Reply to RC2

We extend our gratitude to the anonymous reviewer for her/his positive evaluation of our work. We have carefully considered all the suggestions provided and have addressed them comprehensively in this response. Furthermore, we have also incorporated these suggestions into the updated version of the manuscript.

1) Line 95: What does the word "size" mean? Refers to the dimensions, proportions, or magnitude.

   We agree that the term 'size' could be misleading when applied to earthquake sources. Since the focus is on geometrical properties, we have replaced 'size' with 'dimension', as suggested.

2) Part 3: Is the division of the region into 7 groups and 61 regions based on the current work or previous works? If it has been done in this work, the method needs to be explained.

   We understand the reviewer's concerns. Unfortunately, due to space constraints, it was not possible to summarise the numerous considerations that led to the development of the zonation model. As briefly highlighted in the manuscript, the grouping was done by combining the analysis of seismicity data with seismotectonic considerations.
   For example, statistical data on the distribution of focal mechanisms and the empirical magnitude frequency distributions were analysed together with the characteristics of the main active fault systems (presented in the companion paper to this study) in the context of regional tectonic structures and boundaries.
   To provide a practical example of the construction process, Zone D was found to encompass a tectonic domain that is clearly separated by the stable features of the West Siberian craton (Zone E). As also indicated by the available source mechanisms, Zone D is characterised by a mixed regime, albeit with a dominance of large transpressive fault systems (e.g. Talas-Fergagna fault, Irtysh shear zone) that have influenced the southeastern evolution of the Tianshan Massif (Chen et al. 2022). Towards the south, a change in seismotectonic style becomes evident (Zone G), where the main reverse mechanisms increasingly dominate and large trust systems develop along the suture zone with the former cratonic terrains of the Tarim region (Angiolini et al., 2013). Here, seismic productivity is increasing and large magnitudes have occurred in the past. Further south, a mixed tectonic style is again present (Zone C), while seismicity becomes typical of continental collision (Zone F), with larger and deeper events along the Pamir thrust system (e.g. Murodov 2022). Towards the west, a clear separation between the tectonic styles of the systems at the boundary between the Turan Platform (Zone B) and the Karakum terrains (Zone A) has also been noted along the ideal southwestern extent of the Pamir suture zone (see Ghassemi and Garzanti, 2018 for a comprehensive review).
   We have now expanded the discussion in the manuscript, although an exhaustive description of the entire argument supporting the construction of the zonation model cannot be included due to the limited length.

3) Line 146. How is the depth uncertainty included?

The depth information comes directly from the solutions of the seismological agencies used to compile the homogenised catalogue (more detailed in the article accompanying this article in the same special issue). Unfortunately, the input data often lack the uncertainties related to each individual solution. Nevertheless, the statistical analysis performed in Section 4.1 ("Hypocentral depth distribution") helps us to build the probabilistic source depth distribution model.
In fact, OpenQuake accepts a probability density distribution of the hypocentral depth for each area source, which we then derive from the observations by regularising the depth histogram. Such a distribution is also used to delineate the depth boundaries of the source zones.

4) Line 151. H and L areas are not distinguishable in the map (Figure 2). In this case, as mentioned in the previous question, the discussion of depth uncertainty needs to be included in the analysis and text.

Thank you for pointing out this inconsistency, which is definitely confusing for the reader. Zone L is the lower and smaller zone. We have corrected the problem in the new version of the manuscript.

5) In Figure 3, the H and L zones are not consistent with the text. For example, is the depth of 150 the limit or the depth of 170?

Thank you for spotting the inconsistency. The reviewer is correct. The depth limit was set at 170 km, which is consistent with the depth distribution contained in the model. 150 km is a holdover from an older version of the initial model, the description of which was inadvertently not updated in the manuscript. The correct limit of 170 km has now been inserted.

6) Figure 4. The Gutenberg-Richter (in its logarithmic form) is a linear relationship; why is the fitted curve non-linear?

In contrast to the original (unbounded) formulation, which indeed exhibited linearity, the truncated version of the Gutenberg-Richter relationship (as shown in Equation 1) introduces a dependence on the maximum magnitude parameter (Mmax). This parameter is used to restrict the occurrence of events that exceed this magnitude and effectively exclude them from the cumulative distribution. As the magnitude approaches Mmax, the truncated relationship tends asymptotically towards zero, making its representation inherently non-linear.
The introduction of the truncated Gutenberg-Richter model was historically aimed at preventing the occurrence of 'unphysical' magnitudes (although the definition of what is 'unphysical' remains the subject of ongoing debate). The original Gutenberg-Richter model, since it was not truncated, could theoretically generate magnitudes of arbitrary values, albeit with extremely low probabilities (e.g. M=10).

7) Line 170. Why the (one-side) truncated Gutenberg-Richter relation is used? Why the mmin is not included? In the rest of the text, contradictions can be seen in this field and the double truncated Gutenberg-Richter relationship is used.

The truncated Gutenberg-Richter (G-R) relation is cumulative in exceedance (representing the number of events with magnitude greater than "m") and does not require the inclusion of Mmin as a parameter in its formulation. The lowest truncation, usually labelled Mmin, is not an intrinsic parameter of the distribution itself and can be set arbitrarily without affecting the generality of the model. It is important to emphasise that the inclusion of the lowest truncation does not mean that there are no low magnitude events, but that events unlikely to cause significant damage to the target structures are excluded from the calculation of the hazard integral. Therefore, a generally accepted value for Mmin, e.g. 4.5 (or occasionally 4 in more conservative cases), is considered appropriate for most engineering applications. The importance of Mmin arises primarily when the truncated G-R is converted into a probability density function (PDF) for solving the hazard integral.

8) Line 172. Is Gutenberg-Richter's relation applicable for values lower than completeness magnitude?

Yes, the Gutenberg-Richter relation (G-R) remains applicable for magnitudes below the completeness magnitude, provided that the assumption of occurrence according to the G-R relation is correct. Completeness magnitude, often referred to as incompleteness, does not refer to the inherent nature of the magnitude-frequency distribution itself, but to the seismic catalogue used to calibrate the occurrence model. Events below a certain magnitude threshold may be inadequately represented in seismic catalogues, e.g. due to limitations in the coverage or sensitivity of the seismic network or due to other reporting limitations.
However, it is important to recognise that these events exist in nature. If the G-R relationship is appropriately calibrated, it can accurately predict their occurrence even if they are not fully captured in the calibration dataset.
Another problem arises when the seismic catalogue is incomplete at the upper end, i.e. when the maximum possible magnitude is not included in the historical records, e.g. due to very long return periods that exceed the duration of the available data set. In such cases, a conservative approach is to estimate an upper limit for the largest observed magnitude, taking into account the possibility of generating these larger events. This adjustment can have a significant impact on the truncated G-R relationship, as shown in Equation 1.

9) Line 177. Why is the list square method used? Is the data homogeneous? Considering the age of the countries in the region and the long history, are historical data included or not?

Least squares (LS) and maximum likelihood (ML) are two widely used methods for calibrating the G-R occurrence model. The choice between LS and ML for calibrating the G-R model requires careful consideration of their respective advantages and limitations. While the ML method is often favoured for its ability to fit statistical distributions such as cumulative functions, it may tend to over-fit the lower magnitude range, leading to an underestimation of larger magnitudes, especially in regions with sparse data or low *seismicity. In contrast, LS fitting offers greater robustness and is less prone to overfitting, making it particularly suitable for regions with low seismicity or short catalogues. On the other hand, it suffers the impracticality to be performed on incremental (non-cumulative) magnitude bins to avoid data dependence. The authors have found that LS adjustment provides better results in such regional scenarios, as evidenced by its successful application in other challenging regions, such as Africa.*

Regarding the inclusion of historical data, it is worth mentioning that these data are indeed included in the calibration of the occurrence model. Historical seismic records spanning long periods of time are essential to constrain the long-term seismic activity in the region.

The homogeneity of data across the region is essential for a robust seismic hazard analysis. As part of this project, efforts were made to harmonise and standardise seismic datasets from different countries in Central Asia, taking into account differences in seismic monitoring infrastructure and data recording practises. A detailed description of the homogenised catalogue used in this study can be found in a companion paper in the same special issue, written by the same team (presently under review)

10) Line 187. The minimum magnitude of the Gutenberg-Richter of 4.5 does not match the one- side truncated Gutenberg-Richter relation. There seems to be a problem in the sentence " rate of earthquakes with magnitude greater than 0 "; because minimum magnitude of the Gutenberg Richter's has nothing to do with the potential of failure, but with the completeness magnitude. Therefore, the sentence seems to require revision.

    As already mentioned in this answer, it is important to distinguish between the minimum magnitude parameter (Mmin), which is used in the calculation of the hazard integral and is often selected based on the damage potential of seismic events, and the completeness magnitude, which denotes the lowest magnitude level that is fully represented in the data set. The phrase 'rate of earthquakes with magnitude greater than 0' refers to the productivity parameter of the Gutenberg-Richter relationship, commonly known as the a-value. Although it usually refers to magnitude 0 (the intercept), there are cases in the literature where the a-value is associated with other magnitude levels. This difference in terminology can lead to occasional discrepancies in interpretation, but does not affect the basic principles of seismic hazard analysis.

11) In Table 2, the weights are assigned on what basis?

    The weights assigned to the rupture mechanisms of each source group in Table 2 were determined by comparing the moment tensor solutions, analysed using Kaverina's classification diagrams (e.g. Figure 6), with the distribution of the predominant fault systems in the regions (detailed in the authors' companion paper in this special issue).

    In Group D, for example, there are two main mechanisms, reverse and strike-slip, with a similar number of reported solutions, leading to an initial probability fractionation of 50-50%. However, the definition of the actual strike direction from the mapped faults was ambiguous, leading to the identification of two main families of orientations. As there was no evidence for the dominance of one family over the other, we further split the original 50% probability into 25% to 25%. Similar considerations were made for the other groups.

12) Table 3, what is the source (reference) of the table?

The values in Table 3 are strictly derived from the depth limits defined for the source zones and thus from the seismicity analysis performed in Section 4.1. To define the LSD and USD boundaries, we have in practise allowed the ruptures occurring at the interface between the different depth zones to extend to a certain limit, which is between 15 and 30 km depending on the expected magnitudes. It should be noted that LSD and USD may not be exact values, but conservative limits to avoid the development of ruptures with unrealistic depth extent.

13) Line 304: Seismic coefficient = 0.1, why?

By definition, the aseismic coefficient represents the fraction of the total moment accumulation rate that is not released by earthquakes (i.e. aseismic). The calibration of this parameter is a challenge as there is no consensus within the seismological and geodetic communities on its optimal value and direct methods for its evaluation are currently lacking. Furthermore, the literature on this topic is quite limited.
It would be unrealistic to set the aseismic coefficient to 0, as part of slip is inevitably released by plastic deformation. Conversely, values greater than 0.2 often lead to inconsistencies between the total moment derived from slip rate and that observed from seismic events in the catalogue.
Through comparative analyses of the hazard levels derived from fault models and seismic catalogues, we evaluated different values for the seismic coefficient. A value around 0.1 was found to be workable as it balances the need for realism with the limitations imposed by the available data and modelling techniques. Although this value is a rough estimate, it agrees reasonably well with the empirical observations and ensures consistency between the modelled hazard levels and the observed seismic activity.

14) In the manuscript, the experts' opinion is repeatedly mentioned. What was the mechanism of use and criteria?

In probabilistic seismic hazard analysis, expert opinion often plays a crucial role in modelling, especially when controversial or weakly constrained topics are involved. We have adopted a structured approach that involves soliciting input and assessments from scientists and professionals with expertise in seismic hazard analysis and engineering, particularly from local communities in the target region (partners of the present consortium).
The mechanism for incorporating expert opinion includes holding various targeted meetings, open discussions and special workshops to gather insights and perspectives on various aspects of the study that were considered 'dubious" or contentious. These aspects include characterising the source model, assessing epistemic variability in ground motion prediction and defining the logic tree structure. Once the expert opinions are collected, they are carefully evaluated and synthesised.
This process involved identifying areas of agreement among the experts, addressing conflicting opinions through further discussion and analysis, and finally integrating the experts' insights into the development of the final models and the interpretation of their results.

Overall, the inclusion of expert opinions significantly improved the robustness and validity of our research findings and provided valuable perspectives from practitioners and researchers who were actively involved in the study. Several of these activities were organised as part of this project for the different components of the multi-risk model.

15) In Tables 5, 6 and figure 12, what is the method for assigning weights?

    If sufficient calibration data, such as records of strong motion recordings, are available, weights for ground motion models (GMPEs) can be assigned based on their degree of agreement with observed ground motions. Efficient ranking methods, such as the LLH approach proposed by Scherbaum et al. (2009), are commonly used for this purpose. However, in our case, the available records for the region were not sufficient to perform a robust data-driven evaluation. Therefore, a more conservative approach to assigning weights was required.
    In Tables 5 and 6, two equally weighted GMPE models were selected for the shallow tectonic conditions (Active Shallow and Stable Continental), while only one suitable GMPE was identified for Deep Seismicity. However, the use of a two-level tectonic zonation (as shown in Table 6) allowed us to also consider intermediate tectonic conditions, such as the TRT2, where all four GMPEs of Active Shallow and Stable Continental are considered.
    The reason for maintaining the logical separation between the two steps is to facilitate future changes to the weighting scheme as new data becomes available. This approach ensures flexibility in adjusting the weighting scheme to incorporate additional calibration data or to refine the selection of GMPEs based on improved understanding or progress in the field.

16) In Table 6, most of the weights are 0 and 1, and it doesn't seem that the two-step method (line 380) has much effect on the results.

    In the specific regional context considered in our study, the two-level tectonic zonation becomes particularly relevant when the hybrid buffer region between the active shallow (AS) and the stable continental (SC) tectonic environments is introduced. This hybrid region allows the blending of ground motion models, effectively accounting for intermediate tectonic conditions. While the same goal could have been achieved by directly assigning a weight of 0.25 to each of the four models in a one-level zonation, a two-level zonation provides a level of abstraction to better deal with mixed regions.
    We acknowledge that in the context of our study, with its relatively simple regional ground motion model, two-level zonation may not be strictly necessary. However, we believe that this study serves as an illustrative use case for the methodology and demonstrates its potential utility in more complex scenarios where different mixed environments are present. By applying the two-step approach, we provide a framework that offers flexibility and scalability, ensures adaptation to varying degrees of tectonic complexity, and facilitates future refinements as additional data and insights become available.

17) How does Figure 11 help present the paper?

    The Trellis diagram shown in Figure 11 is a valuable tool for evaluating the performance of selected ground motion prediction equations (GMPEs) for different types of ground *motion intensity measurements, magnitudes, and distances relevant to the analysis. These plots provide a comprehensive visualisation of the performance of each GMPE under different conditions and facilitate the process of model selection, especially in situations where direct ground motion observations are limited.*

By comparing the performance of different GMPEs with the corresponding hazard results from different branches of the logic tree, it also becomes easier to recognise the specific contributions of each model to the overall hazard assessment without the need for disaggregation analysis.

In addition to comparing the mean ground motion estimates shown in the figure, it is worth noting that the variability of the overall standard deviation (not shown in the manuscript for conciseness) is also generally taken into account. This comprehensive evaluation provides a more thorough understanding of the uncertainties associated with ground motion predictions and their implications for seismic hazard analysis.