Responses to Referee #2

We thank the anonymous referee for their careful reading of the manuscript and their constructive and insightful comments. Below, we address each comment in detail and explain how these points will be implemented in the revised version of the manuscript.

RC2–1 The Introduction outlines well-known limitations of catalogue-based PSHA and motivates the use of fault-based and multi-fault rupture approaches. However, the Discussion does not clearly close the loop by explicitly demonstrating how the presented results resolve, reduce, or reframe the issues raised in the Introduction. Yet, the manuscript stops short of articulating what new insight this study provides for that debate, beyond confirming that multi-fault ruptures perform better in this case study.

Author's response (AR): The main challenges in fault-based PSHA applications adopting "relaxed segmentation" frameworks are: (1) the explicit definition of possible earthquake rupture scenarios and (2) the estimation of their associated rates of occurrence. While the approach employed in our application does not tackle or "solve" the physical questions behind the plausibility of multi-fault ruptures in the southern Apennines (that would be the role of physics-based investigations; see next comment), it does provide a systematic framework to structure and explore uncertainties related to their definition (earthquake rupture configurations) and rates within a PSHA context. In this framework, the approach does allow to narrow some of these uncertainties through consistency analyses with available observational data (instrumental seismicity and paleoseismic rates). At the same time, our hypothesis-based setup allows preserving the probabilistic nature of PSHA, as the results are intended to be implemented within a logic-tree scheme. In the revised version of the manuscript, we will place emphasis on discussing how our approach might help to address uncertainties arising from arbitrary decisions on earthquake rupture scenarios and earthquake rate estimates for PSHA.

RC2–2 The Discussion correctly acknowledges that SHERIFS is not a physics-based model and therefore cannot assess rupture compatibility or stress interactions. While this caution is appropriate, it also weakens the strength of the conclusions regarding the feasibility of large multi-fault events (e.g., Mw 7.6). The manuscript would benefit from a clearer distinction between statistical consistency with observations and physical plausibility of rupture scenarios. At present, these two concepts are sometimes conflated, which may lead readers to overinterpret the robustness of the conclusions.

AR: We agree with the reviewer that the distinction between physical plausibility and statistical consistency should be more clearly stated to avoid potential misinterpretations. In the revised manuscript, we will systematically differentiate references to "feasibility" from those referring to "statistical consistency" where appropriate.

RC2–3 Section 4.5 proposes explicit weighting of rupture sets and modelling branches for PSHA logic trees. While this is one of the most practically relevant outcomes of the study, the Discussion does not sufficiently address: the transferability of these weights beyond the study area, whether the proposed weighting scheme is meant as a general methodological template or as a region-specific recommendation. This ambiguity reduces the impact of what could otherwise be a key contribution to hazard modelling practice.

AR: The modelling framework we propose follows a logic-tree structure that can be directly implemented within PSHA engines (e.g., OpenQuake). This is due to the nature of SHERIFS, whose workflow is explicitly designed to explore epistemic uncertainties in data and transfer them to PSHA. However, the weighting of rupture sets and branches in the study remains inherently site- and region-specific because they result from consistency checks with regional benchmarks that are not transferrable (seismicity, paleoseismic data). In the revised manuscript, we will better clarify that the proposed weighting scheme is intended as an application-specific implementation within a general methodological framework.

RC2–4 The distinction between SubArea 1 and SubArea 2 is well motivated and clearly described. However, the underperformance in SubArea 2 is largely attributed to data limitations and simplified fault representations, which raises the question of whether the poorer model performance reflects tectonic behaviour or model-input inadequacy. This distinction should be more explicitly discussed to avoid misinterpretation of the results.

AR: We agree that the interpretation of the lower model performance in SubArea 2 requires careful discussion. While the observed lower seismic activity in SubArea 2 raises the question of whether tectonic behaviour may play a role, the modelling framework adopted in this study is not targeted towards unveiling tectonic behaviour and controls on seismicity. Instead, the model evaluates the statistical consistency of different assumptions with regional observations, which are largely affected by input data. The amount and quality of geological fault data available in Subarea 2 is considerably lower than Subarea 1. That is, most estimates of critical parameters like slip rates are based on rough estimates from long-term geodynamical constraints, while in Subarea 1 these are usually based on reliable fault-specific studies. The underperformance of SubArea 2 is likely highlighting poorly constrained fault parameters in this region. This is evidenced by the consistent misfit between model and seismicity rates for SubArea 2, regardless of model setup. Such issues are common in many fault-based studies and reflect heterogeneities in the state of the art of the region, a topic that is beyond the scope of our work. We will better clarify this point in the discussion of the revised manuscript to avoid misinterpretation of the results.

RC2–5 The conclusion that background seismicity has a limited impact is well supported for the chosen buffer configuration. However, the Discussion itself acknowledges that this result is sensitive to buffer size and expert judgment. As written, the conclusions regarding background seismicity risk are interpreted as more general than justified by the tested configurations.

AR: We tested additional buffer sizes (5, 10 and 15 km; respectively Fig. 1, Fig. 2 and Fig. 3) for the background analysis and the conclusions we extract are the same. A 5-km buffer from fault traces is not optimal because it does not always capture the whole surface projection area of the faults and, as such, might exclude some seismicity related to the faults at study. Conversely, 10 and 15 km buffers show the same result, being the 10-km configuration the preferred one, as it is more constrained to the faults. In terms of background ratios, the broad range of background ratios already explored suggests that such variations are unlikely to substantially affect the main conclusions. These points will be more clearly discussed in the revised manuscript (Sections 2.2.4

and 4.4), where we will further emphasize that our conclusions are valid across buffer size values and ratios tested. Also, the buffer-size test figures, together with a map showing the three buffer configurations (5 km, 10 km, and 15 km) superimposed on the fault system (displayed in Fig. 4), will be included in the Supplement.

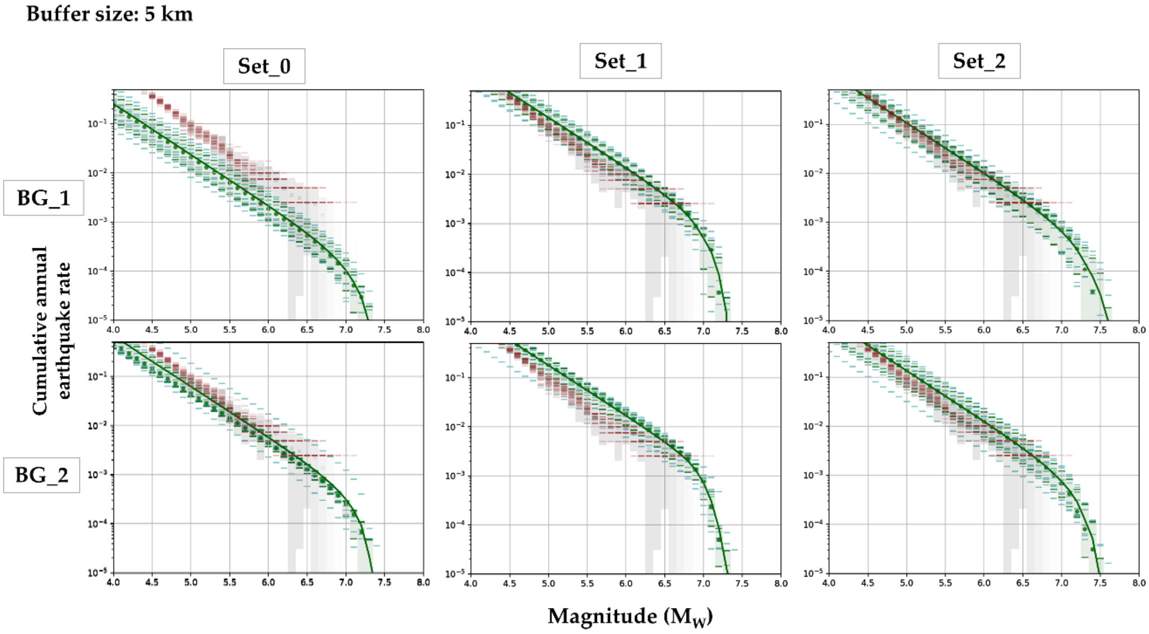

Fig. 1 Buffer size of 5 km from the modelled faults. Columns show the comparison between modelled magnitude-frequency distributions (MFDs; green) and seismicity rates derived from the historical and instrumental regional catalogues within the study area (red), for the three rupture scenarios. Rows correspond to different background seismicity ratio scenarios.

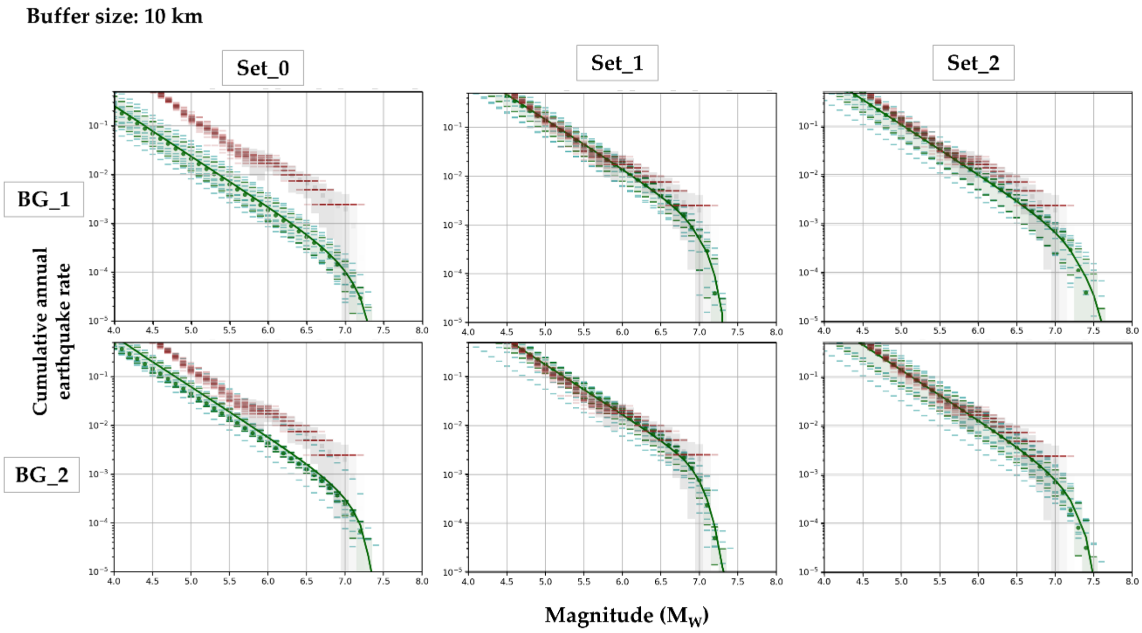

Fig. 2 Buffer size of 10 km from the modelled faults. Columns show the comparison between modelled magnitude-frequency distributions (MFDs; green) and seismicity rates derived from the historical and instrumental regional catalogues within the study area (red), for the three rupture scenarios. Rows correspond to different background seismicity ratio scenario.

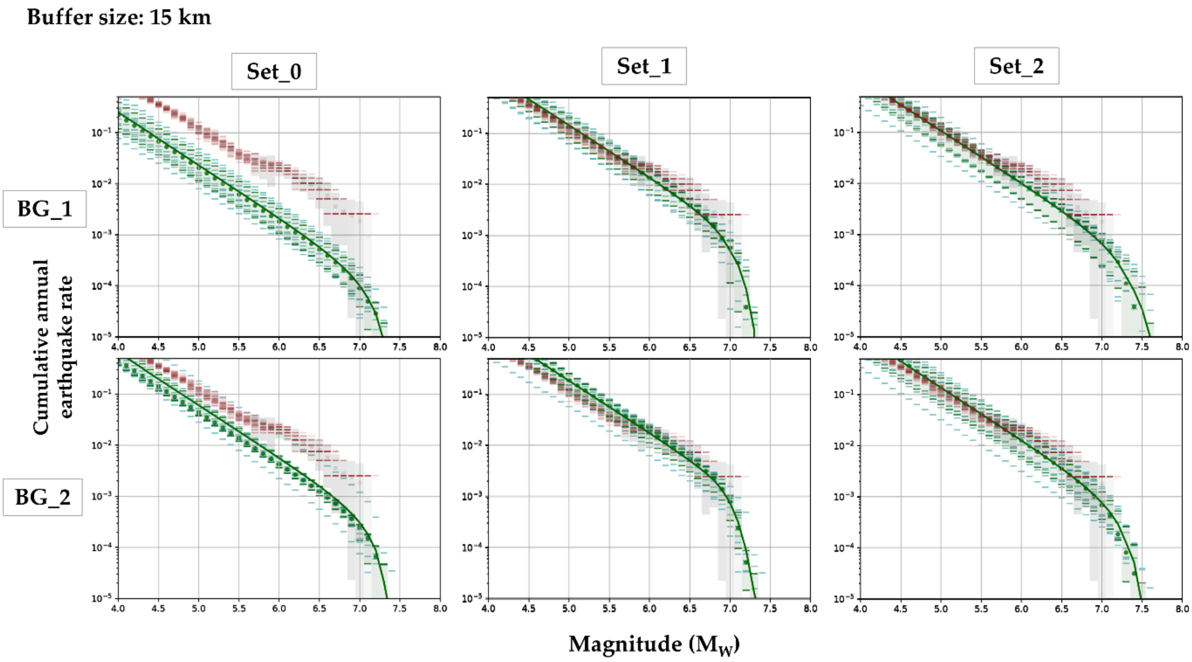

Fig. 3 Buffer size of 15 km from the modelled faults. Columns show the comparison between modelled magnitude-frequency distributions (MFDs; green) and seismicity rates derived from the historical and instrumental regional catalogues within the study area (red), for the three rupture scenarios. Rows correspond to different background seismicity ratio scenario.

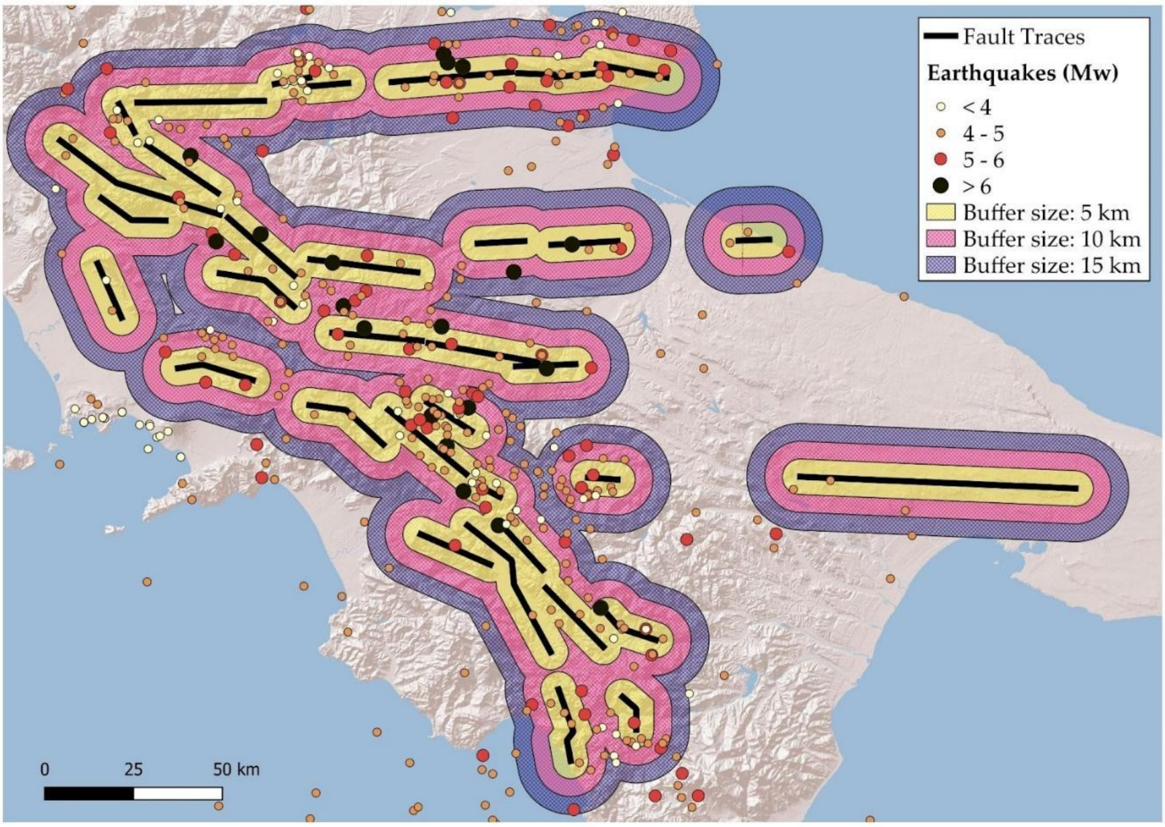

Fig. 4 Geometry of the three buffer configurations (5 km, 10 km, and 15 km) around the faults used to define the background seismicity and to extract the seismic catalogue for the consistency checks. Earthquakes from the CPTI15 v4.0 catalogue (Rovida et al., 2022) are shown for reference.