Assessing Seismic Hazard of the East African Rift: a pilot study from GEM and AfricaArray

Valerio Poggi • Raymond Durrheim • Georges Mavonga Tuluka • Graeme Weatherill • Robin Gee • Marco Pagani • Andrew Nyblade • Damien Delvaux

Manuscript BEEE-D-16-00536 – 1st revision

INTRODUCTION TO REVIEW REPLY

We would first like to thank the two reviewers and the editor for their time in reviewing this manuscript and for the appreciation of our work. We found their comments very useful and fair; these gave us the chance to improve the manuscript, clarify some aspects of our research and trigger new ideas for future developments. As requested, a moderate revision of the manuscript has been undertaken. All the proposed suggestions have been carefully considered and, where necessary, applied. Controversial issues have been explicitly addressed with a (hopefully) comprehensive and detailed reply, highlighting our thoughts and motivations.
All minor comments not directly addressed in this reply, are accepted and implemented by the authors without further discussion. Identified mistakes have been fixed. Please note that all corrections are marked in red on the manuscript, to facilitate the reviewers in the referencing.

Due to rearrangement of figure 1, figure numbering is inconsistent between the current and the previous versions of the paper. For clarity, in the reply we always refer to the numbering of the old version, unless explicitly stated otherwise.

Legend:
- **R1** – Original comments from Reviewer #1
- **R2** – Original comments from Reviewer #2
- **AR** – Author's reply to comments

Reply to Reviewer #1 Comments

Introduction to Reply

R1: The paper is a very good example of seismic hazard assessment, the elaborations sound precise, almost always well documented, and the text is clear and well written.
There are a few aspects that, in my opinion, need more explanation to be fully understood. Moreover, all names cited in the text should be reported in a figure (Fig. 1?) and most figure captions need more description.
In conclusion, the paper needs a moderate revision to be accepted for publication.

AR: We express our gratitude to Dr. Dario Slejko for the appreciation of our work. We present in the following a hopefully exhaustive description of all the modifications we have implemented to face the raised issues, which we discuss point by point.

Detailed Reply

1) R1: Page 4 - Silver is Silveri in the reference list.

AR: The reference has been corrected.

2) R1: Page 4 - Field (2003) is Field et al. (2003) in the reference list.

AR: The reference has been corrected.

3) R1: Page 9 - Fig. 1 merits a larger size and names are difficult to read.

AR: The labels were indeed very difficult to read, we agree. We have therefore increased the font size and the space allocated to Figure 1, by increasing the map size and by moving the magnitude distribution plot in a new figure (now Figure 2, which also includes now an informative plot of the cumulative number of events over time). More importantly, according also to initial suggestion from the reviewer, we have added those labels corresponding to the most significant tectonic features (the rift systems) cited in the manuscript and associated with observed seismicity. However, to avoid overloading of the map, we restricted the labels about African capitals to those city analysed at the end of the paper, which we believe are relevant for the reader.

4) R1: Page 9 - Vilanova (2014) is Vilanova et al. (2014) in the reference list.

AR: The reference has been corrected.

5) R1: Page 10 - Explain in the figure caption the meaning of the red lines inside the zones.

AR: The caption includes now a description of the sub-regions marked as coloured solid lines within area source 7, 10, 11, 12 and 13. Please note that, also in agreement with the suggestions from Reviewer #2, we have further improved the map (now Figure 3).

6) **R1**: Page 10 - Cite the number of the source areas in the description of the 6 tectonic groups.

AR: Grouping of source areas is presented in Table 3 with corresponding IDs. Since the table is referenced at the beginning of section 5, we believe that introducing this information in the paragraph could be somehow redundant. Therefore, we would prefer no to include it.

7) **R1**: Page 12 - Yang and Chen (2008) is Yang and Chen (2010) in the reference list.

AR: The reference has been corrected.

8) **R1**: Page 12 - Explain how you have considered the 2 interpretations of TRM in your source model.

AR: The mixed interpretation relies on to possibilities offered by OpenQuake in modelling ruptures using discrete probability distributions. In such case, instead of modelling such variability as a pure epistemic uncertainty, we implemented a mixed (bimodal) probability distribution rake and dip orientations for the ruptures of the zone. We have now better described the case for the TRM at the end of section 5.3

9) **R1**: Page 15 - You show the b-value for the whole catalogue in Fig. 5a but this estimate seems not used in the computation. What is it worth?

AR: The estimate for the whole catalogue was actually used in our calculations in many ways, although we agree with the reviewer this was not transparent from the manuscript. Primarily, we used such MFD to verify our completeness for the region, subsequently adjusted to match local heterogeneity in spatial distribution of seismicity. Additionally, the b-value from the full catalogue is used as starting model for the inversion of the incremental rates of the different area source groups (we use a steepest-descent like approach for optimization). Moreover, we used the total MFD as a sanity check for comparison against the sum of the MFD computed for each area source, to make sure that the overall rate balance was preserved.

Given that, we nonetheless agree with the reviewer that just presenting the fit for the whole catalogue could be misleading. We have now included in current Figure 7 (formerly Figure 5a) the plots corresponding to the MFD calibrated for each separate tectonic group, whose b-values are then reported in Table 4.

10) **R1**: Page 16 - Explain better your method based on the incremental rates in the text and caption of Fig. 5a. I do not understand what you mean with incremental occurrences: I associate increment to the difference between 2 rates but this is not what I see in Fig. 5a. You say you minimize the residuals between observed rates and truncated MFD: this seems to me the application of the least squares method to the non-cumulative observed rates.

AR: We use the term incremental to describe non-cumulative rates. We kept this naming convention in agreement with the standard notation used in OpenQuake, to avoid

inconsistencies (although I personally acknowledge it is probably not the most straightforward definition). We added the "non-cumulative" definition in the paragraph, for clarity. The caption of Figure 5 (now Figure 7) has also been extended to better describe the figure content.

As the reviewer was pointing out, theoretical incremental rates are computed as the differences between consecutive cumulative rates, while observed incremental rates are simply the number of occurrences (per unit time) in a certain magnitude range.

The reviewer is right a about his last point; the method we implement is basically an adaptation of a non-linear least square approach applied to non-cumulative rates. This is now specified in the manuscript.

10) **R1**: Page 16 - Improve the caption of Fig. 5b: what do the yellow and rose colour mean? What is the difference between the dashed and solid red lines?

AR: We have substantially modified the plot in Figure 5b and then moved it to a separate figure (now Figure 6). The caption has also been improved accordingly.

12) **R1**: Page 16 - Explain better your manual adjustment to identify the complete periods.

AR: We have now described more in detail the procedure of iterative manual adjustment of the completeness periods by substantial rewriting the section (now numbered 6.3.1). The section has also been moved before the discussion on MFD, according to suggestions of reviewer #2.

13) **R1**: Page 20 - The logic tree in Fig. 7 shows different branches only for GMPEs. How are the 2 interpretations of TRM used? How is the incremental rate method applied?

AR: Regarding the interpretation of TRM, we refer the reviewer to point 8 of this reply, where we address the issue about the uncertainty related to such interpretations.

About the use of the procedure of incremental rate fit, no direct representation the related fit uncertainty is given in the logic tree. More precisely, the fitting method we implemented is potentially capable to solve uncertainty on *a*- and *b*-values, given that a sufficiently reliable estimated of the error associated to the incremental rates is provided in input. However, not to overload the computation and given the explorative nature of the presented model, we decided not to include such type of uncertainty in our model, while better focusing on the tectonic regionalisation of the GMPE (though regional groups) and the variability of the maximum magnitude.

We have now extended the caption of Figure 7 (now Figure 9) to better present its content, and clarifying as well some aspects on source model uncertainty in section 7.2.

14) **R1**: Page 22 - Only the 10% PoE is reported in Fig. 9 not different PoEs.

AR: This is correct. Although we tested other PoEs, it was our decision to only include and discuss maps for the 10%PoE given the already considerable length of the manuscript. We acknowledge that our choice was not clearly presented in section 8.2.2 (particularly in the first sentence, which was definitely misleading). The paragraph has then been modified accordingly.

15) **R1**: Page 24 - Specify the PoE in the text and caption of Fig. 10 and cite the overlapping curves in 2 panels.

AR: We have now extended the caption of Fig. 10 (Now Fig. 12) by adding the missing information on the PoE used for the calculation and, as suggested, by emphasizing the matching of the mean and 0.5 percentile curves in the example plots.

Reply to Reviewer #2 Comments

Introduction to Reply

R2: The paper presents a concise, well-written and important study on the PSHA for Eastern Africa. The study region encompasses both intraplate areas and areas of active tectonics, associated with the Easter African Rift System. The nature of the underlying data poses significant challenges which the authors thoroughly describe and discuss. The results are presented in an appropriate way, including a hazard map at 10% probability exceedance in 50 years, and both hazard curves and uniform response spectra for four African Capitals (Addis Ababa, Kampala, Nairobi and Bujumbura).

There are, in my view, two main concerns respecting the PSHA presented. The first is related to the source zones model. […]

The second issue is methodological and concerns the completeness analysis presented […] Therefore, I strongly believe that completeness should be addressed before estimating activity parameters and not simultaneously.

AR: The reviewer raised a number of important issues that were certainly relevant to improve the clarity and transparency of the present work to the scientific community. It is our hope to have exhaustively addressed all this issues in this reply and within this new revision of the manuscript.

In particular, the issue about epistemic variability of source zonation has been addressed at point 14, where we acknowledge its relevance and the need to further better present to the reader the unavoidable limitations of the present work. As indicated, the manuscript has been modified accordingly.

The issue about completeness analysis is then addressed at point 9. The reviewer is right and there was indeed a lack of clarity in the paper. We DO perform completeness analyses before estimating activity, and only later the process is iterated to converge to an optimum solution. As we better explained in the reply, the sections have been rearranged to better present this idea.

Detailed Reply

1) **R2**: Page 7, section 4.2 - "*By mapping the activity period of the different seismological agencies over time, we identified five main time intervals and adopted a different agency prioritization scheme for the selection of the best available location within each*"
How do the authors perform such assessment? Is it by the quantity of data published by each agency?

AR: The selection was quite trivial for some periods such as the historical (<1900) and the pre-instrumental (1901-1959). For the remaining three periods, however, the

selection was more complex, primarily because of the substantial lack of available information (metadata) about some African networks and then because of the need to implement selection/priority rules applicable at regional scale. When possible, we made use of available information from online literature and reports (e.g. quality of the location solutions, period of activity of the network, overall spatial coverage). In some uncertain cases (such as the network BUL – Goetz Observatory, Zimbabwe) we had to directly contact local experts. As specified in the manuscript, often we had to just rely on the analysis of the spatial-temporal distribution of located events (see following figure A of this reply to for an example picture of agency-specific analysis we performed). We have added a sentence in section 4.2 to complement the description of the selection criteria.

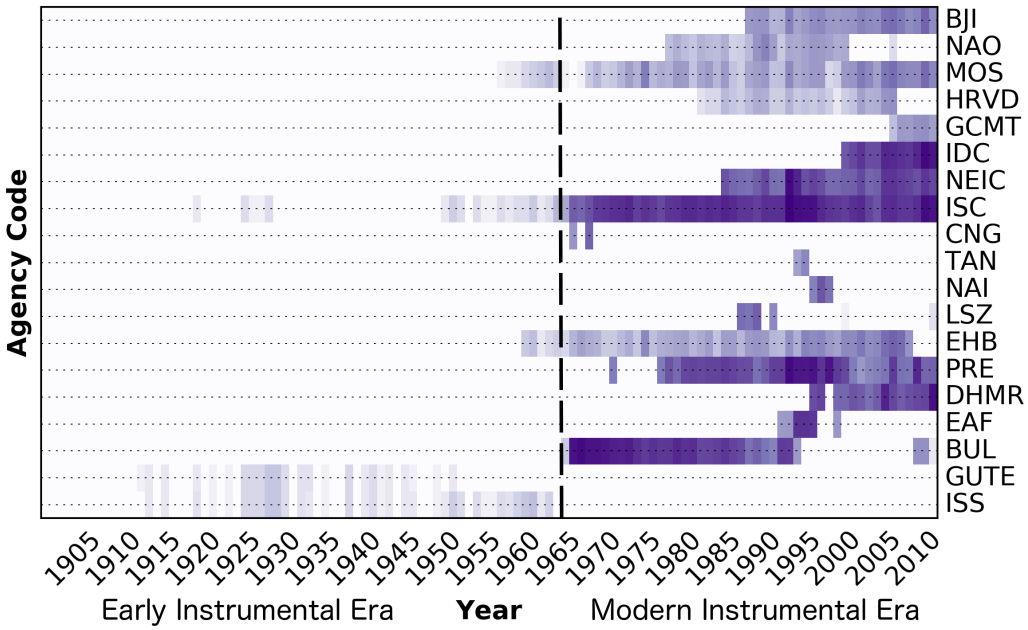
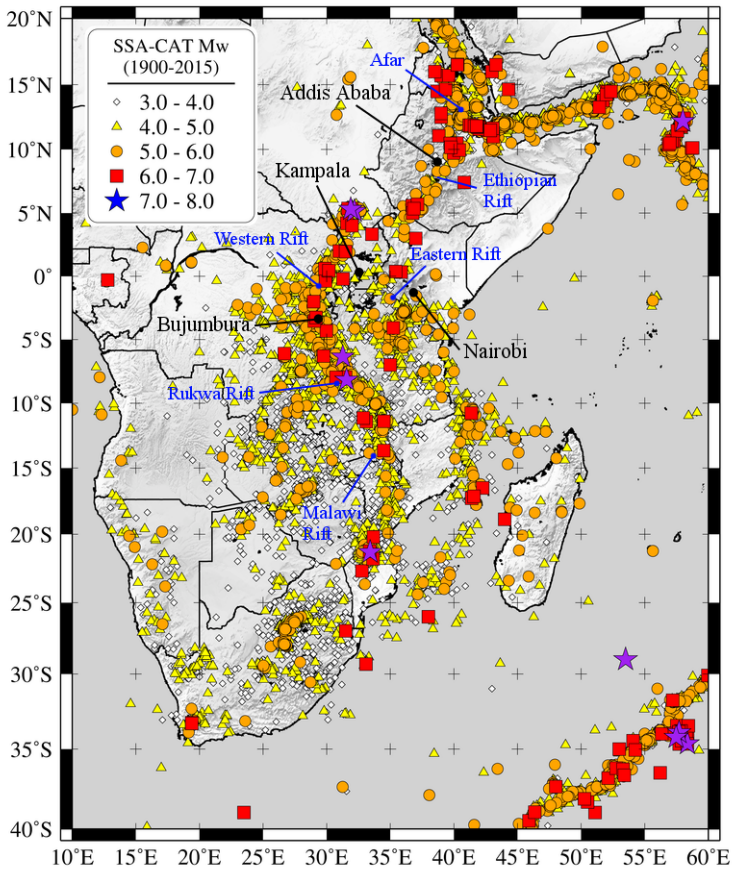


Fig. A - *Time histogram showing the density (relative number, normalised) of events reported for some selected agencies of the ISC-REV bulletin which are potentially relevant for SSA hazard model.*

2) **R2**: Page 9, Figure 1 - In the caption, SSA-GEM should be used instead of SSA. The location names use a very small font and are difficult to read. In addition, I believe there is too many colours in the figure (it's a bit distracting). I would suggest the authors to use a grey shaded topographic map instead of the coloured one (but this is of course a matter of personal preference).

AR: We agree with the reviewer, the labels are indeed very difficult to read. Also in agreement with the issue #3 raised by the same Reviewer and the suggestions from Reviewer #1, we have refactored the map by better selecting the presented location names, using also a larger and more visible font. We now present the location of analysed African capitals and a selection of the most relevant tectonic information (rift systems) cited in the manuscript, which is useful to orient the reader through the main tectonic domains of the EARS.

The figure was also increased in size, by moving the accessory plot (magnitude-time distribution) in a separate figure (now Figure 2). The caption was also modified according to suggestion. About colour, however, we would like to keep the presented style, as we were not fully satisfy by a grey scale for the landscape (see the result below):



3) **R2**: Page 9, section 5 - SSZ where defined on the basis of seismicity and tectonic/geological information. The seismic catalogue used is illustrated in Figure 1. However, the underlying tectonic information on which the zonation is partially based is not available to the readers. I strongly suggest the authors to include a figure depicting a simplified tectonic/geological map. This figure should include the names of the tectonic features that are described in the text (e.g., Main Ethiopian Rift, Afar Depression, etc.) to supply guidance to the readers not well acquainted with regional details.

AR: The suggestion of the reviewer is well received, as there is a clear need to represent the main tectonic features that are relevant for the area. We were initially thinking of including a separate tectonic map, but we realised this would have consumed considerable space, given the present length of the manuscript. Alternatively, we underwent a reorganization of the Figure 1 and Figure 2 (now Figure3).

In particular, we have now included in Figure 1 a selection of the most significant toponyms, useful to orient the reader though the manuscript (see also reply to previous Reviewer's comment). Moreover, Figure 2 (now Figure 3) includes now the fault traces from the database of Macgregor (2015), which have been extensively used for the characterization of the area source model, together with the SSA-GEM catalogue. Caption has been modified accordingly.

4) **R2**: Page 9, section 5- Do the authors assume that no earthquakes will occur outside of the delineated seismic source zones? This is a rather strong assumption as there are many examples of M6 earthquakes having occurred in regions with no previous significant seismic activity. Usually a background source is used to account for this type of unexpected events in regions of sparse seismicity. This background source would raise the hazard in regions of low to moderate seismicity (see for instance Frankel et al., 1996). It would probably have no effect on the hazard curves and uniform response spectra calculated for the selected cities.

AR: As the reviewer was pointing out, our mandate was to focus mostly on the evaluation of the seismic hazard over the different EARS segments, which are not sensibly influenced by the low hazard of the outside stable crust. Moreover, calibrating a proper background seismicity was in this context quite difficult, given the insufficient amount of (complete) data needed to properly constrain the MFD for the background.

Nonetheless, we acknowledge that the definition of background source model for the stable crust is a necessary following step. We are indeed presently working on the integration of the EARS model into a wider hazard model for Africa continent, which will also include such feature. Analysis is however still in progress.

5) **R2**: Page 10, Figure 2
The seismic catalogue presented in this figure is the declustered SSA catalogue depicted in Figure 1? In page 6, section 4.1.5 the authors state that local catalogues have been used to define potentially seismogenic structures… Is this type of data included in Figure 2? Please specify catalogue and the magnitude threshold.

AR: The catalogue presented in the background of Figure 2 (now Figure 3) is indeed the original SSA-GEM (not-declustered) catalogue, presented here from magnitude 3 and above. Indeed, the SSA-GEM catalogue includes the events from the local catalogues listed in section 4.1, although a magnitude threshold filtering was applied.
As the reviewer suggested, we have added this information in the caption.

6) **R2**: Page 10, section 5 - "we further define sub-regions with higher observed seismicity"
By doing this the authors assume that the pattern defined by the seismicity is stationary. That assumption is debatable as the catalogue is probably non-representative of the long-term behaviour of the seismogenic processes (see for instance, Swafford and Stein, 2007, for a discussion on this issue).
In case the seismicity is though to be stationary a smoothed seismicity approach would probably be better suited to represent it. I would encourage the authors to include some discussion on this issue to in the paper.

AR: The reviewer is correct about the potential advantage of using a smoothed seismicity model, and as a matter of fact we have considered this option in the initial phase of this study. However, we found it inconvenient for two reasons. First, a smoothed seismicity model is better suited for regions with at least a moderate seismicity. In the case of EARS the occurrence is sparse enough to produce considerable "gaps" when such a model (e.g. a Frankel approach) is implemented. Furthermore, a smoothed seismicity model cannot be simply conjugated with tectonic evidences (e.g. mapped faults), which we also target in our study and were then used to delineate the shape of the area source zones and subzones.
According to reviewer suggestion, we have extended the discussion in section 5 to better present our rationale.

7) **R2**: Page 15, section 6.2 - "The overall strike distribution was calibrated by performing statistical analysis on the outcropping fault structures available from the database of Macgregor (2015)". What was the procedure for source zones that did not enclose any active faults (stable regions)?

AR: When direct constraints from outcropping faults were not available (or not sufficiently reliable) we based our judgement on distribution of moment tensor solutions and tectonic regime descriptions from literature. In those cases were also this last was not available, we allows an isotropic (equal-probability) distribution of rupture orientation.
We have added such description a the end of section 6.2

8) **R2**: Page 16, section - "In addition to using standard and well-established approaches (e.g. Weichert's maximum likelihood method; Weichert, 1980), we tested an alternative strategy".
Did the authors use both approaches? How do both approaches compare? The Weichert's approach is known to be particularly well suited to the particularities of seismic catalogue data (e.g., varying observation periods)...

AR: Yes, we compared the approaches (this is now specified in section 6.3.2). The match is perfect for synthetics catalogues and reasonably good for large catalogues with a well defined completeness. Deviations were progressively more noticeable (although not dramatic) with short catalogues and particularly with badly calibrated (usually underestimated) rates for incomplete long return period events. We target to present the result of such comparison in a separate publication, to provide a more exhaustive discussion on the matter.

9) **R2**: Page 16, section 6.3.2
"This method, however, proved to be unstable, giving potentially erroneous results in the case of sparse and irregular data coverage, as it is unfortunately the case for Sub-Saharan Africa. As subsequent refinement, therefore, we manually adjusted the completeness estimates by iterative comparison of the corresponding magnitude-frequency distribution."
It would be probably better using the traditional Stepp (1972) method instead of the unsupervised Stepp method. There seems to be enough data to get reliable results with the Stepp method.

AR: As we mentioned, we indeed used the Stepp method (1971, when referring to the PhD thesis or 1972, when referring to the subsequent publication). We are not aware of an original and an unsupervised version of the algorithm.
As we stated in the manuscript, we were not fully satisfied by the results of the Stepp method (it proved to provide a rather irregular GR), although the first round of evaluation was actually done with it. We felt that a further refinement was then necessary. This was performed by manually adjusting the completeness while visually inspecting the rate variation over time and (most importantly) by minimizing the residuals between observed rates and the theoretical GR relation, assumed representative of the seismicity for the area.

R2: The authors state that the analysis was performed for each source zone? Are there any reasons to suspect that different geographical areas within the study region would have different completeness periods?

AR: Yes, this is necessary to account for non-uniform spatial distribution of seismicity, which would lead to coarse error in calibrating of local MFD if disregarded. It is important to mention that, although the completeness is varied between zones (to the less possible extent, clearly), the total moment computed from the different zones should match the one obtained from the whole region. We actually did this test with a positive result, which gave further confirmation on the correctness of our analysis.

R2: I strongly believe that the completeness analysis should be performed independently from the recurrence parameters, otherwise no guarantee can be given that the recurrence values are meaningful.

AR: We agree on the fact that the completeness analysis should be performed before any attempt to calibrate an MFD. This is actually what we did, although we did not present it properly in the paper, since the completeness section was located after the MFD calibration section. We have now rearranged these sections for clarity, also providing further clarification on the procedure.

R2: In fact, the completeness results seem a bit suspicious to me. For instance, in Figure 5b one can read that for M-5.5-6.5 the catalogue is considered complete since 1920. However, looking at Figure 1b, this doesn't seem correct for M5.5-6.0 (there is none earthquake within this magnitude range between 1920 and 1930, and always a few after that decade)...

AR: We also observed this. However, extending the completeness to 1920 provides the best estimate of occurrence rates for this magnitude range, compatible with a more regular GR-relation (with lower fit residuals). For that reason, we have interpreted such "hole" as a simple temporal fluctuation of the Poissonian process, unfortunately occurring at the lowermost edge of the investigation period. A similar behaviour is also visible around 1970, which is definitely within the complete period.

10) R2: Page 17, section 6.3.3

"This approach would be an intermediate approach between standard distributed and smoothed seismicity models."
I don't totally understand this statement because what characterizes the smoothed seismicity method is that the seismicity rates are smoothed-out. In the approach assumed by the authors the rates are concentrated inside the sub-source zone... If I understood correctly this approach is similar to that of a having two regular source zones, one of witch is connected but non-simply connected in the mathematical sense (e.g. some source zones have a hole).

AR: The statement is referred to the preceding sentence in the manuscript. In other words, by progressively increasing the number of overlapping sub-layers (only two were used in our case), the occurrence rates per unit area would smoothly vary along the whole source zone in a way that is "similar" to a gridded-like seismicity approach. Please also note that the overlapping layers have no holes. Each layer simply adds a rate contribution to a specific sub-region of the zone, by keeping nonetheless the total rate balance unmodified. The original sentence has been modified to better present these considerations.

11) **R2**: Page 17, Table 4 - Other source zones' characteristics that are used for PSHA calculation should be included (e.g., source mechanism, source depth).

AR: We avoided including this type of information in the table (or in a separated table, as we were originally testing) because of the rather complex way these quantities are parameterised (as discrete probability distributions). Here is an example for zone 8:

```
...
<nodalPlaneDist>
    <nodalPlane dip="60.0" probability="0.125" rake="-90.0" strike="150.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-90.0" strike="330.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-90" strike="10.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-90" strike="190.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-45.0" strike="150.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-135.0" strike="150.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-45.0" strike="330.0"/>
    <nodalPlane dip="60.0" probability="0.125" rake="-135.0" strike="330.0"/>
 </nodalPlaneDist>
 <hypoDepthDist>
    <hypoDepth depth="5.0" probability="0.17177915"/>
    <hypoDepth depth="15.0" probability="0.35582822"/>
    <hypoDepth depth="25.0" probability="0.32515337"/>
    <hypoDepth depth="35.0" probability="0.14723926"/>
 </hypoDepthDist>
```

It has to be noted, however, that such information will be still available to the reader from the source model input files (in .xml), which will be openly available though the GEM platform. We are presently also evaluating the possibility of including the model input files (in XML format) as electronic supplement of this paper, if the journal allows. This would definitely make the model more transparent and accessible, also in agreement with reviewer's request for clarification.

12) **R2**: Page 23, Fig 9 - The location names use a very small font and are difficult to read.

AR: The names of the African capitals have been removed from the plots, as it would have been difficult to make them more visible for this picture size. Nonetheless, as we already mentioned, the most important toponyms are now better presented in Figure 1 for referencing.

13) **R2**: Page 24, Fig 10 - The authors should include in the caption the probability level for which the uniform hazard spectra is represented.

AR: Corrected according to suggestion

14) **R2**: The hazard results look robust, although the associated epistemic uncertainty seems lower than expected for the type of underlying data available. The uniform hazard spectra calculated for European cities (Woessner et al., 2015, their figure 21) shows significantly higher differences between the 15% and 85% percentiles. As pointed out previously in this document, some sources of epistemic uncertainty, in particular the uncertainty respecting seismic source zonation, were not accounted for. I believe that an alternative seismic source model should be considered to better quantify the epistemic uncertainty. That being said, I acknowledge that the authors clearly stated in the paper that their results should be considered as a starting point and not as final product. I would, at any rate, encourage the authors to discuss the extent to which they consider that the epistemic uncertainty is correctly captured in the presented PSHA.

AR: We agree with the reviewer. There is certainly a need for better exploring the epistemic uncertainty associated with the source model, with a particular focus on alternative zone geometries. As the reviewer noted, this limitation is related to the quite focused initial goals of the project. The proposed model should be therefore considered just "explorative" at this stage. Presently, it has been produced with the consensus of a restricted pool of experts. Nonetheless, being the model freely and openly accessible (according to GEM's philosophy and mandate), we believe that, with progressive integration and feedback from the wider community of African scientist, the model will substantially expand and improve in future.
According to suggestion, we have extended the discussion in section 7.2 (Source Model Uncertainty), better presenting now the limitations of the present model and its associated uncertainty.