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

by Laurentiu Danciu
Swiss Seismological Service
ETH Zurich

General Comments

The manuscript provides a procedure to integrate active faults in a regional seismogenic source model for Italy. A database of active faults was compiled and fully parameterised for use together with observed seismicity (instrumental and historical) to forecast the spatial and temporal distribution of future seismicity. Earthquake recurrence models of the delineated active faults are model by two magnitude-frequency distributions: either a Characteristic Gaussian (CHG) or Truncated Gutenberg-Richter (TGR). Additionally, the seismicity off faults is described by a smoothed seismicity using a complete earthquake catalogue of the region. The two models are complementary not independent, thus the earthquake rates account for double-counting of earthquakes assigned to faults above specified threshold magnitude. Further, a novel weighting function to correct the earthquake rates in vicinity of fault sources is proposed and used. The resulting two seismic sources are eventually combined in a mixed source model representing the suitable activity rates in time and space. The authors conclude with a sensitivity analysis evaluating the impact of the two models of earthquake recurrence rates on the total seismic hazard.

The use of active faults in seismic hazard assessment has become extensive in the last decades due to efforts of data compilation and analysis. Active faults provides the information to extend the observational time of large magnitude earthquakes which often is not captured by the existing catalogues of observed seismicity. The current manuscript provides a step forward into this direction. The combination active faults and smoothed seismicity is not a novel procedure but rather state of practice. Overall, the manuscript is relatively well written, there are several misleading parts to be improved, highlighted in my detailed comments. The structure of the manuscript is consistent with the procedural steps and no major changes are required. The figures, tables and supplemental materials are clear and appropriated. There are some key references missing but this is not necessarily a criticism. The conclusions appear appropriate with the proposed procedure and analysed content. My comments follow the structure of the manuscript and summarised below:

1. First and foremost the authors should be clearly state that this is not an update of the seismic hazard model of Italy, and that the purpose of the study is to integrate the active faults in a hazard calculation. Moreover, the resulting seismogenic model presented in this study has limitations, such as the use only of shallow faults, but not the subduction and volcanic sources.

   To clearly state that our model is not aimed to update seismic hazard model of Italy, we will add at line 70 the following statement: "In conclusion, even if the main purpose of this work is to integrate the active faults in a hazard calculation for the Italian territory, this work does not represent an official update of the seismic hazard model of the Italy".
   About the use of only shallow faults, but not the subduction and volcanic sources, we will more clear introduce this issue in the manuscript. In any case in this paper we want to focus on the

impact of the integration of faults and earthquakes data, without the assumption to be complete in terms of fault database, but on the contrary suggesting a way to integrate two incomplete database in the best way, without throwing data.

2. A definition of active fault in the context of the study must be introduced. The literature distinguishes between active faults in geological time, i.e. Quaternary or Neocene, capable of future reactivation. Moreover, the slip rate assumptions must be discussed. It is well accepted that large variability are associated with the slip-rate values, and some portion of slip-rate can be aseismic. Extension of this discussion must be introduced in the context of this study.

   We agree that a definition of active fault in the context of the study is necessary. We will add at line 82 a phrase as: "For seismic hazard assessment an active fault is a structure that has evidence of activity in the late Quaternary (i.e. in the past 125 kyr), a demonstrable or potential capability of generating major earthquakes and capable of future reactivation (see Machette, 2000 for a discussion on terminology). The evidences of quaternary activity can be geomorphological and/or paleoseismological, when activation during instrumental seismic sequences and/or association to historical earthquakes are not available".
   We will also extend discussion about slip rates assumptions for PSHA. In particular, we will more clear to state that we are assuming that slip-rates used are representative of seismic movements (no-aseismic factor). We think that investigating the impact of this assumption could be an issue of uncertainty-focused paper, for example by differentiating aseismic slip factor in respect to different tectonic contests.

3. Further, the authors are aware of the 2013 European Seismic Hazard Model (ESHM13, Woesner et al 2015) developed within the SHARE Project. It might be worth discussing the two approaches side by side, as the ESHM13 is the first reference model to introduce active faults for Euro-Mediterranean Region.

   We prepared a new figure (Figure S1 in supplement) to compare our model, FSBG model proposed by SHARE and the Italian seismic hazard map MPS04, using the same GMPE's. A discussion about this comparison will be added in the "Results and Discussion" chapter.

4. There are several procedural steps that are not well explained in the document, such as the estimation of the activity rates for faults. Albeit, the main focus of the procedure is to implement active faults to seismic hazard, the activity rates are yet described as input to the FiSH code and the segment seismic moment conservation. In my opinion this is not enough. The key elements and assumptions for computing the activity rates of active faults needs more attention, supported with discussions of the sensitivity of the input parameters, i.e. the effect of slip rates to earthquake recurrence rates.

   In order to explain more in detail the segment seismic moment conservation, we will modify part of the text by adding the following paragraph:
   "… which divides the seismic moment that corresponds to Mmax by the moment rate given a slip rate:

$$T_{mean} = \frac{1}{Char\_Rate} = \frac{10^{1.5M_{max}9.1}}{\mu VLW}$$

where Tmean is the mean recurrence time in years, Char_Rate is the annual mean rate of occurrence, Mmax is the computed mean maximum magnitude, μ is the shear modulus, V is the average long-term slip rate, and L and W are the geometrical parameters of the fault, along-strike rupture length and down dip width respectively."
Moreover, to explain how magnitude frequency distribution of TGR is computed we will state that: " For MFD, the b-value is constant and equal to 1.0 for all faults, obtained by the interpolation of the earthquakes in the CPTI15 catalogue, as the events on the single sources are insufficient for statistics. However the a-values have been computed by Activity Rate FiSH code, balancing the total expected seismic moment rate with the seismic moment rate that was obtained by the pair Mmax and Tmean, evaluated by the fault geometry and the slip rate of each individual source (details in Pace et al., 2016)."

5. The role of each magnitude frequency distribution (MFD) for each fault is not clear as described in the current version. One might expect a logic tree of the two MFDs. This aspect needs to be emphasised in the introduction.

   Thanks for your suggestion. We'll clarify in the Introduction our choices, explaining that the TGR and CHG MFD are here used as end members, in order to explore the epistemic uncertainties, and we consider the Mixed model as a sort of an "expert judgment" model, useful for comparison analysis. As our model is not aimed to update seismic hazard model of Italy, we don't think we need to use a logic tree approach to produce a weighted model.

6. Maximum magnitude assigned to each fault based on empirical magnitude scaling relationships do not account for uncertainties of the fault size (subsurface length or area). From the current version of the manuscript it is not evident the error associated to the fault size in the fault dataset.

   In our work, the error associated to the fault size was not taken into account because there are no indications to quantify these errors from the published data used to obtain the active fault database. The error associated to the Mmax of the fault sources is only based on the errors of the used empirical relationships and observations.

7. Also, one can argue that more recent magnitude scaling relationships can be used (e.g Leonard et al 2010) but for those used, the role of aleatory uncertainty must be mentioned and quantified herein. The authors should describe the procedure implemented in the FiSH code because not everyone has access to that manuscript.
8. Five maximum magnitude values are described as being assigned to each fault. The way these five values are implemented in the final computational model is not clear. Are these values modelled in a logic tree?

   We will add a description of the procedure to estimate Mmax for faults after summarizing what has been done in Pace et al. (2016) FiSH code: "Because all the empirical relationships and observations are affected by uncertainties, a first code (MB) is designed to take these factors into *account and return a maximum magnitude value and a standard deviation. The uncertainties in the empirical scaling relationship are taken from the studies of Wells and Coppersmith (1994), Peruzza and Pace (2002) and Leonard (2010). Currently, the uncertainty in magnitude from seismic moment is fixed and set to 0.3, whereas the uncertainty in Mobs is defined by the catalogue. To combine the maximum magnitudes, MB draws a probability curve for each magnitude estimate by assuming a normal distribution. It is possible to define the number of standard deviations (σ) for truncating the normal distribution of magnitudes at both sides. MB successively sums the probability density curves and fits the summed curve to a normal distribution to obtain the mean of the maximum magnitude Mmax and its standard deviation. Therefore, Mmax represents an evaluation of the maximum rupture that is allowed by the fault geometry and the rheological properties".*

9. A sensitivity analysis to the choice of the maximum magnitude may be necessary to explain the effect of maximum magnitude for the TGR. For the same slip rate increase of the maximum magnitude will result in a decrease of the recurrence of small events. This effect is due to the fact that the largest earthquake accounts for most of the seismic moment and this requires the subtraction of small events to maintain the seismic moment balance.

   We agree with the topic here raised by the reviewer. Actually, the impact of uncertainties in Mmax and slip rate into PSHA is an important question, but we think it deserves a more extensive work to be exhaustively pointed out. We prepared a figure to show how varying these two parameters the seismicity rates can be distributed following a TGR model (Figure S3 in supplement). In our paper only the central values of the shown MFDs has been used. It is clear the final PSHA is substantially modified when Mmax and slip rate are changed, but, for the purpose of our work, this aspect is out of topic. We are exploring these (and other) aspects of fault-based approaches but, again, to be at least sufficiently analysed, they should be ingredients for a new work.

10. In a general way, the characteristic model implies a recurrence rate estimated on large past large-magnitude earthquakes recognised from past geological record and the time interval between events can be measured. How many of the faults have a geological record long enough to characterise the recurrence of the large magnitude events? In the current version of the manuscript the historical events are linked to the faults, thus the long-term representation of the fault activity is questionable.

    Thanks for your comment, we were not clear in explain how the mean recurrence times (Tmean) of the characteristic earthquake have been calculated. Similarly to TGR MFD we evaluated Mmax and Tmean by the fault geometry and the slip rate (not with the observed occurrences) of each individual source and we calculated the total expected seismic moment rate (eq. in the answer to comment 4). Then, we partitioned the total expected seismic moment rate in a range given by Mmax +- 1 standard deviation following a Gaussian bell distribution. We'll improve the manuscript to better explain this concept.

11. Slip rates are averaged over successive geologically recognised earthquakes and prone to error in measurements, hence the uncertainties of the slip-rates needs to be quantified.

    Uncertainties in slip rates estimates are given in the seismogenic sources database in the appendix. For our PSHA model we used the central value of the slip rate range given for each *fault. We are assuming that this value is representative of the average long term behaviour of the fault. Unfortunately, the state of the art of the knowledge of slip rates in Italy cannot allows to resolve a more detailed analysis of slip rate. However, varying slip rates in the currently range of uncertainty (as published in the papers cited in the appendix), we produced the figure S3 (in the supplement) to show the impact of these uncertainties on the activity rates.*

12. When combining active faults and background seismicity, it is mandatory a comparison of the seismic productivity (CHG and TRT) of the faults with the gridded seismicity in the vicinity of faults. Without such comparison it is difficult to assess the performance of the models.

    Thanks for your suggestion. We attach in supplement a new figure (Figure S2) showing and comparing the summed MFD's of the fault source inputs (TGR, CHG, Mixed), the distributed source input, the final model (distributed + fault) and the CPTI15 catalogue. This new figure shows, in a sector of Italy where the faults are well defined, the behaviour of the activity rates as derived by our approach. The new figure we'll be included in the revised version of the manuscript.

13. Generally, evaluating the performance of seismogenic sources based on seismic hazard estimates is not recommended. The hazard estimates based on active faults only is misleading, as the active faults are incomplete in space, and not treated as independent models. Thus the model performance may be evaluated at the level of seismicity rates comparison, not for hazard estimates.

    Thanks for your comment, we agree it is important, in order to evaluate the performance of different seismic models for seismic hazard, a direct comparison of seismicity rates. For this reason we'll add in the manuscript the figure above described (Figure S2 in supplement). In any case we think it is interesting to show the impact of different seismogenic sources also in terms of seismic hazard maps.

14. The authors should state clearly that a suitable seismogenic source model combines the active faults and the gridded seismicity as mixed model.

    As also commented later, we agree that a model should include faults and distributed sources. We will clearly state that the mixed fault source is obtained by our judgment on the MFD assigned to each single fault, and that the mixed model combines this fault source input with the distributed sources input.

Section Specific Comments

L50:51: "In Europe, a working group…" In Europe, within the SHARE project (Giardini et al 2010) has introduced the use of active faults at the region level for the first time. I am surprised that the authors do not refer in their study to the fault source models for Italy, the DISS (Database of Individual Seismogenic Sources). What are the main similarities and differences between the two dataset? The authors may consider adding a reference and a discuss the two datasets to avoid confusion.

We mentioned SHARE project in our manuscript at line 58, and a new figure (S1 in supplement) compares the results. About the DISS, we will at line 84: "Although for the Italian territory there is already a database that contains the results of the investigations of the active tectonics during the past 20 years (Database of Individual Seismogenic Sources, DISS, http://diss.rm.ingv.it/diss/), made by three main categories of seismogenic sources: individual seismogenic sources, seismogenic areas, macroseismic sources, it does not work well to elaborate a PSHA model using individual seismogenic sources, as in this work. In fact, the DISS Authors (Basili et al., 2008) say that the individual seismogenic sources database cannot guarantee the completeness of the sources themselves and are not meant to comprise a complete input dataset for probabilistic assessment of seismic hazard. For this reason, we are not restricted to just use of the DISS, but trough a synthesis of published works over the last twenty years (see supplements for complete references) we defined a database as complete as possible, in terms of individual seismogenic sources, and parameters to have input dataset for PSHA."

L63: 66 The uniform seismotectonic sources of the Italian hazard described by Stuchi et al (2011) are delineated considering the fault information where and when available. The more realistic pattern of ground motion due to faults it is questionable, because an area source delineated to describe a group of faults, it will produce a similar pattern with the individual faults. The major benefits of using the active faults is to extend the observational time to capture the recurrence of large magnitude events. The local pattern due to fault location might be controlled by other factors such as hanging wall, upper seismogenic depth, style of faulting. However, these effects are not evident if an inappropriate ground motion model is selected. Thus the seismic hazard pattern depends on both seismic source representation and ground motion models.

We will modify from line 65: "…in order to obtain more detailed patterns of ground motion, extend the observational time to capture the recurrence of large magnitude events, and to improve the reliability of seismic hazard assessments." Moreover, we will add a new figure (Figure S1 in supplement) to compare the MPS04 and our PSHA model

L72. The term models is misleading. A source model implies a complete source representation in space and time aimed at describing the seismogenic potential of the region. In the current context, the active faults are incomplete in space, they are not describing all the tectonics of the region - not volcanic, subduction or deep seismicity reported for the Italian territory. It has to be specified that these are individual seismic sources, but not independent models. The procedure proposed here is aiming at creating a "model" for an exercise of seismic hazard evaluation. Moreover, if the goal of the work is to provide a robust seismic hazard estimates, then the authors resolve the issues of model independence and completeness as well as to capture the epistemic uncertainties in the mixed source model.

We agree with your comment, and so following your suggestion we'll remove the term "model" when we describe the fault source geometry, while we'll maintain the term "model" when we combine fault and distributed sources for the seismic hazard evaluations. In any case we want to highlight that the main aim of this work is how to combine fault and distributed sources in order to take into account and possibly overcome the incompleteness of the fault source database, without throwing data. We will add in the manuscript a phrase explaining our choices.

L120: The time scale is a key aspect to evaluate the long-term representation of the seismic productivity of active faults. If a fault has moved in the recent geologically time , i.e Holocene, it might be considered as seismically active, if it moved in the far-off geologic time and has not moved again since then the fault might be judged to be an inactive fault. Hence, it might be of interest to specify the time scale and the definition of active faults on the present investigation. Yet, as mentioned before there is need to clarify the definition of fault activity or non activity.

Please see our comment above on active fault definition (remark n. 2).

L131:135. The slip rate values for some faults are very low. Values of 0.3 mm/year are extremely low and the movement on these faults could also takes place as creep. Is the aseismic factor adjusting the slip rates? Are these slip-rates supported by historical seismicity observations, geological investigations and /or paleoseismicity studies?

These slip-rates are supported by historical seismicity observations, geological investigations and /or paleoseismicity studies as reported in the supplement files. Moreover we are assuming that the used slip-rates are representative of seismic movements (no-aseismic factor), as discussed above (remark n.2).

L152: The name could be “Segmentation rules for delineating (or aggregating) fault sources”

Thanks for the suggestion, we will modify it.

L199: The role of aspect ratio must be discussed in greater extend than currently version. The extension along-strike dimensions of the faults seems to be constrained by this parameter.

We will rephrase from line 199 as: ”…by 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 as derived by Peruzza and Pace (2002).”

L191: There are five Mmax values for each fault. How is the Mmax modelled in the hazard calculation?

Please, see the comment to remark n. 8

L202: Introduce and explain the “segment seismic moment conservation”? The key assumptions and the input parameters of the recurrence rates must be described. Characterisation of the active faults is a key aspect of this approach, thus it requires more description. As mentioned before, the effect of maximum magnitude must be discussed. In the case of seismic moment balance, for a constant slip rate, the recurrence rates of small events are decreasing with increased magnitude.

We will introduce and explain better this issue. Please, see the replies to remarks n. 2, 4, and 9.

L207:211: What is the rationale of the two MFDs? It is not evident why the two recurrence models are selected? In a general way, the characteristic earthquake is used to define an earthquake of a given magnitude and well identified recurrence time by geological evidences. The fault sources used here do not qualify for such model, for various reasons including the way they are constructed by linkage of various segments. A characteristic model will be appropriate for use on individual segment rather than a long composite fault. See discussions of Kagan (1993), that clearly states that the evidence of the characteristic earthquake hypothesis can be explained either by statistical bias or statistical artifact. Thus, it will be of great interest for the readers to specify the assumptions for the two MFDs.

We agree that it is difficult to define an appropriate MFD (e.g. characteristic earthquake) for individual source using the available geological data, and important project as UCERF3 didn't solve the same doubts. In any case our fault source database have been developed to be representative of the maximum single earthquake rupture, and not long composite faults, by using restrictive segmentation rules described in chapter 2.1.2. Moreover, the two MFD are used as end members, in order to explore the epistemic uncertainties, and we consider the Mixed model as a sort of an "expert judgment" model, useful for comparison analysis.

L278: the number of Voronoi polygons is not clear to me. There are 3 to 50 polygons across the entire region? Each polygon is tectonic dependent? Please clarify.

We will modify the manuscript from the line 276: "… the Voronoi tessellation of space without tectonic dependency. The whole Italian territory has been divided into a grid with a longitude/latitude spacing of 0.05°, and the centres of the grid cells represent the possible centres of Voronoi polygons. We vary the number Voronoy polygons, Nv, from 3 to 50, generating 1000 tessellations for each Nv."

L286: Who is parametrised the depth and the maximum magnitude for gridded seismicity? Are these parameters treated as aleatory or epistemic?

The parameters have been taken from SHARE project, as written at lines 285-291. We did not explore the variability of these parameters.

L382: For the purpose of an exercise one GMPE might have been justified. However, the focus of the study should be the comparison of the earthquake recurrence rates not the hazard estimates.

We believe that the use of these GMPE's is correct, as they have been developed for Active Crust regions. Comparing model in terms of rates is for sure a valid approach. However, as the aim of our work is a PSH model, we believe that comparing different model (using the same GMPE's) can be useful. In any case we'll add in the manuscript a figure comparing the results also in terms of activity rates (Figure S2 in supplement).

Cited papers

Basili, R., G. Valensise, P. Vannoli, P. Burrato, U. Fracassi, S. Mariano, M. M. Tiberti, and E. Boschi. 2008. 'The Database of Individual Seismogenic Sources (DISS), version 3: Summarizing 20 years of research on Italy's earthquake geology', *Tectonophysics*, 453: 20- 43.

Leonard, M. (2010). Earthquake fault scaling: Self-consistent relating of rupture length, width, average displacement, and moment release. *Bulletin of the Seismological Society of America*, *100*(5A), 1971-1988.

Machette, M.N., 2000, Active, capable, and potentially active faults; a paleoseismic perspective, *J. Geodyn.* **29**, 387–392.

Pace, B., F. Visini, and L. Peruzza. 2016. 'FiSH: MATLAB Tools to Turn Fault Data into Seismic-Hazard Models', Seismological Research Letters, 87: 374-86.

Peruzza, L., and B. Pace. 2002. 'Sensitivity analysis for seismic source characteristics to probabilistic seismic hazard assessment in central Apennines (Abruzzo area) '. Bollettino di Geofisica Teorica ed Applicata 43, 79–100.

Wells, D. L., and K. J. Coppersmith. 1994. 'New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement', Bulletin of the Seismological Society of America, 84: 974-1002.

Woessner, J., D. Laurentiu, D. Giardini, H. Crowley, F. Cotton, G. Grunthal, G. Valensise, R. Arvidsson, R. Basili, M. B. Demircioglu, S. Hiemer, C. Meletti, R. W. Musson, A. N. Rovida, K. Sesetyan, M. Stucchi, and SHARE Consortium. 2015. 'The 2013 European Seismic Hazard Model: key components and results', Bulletin of Earthquake Engineering, 13: 3553-96