The NHESS manuscript “Modeling Seismic Hazard and Landslide Potentials in Northwestern Yunnan, China: Exploring Complex Fault Systems with multi-segment rupturing in a Block Rotational Tectonic Zone” by Cheng et al. focuses on forecasting earthquake activity on the complex northwestern Yunnan fault system. This paper is generally well written and logically organized. The authors have broadened the scope of this study by also mentioning implications of their modeling results to ground-motion assessment, regional landslide hazard, and local tectonics. These ancillary topics are treated superficially, but the core modeling methodology is well founded. However, characterization of potential ruptures needs to be broadened and better justified (see Comment 1). Major comments are included below, as well as some minor details that should be easily addressed by the authors.

Major comments:

1. The authors develop four models of multi-segment and multi-fault rupture combinations based on “the segmentation model and fault rupture behaviors”, informed largely by historical earthquake ruptures. Given the limited record of finite-rupture observations, this is prone to a great deal of bias [see Stein et al., 2012]. A more objective method is to evaluate all possible segment combinations for a given fault and establish “plausibility filters” (as suggested in Section 4.1) for multi-

fault ruptures [Field et al., 2014]. Then, the results from SHERIFS can be evaluated against the historical record for verification. At minimum, more explanation is needed in Section 2.2 to firmly establish the authors' preferred combinations and perhaps include more possibilities for multi-segment/multi-fault rupture.

Thanks for your comments. I agree with your opinion of the explanation for rupture combinations. For this work, we did not consider the rupture combinations with step width of 5+ km and the strike difference ⩾28° between the linked segments. We modified the words in section 2.2 to make the words more reasonable.

2. Uncertainty analysis of the model results is not well described and perhaps incomplete. For example, it is unclear whether uncertainty in fault slip rates, which is detailed in Section 2.1, the regional MFD parameters and the M-A relations are all propagated through to the results.

   Thanks for your suggestion. We added the words in section 3.1.

3. In addition, evaluation of the model results is based on NMS ratios, rather than rigorously establishing quantitative prediction errors or goodness-of-fit metrics.

Thank you for your comment. We chose to use NMS ratios for evaluating the model results due to their practical utility in our context. NMS ratios offer a straightforward method to assess model performance relative to a baseline and reflect the goodness-of-fit metrics, as seen in right panel in figure 6. The iteration process focuses predominantly on the fault slip rate, with the remaining portion accounted for by the NMS, thus providing an integrated view of model performance.

4. Description of the PGA calculation is cursory, and it is unclear whether source of uncertainty other than the GMPEs are included.

Thanks for your comment. We added the model limitations in section 4.1.

4.1 Model Limitations and Mitigation Measures

Our seismic hazard modeling for NWYR represents our current understanding of average earthquake hazards in the region based on available data. The results are affected by numerous epistemic and aleatory uncertainties inherent in seismic hazard modeling processes, including the MFD, fault geometry, fault type, slip rate, and variability in GMPEs. Mitigating the impact of these uncertainties is critical for accurate seismic hazard assessment.

The MFD relationship, calculated from historical earthquakes, is essential for determining seismicity rate ratios across different magnitude bins. The deflection of the MFD directly influences the distribution of the modeled seismicity rates. In this study, we chose the G-R relationship over the Y-C relationship due to the regional fragmented tectonic environment. The calculated *b*-value of 0.96 aligns closely with the expected value of 1 found in seismically active regions (Pacheco et al., 1992). To derive earthquake magnitudes on fault segments, we employed rupture scaling relationships based on historical rupture parameters of earthquakes in China as proposed by Cheng et al. (2020), ensuring consistency with unique tectonic characteristics. Achieving more precise MFDs and rupture scaling laws necessitates further refinement in methodology and the use of reliable catalogs specific to the study area.

For fault geometry, type, and slip rates, we relied exclusively on recent field investigation data. In compiling fault rupture models for NWYR, we analyzed these geological data under a unified tectonic stress field, ensuring coordinated fault system movements. The variability in GMPEs is complex, influenced by factors such as earthquake rupture characteristics, seismic wave propagation, and site conditions. Consequently, we incorporated Quaternary sediment site amplification effects on PGA values. Addressing basin effects on ground motion requires dynamic simulations to achieve more precise results.

5. Similarly, uncertainty associated with the landslide hazard analysis is incomplete. See for example, Wang and Rathje [2015]. It is even unclear in this analysis what the parameters of the hazard calculation are (e.g., exposure time, probability model, etc.).

We added the words to explain the landslide hazard analysis as follows:

We used a logistic regression model, well-regarded for its robust performance in machine learning. Unlike previous models (e.g., Nowicki et al., 2014; Wang and Rathje, 2015; Parker et al., 2017) for calculating earthquake-triggered landslide hazards. Our model directly assessed the absolute probability of landslide occurrence, represented as the percentage of the landslide area within a region relative to the total area of the region (Shao et al., 2020). As a result, our hazard estimates have a true probabilistic meaning, reflecting the actual probability of landslide occurrence rather than being merely a formal expression of probability. We then calculated the probabilistic seismic susceptibility for a specific point in time within the study area, which produced a probabilistic PGA distribution map. By using this probabilistic PGA map as input for our model, we can estimate the corresponding probability of earthquake-triggered landslide occurrence. We employed these steps as the basis of our approach to calculating the probability of such landslides.

Minor comments:

(6) L20: Specify “ductile flow of the lower crust” to be clearer.

Thanks! We revised it to “ductile flow of the lower crust with low shear-wave velocity”.

(7) L32 and throughout: “averagely” -> “on average”.

Revised.

(8) L65: Is the “low velocity belt” delineated by the faults located in the lower (i.e., ductile region) or upper crust (i.e., the host rock of the faults)?

Revised. We mean that the low velocity belt with lower-crust flow

(9) L106: Unclear what the “pre-earthquake period” refers to.

Revised.

(10) L108, 112: “errors”-> I think you mean “uncertainty”.

Revised.

(11) L151-153: Indicate some brief description of GMPEs and site conditions used, as this is key to PGA estimates.

Thanks to your advice. We added the words for GMPEs and site conditions in Line 153 and Line 157.

(12) L193: Reference Figure 2.

Thanks! We added it in Line 199.

(13) L303 and throughout this section: “integrated”->”included” or similar.

We revised it in Line 303.

(14) L313-316: This seems like conjecture.   Any evidence to support this inference?

The fieldwork in this region is relatively scare, as the rugged and uneven terrain. Here, we modified the words from “hinder” to “strongly impacted on” in Line 342.

(15) L336 and throughout this section: Need to distinguish the regional MFD (input to model) from the on-fault MFD (output).

Thanks! We added the words in Line 378 and Line 395.

(16) L357, 434: The Wells and Coppersmith (1995) relations are dated at this point. Better to use, for example Leonard [2010], or a similar recent

study as an alternative to Cheng et al. (2020). See summary by Stirling et al. [2013].

We selected the scaling relationship of Cheng et al. (2020) because it is specifically developed for earthquakes in mainland China, making it more regionally appropriate for our study. By comparing this with the well-established scaling relationship of Wells and Coppersmith (1994), which is based on a global dataset of both interplate and intraplate earthquakes, we aim to assess whether regional-specific models offer improved accuracy over more generalized, globally applicable models.

The words to explain the reasons in the context are as follows: For the rupture scaling relationships, most of them are developed for plate boundary regions (Stirling et al., 2013). In this study, we selected a regression scaling relationship based on a dataset of earthquakes from mainland China (Cheng et al., 2020) and compared the results with the widely used rupture scaling relationship of Wells and Coppersmith (1994), which incorporates global data from both interplate and intraplate earthquakes.

(17) L406: Shouldn't some goodness-of-fit metric be used then?

Thank you for your insightful suggestion. We agree that incorporating a quantitative goodness-of-fit (GOF) metric would provide a more rigorous assessment of model performance. However, in this study, the NMS ratios not only reflect the regression fit but also clearly indicate which segments have lower NMS ratios. This makes it easier to identify which segments deviate from the modeled seismicity rates, providing valuable insights for comparison. Nonetheless, we will consider adding a GOF metric to complement the NMS ratios and provide a more quantitative evaluation.

(18) L448: There hasn't been any explanation on how these prediction intervals are calculated. Please include a detailed description, particularly which sources of uncertainty this pertains to.

Thank you for your insightful comment regarding the calculation of prediction intervals. We added the words in the 3rd paragraph in section 3.1, Methodology.

"In these steps, the b-value from historical earthquakes, the rupture scaling law of the faults, and the fault slip rates are typically accompanied by significant uncertainties. SHERIFS used the random sampling method to explore the uncertainty bounds. The rates are derived while examining uncertainties related to earthquake magnitudes, the duration of the completeness period, and the low number of observed earthquakes for larger magnitudes, using a Monte Carlo approach (Chartier et al., 2021). For each branch of the logic tree in the random sampling, it generates a corresponding number of models that match the total count of random samples. For each model, the slip-rate value is selected uniformly within its uncertainty bounds, scaling law parameters are chosen independently from a Gaussian distribution within their error bounds, and the b-value is picked from the user-defined range. All these uncertainties propagate to the final step of calculating seismicity rates with uncertainties. ”

(19) L463: “branches” of what?

Thanks! We revised it to “branches of GMPEs”.

(20) Figures: Font size is very small, to the point where the labels and numbers are unreadable.

Thanks! We revised the labels and numbers to be readable.

References cited in review

Field, E. H., et al. (2014), Uniform California Earthquake Rupture Forecast, version 3 (UCERF3)--The time-independent model, Bull. Seismol. Soc. Am., 104, 1122-1180.

Leonard, M. (2010), Earthquake fault scaling: Self-consistent relating of rupture length, width, average displacement, and moment release, Bull. Seismol. Soc. Am., 100(5A), 1971-1988.

Stein, S., R. J. Geller, and M. Liu (2012), Why earthquake hazard maps often fail and what to do about it, Tectonophys., 562-563, 1-25, doi:https://doi.org/10.1016/j.tecto.2012.06.047.

Stirling, M., T. Goded, K. Berryman, and N. Litchfield (2013), Selection of Earthquake Scaling Relationships for Seismic-Hazard Analysis, Bull. Seismol. Soc. Am., 103(6), 2993-3011, doi:10.1785/0120130052.

Wang, Y., and E. M. Rathje (2015), Probabilistic seismic landslide hazard maps including epistemic uncertainty, Eng. Geol., 196, 313-324, doi:https://doi.org/10.1016/j.enggeo.2015.08.001.