<div align="center">**Review Comments on:**</div>

<div align="center">*From rockfall source areas identification to susceptibility zonation: a*
proposed workflow tested in El Hierro (Canary Islands, Spain)</div>

General comments

The manuscript under review presents a well-conceptualized and executed study that aims to analyze how different approaches to defining source areas can influence the accuracy of rockfall modeling, using a methodological experiment conducted on the island of El Hierro (Canary Islands, Spain). Although the topic of rockfall susceptibility modeling is common in the literature, this study makes an interesting contribution by highlighting the critical importance of source area definitions. The experiments conducted strongly support the prioritization of probabilistic approaches for identifying source areas at a regional scale, and the manuscript argues effectively for the benefits of supervised classification in susceptibility mapping over unsupervised methods. The study is not only of scientific interest but also holds practical value for local managers and stakeholders. The manuscript is generally well-structured and, for the most part, easy to follow. However, the methodological section requires clearer explanations and the inclusion of some missing details, which will be addressed in the specific comments. Overall, I believe that this manuscript has the potential for publication once these comments and corrections have been addressed.

Specific comments

1. In the introduction, the authors talk several times about deterministic, statistical and probabilistic approaches. I suggest to add a few lines explaining the basic differences between these three methodologies in order to ensure that the reader understands correctly what they are trying to explain when they use such a term.

2. In section 2.1 the authors offer a good overview of the geographical and geological settings of the study area. However, there is no reference to Figure 1, where the reader can actually locate the many locations mentioned in the paragraph.

3. In section 2.2 the authors list some sources of information used to define rockfall source areas, among which there is something cited as "*some geomorphological information*". I find this phrase too ambiguous and it should be more specific. What exactly did they use?

4. In section 2.2 there is the weak point of the paper. If I have well understood, some crucial steps of the analysis are dependent on the available rockfall inventory. For instance, the ECDF model is built on data obtained within the mapped source areas; so is for the training and validation of the probabilistic model (logistic regression); and the supervised classification approach is fed by the rockfall deposition zones previously mapped. Notwithstanding, the only information provided about such an inventory is that they are "*areas affected by rockfalls where we have identified detached boulders by field investigation*". It is not clear if source areas and deposition areas are independent polygons or not. There is no extra information about the number of the mapped rockfalls and the period in which the field survey was carried out. Furthermore, later in section 3.4 the authors mention two different inventories, but there is no information about what the origin of these data is. In my opinion this is one thing to be improved in the revised version.

5. In section 3 the authors make reference to Figure 2 a couple of times. I'll leave the decision to the authors, but from my point of view it is strange to mention the main results in the methodology section.

6. In section 3.1.1 the authors argue some slope angle cut values used in the literature as a threshold, but they do not specify which is the one applied in their study. This is only clarified in section 4.1 (i.e. slope threshold = 40º). This should be clearly specified in the methodology.

7. In section 3.1.3 there are some confusing explanations. It is not clear if the probabilistic model has been done merging the three outputs of the logistic regression, discriminant analysis and quadratic analysis; or instead, the authors just selected the better performing among them. Another important information is missing: the training and validation sample proportions. For the sake of the comprehensiveness of the paper I suggest to improve this section and to provide more details.

8. In section 3.2 the authors mention the need of three coefficient maps in order to run STONE, and that the values of such "*coefficients were estimated considering different lithological/geotechnical categories reported in the geotechnical map of El Hierro and selecting values reported for similar lithologies in the literature (Alvioli et al., 2021; Guzzetti et al., 2003; Mateos et al., 2016; Sarro et al., 2020)*". Further than that, I find compulsory to specify the coefficient values applied in the study, in order to facilitate the reproducibility of the experiment.

9. In section 3.4 the authors introduce two validation tests that are not so common in landslide susceptibility evaluation tests: (i) 2D hexagonal bin count heat maps and (ii) distribution of average susceptibility values within circular buffers (i.e., violin plots). I appreciate the effort made by the authors to include innovative validation proves. However, I believe that some extra explanations in the methodology section about how one should interpret this kind of plots, together with additional references, would improve substantially the quality of the manuscript.

10. In section 4.1 I was expecting the validation results of the probabilistic approach applied to generate the PROB$_{RSA}$ map, since in section 3.1.3 the authors state that "*Specifically, contingency matrices and plots along with model sensitivity, specificity, Cohen's kappa indices and ROC curves with the corresponding area under curve (AUCROC) values, were used to compare the observed and modelled source areas and to explore quantitatively the performances of different model configurations allowing the selection of the best model and the corresponding probabilistic source area map*". In my opinion these are very relevant results that need to be shown up.

11. I strongly suggest improving the writing of Section 4.2. The argumentation was difficult to follow. Since this section discusses the core results, it is important to present it as clearly as possible. Therefore, I recommend dedicating additional effort to ensure clarity in this crucial part of the manuscript.

12. Section 5 correctly synthesizes the presented results and draws conclusions that are well supported by the evidence. However, to enhance this section, I would appreciate a more in-depth discussion on the implications of the findings. For instance, does this mean that every rockfall susceptibility analysis should utilize the PROB$_{RSA}$ approach for identifying source areas, in combination with STONE and the ECDF classification method? Additionally, while STONE, like many other rockfall simulation software mentioned by the authors, is effective, it does not account for certain relevant factors in fall trajectories, such as the initial size of the

detached boulder or other complex mechanical aspects. A brief discussion of the limitations and advantages of this tool would be valuable for readers to consider.

Technical corrections (*a compact listing of purely technical corrections*)

Page 1 – Line 17: "*A morphometric firstly approach establishes a slope angle …*" Please verify if the sentence is grammatically correct.

Page 2 – Line 48: "*Rockfalls simulation models …*" Shouldn't be Rockfall (singular)? Pease verify.

Page 2 – Line 49: "*sources areas…*" Shouldn't be source areas? Pease verify.

Page 2 – Line 59: "*dataset …*" datasets (plural)?

Page 4 – Line 94: "*The Canary Islands is a volcanic archipelago …*" is or are? Pease verify.

Page 4 – Line 119: you use both *modeling* and *modelling*. Please chose one forms and be consistent.

Page 5 – Line 130: "*(BDMoves) …*" Is it a citation? In such case, the reference is missing in the list. If not, please provide some more info about that because it is not a convention.

Page 6 – Line 158: "*For the first statistical identification …*" I don't understand why it is THE FIRST

Page 6 – Line 164: "*…denotes the CDF of a random…*" Do you mean ECDF*?*

Page 6 – Line 173 & 176: CDF$_{RSA}$ or ECDF$_{RSA}$*?*

Page 6 – Line 182: "The model uses in input morphometric …" Remove in.

Page 7 – Line 203: "…employing in input…" as input? Please verify

Page 7 – Line 203: "…the three source areas maps…" source area maps?

Page 8 – Line 230: "*The resulting map is probabilistic with values ranging from 0 to 1 and shows a probabilistic estimation…*" too much probabilistic.

Page 8 – Line 230: "*…three source areas maps…*" source area maps.

Page 8 – Line 230: "*…ECDFs graphs…*" ECDF graphs.

Page 9 – Line 277: The first two sentences are redundant with the previous paragraph. Better to remove.

Page 10 – Line 283: "*Furthermore, Table 2 shows…*" Its Table 1 I guess.

Page 10 – Line 286: "*proposed by (Rossi et al., 2020)) and classifies…*" Correct citation

Page 10 – Line 290: "*The output of run-out simulation…*" runout.

Page 10 – Line 295: "*(Figure 1 in (Rossi et al., 2020)) reveals*" (Figure 1 in Rossi et al. (2020)) revealed that the rockfall trajectories

Page 10 – Line 301: the "hard soil" class … quotation marks show different format. Revise in the complete manuscript.

Page 11 – Line 327 & 328: "*…the model with the best performance is obtained by using the PROBRSA source areas (AUCROC=0.88), followed by the CDFRSA*"

(AUCROC=0.84)…" You should add the AUROC value of ST_{RSA} to this paragraph.

Page 11 – Line 341: "…source areas of increasingly complexity…" of increasing complexity?