Dear Editor,

please find enclosed our answers to the three reviews, as well as the revised manuscript with tracked changes.

We provide a response to all comments raised. Reviewer comments are in black. Our answers are in magenta. All modifications in the manuscript are in magenta.

We add an Appendix with 2 figures. We also now add as an electronic supplement the coordinates of fault sections, as requested by the reviewers.

We are grateful for these constructive reviews that help improve the manuscript.

Sincerely,

Sarah El Kadri and co-authors

--

RC1: 'Comment on nhess-2024-184', Anonymous Referee #1, 09 Dec 2024 reply
Dear editor and authors,
I reviewed your manuscript "Implementation of an interconnected fault system in PSHA, example on the Levant fault".
The article aims at improving PSHA for the Levant fault system including different scenarios about possible multiple activations of different faults by introducing connectivity between different seismogenic sources.
The manuscript is well written, clear and of wide and specific scientific interest and definitely deserves publication in NHESS after minor review. I reported my specific comments, suggestions and (minor) requests on the attached pdf for the sake of simplicity.
Thank you for considering my review

--
We have copy-pasted here all comments raised inside the pdf:
L18
Middle East?
Ok, we made the modification.
L47
List in the correct order
Ok, we made the modification.
L67
Add a white space
done
L88
In this manuscript you are looking for a model able to allow fault interaction, so that already observed magnitudes can be explained. However, calibrating your connectivity assuming that they are the maximum possible ones may introduce a bias. The case of Kaikoura you cited above is paradigmatical: probabilistic hazard models set up before its occurrence gave very low chances of M 7+ earthquakes in that area just because it seemed unlikely that fault systems on site could undergo cascading interactions.
We understand the concerns of the reviewer. Nonetheless, the magnitude-frequency distribution must be bounded in the upper range with a maximum magnitude that corresponds to the maximum earthquake that could occur along the fault system. Building the source model for PSHA, decisions must be taken on maximum magnitudes. In the present study, as the source model is made of faults, maximum magnitude is estimated

applying scaling relationships as well as considering maximum observed magnitudes in comparable strike-slip fault systems worldwide. As there is considerable uncertainty on this parameter, we explore different values. We select values that lead to a magnitude-frequency distribution at the scale of the system that is compatible with the deformation rates (moment-balanced magnitude distribution).

Figure 1
Please, can you add the reference to the fault dataset you used for realizing this figure?
Figure 1a : we completed the caption with "fault map modified from Daeron et al. (2007)"
Figure 1b : as explained in the text (Section 2), in the present study we have reanalyzed satellite images along the whole fault system to define the segmentation. This set of sections is new. We have now added the set of sections in an electronic supplement.
L133, gap = ? Do you mean its width? length?
It is the width of the pull apart. Text has been corrected.
L145
I suppose that in your model there is always a finite probability, even though tiny, that a rupture may pass through a barrier. So, the difference between a strong and a weak barrier is just the value of such probability.
In the SHERIF algorithm, the probability is either 0 or 1. If the gap between two sections is smaller than the maximum jump allowed, the rupture can pass through the barrier.
L155
Fletcher
Done
Figure 2
Add letters for referring to subplots, e.g., A, B
Done
Specify the period (e.g., since ...)
the period depends on the magnitude, we have slightly modified the sentence : "instrumental events from global datasets (circles, magnitude larger or equal to 4.1 *in the instrumental catalog starting in 1900*, see Section 6)"
L188
Specify your datasets for this research here
We have modified the sentence : "the set of fault sections' traces with extension at depth (dip angles and widths), *displayed in Fig. 1b and described in Table 1 (as well as an electronic supplement)"*

L190-191
75° : why? Justify this value shortly
"maximum distance between sections that a rupture may jump" : Explain how did you retrieve or infer such data
We have modified the paragraph as follows: "the geometrical rules for a section to be able to break with its neighboring sections: the maximum azimuth between two adjacent sections (here we use 75°; *Milner et al. 2013 used 60°*) and the maximum distance between sections that a rupture may jump *(see e.g. 5 km in Milner et al. 2013, 15 km in Milner et al. 2022)*"; adding these two references:
Milner, K., Page, M.T., Field, E. H., Parsons, T., Biasi, G., and Shaw, B. E., 2013. Defining the inversion rupture set via plausibility filters, U.S.G.S Open-File Report 2013-1165, Uniform California Earthquake Rupture Forecast Version 3 (UCERF3) - The Time-Independent Model, Appendix T, 14 pp.
Milner, K. R., B. E. Shaw, and E. H. Field (2022). Enumerating Plausible Multifault Ruptures in Complex Fault Systems with Physical Constraints, Bull. Seismol. Soc. Am. 112, 1806–1824, doi: 10.1785/0120210322

L194

Gutenberg-Richter distribution : I agree that the GR is a good choice, at least for a first order approximation. However, I would like to add a note: in this study you are focused on the largest magnitudes (6-8). In this range, the best-fitting functions for the frequency-magnitude scaling is a TAPERED-GR. Perhaps, even this choice may be considered. What do you think about it? If you agree, you could add a few lines on this topic and its possible impact on your results.

Thank you for raising this point and for the suggestion. This is still unclear which shape in the upper magnitude range is best adapted to observations, especially for a fault system. We could use a tapered Gutenberg-Richter (e.g. Pareto), we would have to take a decision on the corner magnitude rather than on a firm maximum magnitude. Rates in the upper magnitude range would decrease more rapidly than with a truncated Gutenberg-Richter, and the model would not be strictly bounded anymore. We prefer to use a more common Gutenberg-Richter (as Danciu et al. 2017 in the EMME Middle East project, or Danciu et al. 2024 in the ESHM20 fault model at the scale of Europe).

We have added the following sentence in the text (Section 3):

"A tapered Pareto distribution (Kagan, 2002) could be used rather than a truncated Gutenberg-Richter, as in the fault models built for the 2019 Italian seismic hazard model (Visini et al., 2021). This distribution includes a bending of the recurrence model from a magnitude called the corner magnitude, rather than a sharp cutoff at a maximum magnitude in the truncated distribution. The tapered distribution usually leads to a stronger decrease of seismic rates in the upper magnitude range with respect to the truncated Gutenberg-Richter. For a fixed moment rate budget, a decrease of rates in the upper magnitude range would lead to an increase of rates in the moderate magnitude range. Using a tapered Pareto rather than a truncated Gutenberg-Richter could lead to slightly different results, but would not impact the main findings of the present study. "

L197

"an estimate for the maximum magnitude in the system": See my comment above. I think that this choice is a weak point of your model: your map would like to provide information about ground motions, but the final result is strongly affected by the choice of the maximum magnitude: I understand that you calibrated your connectivity on the base of the maximum observed magnitude. However, you should take into account of the possibility of larger unprecedented events.

We understand your concern, but when magnitude-frequency distributions are moment-balanced, i.e. based on slip rates, the maximum magnitude is always a key parameter. Nonetheless, we are not 'calibrating' the connectivity. There are potential barriers in the fault system (gaps pull aparts, jogs, ..). SHERIFS is run assuming that these barriers are firm, or alternatively that ruptures can pass these barriers. We test different maximum magnitudes (based on observed earthquakes in other strike-slip fault system worldwide) and quantify the aseismic deformation obtained. Maximum magnitudes (and magnitude-frequency shapes) that lead to a too high aseismic deformation are discarded. Maximum magnitude 8.1 is not favored because in this case 11% of the deformation has to be aseismic.

This article is an exploratory study where we test the application of SHERIFS on the Levant fault system. If we had to use these results for a probabilistic seismic hazard study aimed at delivering hazard levels (not just exploratory study), we would build a logic tree for the source model. Alternative Mmax values would be considered, the 8.1 Mmax choice would be attributed a lower weight than the 7.9 choice. To make this clear, we have added a sentence in the text regarding this aspect at the end of Section 5.2:

"This is an exploratory study aimed at understanding how the algorithm SHERIFS works. In a probabilistic seismic hazard study aimed at delivering seismic hazard levels for a country, we would populate the source model logic tree with these alternative models to cover the epistemic uncertainty (attributing larger weight to the model associated to the lowest aseismic deformation). "

L204-208

This part is not clear: you assume that magnitudes are power law distributed, then...?

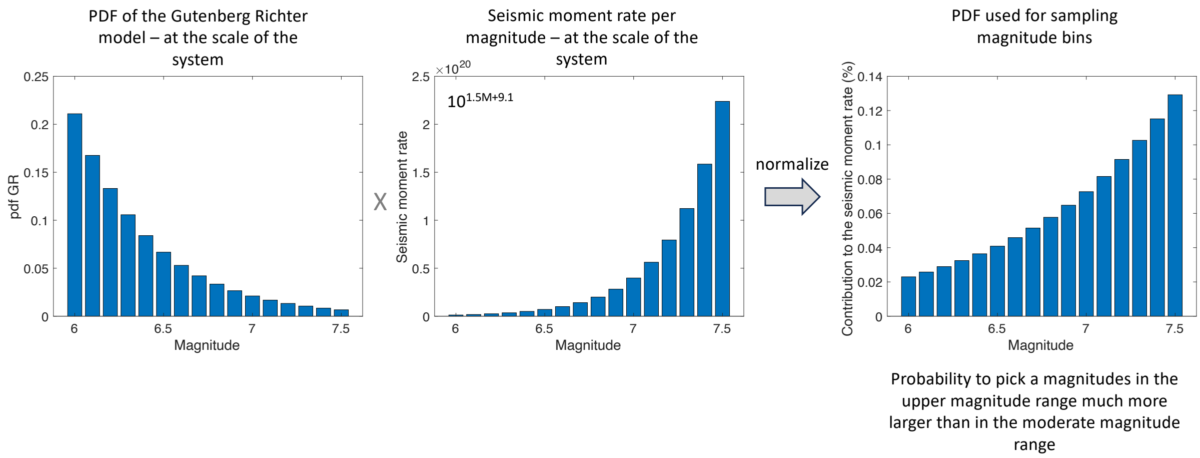
The reviewer refers to this paragraph:

"Based on the hypothesis that earthquake rates follow a Gutenberg-Richter distribution, a probability density function (PDF) for the magnitude is built, corresponding to the relative contribution of the magnitude bins in terms of moment rates within the system (Fig. 3, see also Chartier et al. 2017, 2019). The exponential decrease of rates with increasing magnitudes is compensated by the huge increase in moment rate with magnitude. Using this pdf to sample magnitudes, large magnitudes are picked much more frequently than low magnitudes."

Chartier et al. (2017, 2019) defined how the pdf in magnitude, used to pick the magnitude during the iterative process, is built. We acknowledge that this is not straightforward, but this is inherent to the SHERIFS algorithm.

We have modified the § as follows : "Based on the hypothesis that earthquake rates follow a Gutenberg-Richter distribution, a probability density function (PDF) for the magnitude is built, corresponding to the relative contribution of the magnitude bins in terms of moment rates within the system (see Fig. 1 in Chartier et al. 2017). The Gutenberg-Richter model delivers probabilities of occurrence that decrease with increasing magnitude according to an exponential. In SHERIFS, these probabilities are multiplied by the corresponding moment rates, then normalised, to obtain the final probability density function used to sample magnitudes in the iterative process. The exponential decrease of rates with increasing magnitudes is compensated by the huge increase in moment rate with magnitude. In the final PDF, probabilities increase with magnitude (step 1 in Fig. 3). Using this PDF to sample magnitudes, large magnitudes are picked much more frequently than low magnitudes. "

We add the following Figure in an Appendix (Fig. A1):



L236

Why? Aren't you drawing magnitudes from a GR law? Large magnitudes should be less frequently picked than smaller ones to be coherent with a GR distribution. please, clarify

This is a specificity of SHERIFS' algorithm. As explained above at each iteration the magnitude is picked in a pdf in magnitude, with the probability representing the relative contribution of the magnitude bins in terms of moment rate within the system. Thus, this pdf delivers probabilities that increase with magnitude. Large magnitudes are picked more frequently than moderate magnitudes. This set of magnitudes/ruptures is not a synthetic catalog. Rates are associated to magnitudes/ruptures so that eventually the magnitude-frequency distribution at the scale of the system fits a Gutenberg-Richter distribution (Figure 4, second row). We fully agree with the reviewer that this is not straightforward nor intuitive, but this is inherent to how SHERIFS was designed.

Please see page 1860 in Chartier et al. (2017) : "To convert the target MFD, expressed in terms of rate of earthquakes, into moment rates (Fig. 1b). This target MFD will be used to pick the magnitude bin on which an increment dsr will be spent. Notice that the formulation in terms of moment rate implies that greater magnitudes are more likely to be picked."

L249
"we set the maximum jump to 10km": Why not to set a smooth cutoff, for instance, a sigmoid function with various steepnesses to model the upper jump? I think that the final results will be more reliable. Using a probabilistic weight for rupture jumps would allow the occurrence of very large earthquakes are will make your model more flexible and less prone to biases due to the limited set of data on which they are based.
This is true. Using a smooth cutoff function, such as a sigmoid function with adjustable steepness, could offer a more realistic approach to modeling upper rupture jumps. However, in the present study we apply SHERIFS as implemented originally by Chartier et al. (2017, 2019).

Figure 4a
Measure units of the cumulative annual rate? It is a magnitude-frequency distribution, so it is the usual cumulative annual seismic rate.
The red line suggests that events larger than m 7.5 may actually occur in the long term. Why do not allow such a possibility?
The magnitude-frequency distribution allows earthquakes with magnitude 7.5, at maximum.

L368-370
"Overall, using the interconnected fault model, the hazard levels decrease along the main strand (from ~0.7-0.8 to ~0.5-0.6g), but increase along the secondary faults (from ~0.4 to ~0.5g), with respect to the classical implementation. "
Perhaps, the PGA is not the best observable to highlight differences between the two models. I would suggest to add another parameter to include the spatial extention of high ground acceleration during the same event. I would expect that during a large event, the extension of the strongly shaken region is larger, not only because even secondary faults can be activated at the same time of adjoining segments.
This map displays acceleration levels for a given mean return period. Earthquakes of different magnitudes at different distances for each site contribute. These probabilistic seismic hazard maps are not scenario maps that would display the ground motions produced during one single event. They can't be analysed as if they correspond to specific earthquake scenarios.

L387
"A realistic fault model for the Levant fault system: full connectivity and M_{max} 7.9" :
Good! But why not even higher?
i suppose because the size of the largest seismogenic zone cannot be larger than that needed for a M 7.9. However, please, explain in detail the reason for your choice.
We decided to test M7.9 after analysing the earthquakes that occurred worldwide, along comparable strike-slip plate boundary faults. Next, in Section 5.2 we test maximum magnitude 8.1.

L394
"we test two potential maximum magnitude: 7.9 and 8.1"
why?
We have modified the introduction text of Section 5:
"Reviewing other major strike-slip fault systems worldwide and the largest earthquakes they have generated (e.g. the M_W 7.8 1906 earthquake on the San Andreas, Yeats et al. 1997; 2002 M_W 7.9 Earthquake along the Denali fault in Alaska, Eberhart-Phillips et al. 2003; or the recent 2023 M_W 7.8 earthquake on the East Anatolian fault, Zhang et al. 2023), we believe

magnitudes larger than 7.5 could occur along the Levant fault system. Thus, the source model for PSHA must include the possibility for large events, and therefore we test two potential maximum magnitudes: 7.9 and 8.1 *(magnitude 7.9 because it is the maximum magnitude observed on a strike-slip fault system, magnitude 8.1 to allow a larger earthquake than observed)*."

L440-441
"We believe that a 5% percentage of aseismic deformation is more realistic than 9 or 11%, for the Levant fault system"
This choice should be motivated based on a quantitative analysis.
5% and 10% are both small values of aseismic deformations, but with a pretty large impact on the hazard. So, please, provide a more detailed justification of your assumptions.
We have added the following sentence: "Based on interferometric time-series analysis of satellite radar images, Li et al. (2024) have shown that no significant aseismic slip can be measured anywhere along the entire system."
Li X., Jonsson S., Liu S., Ma Z., Castro-Perdomo N., Cesca S., Masson F., Klinger Y., Resolving the slip-rate inconsistency of the northern Dead Sea fault, Sci. Adv., 10, eadj8408, 2024.

Figure 11
In your work you assumed b-value = 1. Why do not you allow possible b-value variations? Increasing the connectivity of the fault network may produce a decrease of the b-value. Even though I understand that additional simulations are beyond the goal of the present article, it would be nice to include a short discussion on this topic, if you agree.
This is true, only one estimate for the b-value is currently considered. In a previous article (El Kadri et al., Bulletin of Earthquake Engineering, 2023), we implemented a moment-balanced fault model and showed that the uncertainty on the b-value has a negligible impact on the hazard estimates. In this 2023 article, the source model was based on a classical implementation of faults (isolated segments). We believe that the conclusions should also hold for an interconnected fault system as implemented in SHERIFS, but we would need to perform the test to be certain. We have added the following sentence in the text end of Section 3 :
"Additionally, we use a b-value equal to 1. The choice of the b-value may impact the seismic rates obtained, however El Kadri et al. (2023) have shown that using moment balanced magnitude-frequency distributions with b-values within a reasonable range (0.85-1) has little impact on hazard estimates. "

L512
The catalog is not complete below a certain completeness magnitude, which I suppose being larger than 4.1. This conclusion seems coherent with your description Only events above the completeness should be included in the analysis. If you need all of them because their number is already limited, please, discuss the possible bias this choice may entail.
We mention 35 events in the instrumental catalog with magnitude between 4.1 and 6.1, in total, before considering the periods of completeness. Considering only earthquakes within periods of completeness, 20 events remain.

L531-532
"We estimate periods from the ISC-GEM catalog at the global scale: magnitudes larger or equal to 5.6 are considered complete since 1965, and magnitudes larger or equal to 6.1 since 1925. ":
My check returns Mc = 5.3 since 2001, 5.5 since 1991, 5.8 from 1965 to nowadays and 6.2 since 1925. However, they are global averages: what about the regional completeness?
We use 5.6 since 1965, there might be a 0.2 magnitude degree uncertainty.

For such a level of magnitude in the ISC-GEM catalog, the completeness estimated at the global scale should be approximately the same at the regional scale (at the scale of the Middle East).

Figure 14
Could you explain the blue bar in the caption, please?
We have added the following sentence : "Blue bar : mean recurrence times inferred from paleoseismic trenches (Lefevre et al. 2018)."

Conclusion
I think that conclusions would be more effective if made shorter: move part of their content in discussions.
We have moved this paragraph to the end of Section 3:
"In the SHERIFS iterative process, magnitudes are sampled in a PDF at each iteration and associated to a combination of segments (with area matching the magnitude). At the scale of the system, the summed seismic rates follow a Gutenberg-Richter magnitude-frequency distribution (or another MFD shape). However, the set of ruptures and associated rates does not constitute a synthetic catalog (Chartier et al. 2019)."

L611
It should be recognized that several recent models are now available allowing interactions of fault segments, e.g., for California etc.
add a few lines recognizing previous advances in this field, please
Sure, there is an entire paragraph on this in the introduction:
"Therefore, several methods have been developed to take into account these complex ruptures into hazard models. In 2014, the Working Group on California Earthquake Probabilities (WGCEP) developed a new inversion-based methodology called the "Grand inversion", to relax fault segmentation and incorporate multifault ruptures in the Uniform California Rupture Forecast (UCERF, Field et al., 2014; Page et al., 2014). Subsequently, Chartier et al. (2017) implemented the SHERIFS (Seismic Hazard and Earthquake Rate In Fault Systems) algorithm, a method to relax fault segmentation which is simpler than the UCERF framework and that requires less input parameters. Additional algorithms were also developed, such as the integer-programming optimization by Geist and ten Brink 2021, or the SUNFiSH approach by Visini et al. (2020). "
Besides, as indicated above, in Section 3 we have added the following two references related to UCERF3 in California:
Milner, K., Page, M.T., Field, E. H., Parsons, T., Biasi, G., and Shaw, B. E., 2013. Defining the inversion rupture set via plausibility filters, U.S.G.S Open-File Report 2013-1165, Uniform California Earthquake Rupture Forecast Version 3 (UCERF3) - The Time-Independent Model, Appendix T, 14 pp.
Milner, K. R., B. E. Shaw, and E. H. Field (2022). Enumerating Plausible Multifault Ruptures in Complex Fault Systems with Physical Constraints, Bull. Seismol. Soc. Am. 112, 1806–1824, doi: 10.1785/0120210322

L682
Add fault and slip rate data you used in this research, please
Slip rate estimates are reported in Table 1 in the manuscript.
We now add as an electronic supplement the new set of fault sections.

RC2: 'Comment on nhess-2024-184', Anonymous Referee #2, 07 Feb 2025 reply
Review of the paper "Implementation of an interconnected fault system in PSHA, example on the Levant fault" by

Sarah El Kadri, Celine Beauval, Marlene Brax, and Yann Klinger
The manuscript presents an analysis of the Levant Fault System (LFS), a 1200 km-long left-lateral strike-slip fault, with the aim of improving regional Probabilistic Seismic Hazard Assessment (PSHA) models. By considering the interconnected nature of the LFS, the study challenges traditional approaches that treat faults as isolated segments. Building upon El Kadri et al. (2023), who developed a seismic hazard model for Lebanon using classical fault segmentation and a moment-balanced recurrence model, this research explores how fault connectivity affects source modeling and hazard estimates. It examines various rupture scenarios, both single and multi-segment, using the SHERIFS and OpenQuake Engine software.
The study provides valuable insights, but the structure of the manuscript could be improved for better clarity and coherence. One key concern is the inconsistency in the treatment of maximum magnitude (Mmax). While the study initially assumes Mmax = 7.5, different values are later tested without a clear rationale. A more structured approach would be to incorporate multiple Mmax values from the outset, considering both exponential and characteristic models within the interconnected fault system. This would enable a clearer comparison with observed seismicity rates and better illustrate model and parameter uncertainties, which are central to the study's objectives.

As explained in the manuscript, we start with an Mmax of 7.5 (Section 4), to enable a comparison between the classical implementation of faults and an interconnected fault model. As 7.5 is the mean Mmax obtained considering isolated faults, we have to start with a comparison relying on this value.
Next, Section 5 tests larger Mmax values.
Following RC3 comments, we have renamed the sections 5, 5.1, 5.2, 5.3 so that the approach can be easily followed. We eventually calculate hazard with our preferred model associated to an Mmax that leads to the lowest aseismic deformation. The present study aims at understanding how the SHERIFS algorithm works, the aim is not a systematic exploration of uncertainties. In a complete PSHA study these alternative Mmax would populate the source model logic tree. We have added the following sentences in the text, at the end of Section 5.2:
"This is an exploratory study aimed at understanding how the algorithm SHERIFS works. In a probabilistic seismic hazard study aimed at delivering seismic hazard levels for a country, we would populate the source model logic tree with these alternative models to cover the epistemic uncertainty (attributing larger weight to the model associated to the lowest aseismic deformation). "
We have modified the introduction text of Section 5:
"Reviewing other major strike-slip fault systems worldwide and the largest earthquakes they have generated (e.g. the M_W 7.8 1906 earthquake on the San Andreas, Yeats et al. 1997; 2002 M_W 7.9 Earthquake along the Denali fault in Alaska, Eberhart-Phillips et al. 2003; or the recent 2023 M_W 7.8 earthquake on the East Anatolian fault, Zhang et al. 2023), we believe magnitudes larger than 7.5 could occur along the Levant fault system. Thus, the source model for PSHA must include the possibility for large events, and therefore we test two potential maximum magnitudes: 7.9 and 8.1 *(magnitude 7.9 because it is the maximum magnitude observed on a strike-slip fault system, magnitude 8.1 to allow a larger earthquake than observed)*."

Additionally, some assumptions are presented without sufficient justification, which could lead to confusion. Providing clearer explanations would enhance the readability and impact of the manuscript. A major revision is recommended to improve the coherence, transparency, and overall clarity of the study.

- The segmentation of the fault system into 43 sections provides valuable insight into its complexity. However, it would be helpful to clarify how this segmentation process influences the final results. Specifically, how do variations in segment length (min 4km max 40km) and number affect earthquake rate estimates and hazard assessment?

  We test only one segmentation and unfortunately we can't evaluate what would be the impact on the results of different ideas of the segmentation. Often we tend to oversample our fault system by using quite short sections. In this way we get closer to a continuous system where rupture could start and stop anywhere, which is not realistic because fault geometry plays a role.

  We have added in the Appendix (Fig. A2) the histogram of fault section lengths, so that the reader can appraise the lengths of the 52 sections.

- The hypothesis that strike-slip fault segments are activated together with thrust fault sections (GF, MF, MLT) requires further clarification. Given the differences in kinematics and orientation between these structures, on what basis is this hypothesis supported? Are there geological, paleoseismological, or Coulomb stress analysis results that justify this assumption?

  Although there is no specific evidence from paleoseismological work that the strike-slip sections of the LFS did rupture together with some thrust sections, this possibility cannot be ruled out. Documented examples of partitioning during recent past ruptures are numerous with parallel thrust and strike-slip rupture (e.g. the 2008, Mw7.9 Wenchuan earthquake, Xu et al. (2009)) or parallel normal and strike-slip rupture (e.g. the 2001, Mw 7.8 Kokoxili earthquake, King et al. (2005)) breaking together.

- The manuscript considers earthquakes up to Mw 8.1, despite paleoseismic records not exceeding magnitude 7.5. Could the authors justify the use of models that incorporate larger magnitudes?

  We test a maximum magnitude of 7.9 and 8.1.

  Checking the historical earthquake catalog by Brax et al. (2019, Lebanese region), taking into account the magnitude estimate from macroseismic intensities and adding the uncertainty yield magnitudes of 7.7 (363 May 19), 7.6 (1170 June 29), or 7.8 (1202 May 20).

  Some authors believe that in the long term, magnitudes 8 are possible along this fault system, e.g. Lu et al. (2020) adopt a maximum magnitude of 8.0 for the fault system (Y. Lu, N. Wetzler, N. Waldmann, A. Agnon, G. P. Biasi, S. Marco (2020), A 220,000-year-long continuous large earthquake record on a slow-slipping plate boundary. Sci. Adv. 6, eaba4170).

  We cannot rule out the possibility for a magnitude 8-8.1, and we test this value.

- The study states that the interconnected fault model leads to increased hazard along secondary faults and lower hazard along the main strand compared to the classical implementation. However, variations in hazard levels are influenced not only by segmentation and connectivity assumptions but also by the annual probability of exceedance and return periods of individual ruptures, which vary with the maximum magnitude considered. It would be useful to clarify how these factors influence the observed differences in hazard estimates between the classical and interconnected models.

Indeed, the precise distribution of earthquake magnitudes along the fault system controls the hazard levels. To this aim, we calculate participation rate of sections, to actually plot in space seismic rates. Our present study is the first article implementing SHERIFS that does this effort. Currently, we stress in the text that hazard levels can be understood based on the distribution in space of magnitude occurrence rates:

Section 3 (Figure 4, Figure 5) : "SHERIFS' algorithm delivers a set of sections and sections' combinations (ruptures) with associated magnitudes and occurrence rates. In previous applications of SHERIFS, no information is provided on the obtained distribution of rupture magnitudes in space. Knowing how seismic rates are distributed in space is key to understanding the geographical pattern of hazard levels. In PSHA, at a site, ground-motion exceedance rates are calculated by multiplying rates of ruptures with the probabilities that the ruptures produce an exceedance of the ground-motion levels at the site. Ruptures close to the site will contribute more than ruptures away from the site. In the present study, we aim at understanding the exact distribution in magnitude and space of the ruptures, and its link with hazard levels."

Section 4.3 : Figures 6 and 7

Section 4.4 : "In the interconnected model, hazard levels are no longer uniform within a fault, they vary significantly depending on the location of the site along the fault. They are highest along the southern part of the Yammouneh fault, as well as along the southern part of JVF, and northern part of Araba fault, corresponding to the sections with the highest rates in the moderate magnitude range (Figs. 5a and 5b, rates for magnitudes 6 and 6.5). These higher hazard levels can be explained by the observation that moderate magnitudes often control hazard estimates at 475 years return period, when a Gutenberg-Richter model is used (e.g., El Kadri et al. 2023).

For sites above the dipping Mount Lebanon Thrust, the interconnected fault model delivers hazard levels much higher along the southern part than in the north. The northern sections of Mount Lebanon Thrust are involved in more large magnitude ruptures than the southern sections, as they may break with segments from the Missyaf and Yammouneh faults. Southern sections cannot rupture with the Roum fault when the maximum jump is set to 10km and as a consequence, annual rates of moderate magnitudes are higher in the south resulting in higher hazard. "

- For a maximum magnitude of approximately 7.9, the study states that the LFS fault system must be fully released. At a 475-year return period, PGA values reach around 0.3g within 20 km of the faults, with peak values of approximately 0.5g along specific sections. However, these relatively low PGA values appear inconsistent with an M7.9 earthquake, as recent smaller earthquakes have recorded PGA values close to 1.0g at similar distances. Could the authors clarify this discrepancy? Additionally, would it be beneficial to present hazard maps with a 2% probability of exceedance in 50 years to better capture the strong shaking potential of rare M7.9 events?

The seismic hazard map displays the acceleration levels that have a 10% probability of being exceeded at least once in the next 50 years, taking into account contributions from all earthquakes (up to maximum magnitude 7.9). This map does not display the ground motions that would be obtained if an earthquake of magnitude 7.9 would occur (this would be a scenario map, the worst case). It is not possible to apply deterministic reasoning on a probabilistic seismic hazard map. Hazard levels will increase if considering a lower probability of exceedance (e.g. 2% over 50 years, mean return period 2475 years).

- Lines 63-67 discuss how recent earthquakes have demonstrated that ruptures can propagate across geometrical discontinuities, leading to larger-than-expected magnitudes. It may be helpful to include the 2023 Mw 7.8 and Mw 7.5 Kahramanmaraş, Turkey earthquakes, which ruptured a complex fault system just north of the study area. These events further support the need for an interconnected fault model in PSHA.

  The Mw 7.8 event occurred along a fault that was included in up-to-date source models for PSHA (ESHM20, Danciu et al. 2024). Such long ruptures were included in the ESHM20 model and did not come as a surprise. See e.g. Weatherill et al. EGU 2023, https://doi.org/10.5194/egusphere-egu23-17610. Nonetheless, such doublet events are not modeled in present time-independent source models for PSHA.

- The discussion on the Grand Inversion in UCERF and the SHERIFS algorithm is useful, but a brief comparison of their key differences would improve clarity. The explanation in lines 84-88, describing how SHERIFS constructs rupture scenarios, assigns occurrence rates, and distributes seismic moment, could also be better structured and clarified.

  The section describing the SHERIFS algorithm has been modified following comments from RC1.

  The Grand Inversion in Uniform California Earthquake Rupture Forecast (UCERF3, Field et al., 2014) relies on a completely different methodology. The rates of all earthquakes are solved for simultaneously and from a broad range of data (slip rates,regional MFD, paleoseismic event rate constraints), using a system-level inversion. The inverse problem is large and underdetermined, so a range of models is obtained using a simulated annealing algorithm.

- Lines 161-163 introduce the term "seismic gaps" without a clear definition. Are the authors referring to a lack of recent seismicity, barriers to rupture propagation, or zones with unknown fault connectivity? Clarifying this would improve the interpretation of fault segmentation and its impact on hazard assessment.

  Gaps are potential barriers to rupture propagation. The first sentence in the article that refers to gaps is the following in the introduction : "A number of earthquakes in the last 30 years have shown that ruptures can jump over some geometrical discontinuities, such as gaps or steps in the fault system, that were previously considered as major obstacles to rupture propagation."

  Then in section 2 we clearly relate gaps to pull aparts : "The LFS mostly exhibits transtensional features, such as the significant pull-apart structures of the Gulf of Aqaba, the Dead Sea (gap width ~14km), and the Ghab pull-apart (~11km). "

  However we also stress that gaps could be smaller than they currently appear in map-view (last sentence of Section 2) : "In the present work we test different levels of connectivity, allowing progressively larger jumps for ruptures. Nonetheless, it is important to keep in mind that within these discontinuities, substantial uncertainty exists regarding the presence of secondary faults connecting neighboring faults. Hence, these gaps might be smaller than they currently appear in map-view."

Similarly, in lines 200-202, the manuscript states that if section lengths are too heterogeneous, the algorithm subdivides longer sections to homogenize segment lengths. What criteria define heterogeneity? Is there a maximum allowable section length or a statistical or geological basis for these subdivisions? Furthermore, does this segmentation process impact rupture connectivity assumptions and earthquake rate estimates?

There is three rules in SHERIFS for subdividing long sections:

- Maximum Section Length (40km)
- Maximum Number of Sections (6)
- Local Changes in Fault Azimuth (Orientation)

But in our case only the first one is used “Maximum Section Length (40km)”. So, if the fault length is shorter than 40km, it is not divided further. If the fault is longer, it is split into multiple sections, each not exceeding 40km.

We have modified the text as follows: “nine tectonic sections with length larger than 40km are arbitrarily subdivided into two sections” (Section 3).

- Lines 205-208 suggest that large magnitudes are picked more frequently than smaller magnitudes, which contradicts the Gutenberg-Richter model, where lower-magnitude earthquakes should be more frequent. Could the authors clarify how magnitudes are sampled? Is there a weighting applied to moment rates that modifies the expected earthquake rate distribution?

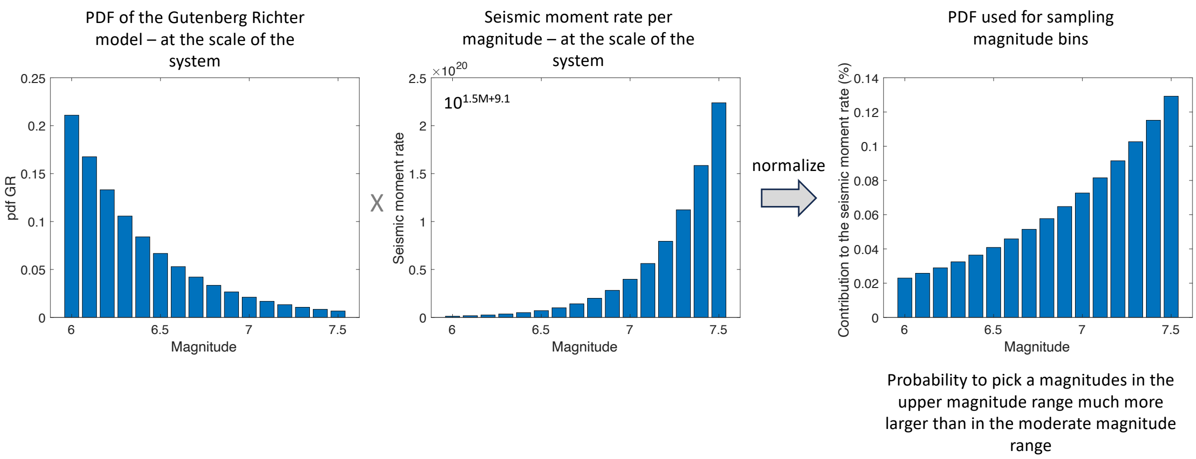
Chartier et al. (2017, 2019) defined how the pdf in magnitude, used to pick the magnitude during the iterative process, is built. We acknowledge that this is not straightforward, but this is inherent to the SHERIFS algorithm.

We have modified the § as follows : “Based on the hypothesis that earthquake rates follow a Gutenberg-Richter distribution, a probability density function (PDF) for the magnitude is built, corresponding to the relative contribution of the magnitude bins in terms of moment rates within the system (see Fig. 1 in Chartier et al. 2017). The Gutenberg-Richter model delivers probabilities of occurrence that decrease with magnitude according to an exponential. These probabilities are multiplied by the corresponding moment rates, then normalised, to obtain the final probability density function used to sample magnitudes in the iterative process. The exponential decrease of rates with increasing magnitudes is compensated by the huge increase in moment rate with magnitude. In the final pdf, probabilities increase with magnitude (step 1 in Fig. 3). Using this pdf to sample magnitudes, large magnitudes are picked much more frequently than low magnitudes. “

This is a specificity of SHERIFS’ algorithm. At each iteration the magnitude is picked in a pdf in magnitude, with the probability representing the relative contribution of the magnitude bins in terms of moment rate within the system. Thus, this pdf delivers probabilities that increase with magnitude. Large magnitudes are picked more frequently than moderate magnitudes. This set of magnitudes/ruptures is not a synthetic catalog. Rates are associated to magnitudes/ruptures so that eventually the magnitude-frequency distribution at the scale of the system fits a Gutenberg-Richter distribution (Figure 4, second row).

See page 1860 in Chartier et al. (2017) : “To convert the target MFD, expressed in terms of rate of earthquakes, into moment rates (Fig. 1b). This target MFD will be used to pick the magnitude bin on which an increment dsr will be spent. Notice that the formulation in terms of moment rate implies that greater magnitudes are more likely to be picked."

We add the following figure in the Appendix (Fig. A1).



- The slip rate increment is described, but its definition and application remain unclear. Is it sampled from a probability density function or applied in fixed steps? How does it relate to long-term slip rate constraints?

  Section 3: "The moment rate is distributed through an iterative process over magnitudes and associated sections or sections combinations. In a preliminary step, the algorithm establishes all possible ruptures, or section combinations, and associates earthquake magnitudes to these ruptures by applying the area-magnitude scaling relationship. Then, an iterative process starts (Figure 3) where at each iteration, the same amount of slip rate is spent (called 'dsr'). "

  Lines 237-240 also appear contradictory, as they state that the Gutenberg-Richter model is used while indicating that large magnitudes are picked more frequently. Since the maximum magnitude assumption (Mmax = 7.5) strongly influences hazard estimates, how does this assumption impact the b-value and the overall shape of the magnitude-frequency distribution?

  Please see above.

- Figure 5 appears before Figure 4, and in lines 257-266, Figure 4 is referenced as illustrating the process at three different steps, but this does not seem to be the case. Did the authors mean Figure 3 instead?

  We are sorry for this mistake. Figures are now numbered correctly.

- The discussion on aseismic slip (9%) suggests that it remains constant, but does it vary with Mmax choices?

  Considering Mmax 7.5 and maximum jump of 10km leads to 9% of aseismic deformation (Section 4). Next, considering larger Mmax, and different maximum jumps lead to different among of aseismic deformation (Section 5). Every new run, with new Mmax, leads to a different amount of aseismic slip (Fig. 11). The discussion on which Mmax/associated MFD is the most appropriate relies on this amount of aseismic slip.

Additionally, in lines 440-442, the choice of 5% aseismic deformation is mentioned, but the rationale behind this assumption is unclear. Could the authors clarify the motivation for this choice? If a tapered Gutenberg-Richter model was used instead of a truncated Gutenberg-Richter model, how would it impact the slip distribution and moment balance?

We have added the following sentence: “Based on interferometric time-series analysis of satellite radar images, Li et al. (2024) have shown that no significant aseismic slip can be measured anywhere along the entire system.“

Li X., Jonsson S., Liu S., Ma Z., Castro-Perdomo N., Cesca S., Masson F., Klinger Y., Resolving the slip-rate inconsistency of the northern Dead Sea fault, Sci. Adv., 10, eadj8408, 2024.

To evaluate the exact impact of a tapered distribution we would need to perform the test. Nonetheless, we would obtain the same tendencies (a larger corner magnitude Mc would lead to lower rates in the moderate magnitude range).

Figures 14 and 15 appear to be important as they provide a comprehensive comparison. The figure 14 include two different models, characteristic and exponential, and it seems that the characteristic interconnected model with 14% fits well with observed events for M > 6.0. Could the authors comment on this observation?

We comment on these results as follows : “To know if a characteristic Youngs and Coppersmith (1985) distribution would be more compatible with observed rates, we run again the algorithm with an M_{max} 7.9, full connectivity, and a characteristic earthquake model. The model obtained is roughly consistent for magnitudes larger or equal to 7.1, but strongly underpredicts rates for magnitudes larger or equal to 6.6 and 6.1. Fourteen percent of the total slip rate is not used and considered aseismic, which is not realistic. ” We believe that this model strongly underpredicts rates for magnitudes between 6.1 and 7.

Additionally, the figure captions and curve definitions are not clearly described, making them difficult to interpret. Improving these descriptions would enhance clarity and ensure a more accurate understanding of the presented results.

We have corrected the caption of Figure 15.

RC3: 'Comment on nhess-2024-184', Anonymous Referee #3, 10 Feb 2025 reply
It is a pleasure from time to time to find a paper well done as the one by El Kadri et al. dealing with PSHA in the Levant fault system relaxing the assumption of isolated fault segments. The authors provide evidences that many ingredients are affected by large uncertainties, some choices are subjective, but the "exploration" character of the study is clearly stated in the text.
True, thank you for underlying that important aspect.

My comments are therefore to be intended as suggestions to the authors, not mandatory requests of changes.
1) chapter 2: the observations on trenches and earthquake catalog are mentioned in Fig. 2, but the description is given in chap. 6.1-6.2. Consider to anticipate part of these subchapters in chap. 2 (also fig. 13), leaving the comparison with models in chap 6.

We have thought about the suggestion, but we prefer to keep the presentation of earthquake catalogs in Section 6. We have simply modified this sentence in Section 2: "Lefevre et al. (2018) has summarized the known history of major earthquakes along the southern fault section, between the Gulf of Aqaba and the Sea of Galilée, over the last ~1200 years, based on tectonic, paleoseismic, and historical data (see trench sites in Fig. 2). "

2) chapter 3: the description of the SHERIFS steps is always an headache, quite intriguing the transformation of a classical G-R in a pdf of the relative contribution of mag bins in the moment rates within the system. I suggest to clearly state that the moment rate for a section is L*W*mu*sliprate (according to the values listed in Tab.1), and the global budget of the system is the summation of the moment rates of all the sections.
OK, we have added this equation in Section 3.

The intrinsic problem of modelling "independent" events from the statistical point of view, or full earthquake sequences remains open, and probably deserves to be mentioned, or commented when the "acceptable" aseismic budget is mentioned.

We understand that the reviewer refers to the moment rate of clustered events (foreshocks, aftershocks). In theory, the source model built is made of mainshocks, so indeed we do not count the moment rate associated with clustered events. However, when applying one of the classical declustering method to an earthquake catalog (e.g. Gardner and Knopoff 1974, Reasenberg 1985) the aftershocks usually represent a very small percentage of the total moment rates (see e.g. Marinière et al. 2021, example source zone with 52% event with M>=4.5 identified as aftershocks, representing only 0.8% of the total moment rate).
We have added the following sentence in the text:
"Using the term 'non-mainshock slip' may imply that this slip could correspond to aftershocks that are not modeled, however aftershocks usually represent a negligible fraction of the total moment rate (see e.g. Marinière et al. 2021)."

3) chapter 4: the distribution in space of the magnitude rates, given in fig. 5 is interesting, even if I am not sure that the "participation rate" of a section to a rupture can be considered a truly annual rate. Please specify what happens to those sub-sections in Tab. 1 (e.g 1-3, 43, 32-34) not allowing M>=6.0; is their modelling limited to a participation rate in bigger ruptures?
Yes, they participate to larger ruptures.

Then, considering fig. 6a the difference in annual rates of moderate magnitude is not distinguishable as stated in lines 323-324. I wonder if the moment budget is fully preserved, as stated for the results showed in Fig. 8

In the classical implementation, all the moment rate budget is spent in earthquakes, whereas in the interconnected fault model, in this run (Mmax 7.5, maximum jump 10km), 9% of the moment rate budget is not spent in earthquakes.

4) chapter 5: I suggest to change some titles in this chapter, as it is really difficult to state which model is the most realistic, given the uncertainties in paleoseismic magnitudes, incompleteness of the catalog, evaluation of the acceptable aseismic budget (considering also that the modelling is limited to M>=6, and some seismic moment budget could be spent by small ruptures too). Titles such as for example: "5. Sensitivity tests on Levant fault system; 5.1 Test with Mmax 7.9 and full connectivity; 5.2 Selection among different hypotheses" are more neutral and open to alternative interpretations (e.g. the char eq model, instead of G-R, that in my opinion could accomplish better the absence of moderate magnitude in the last century.
OK, we have modified the titles as suggested.
"5 A realistic fault model for the Levant fault system: full connectivity and M_{max} 7.9"
Modified, now "Testing different maximum magnitudes for the Levant fault system"
"5.1 Test with M_{max} 7.9 and need for full connectivity"
kept
"5.2 Selection of the most realistic model among models tested"
Modified, now "Selection of the model with lowest aseismic deformation"
"5.3 Hazard levels associated to our preferred fault model (M_{max} 7.9 and full connectivity)"
kept

5) chapter 6: I suggest to repeat in this chapter some basic data/assumptions used, e.g. at line 485, after "of occurrences" add "grounded on slip rates, and fault surfaces ad given in Tab. 1." and at line 487, after "given shape" the sentence "i.e. a G-R truncated model with b-value=1".
Ok, sentence modified
Then you cannot say that paleoseismic data were not used to derive the model, as I presume that slip rates of fault segments are mainly controlled by these data.
This is not the case for trenching across strike-slip faults, where slip rates based on trenching are in general dubious. Here, most of the long-term slip rate is not established based on paleoseismology but rather on geomorphologic data. Long-term slip rates can be inferred from paleoseismology only when a long time series for earthquakes with measurement of slip per event is available (within the LFS, only available for the trench in the Jordan Gorge section, Wechsler et al. 2018).
We have modified the paragraph as follows:
"The earthquake forecast delivers a magnitude-frequency distribution at the scale of the fault system that follows a given shape, here a Gutenberg-Richter model with a b-value of 1. This magnitude-frequency distribution is moment-balanced with the long-term slip rates. Long-term slip rates on strike-slip faults are mainly established from geomorphologic data (see El Kadri et al. 2023). Slip rates can be inferred from trenching only if a long time series of earthquakes with a measurement of slip per event is available, which is the case only for the Jordan Gorge section from 3D trenching (Wechsler et al. 2018). Both the earthquake catalog of the region and the available paleoseismic data were not directly used to derive the model; these observations can be compared with the earthquake forecast."

I also suggest to avoid "regularly" (at line 515),
OK, sentence has been modified : "a number of destructive earthquakes with magnitudes larger or equal to ~6.5 occurred in the last 2000 years in the region"
stress the uncertainties in assigning magnitude in trenches (the Central Italy sequence in 2016 is a lesson on how many events of 6<m<6.5 can generate a rupture of a single M=6.7).

True, the width of the grey boxes highlight these uncertainties in magnitude, "Daeron et al. (2007) evaluated a characteristic coseismic slip of about 5.5m, which according to Leonard (2014) corresponds to an interval of magnitude 7.4 to 8 (extension of the grey box on the graphic)."
The standard deviations assigned to M at line 535 are underestimated, in my opinion;
We have used 0.3 standard deviation for historical events and 0.1 for instrumental events. This is true that it is a lower bound for uncertainties on a historical catalog and an instrumental catalog based on global data. We believe the mean rates obtained would be the same if using e.g. 0.5 and 0.2.
I suggest to be less sharp also in the description of the model fit vs observations given in Fig. 14.
We did not try to derive more conclusions from the comparison. The comparison is important but bears large uncertainties, highlighted in the text.

Here some technical, minor corrections:
a) line 27 and in the following text: I suggest to avoid the use of "secondary" faults, as it can make confusion with the terminology used in displacement hazard. I suggest to use the terms "splays" or "branches" that clearly address the departure from the main fault strand.
We agree that it could be confusing in some sections of the text and with few exceptions where there is no ambiguity we changed "secondary fault" for other vocabulary according to the context (branch, splay fault).
b) line 281, Figure 3 is figure 4, correctly mentioned in the text
corrected
c) line 321, figure 4a is is figure 6a
corrected
d) line 341, figure 4 is figure 6,
corrected
I suggest to change the color of the dash gray line in frame (a) with black
Done
e) add a scale bar in both Fig. 8 (line 382) and Fig. 12 (line 479)
Done
f) at line 528, after "363" add "A.D.".
corrected
At line 540 add "known" before "fault".
corrected
g) lines 601-603, Solid orange and dashed lines are yellow and orange solid lines, if I have correctly understood
We have corrected as follows: "Red line: fully interconnected model with M_{max} 8.1 for the Gutenberg-Richter system MFD. "

Congrats to the authors.
Thank you for your feedback, and for this constructive review.