A note to the editor: below, we have copied the reviewer comments in black non-italicised text, and replied to them in turn in blue italicised text. Where appropriate, we directly quote from the revised manuscript in red non-italicised text. Line numbers refer to the clean version of the manuscript.

We have also taken the decision to revise the name of the dataset presented in this study from the Malawi Seismogenic Source Database (MSSD) to the Malawi Seismogenic Sources Model (MSSM). This reflects that many of the attributes included in this dataset are taken from subjective decisions and modelling. Hence, they should not necessarily be described as 'data.'

Reviewer 1

General comments:

In the introduction, the authors give a good description of the use of fault databases in the framework of seismic hazard and risk assessment. While the results of this paper, in the shape of a seismogenic source database, are a major component of seismic hazard assessment, there are not seismic hazard results themselves. Therefore, my opinion is that the title of this paper should be modified and the words "seismic hazard" should be removed, since no hazard results are presented in the paper.

In order to allow the results of these study to be used in a PSHA study, the way the weighting of the different source types should be done needs to be better discussed. It is not clear how the earthquake rate from the different source types should be combined. Should the weighting be done in order to fit a given MFD for the entire system? I would be useful if the authors added a section in the discussion part of the article to clarify this point. If possible, a comparison of their computed earthquake rates with the rates calculated using the earthquake catalogue could be added.

We thank the reviewer for their comments and agree that since we do not present any seismic hazard results in this paper, we should remove these terms from its title. Hence, the title of this study has been revised to:

'Geologic and geodetic constraints on the magnitude and frequency of earthquakes along Malawi's active faults: The Malawi Seismogenic Sources Model (MSSM).'

We agree too that how different source types are weighted is a significant source of uncertainty when incorporating the MSSM into probabilistic seismic hazard analysis (PSHA). In this case, we note that we are currently exploring this topic in a subsequent study in which the MSSM is being used to perform a PSHA for Malawi (Williams et al 2022a). This study is currently in review; however, a preprint of it can be found here:

Williams, J., Werner, M., Goda, K., Wedmore, L., De Risi, R., Biggs, J., ... & Chindandali, P. (2022a). Fault-based Probabilistic Seismic Hazard Analysis in Regions with Low Strain Rates and a Thick Seismogenic Layer: A Case Study from Malawi. https://doi.org/10.21203/rs.3.rs-1452299/v1

In essence, however, we used the MSSM to simulate stochastic earthquake event catalogs with a 2-million-year duration, and in which earthquakes occurred randomly on each MSSM source following a memoryless Poisson process. In this study, we generated catalogs for all possible source type weighting combinations at increments of 0.1, and with the constraint that the weighting of any source type must be ε0.1 (n=36). We then selected the combination that, for the magnitude range 6-7.6, produced a catalog with the closest b-value estimated for Malawi from instrumental seismicity (1.02; Poggi et al 2017). Following this, we selected a weighting of 0.6-0.3-0.1 for section, fault, and multifault sources respectively, which is *qualitatively consistent with the inference that there must be relatively frequent small magnitude section source events to maintain a b-value ~1. This is discussed further in Section 4.4 and Figure 7 in Williams et al 2022a.*

In Williams et al 2022a, we also provide an analysis for how the moment rate ($\dot{M}_0$) of the synthetic earthquake catalogs generated using the MSSM compared to the instrumental catalog $\dot{M}_0$ in Malawi. This analysis concludes that although the $\dot{M}_0$ of the synthetic catalogs are 2x higher than the observed $\dot{M}_0$ in Malawi, this can be accounted for by the incomplete nature of the instrumental record in Malawi, which in turn is indicative of the low slip rate of its faults, the inference that they may be locked and/or host clustered earthquakes, and that the instrumental catalog has only a short (<60 years) duration (see section 4.6 in Williams et al 2022).

We note that given that the reviewer has raised these points on this study about the MSSM, not the PSHA, we had considered whether some of the above discussion could be incorporated into this current study. However, we concluded that this would exacerbate the length of this study, and that describing the MSSM and PSHA separately remains the best strategy for presenting this work. Nevertheless, we have now discussed some of the above points and signposted our ongoing PSHA for Malawi in Sect. 5.3 (Lines 543-544):

As 'section,' 'fault,' and 'multifault' sources are mutually exclusive, in future, weightings could be assigned to each source to indicate their relative likelihood. Future updates to the MSSM may also consider that

And (Lines 561-563):

Reconciling seismic and geodetic moment rates in Malawi, weighting different source types, and allowing sources in the MSSM to exhibit a more diverse set of earthquake ruptures, are being considered in an ongoing new fault-based PSHA for Malawi (Williams et al., 2022a).

In either case, our PSHA shows the importance of using geologic and geodetic data to constrain the activity of faults in Malawi, and hence the importance of the MSSM's development.

Poggi, V., Durrheim, R., Tuluka, G. M., Weatherill, G., Gee, R., Pagani, M., ... & Delvaux, D. (2017). Assessing seismic hazard of the East African Rift: a pilot study from GEM and AfricaArray. Bulletin of Earthquake Engineering, 15(11), 4499-4529.

Specific comments :

Line 169 – I suggest modifying the term "statistical treatment" by "exploration"

Corrected (Lines 171)

Line 218 – Simplifying the surfaces of the faults is a potentially impactful hypothesis in terms of hazard assessment. The change in the surface can both affect the moment rate estimate for the fault and the distance taken into account in the GMPEs. While it is possible that the complexity observed in the fault trace might not be present at depth, the straight line is the other end-member of the possibilities for the fault surface. Why not let the final user of the database, the hazard modeller, choose the level of simplification to be applied? Especially since modern PSHA codes can now handle rather complicated geometries.

We agree that the way in which we simplify the fault geometries from the Malawi Active Fault Database (MAFD; Williams et al 2022b) into the MSSM is non-unique. However, many of the attributes that are assigned to sources in the MSSM are calculated using these simplified geometries (e.g., earthquake magnitude, recurrence interval in equations 4 and 5). Hence, if a final user wishes to change the source geometry, for consistency, these attributes will also need to be revised. For the MSSM to incorporate multiple interpretations of the MAFD would be impractical.

We have therefore emphasized that alterative interpretations of the MAFD for seismic hazard assessment are possible, and since the MAFD is openly available (Williams et al 2022b), final users are welcome to choose their own level of source geometry simplification, although this will require changing other source attributes (Lines 222-227):

Should a MSSM user want to consider alternative fault source geometries using the MAFD, this database is also readily available (Williams et al., 2021b). However, since other attributes in the MSSM (e.g., magnitude, recurrence interval) are contingent on the source geometries we define, other interpretations of source geometry will require that these attributes are also revised. In instances when accurate fault traces are required (e.g., assessment of surface rupture hazards), the MAFD should be used in preference to the MSSM, as in these cases, the MSSM's simplified geometries will not be realistic.

Williams, J. N., Wedmore, L. N. J., Scholz, C. A., Kolawole, F., Wright, L. J. M., Shillington, D. J., Fagereng, Å., Biggs, J., Mdala, H., Dulanya, Z., Mphepo, F., Chindandali, P. and Werner, M. J (2022b).: The Malawi Active Fault Database: an onshore-offshore database for regional assessment of seismic hazard and tectonic evolution, Geochemistry, Geophys. Geosystems, 23(5), e2022GC010425, doi:10.1029/2022gc010425, 2022b.

Line 230 – Would it be possible to add the uncertainty on the dip in the database ? This parameter can be source of large uncertainties in the hazard levels, and since the knowledge of the dip is not uniform in the system, adding the uncertainty on each fault could be useful.

We have followed the reviewer's recommendation in the revised manuscript and added dip uncertainties as attributes in the MSSM (Table 1). It should, however, be noted that our model of source geometry only considers the intermediate dip estimate (Lines 237-239):

The moderately-steeply dipping (40-65º) planar faults indicated by these studies are also used to bound the dip for MSSM sources where no direct dip measurements are currently available (Table 1), and this uncertainty is incorporated into the slip rate calculations (Fig. 5). However, this uncertainty is not incorporated into the MSSM geometrical model, which considers only an intermediate dip estimate of 53º for these sources.

Line 278 – Simplifying the fault system by removing splay faults also implies to consider that the whole deformation is accommodated by the main fault. Since the metrics used in GMPEs don't usually take into account such details, the impact on the hazard would probably be minimal or within the simplification already made when using a GMPE. However, the impact of the simplification on the deformation should be commented in the text.

We have now added a sentence to clarify how we interpret removing splays influences slip rate estimates in the MSSM (Lines 277-279):

In these cases, the slip rate assigned to the simplified MSSM source represents the cumulative slip rate of the main fault strand and its smaller splays.

Line 441 – A point is missing.

Corrected (now line 451)

Line 455 – Some underlaying assumptions behind these results should be stated here, even if there are discussed later in the article. These recurrence intervals are obtained assuming that the slip-rate is fully seismogenic. It is also assumed that each source can only host on magnitude (for one branch of the logic tree), but other magnitude frequency distributions could be possible.

We agree and have incorporated this comment into the following text in Section 4.1 (Lines 465-469):

If earthquakes in Malawi occur only as 'section' type events, then their recurrence intervals are ~500-30,000 years. Alternatively, if they only occur on fault and multi-fault systems, recurrence intervals are ~1,000-40,000 years (Table 3). In reality, earthquakes in Malawi likely occur as a combination of section, fault, and multifault events, and so these recurrence interval estimates are a minimum estimate, and furthermore they assume that the MSSM sources do have any component of aseismic deformation. We discuss this further in Sect. 5.3.

Table 3 – In this table, it is not very clear if the values are for one specific fault or for the system as a whole. If it is for the system as a whole, can the different lines be read together? For example, is the table saying that the mean recurrence of a M6.8 earthquake is 10900 years? The legend of the table should be better detailed.

The reviewer is correct to point out that Table 3 was not described in sufficient detail. In the revised manuscript, we clarify in the table caption that the values we provide are calculated from the intermediate estimates of all MSSM source for the given type (e.g., the section magnitudes minimum, mean, and maximum values considers the magnitudes of all section sources in the MSSM):

Table 3: Range and mean of selected attributes in the MSSM. The reported values are calculated by considering the intermediate estimates from all MSSM sources for the given type. The analysis of recurrence interval intermediates assumes that each source ruptures only in the given type (Lines 473-475).

Line 465 - The 5% threshold is probably too severe for this type of analysis. For some fault the two distributions are very similar, and the difference are minimal, sometime affecting only the width of the distribution, but the mean values are similar. The discussion in the following paragraph is probably more useful in order to understand the difference between the slip-rate estimates.

We acknowledge that it is surprising how many of the t-tests reject the null hypothesis that the two distributions come from probability distributions with the same mean but unequal variances. Our interpretation of this result is that since the variance of each slip rate distribution is high, many (10,000) Monte Carlo simulations must be run to achieve stable results. In other words, if fewer simulations are run, the result of the t-test changes each time we perform the analysis. This large number of simulations entails that the p-value is very sensitive to even small differences in the mean between the two distributions, and hence the *null hypothesis is rejected in cases when the mean values of each distribution qualitatively appear to be quite similar.*

Given the difficulties of using a t-test, in the revised manuscript for this comparison, we have instead used a two sample Chi Square (χ^2) test. In this case, instead of comparing the distributions to see if they their mean values are the same, we are testing if samples drawn from the probability distributions themselves are distinct (for a 95% significance level, lines 429-430):

we performed the following statistical tests to test how well these independent estimates of fault slip rates compare: (1) a two sample Chi Square (χ^2) test that 600 slip rates randomly drawn from the slip rate distributions are distinct at a 95% confidence level,

In this new analysis, only two of the assessed faults have samples that are distinct from one another. Hence, it indicates that the slip rates derived from the systems-based approach are reasonable in the context of those derived independently from the offset reflector. As discussed in the following comment, we now report p-values of each test in Fig 8 in the revised text.

Figure 8 - The authors should add indexes to these figures, so each individual fault could be identified on the map in figure 2 and in the database. Additionally, the t-test result value could be added to the figure, helping to understand the reason why one is accepted and not the others.

We agree, and have made the necessary revisions to Fig. 8.

Reviewer 2

The manuscript by Williams et al.: "Geologic and geodetic constraints on the seismic hazard of Malawi's active faults: The Malawi Seismogenic Source Database (MSSM)" represents a comprehensive contribution to parametrize seismogenic sources in this section of the EAR and helps assessing the resulting hazard.

The steps for building the database are clearly illustrated and uncertainties explored in details.

This database extends the previous database available only for the southern part of the rift (SMSSM) to the whole Malawi rift (south, central and north), and increases the estimates of source parameters by adopting an updated geodetic model which results in a useful reduction of parameters uncertainties. I find particularly interesting the comparison between system-based and geologic-based (the offset of a 75-ka seismic reflector in Lake Malawi) estimates of slip rate and recurrence, which offers confidence in adopting the system-based approach elsewhere (central and northern sectors) where geologic information is scarce. I also agree with the possibility of very large (>7.5 Mw) but infrequent extensional earthquakes in this strong and thus elastically thicker continental crust although the hazard is clearly posed by intermediate and more frequent earthquakes.

In summary, the presented compilation poses a strong basis for future detailed studies aiming at more detailed filed and geophysical characterization of fault geometries and segmentation patterns and of estimations of aseismic release on some faults. I have no observations on the manuscript structure and arguments. Two typos are indicated below.

We acknowledge and thank the reviewer for their positive comments on our study.

Line 138: invert “lower aseismic crust” with “aseismic lower crust”

Corrected (Line 139).

Line 379: “and there a range”, correct with “and there is a range”

Corrected (Line 389).