Italy" by Alessandro Valentini, Francesco Visini & Bruno Pace

Main comments

This manuscript describes an approach to model seismic hazard in Italy using a combination of active fault data and gridded seismicity based on the instrumental and historical earthquake catalog. A database of active faults has been compiled, and important historical earthquakes have been assigned to their causative faults. Two models are considered for the magnitude-frequency distributions (MFDs) of the faults, either a truncated Gutenberg-Richter (TGR) MFD or a characteristic Gaussian (CHG) MFD. The gridded source model accounts for off-fault seismicity, and its MFD is computed in a way that it is complementary to the MFD of the fault source model (using a threshold magnitude, avoiding double-counting of earthquakes assigned to faults, and an additional weighting function that reduces gridded seismicity in the vicinity of faults). The authors explore the impact of the two MFD models, as well as the contribution of fault sources and gridded seismicity to the total hazard. They also define a preferred source model, in which the most appropriate MFD model for each fault is selected.

The approach to model fault sources is state of the art, and the integration of fault sources and gridded seismicity contains some innovative elements. The manuscript is mostly well written (with some exceptions, which are pointed out in the detailed comments below), the figures are clear, and the references are appropriate. The conclusions are supported by the results.

However, a number of improvements need to be made before the manuscript can be published. Below, I have listed a number of detailed comments. I summarize my main comments here:

- A major shortcoming is that the paper does not contain any reference to other published fault source models for Italy, notably DISS (Database of Individual Seismogenic Sources, http://diss.rm.ingv.it/diss/). At the very least, the authors should indicate how their fault source model relates to DISS, and what are the main differences (concepts and/or data).
- Although the authors refer to the SHARE project, and even use certain aspects of it, they do not compare their results to the fault-based hazard map (FSBG model) created in this project.
- Similarly, although a comparison with the current national hazard map is described in general terms, this comparison is not shown.
- In my opinion, it is also essential to show the summed MFDs of the different source models, and comparing those to each other and to the observed MFD based on the full catalog. Without this information, it is not possible to evaluate the performance of their model. Notably, it is indicated that the rate of M 5.5-6.0 earthquakes in the TGR end member is higher than in the CHG end member, but this is not shown.
- I have some doubt whether maximum magnitudes are correctly modelled, as it is indicated at some point that an earthquake assigned to a fault could have a magnitude larger than the magnitude range in the MFD for that fault, which should not be allowed.

- To improve clarity, the authors should more clearly explain in advance what they intend to do. Two main cases are:
  - They first describe the fault-source model and the distributed source model, and only later explain that these are not independent models, but are complementary, together accounting for all seismicity in Italy;
  - They first show hazard maps produced with the TGR and CHG MFD models, but only later explain that these are two end members, and that their preferred model is the Mixed model, in which a particular MFD model is assigned to each fault.

Recommendation

I recommend that this manuscript can be published after moderate revision.

Detailed comments

Abstract

L. 11: "many of active faults" → "many active faults"

L. 13-14: "modelled on a number of seismotectonic zones and the occurrence of earthquakes is assumed to be uniform" → "modelled <u>using</u> a number of seismotectonic zones <u>in which</u> the occurrence of earthquakes is assumed to be uniform";

L. 24-25: "earthquakes" → "magnitude-frequency distributions" (2x)

L. 30: "the spatial pattern of our model is far more detailed" → "the spatial pattern of the hazard maps obtained with our model is far more detailed". Unfortunately, this is not demonstrated in the paper, as there is no direct comparison with other hazard maps.

1. Introduction

L. 42: "in seismic hazard estimations": may be left out, it is obvious from the previous lines.

L. 52: "Combining seismic hazards from active faults with background sources" → "Combining active faults with background sources". I also note that the plural "seismic hazards" is used in other places in the manuscript, but it should be singular, as the paper deals with only one type of seismic hazard, namely ground-motion seismic hazard.

L. 58: Add reference for the SHARE FSBG model.

L. 62: "uniform on" → "uniform in"

L. 65-66: "…, obtaining more detailed and possibly more realistic patterns of ground motion, in order to improve the reliability …" → "…, in order to obtain more detailed and possibly more realistic patterns of ground motion, and to improve the reliability …"

L. 69: "can also be give insights" → "can also give insights"

2.1 Fault Source Model

L. 90-91: Add degree symbols to all dip and rake values.

L. 92: “thrust faults could be considered in a future study”: Is there a particular reason for not including thrust faults in the present study? And for which areas in Italy will this have the largest impact?

L. 93: “seismogenic thickness” → “seismogenic layer” (a thickness has no upper and lower boundaries)

L. 101-102: “Slip rates control fault-based seismic hazards … and provide a time scale …”: Strange phrasing. Slip rates do not provide a time scale. I’m not sure whether the authors mean to say that slip rates may be measured over different time scales or that slip rates may vary through time or both.

L. 112-124: This paragraph discusses slip rate variability through time, and states that slip rates have been determined for different time scales. However, (1) it is not clear how this time variability is handled in this study (it is not mentioned anymore further in the paper), and (2) Table 1 only lists minimum and maximum slip rates, without indication of the corresponding time scale. Is the time scale the same for all faults in this table?

L. 141: “the function with the lowest log-likelihood”: Shouldn’t this be the highest log-likelihood? Usually, one seeks the maximum likelihood, not the minimum likelihood.

L. 145-150: Is this an appropriate way to determine the overall standard deviation of the slip rate distribution in an area? I think it would be more appropriate to apply the Central Limit Theorem. If you consider each fault slip rate (x) as a sample from a population with mean μ and standard deviation σ, then μ can be found as $\mu_{\bar{x}}$ (mean value of the sample means), and σ as $\sqrt{n}\sigma_{\bar{x}}$ (with n the number of samples and $\sigma_{\bar{x}}$ the standard deviation <u>of the sample means</u>).

L. 160: “a fault source is …” → “a fault source is considered as …”

L. 166-169: there seems to be overlap between criterion ii (sharp bends) and criterion iv (bending ≥ 60°).

L. 171: “seismogenic thickness” → “changes in seismogenic thickness”

L. 180: “thinnest ST” → “smallest ST”. Can you comment on the small ST value of 2.5 km? Is this in a volcanic zone?

L. 181: “Observed maximum magnitude data have been assigned to 47 fault sources”. Is this based on Table 2?

L. 197-198: “a value that corresponds to the maximum observed magnitude (Mobs)”. I’m not convinced it is correct to consider Mobs as one of the possible Mmax values, and treat it the same as the other estimations. In fact, the only thing we know for sure about Mmax is that it cannot be lower than Mobs. For that reason, Mobs is often used as a lower truncation of Mmax distributions (e.g., EPRI method for Stable Continental Regions). Not doing this can have strange consequences, as in lines 442-444, where it is stated “If an earthquake assigned to a fault source has a magnitude lower or higher than the bell curve of the CHG model distribution, …”. However, the second case (observed magnitude higher than modelled Mmax distribution) should not be allowed in the PSHA model!

L. 199: “modifying the along-strike dimension if the rupture length exceeds the length predicted by the aspect ratio relationships”. This is not very clear. Maybe rephrase as “reducing the fault length if the aspect ratio (W/L) is smaller than indicated by the relation between aspect ratio and rupture length for observed earthquake ruptures in the Abruzzo (Peruzza & Pace, 2002)”.

L. 202: “we use the criterion of “segment seismic moment conservation””: is this a criterion or a concept, and can you briefly describe what it implies?

L. 206-207: “we use two magnitude-frequency distributions” → “we use two magnitude-frequency distribution <u>models</u>”. I also recommend introducing the acronym MFD here, as the term is used frequently in the remainder of the manuscript.

L. 208: “Gaussian bell curve centred on the Mmax”: Perhaps it is worth mentioning that this Gaussian curve applies to the incremental MFD values, not to the cumulative MFD values that are shown in Fig. 2c.

L. 209-211: It is not explained how the a- and b-values are determined for each fault when the TGR model is used. I assume this is done with the FiSH code, but it would be good to briefly describe the underlying concept (relation with slip rate).

L. 213: “the two above described magnitude-frequency distributions” → “the two MFD models described above”.

2.2 Distributed Source Model

L. 218: “a number of earthquakes in the historical catalogue”: Perhaps “important earthquakes”?

L. 228: “events not considered the mainshock” → “events not considered as mainshocks”.

L. 233-234: “If the causative source of an earthquake is known, the impact of that earthquake does not need to be included in the seismicity smoothing process” → “If the causative fault of an earthquake is known, that earthquake does not need to be included in the seismicity smoothing procedure”. It should be explicitly mentioned before that the fault and distributed source models are conceived as <u>complementary</u> source models, not as alternative source models (competing models in a logic tree). In the latter case, they should be independent.

L. 244: “where *n* is the cumulative rate of earthquakes *ni*” → “where n_i is the cumulative rate of earthquakes”

L. 263: I think the * symbol in the equation should be left out. If I understand correctly, rather than a multiplication, $\lambda(i_x, i_y)$ represents the seismicity rate in grid cell (i_x, i_y).

L. 265: “jy” → “i_y”

L. 276-278: I don't understand the description of the Voronoi partition procedure: if the Italian territory is divided in a grid with 0.05° lon/lat spacing, then how can the number of grid cell centres be varied? Perhaps the centres of the grid cells represent the possible centres of Voronoi polygons, and you vary the number of Voronoi polygons from 3 to 50, for each case drawing 1000 random subsets of Nv grid cell centres?

L. 297: "β = 2/3 b": I think this should be "$\beta = \mathrm{b}.\ln(10)$", which is ~2.3 b.

2.3 Combining fault and distributed sources

L. 299-300: It would be better to describe this concept before the two source model components are described (see general remark).

L. 307: Add some statement that this assumption is explained in more detail in the following paragraphs.

L. 311: "located across strike" → "arranged across strike"

L. 315: "neighbour faults" → "neighbouring faults"

L. 322: "has also shown similar" → "has shown similar"

L. 324-325: "from … to … to…" → "from … over … to …"

L. 330: "the spatial distribution … are" → "the spatial distribution … is"

L. 333: "appear to control to" → "appear to control"

L. 336: "Earthquakes" → "Large earthquakes"

L. 338-340: Is this valid for all types of faults or only for dip-slip faults?

L. 342: "distance along strike": I think this should be "across strike".

L. 352-353: "considerations to source models for fault sources and distributed seismicity" → "considerations to combine fault and distributed seismicity source models"

L. 360: Perhaps add that it is a linear function.

L. 363: Write the equation more completely:

$$Pe = \begin{cases} 0 \ (d \leq 1km) \\ \frac{1}{d} \ (1km < d \leq d_{max}) \\ 1 \ (d > d_{max}) \end{cases}$$

However, there is still a problem with the second line, which does the opposite of what is intended (going to 1 as d increases): instead of 1/d it should be d/d_{max}…

L. 366-367: What is the rationale for varying d_{max} in function of slip rate?

L. 369-371: This is hard to understand. Maybe rephrase as "Because we considered two fault source models, one using only TGR MFDs and the other only CHR MFDs, and because the MFDs of distributed seismicity grid points in the vicinity of faults are modified with respect to the MFDs of these faults, we also obtain two different models of distributed seismicity."

In my opinion, it is also necessary at this point to show the summed MFDs of the different (sub)models, i.e. summed MFD of the TGR fault source model, of the CHR fault source model, of the TGR distributed source model, of the CHR distributed source model, and of the combined TGR and CHR source models.

L. 380: "in future" → "in the future".

3. Results and discussion

L. 382: "designed under the traditional Poisson hypothesis": Rephrase

L. 386: "well-known": this is not the most relevant property for choosing OpenQuake. Perhaps widely used, open-source, tested, …?

L. 390: "Bindi et al. 2014" → "Bindi et al. (2014)"

L. 391: "all GMPE" → "all GMPEs"

L. 401: "estimated by" → "obtained with"

L. 402: Explain more explicitly that the TGR and CHG fault source models are end members that are only used to explore the epistemic uncertainty, and that in the preferred fault source model a choice is made between the two MFD models for each fault.

L. 403-404: "Although both models have the same amount of seismic moment release": this has not been demonstrated.

L. 405-406: "all faults exhibit a 10% probability of exceedance in 50 years": Incorrect phrasing → "all faults contribute significantly to the seismic hazard".

L. 409-411: "The rates of earthquakes with magnitudes between 5.5 and approximately 6, …, are generally higher in the TGR model than in the CHG model": Please demonstrate by showing the summed MFDs.

L. 416-417: "threshold" → "lower value" (2x)

L. 424: "the fault source models contribute to …" → "the fault source models contribute <u>most</u> to …"

L. 434: "hazard maps with PGAs" → "hazard maps for PGA"

L. 436-437: "applying a magnitude frequency distribution to each fault" → "selecting the most appropriate MFD model (TGR or CHG) for each fault".

L. 443: “a magnitude lower or higher than the bell curve” → “a magnitude lower or higher than the magnitude range in the bell curve”. See also my remark at lines 197-198: a higher magnitude should not be possible!

L. 448: “have been not yet mapped” → “have not yet been mapped”.

L. 448: “The magnitude-frequency distribution” → “The MFD model”.

L. 455: “sources that have been modelled using a magnitude-frequency distribution”: all sources are modelled using an MFD → “sources that have been modelled using one or the other MFD model”.

L. 462: “a particular magnitude-frequency distribution” → “a particular MFD model”.

L. 465: “higher than those the CHG” → “higher than those of the CHG source model”.

L. 467: “seismicity rate model” → “MFD model”.

L. 468-471: It has not been explained exactly how the TGR MFDs have been constructed. See my remark at lines 209-211.

L. 475: “generally decrease with increases in the exceedance probability” → “generally decreases with increasing exceedance probability”.

L. 505: Perhaps replace “TGR model” with a brief description like you do for the CHG model in the following line.

L. 513: “tests to the compatibility” → “tests of the compatibility”.

L. 524: “The strength of our approach is” → “The strength of our approach lies in”.

L. 527: “seismic hazards” → “seismic hazard”.

4. Conclusions

L. 538-539: “110 faults with 86 fault sources” → “110 faults grouped into 86 fault sources”.

L. 542-543: “two magnitude-frequency distributions” → “two MFD models”.

L. 552: “overlapping the magnitude-frequency distribution” → “overlapping with the MFD”.

L. 558-559: “pattern similar to that of the current national maps at the national scale, but some significant differences in hazard are present at the regional-to-local scale”: this has not been discussed in the main text. It would be instructive to show both maps side by side and describe the comparison in some more detail in §3.

L. 560: “different magnitude-frequency distributions” → “different MFD models”.

L. 563-565: See my comment for lines 409-411. It would also be interesting to compare the summed MFDs to the observed MFD based on the full catalog, to see which of the two MFD models is closer to the observations in this particular magnitude range (M 5.5 to ~6.0).

L. 568: “high values of probability of exceedance” → “high probabilities of exceedance”.

L. 569: “equal the fault model one” → “equals that of the fault model”.

L. 570-571: “applying a magnitude-frequency distribution to each fault” → “selecting the most appropriate MFD model for each fault”.

Figure captions

Fig. 2, L. 824: “modified by” → “modified from”?

Fig. 2, L. 829-830: “the maximum subsurface fault length (MRLD, red line) and maximum rupture area (MRA, cyan line) …” → “MRLD (red line) and MRA (cyan line) correspond to estimates based on maximum subsurface fault length and maximum rupture area …”

Fig. 3, L. 841: “in correspondence of” → “corresponding to”

Fig. 6, L. 855: “versus c” → “versus correlation distance c”

Fig. 9 : Explain acronym "poe"

Fig. 12: How are the contributions of the component source models computed? The perfect symmetry between the contributions of the fault source model and the distributed source model gives me the impression that they do not correspond to the contributions one would obtain from a deaggregation.