Dear Prof. Chan, thank you for your comments and corrections. Please, find in this document the response to the questions and concerns for the second version of the manuscript.

Major concerns:

1) **1D Gaussian function: I am puzzled as to why authors have proposed a new smoothing method based on a one-dimensional distance framework. This assumption might distort the distribution of seismic activity. Consider Figure 1d, for instance; the method presupposes a uniform seismic rate around the target site's periphery (full ring), even though there are no seismic sources to the southeast or northwest. Furthermore, there are already several established smoothing methods utilizing two- dimensional (e.g., Frankel, 1995; Woo, 1996) and even three-dimensional (e.g., Chan, 2016) distance frameworks. The rationale behind introducing this new approach remains unclear to me.**

Response:

We apologize for the confusing explanation; the Gaussian function in this work is a generalization of the one presented by Frankel (1995) with the difference that the main parameters are identified using different geophysical magnitudes, and it is not always centred in zero, hence it is not 1D. The presentation of the smoothing function has been changed in order to be clearer:

"*For this work, a modification of the kernel proposed by Frankel (1995) has been used to smooth the gridded seismicity (Eq. 6):*"

As for the Figure 1d the main consideration for the full ring to be created is that the distance from the spatial cell to both seismic sources is close, and also lower than the mean distance in between the sources in the region. For this to happen if the faults are the main seismic sources, it means that the part of the area of study being considered is densely populated with faults, in which case, considering the full ring could account for the seismicity distribution. The main difference with the classic Frankel (1995) approach would be that the weight is greater for the traces of the faults and their surroundings. The Figure 1 has been modified to better suit the explanations in the section.

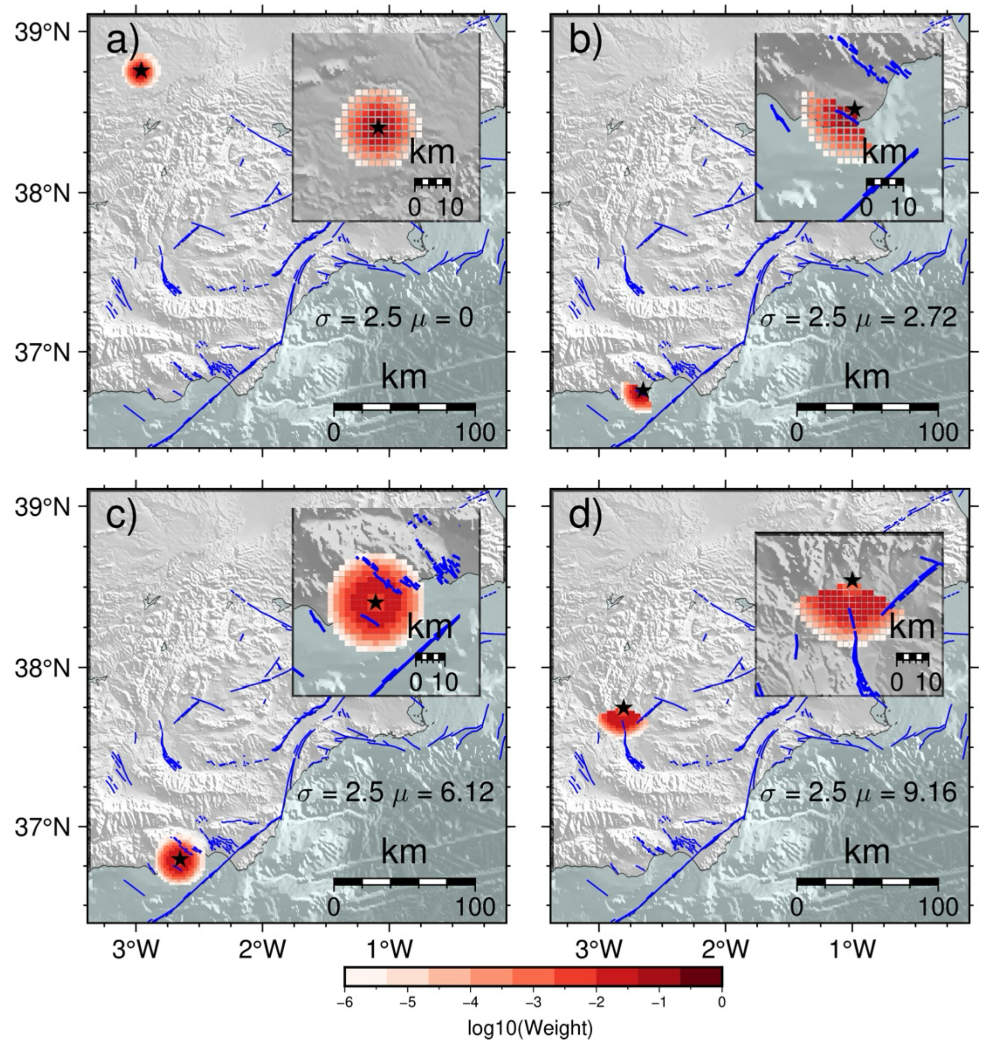


Figure 1. a) Smoothing function for $\mu = 0$. This can happen when either the distance is greater than d_c or when the spatial cell is over the seismic source. b) Smoothing function for $\mu \neq 0$ i.e., when only one seismic source is present and with distance lesser than d_c. c) Smoothing function for $\mu \neq 0$ in the case that several seismic sources are surrounding the spatial grid cell at similar distances. d) Smoothing function for $\mu \neq 0$ when the spatial grid cell is near seismic sources with angular amplitude lesser than 180º. The blue lines show the fault traces. In this example d_c equals 48 km. The stars mark the spatial grid cell considered in each case. A zoom in has been added in each panel to highlight the spatial distribution of smoothing kernel values.

2) **<u>Discussion on credibility of the approach:</u> I am pleased to see the discussion on the results of this time-dependent PSHA, highlighting the impact on seismic hazard from each event. At the same time, I would recommend further discussion on the credibility of this new smoothing approach. When introducing a new forecasting method, establishing its credibility is crucial, and retrospective validation could be an effective means to achieve this.**

We agree with the reviewer of the importance of the discussion on the credibility of the approach. That's why we have applied the smoothing methodology and the time-dependent PSHA to two different countries with different seismic activity. In both cases we have discussed how the exceedance probabilities change before given earthquakes and we have concluded that for high seismic activity regions (Italy) the methodology provides results that may be used for taking decisions before the main earthquake. However, for low seismic activity regions (Spain) the methodology is not so effective (at least with the data that we have). Here probably we will need more time to check if with a bigger earthquake the methodology behaves as in Italy.

In order not to extend the length of the manuscript too much we have clarified this discussion adding an introduction paragraph after section 3

"*3 Case studies*

As explained before, the goal of our smoothing methodology is to test the viability of producing time-dependent seismic hazard results which may be used for taking decisions before the main earthquake. Therefore, now we will present and discuss the results obtained for two different regions with different seismic behaviour. Central Italy (high seismicity) and south-east Spain (low seismicity). We will check the if there are significant changes in the metrics before the occurrence of important earthquakes carrying out a retrospective validation of how useful the results are."

Response:

3) **<u>Validation of the declustering approaches:</u> To validate the declustering approaches, the authors compared the results with the 2011 Lorca earthquake seismic series as defined by Cabañas et al. (2011), detailed in Lines 320-327. However, it is challenging to assert that this series, defined by a previous study, represents the ground truth. In my view, the definition of an aftershock (how to determine if two events are related) is still contentious. I would appreciate some discussion on this topic in the manuscript if possible.**

Response:

We agree with the reviewer's comment. In fact, that's the reason why before explaining the comparison with Cabañas et al. (2011) we discussed the challenges on cluster definition using the works of Zaliapin and Ben-Zion (2020) and Anderson and Zaliapin (2023).

Anyway, we have also added your suggestion starting the sentence as follows:

"In spite of the difficulties in defining the clusters, Cabañas et al. (2011) carried out a detailed study on the 2011 Lorca's earthquake seismic series. This study is the best definition at the moment so we will use it to validate the best algorithm."

Minor Comments:

1) **b-value calculation (Line 227): Are there sufficient events to support the calculation of the b-value? Cases with an insufficient number of events (for instance, fewer than 100, as noted by Aki, 1965) could lead to greater uncertainties in the b-value. It would be beneficial to address this issue in the paper.**

   References:

   Aki, K. (1965). Maximum likelihood estimate of b in the formula log N= a-bM and its confidence limits. Bull. Earthquake Res. Inst., Tokyo Univ., 43, 237-239.

Response:

For the b-value computation the methodology explained in a previous work has been followed (Montiel-López et al. 2023). In this case, in order to ensure that enough events are considered in the analysis and that the b-value uncertainty remains low, a 1-year window has been selected.

Reference:

Montiel-López, D., Molina, S., Galiana-Merino, J. J., and Gómez, I. (2023). On the calculation of smoothing kernels for seismic parameter spatial mapping: methodology and examples, Natural Hazards and Earth System Sciences, 23, 91–106, https://doi.org/10.5194/nhess-23-91-2023

2) **Table 9: Some parameters are also from Table 9?**

Response:

We apologize for the mistake, the cross reference should have pointed to Table 8, this has been corrected so there is no circular reference.

First of all, we want to thank the referee_2 for the helpful comments and corrections that contribute to improve the quality of the manuscript. All of these questions have addressed in an orderly manner:

1) **The article discusses the uncertainty of epicentres, while at the same time seismic faults have no "width" but are defined by lines on the Earth's surface. This approach requires clarification in the text of the article.**

Response:

We agree that faults have been oversimplified by assuming they can be represented by their trace from QAFI database. We have decided to use their trace as most probable location near surface in order to define the maximum probability value for the seismicity smoothing model 1 (mu different from zero). The idea is to give more weight to seismic activity in the proximities of the faults. This has been done as for the two case studies the seismicity is shallow and mainly related to the faults.

Paragraph 1 from section 2.1.3 has been modified as follows:

"*In this section, some examples of how the smoothing kernel works are shown. In this case the seismic sources are faults in a shallow seismicity context, so the trace of such faults has been considered as the location with the maximum probability of having an earthquake. This approach has also been considered for the two case studies in this work. Three main scenarios have considered to showcase the smoothing kernel:*"

2) **Figure 1 - All four panels should be in the same scale. Figure 1c differs in scale from figures a, b and d. It is necessary to correct the parameter-values in the figure with the description in the text. For example, in the model description the parameter "mu" is defined as zero, whereas in Figure 1a it is given as 68.64. Explain the meaning of the colour code in Figure 1. To better understand the method of determining the parameter "mu" from fault locations, consider to add the additional lines to Figure 1. Swap figures 1d and 1c. Axis labelling should be added.**

Response:

Figure 1 has been corrected as there was an erroneous label in Figure 1a. The scales of each subfigure have been checked and a scale bar has been added to make all the subfigures comparable. The colour of each subfigure has been changed so there is only one common colour with its corresponding labelled colour bar. A different python library has been used to create the figures so there is regional context for each of the examples, in this way we hope it is not necessary to label the axis as they refer to the longitude and latitude.

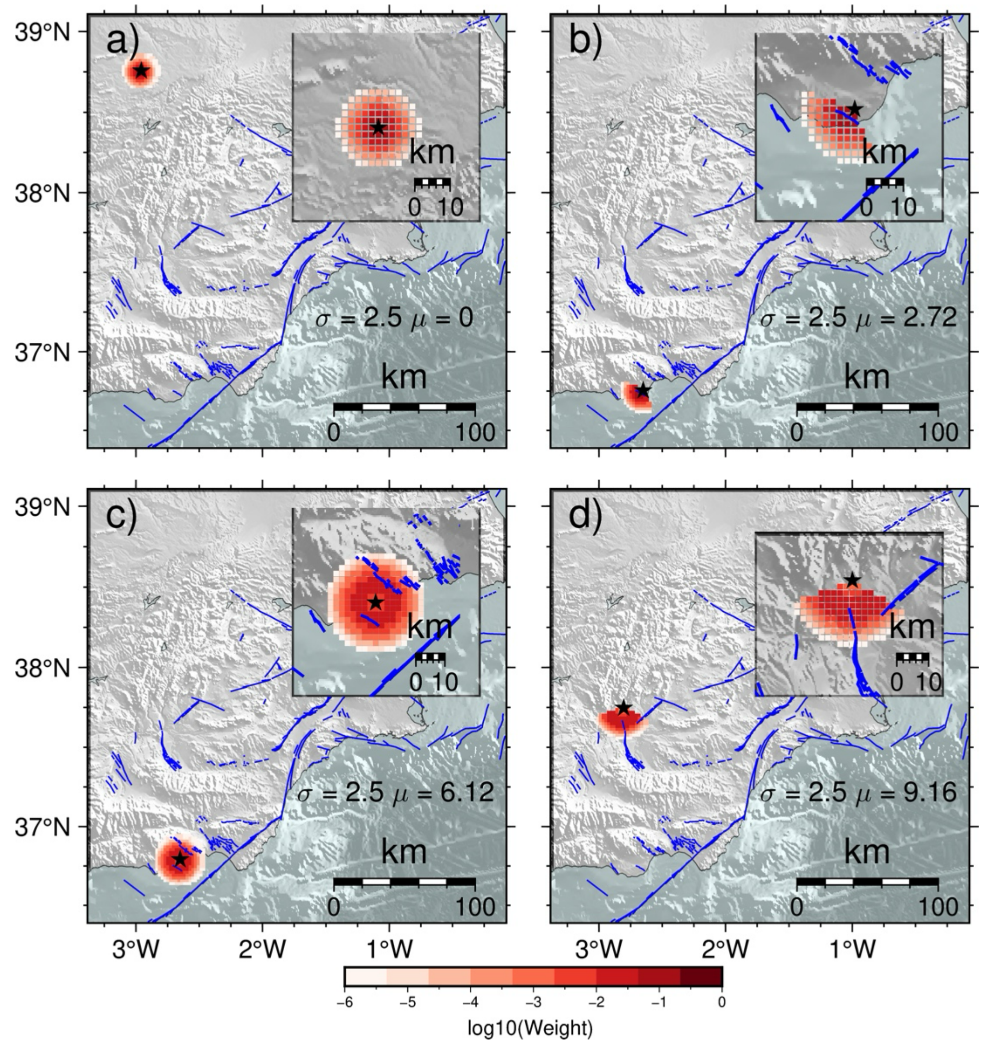


Figure 2. a) Smoothing function for μ = 0. This can happen when either the distance is greater than d_c or when the spatial cell is over the seismic source. b) Smoothing function for μ ≠ 0 i.e., when only one seismic source is present and with distance lesser than d_c. c) Smoothing function for μ ≠ 0 in the case that several seismic sources are surrounding the spatial grid cell at similar distances. d) Smoothing function for μ ≠ 0 when the spatial grid cell is near seismic sources with angular amplitude lesser than 180º. The blue lines show the fault traces. In this example d_c equals 48 km. The stars mark the spatial grid cell considered in each case.

3) Paragraph 2.2 lacks a description of the grid spacing used in the analysis. There should also be a discussion of the choice of grid spacing used.

Response:

We agree with the referee. The cell size of both case of studies is 0.015ºx0.015º which is equivalent to a ~1.5km^2 in surface. As for the choice of the cells the size we selected a size that could depict changes in the b-value and seismic activity rate spatial distribution and allowed for a reasonable computation time. Although for our study the deformation is not high (the longitude and latitude ranges are not wide) we prefer to use the ºxº notation to avoid uncertainty. An example of the variations can be read in Wiemer and Wyss (2002) where the authors the typical values for nodal separation range between 0.5 km and 10 km. In the case of Spain, we computed a grid similar to the one in a previous work in the area of study (Montiel-López et al. 2023). For Central Italy, examples of grid sizes can be seen in Murru et al. (2016) with a 0.025ºx0.025º grid or Gulia and Wiemer (2019) with a 2-km spaced grid, so we compromised to a slightly higher definition (although the higher computation time) in order to use the same grid for both case studies. In general, for both cases the choice is also motivated by epicentre uncertainty for the most recent events in the catalogue (that belong to the studied periods).

The first paragraph in 2.2 has been modified to explain the reasoning behind the choice of the resolution of the grid.

"*First, the spatial grid is defined by creating a rectangle spanning the maximum and minimum longitudes and latitudes of the catalogue with the desired resolution. The choice of the resolution can be motivated by similar studies in comparable tectonic settings or the order of the epicentral uncertainty of the earthquakes in the catalogue. For this work, in the case of Spain the same resolution as in a previous work in the same area by Montiel-López et al. (2023) has been used. In the case of Italy, although Murru et al. (2016) use a 0.025ºx0.025º grid and Gulia and Wiemer (2019) use a 2-km spaced grid for Central Italy, we decided to use the same resolution for both case studies, a 0.015ºx0.0.015º grid.*"

Also in section 3.1.1, a correction has been made in the number of points as it was incorrect in the previous version of the manuscript:

“A spatial cell grid of 0.015ºx0.015º spanning the above longitude and latitude ranges has been created (using 70756 points).”

And in section 3.2.1:

“A spatial grid of 0.015ºx0.015º covers the area of study (using 40401 points).”

References:

Wiemer, S. & Wyss, M. (2002) Mapping spatial variability of the frequency-magnitude distribution of earthquakes. Adv. Geophys. 45, 259–302. doi: https://doi.org/10.1016/S0065-2687(02)80007-3

Murru, M., Taroni, M., Akinci, A. and Falcone, G. (2016) “What is the impact of the August 24, 2016 Amatrice earthquake on the seismic hazard assessment in central Italy?”, Annals of Geophysics, 59. doi: https://doi.org/10.4401/ag-7209

Montiel-López, D., Molina, S., Galiana-Merino, J. J., and Gómez, I. (2023). On the calculation of smoothing kernels for seismic parameter spatial mapping: methodology and examples, Natural Hazards and Earth System Sciences, 23, 91–106, https://doi.org/10.5194/nhess-23-91-2023

Gulia L., Wiemer S. (2019) Real-Time Discrimination of Earthquake Foreshocks and Aftershocks. Nature,574, 193–199. doi: https://doi.org/10.1038/s41586-019-1606-4

4) **188- there is no reference to the section describing the comparison of several declustering algorithms.**

Response:

There was an error in the cross-reference pointing to section 3.2.1 that has been corrected:

“In order to decide which algorithm performs best on the data, a comparison between the RJ, A, and GK74 declustering algorithms has been made using default parameters (see section 3.2.1).”

5) OpenQuake or Openquake use a single caption.

Response:

All the instances in the text have been corrected to OpenQuake be coherent.

6) **299 – 304 The sudden mention of the NN method when discussing the results of the 3 already selected declustering methods is not clear.**

Response:

The mention to the works of the authors is related to the difficulties in assigning events to clusters and how it could affect to the seismicity hazard analysis, as this topic is discussed in the cited works. Nevertheless, the order has been changed so it comes before the table and is mainly focused on the results of the declustering.

"Table 6 presents the number of clusters and the events in clusters for the whole seismic catalogue. The RJ algorithm classifies a total of 652 clusters in the catalogue while GK74 detects 1012 clusters. The A algorithm identifies 1245 clusters. As can be seen, despite the three methods relying on windows for their calculations, there are significant differences in the results, not only in the number of clusters but also in the number of events inside each cluster.

Zaliapin and Ben-Zion (2020) pointed out these problems with the identification of aftershocks and main shocks and proposed an algorithm to discriminate between background and clustered events by randomly thinning a complete catalogue by removing nearest-neighbour earthquakes. Moreover, Anderson and Zaliapin (2023) examine the effect on the hazard estimation when using different declustering thresholds. They conclude that hazard estimates are most sensitive to the catalogue thinning near the aftershock zone, and less sensitive elsewhere.

In spite of the difficulties in defining the clusters, Cabañas et al. (2011) carried out a detailed study on the 2011 Lorca's earthquake seismic series. This study is the best definition at the moment so we will use it to validate the best algorithm. They identified 146 events (including the foreshock, the main shock and the aftershocks) that belong to Lorca's series, from 11 May 2011 until 19 July 2011. With this information, in order to test the performance of the declustering methods, the confusion matrices for each one have been computed. In the area of study, a total of 249 events have been recorded, which means a total of 103 background events should be identified. For this analysis, all the events classified in a cluster different from the one of Lorca series have been considered as background for simplicity.

Figure 9 shows that GK74 method is the most adequate (with a 94.43% mean for the metrics compared with the 92.88% for RJ and a 78.79% for A) and also the one that is able to identify more events belonging to Lorca's series."

Reference:

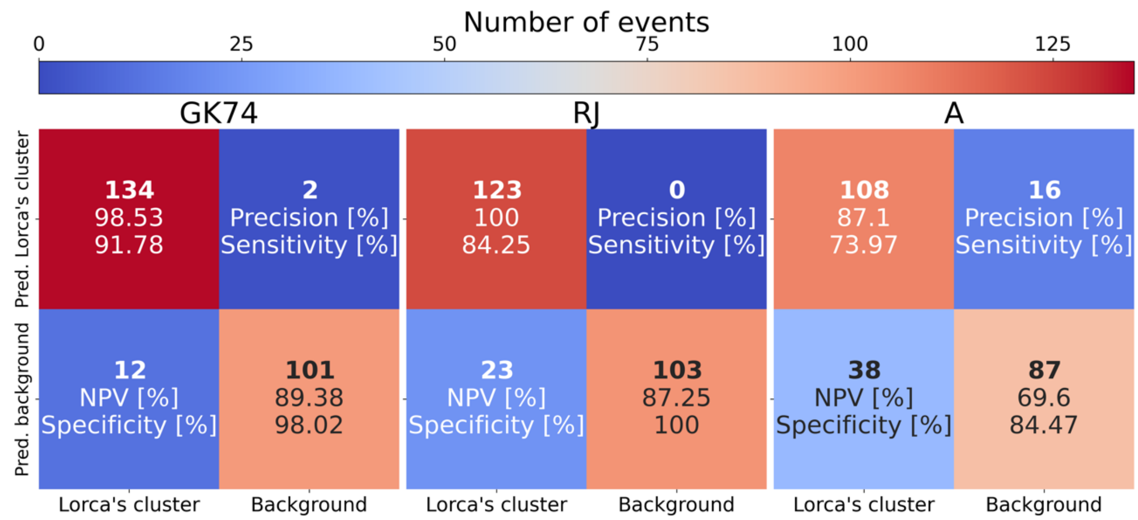
Cabañas, L., Carreño, E., Izquierdo, A., Martínez, J. M., Capote, R., Martínez-Díaz, J., Benito, B., Gaspar-Escribano, J., Rivas-Medina, A., García-Mayordomo, J., Pérez, R., Rodríguez-Pascua, M. A., and Murphy, P. (2011) Informe del sismo de Lorca del 11 de mayo de 2011 (in spanish), https://digital.csic.es/handle/10261/62381

7) **Fig 9 AF – is an abbreviation for Algorithm A? How do the data for Algorithm A from Fig. 9 and Table 6 agree, where 3394 events are listed for the Lorca series? The huge discrepancies between method A and the other two cluster extraction methods remain uncommented on in the article text.**

Response:

There is a typo in the figure, as the label should be A, not AF, which is referring to the Afteran's algorithm. The figure has been corrected. Regarding the discrepancy between the Fig.9 and Table 6 we have revised the results of the declustering and there was an error regarding the A algorithm. A correction has been made in the Figure 9, in Table 6 and in the paragraph that accompanies the results:

Algorithm	**RJ**	**A**	**GK74**
Number of clusters	652	1245	1012
Events in clusters	7143	10167	7552
Events inside Lorca's series	123	196	136



"Table 6 presents the number of clusters and the events in clusters for the whole seismic catalogue. The RJ algorithm classifies a total of 652 clusters in the catalogue while GK74 detects 1012 clusters. The A algorithm identifies ***1245*** *clusters. As can be seen, despite the three methods relying on windows for their calculations, there are significant differences in the results, not only in the number of clusters but also in the number of events inside each cluster.*

[…]

In spite of the difficulties in defining the clusters, Cabañas et al. (2011) carried out a detailed study on the 2011 Lorca's earthquake seismic series. This study is the best definition at the moment so we will use it to validate the best algorithm. They identified 146 events (including the foreshock, the main shock and the aftershocks) that belong to Lorca's series, from 11 May 2011 until 19 July 2011. With this information, in order to test the performance of the declustering methods, the confusion matrices for each one have been computed. In the area of study, a total of 249 events have been recorded, which means a total of 103 background events should be identified. For this analysis, all the events classified in a cluster different from the one of Lorca series have been considered as background for simplicity. Figure 9 shows that GK74 method is the most adequate (with a 94.43% mean for the metrics compared with the 92.88% for RJ and a ***78.79%*** *for A) and also the one that is able to identify more events belonging to Lorca's series."*

8) **The data in Table 9 refers to Table 9. What is meant by this?**

Response:

There was an erroneous cross-reference to Table 8 that has been corrected.

9) **For the Italian catalogue, calculations are from 2005. It is not clear why Table 3 data is given.**

Response:

We agree that this should be clarified. This question is related with the next one. In order to compute the probability of exceedance for a certain strong motion, we consider a background value. For this reason, the catalogue used to compute the background ends in the start of 2007 for Central Italy (Figure 5), and the start of 1990 for Spain (related figures). Hence the need of the different completeness magnitudes shown in the Table 3 and Table 7, as the data prior 2007 and 1990 constitutes the background PGA the changes about which we are computing.

10) **For the catalogue of Spain, calculations are made from 1990. It is not clear how the catalogue data from 1396 (and in Table 7 from 1048) are used in the calculations.**

Response:

This is explained in the previous answer to the question 9).

11) **For the neighbourhood of the L'Aquila earthquake and calculations in the vicinity of the epicentres of the three events in Spain it is not specified how the sizes of these neighbourhoods are chosen ? 200x200 cells - the regular grid cell sizes should be specified in km since the parameter of the epicentral uncertainty is specified in kilometres.**

Response:

Typically, PSHA studies require 300 km of area around the point of interest in order to avoid lack of data and border effects. For the selected point, the rest of the points in the grid are used as point sources that contribute towards the seismic hazard. Nonetheless, for the selected GMPs the distance from this point to the rest are computed and their contribution is weighted accordingly, so the closer points are the ones that contribute the most. There is no preselection of the neighbourhood, the function that computes the median spectral acceleration already depends on the distance, an example of the functional form can be seen in Akkar et al. (2013).

Reference:

Akkar, S., Sandikkaya, M. A., Bommer J. J. (2014) Empirical Ground-Motion Models for Point- and Extended- Source Crustal Earthquake Scenarios in Europe and the Middle East, Bulletin of Earthquake Engineering (2014), 12(1): 359 - 387

12) **In Figure 14 and Figure 15, the differences in the models shown are indistinguishable. A different presentation of the material should be chosen to demonstrate the convergence (or difference).**

Response:

We agree that the there are no distinguishable differences in Figure 14, for the given period, hence the need of defining other metrics that can be useful in low-to-moderate seismicity settings. As for the Figure 15, there is important differences in the models after Lorca's earthquake in 2011, meaning these selected metrics could be useful in the context of moderate earthquakes (as is the case of Lorca's series with 2 earthquakes with magnitude greater than 4.5).

Some modifications in the paragraphs that accompany Figure 14 and Figure 15 have been made:

"*Similar results are obtained for all three locations regarding the monthly variations (Figure 14). Overall, the monthly variations do not show changes preceding relevant earthquakes for this case study. One of the possible explanations is the lack of foreshocks in most of the main shocks. In Lorca earthquake, even though there was a 4.5 Mw earthquake almost two hours before the main-shock, the one-month increments on the computation process are not able to show any change in RC.*

The annual variations (Figure 15), on the other hand, show periods of increased RC before some of the selected earthquakes. An example is seen in Lorca site where a 15% increase is seen before Mula earthquake from June 1998 (the earthquake occurred in February 1999). Another example can be seen in both Murcia and Lorca sites, where a 10% increase can be seen before Aledo earthquake from May 2004 until the earthquake occurrence in January 2005. For this metric, differences *between the three models can be seen. For instance, Model 1t and Model 2t show greater changes after Lorca's earthquake in Lorca's site.”*

13) **429 "Model 1" - which model is this referring to?**

Response:

Model 1 has been corrected to Model 1t to be coherent with this manuscript’s notation. This correction has also been applied to the labels in Figure 14 and Figure 15.

14) **441- 442 The sentence needs to be rewritten as PSHA - Probabilistic Seismic Hazard Analysis cannot be "high in the region..." or "continuous increase...".**

Response:

We agree that PSHA cannot be addressed in terms used in the aforementioned sentences. The text has been modified because it referred to the change in the probability of exceedance:

“*Finally, in the case of south-eastern Spain, the relative change in the annual probability of exceedance kept high in the region after Mula earthquake and did not decrease until the occurrence of the Lorca earthquake. However, the continuous increase in this parameter in Vera after the Lorca earthquake cannot be directly related to a potential upcoming earthquake similar to the one from Lorca. Therefore, more time and data are needed to confirm this*.”

First of all, we thank referee_3 for the comments and suggestions. Please, find in the document in an orderly manner the answers to the comments:

- The main message of the paper is unclear, if not wrong. From one side the authors seem to focus on classical long-tern PSHA that is used for building code purposes (return periods of 475 years, which corresponds to exceedance probability of 10%in50 years). Of course, PSHA can be rescaled to shorter forecasting time windows, but, if it is used for land use planning the time windows remain long. Conversely, they also address the short time scales of the operational earthquake forecasting (OEF) whose forecasts are updated daily (if not more frequently during a sequence) because they are used to manage earthquake sequences. In other words, the updating time window and the forecasting time window are both very important for practical applications and should be coherent with the use of the models (land use planning or emergency management).

We agree with the concepts of PSHA and OEF you have explained, however the main message of the paper has been misunderstood by the reviewer. We proposed a smoothing methodology that can be used to compute first a long-term background PSHA which will be later compared with a short-term time dependent SHA so we will investigate metrics that can be used to take decisions before the main earthquake (similarly to OEF). We think that the changes introduced in the manuscript, also with the help of the other reviewers, have helped to clarify this.

- Many recent papers, that are not cited here, have already addressed the inclusion of earthquake clustering in PSHA. Besides quoting the relevant literature on this point, it would be interesting to explain why the authors think that their procedure would be better.

We do not pretend to conclude which is the best procedure in the definition of earthquake clustering but to compare several procedures and choose the one who are closer to the better definition of clustering. That's the reason why before explaining the comparison with Cabañas et al. (2011) we discussed the challenges on cluster definition using the works of Zaliapin and Ben-Zion (2020) and Anderson and Zaliapin (2023).

Anyway, we have improved the paragraphs as follows:

"*In spite of the difficulties in defining the clusters, Cabañas et al. (2011) carried out a detailed study on the 2011 Lorca's earthquake seismic series. This study is the best definition at the moment so we will use it to validate the best algorithm.*"

- **The method proposed by the authors estimates the earthquake rate and b-value from a short-time window (for instance during a foreshock sequence) and then it extrapolates them to the next decades (classical time scale of long-term PSHA). This is clearly wrong, because both parameters have a large variability on the long-term forecasting time window of classical PSHA (decades) and they cannot be considered stationary in such a time window.**

Current methods to incorporate earthquake clusters in PSHA consider that the time evolution of the earthquake clustering rapidly decays with time and it is much shorter than the forecasting time window.

The reviewer has misunderstood the concept. As explained before compute first a long-term background PSHA which will be later compared with a short-term time dependent SHA. For the short-term SHA the a and b parameters have been obtained using a time window of enough time to assure we will be able to get reliable results.

- **As said before, many recent PSHA initiatives account for aftershocks and foreshocks in different ways. Despite the differences among these procedures, all of them agree on the fact that declustering has to be necessarily applied for reducing the spatial bias. This is not done in this paper. In essence, if no declustering is applied, it is tacitly assumed that earthquake clusters that will occur in the next future will have preferentially the same locations of the past clusters. This has no scientific basis according to the present state of knowledge (earthquake clusters may happen everywhere).**

We agree with the reviewer. That's why our smoothing methodology investigates different approaches to obtain the gridded seismic activity rate.

- Very often the b-value varies with time due to catalog incompleteness, which is pervasive after a major shock and it can then induce to string bias in hazard calculations. It is not clear if the authors address this important issue properly.

We agree with the reviewer. That's why we have addressed the importance of a using different magnitudes of completeness when obtaining a and b.