Referee Report

The paper "From rockfall source areas identification to susceptibility zonation: a proposed workflow tested in El Hierro (Canary Islands, Spain) presents three approaches to identifying rockfall source areas and compares results of rockfall trajectory simulation made with STONE-Software based on these three different estimates highlighting the influence of the source area definition on the modeled distribution of rockfall trajectories.

General comments:

The manuscript is currently undergoing a "major revision" process, and I sincerely acknowledge the authors' efforts to address previous reviewer comments and improve certain sections.

However, considering all these efforts, I must state that the paper still faces significant challenges in demonstrating sufficient novelty. The methods applied have already been used in the same study area, and apparently, no new datasets have been introduced to advance the state of knowledge. Additionally, the frequent references to the authors' previous work diminish the standalone value of the current study, making it heavily reliant on prior publications. Consequently, the only new contribution is the comparative analysis.

Unfortunately, the relevance and impact of this comparative analysis are limited. The presentation of results raises some questions, partly due to the inconsistent introduction of methods and the superficial description of workflow steps and data used. Moreover, the study lacks a robust uncertainty assessment, which is essential for convincingly demonstrating the advantages or broader applicability of the proposed approaches. The conclusions remain overly general and fail to provide significant novelty or actionable insights.

While portions of the paper are well-written, some statements are overly simplistic and require greater precision and clarity. Inconsistent usage of established technical terminology in the text should also be addressed to ensure coherence and improve the overall quality of the manuscript.

The authors should consistently use established terminology to strengthen the paper, provide a more detailed and transparent presentation of the methods and data, and streamline the workflow with clear and consistent definitions of its components (e.g., susceptibility analysis and runout zonation). Additionally, the discussion should include a rigorous assessment of uncertainties and present a more explicit interpretation of the results. Addressing these issues, the paper could add value as a rigorous and insightful case study.

In the following, I exemplarly describe a few identified issues. For more, please see the specific comments below.

Inconsistent usage of defined terminology

A key issue throughout the manuscript is the inconsistent definition and usage of the term rockfall susceptibility. On page 2, lines 45 and following, the authors adopt the widely accepted definition that susceptibility represents the "spatial occurrence of slope failures, given a set of geo-environmental conditions." This definition primarily pertains to the identification of rockfall source areas. However, in later parts of the study, this definition is inconsistently extended to include rockfall trajectories simulations. For instance, on page 8, line 240 and following, the manuscript states: "To derive rockfall susceptibility maps, the trajectory values can be classified using different systems…". This interpretation extends beyond the original definition, creating confusion.
Further ambiguity arises in the validation section, where the authors compare the modeled susceptibility to "observed susceptibility values," which they describe in terms of rockfall deposits. This

formulation is problematic for two reasons. First, susceptibility, by definition, cannot be directly observed or measured; it represents the potential for slope failure under specific conditions and can only be indirectly assessed. Second, rockfall deposits do not represent the areas where mass separation occurred. Instead, deposits are associated with the final resting positions of rockfall material, where susceptibility is inherently very low due to the minimal or even absent potential for further movement. This misrepresentation diverges from the initial susceptibility definition and adds to the confusion.

It is crucial to address these discrepancies to ensure clarity and alignment with established terminology. While valuable, the inclusion of runout zonation is not typically part of susceptibility assessments. A clear distinction between susceptibility analysis and runout modeling should be made, with appropriate terminology used consistently throughout the manuscript.

Incompleteness and ambiguity in the provided information

As noted earlier, in several sections, the authors direct readers to another publication (a previous study covering the same area) for detailed information. However, a scientific paper should ideally be self-contained, providing sufficient detail to be understood independently rather than functioning as an appendix to earlier publications.

In particular, details about the parameters and inputs used in the study must be explicitly presented. This includes information about the mapped rockfall source areas, as readers are left without understanding how many there are or how they are distributed across the study area.

To improve clarity and completeness, the authors should include a concise introduction to the datasets used. This could be effectively presented in a table, clearly listing the data sources and providing appropriate references if the datasets were derived from previous studies. This approach would significantly enhance the accessibility and transparency of the paper.

Methodology not transparently described

Another inconsistency lies in presenting the results for unsupervised and supervised ST_{RSA}. Neither method is appropriately introduced in the methods section, leaving the reader unclear about their implementation and distinction. Given the simplicity of the ST_{RSA} approach, it is unclear what the "supervised" component entails. Does it involve setting a threshold at the lowest slope angle where a rockfall source area is identified? This can be inferred from the brief mention on page 5, lines 130ff, where the authors state that supervised ST_{RSA} requires the location of source areas. However, instead of piecing together information from various parts of the manuscript, the methods section should clearly and comprehensively describe the applied methods. A detailed explanation of unsupervised and supervised ST_{RSA} approaches is essential for readers to understand and evaluate the study entirely. The same applies to CDF_{RSA}; however, it applies to the unsupervised part. While a definition of an empirical cumulative distribution function (ECDF) over a set of observations is clear, the unsupervised part is unclear. By definition, an ECDF calculates the proportion of observations less than or equal to a given value, requiring actual data as input. Without observations, the fundamental basis for constructing an ECDF is absent. Suppose the authors propose an "unsupervised" ECDF for rockfall source areas. In that case, they need to clarify how it is derived, what assumptions are made, and what data (if any) are used as a substitute for direct observations. Without this explanation, the method risks being both conceptually and terminologically misleading.

Deterministic or probabilistic runout simulation framework?

There is some confusion regarding the use of deterministic rockfall models, such as STONE, that involve multiple simulations for the same cell (e.g., up to 10 simulations for cells with unity probability). If the algorithm is inherently deterministic, meaning it always produces the same output for unchanged input parameters, it is unclear what additional value these multiple simulations provide.

In a deterministic framework, running multiple simulations for a single cell without varying the input parameters appears redundant, as the outputs will be identical across runs. This raises the question: what is the purpose of specifying multiple simulations per cell in this context?

A possible explanation is that the framework might not be purely deterministic if the input parameters are intentionally varied to account for uncertainties or randomness in the rockfall process. For example, there is variability in release conditions such as initial velocity, angle, or uncertainty in other inputs (e.g., coefficients of restitution or friction). If such input variability is indeed part of the model framework, then the deterministic nature of the core algorithm would still apply to individual simulations, but the overall approach would effectively become probabilistic. In this case, explaining how these variations are implemented is crucial: Which parameters are varied, and why? What ranges or distributions are used for these parameters? Without such variations, it remains unclear why multiple deterministic simulations for the same cell are necessary. The authors should clarify whether the input parameters are varied and, if so, how this variability is introduced into the modeling process.

Results

The consistency of the reported results raises some questions. The authors indicate that the simple slope threshold (ST_{RSA}) method identifies 727,603 pixels as potential rockfall sources when applying a 40° slope angle threshold. As I understand it, this encompasses all pixels with a slope angle exceeding 40°. In comparison, the ECDF (CDF_{RSA}) method identifies 1,628,048 pixels with a non-zero probability. Furthermore, on page 7, line 195, the authors state that in the CDF_{RSA} method, slope values associated with a classification of 0 (or nil probability) do not exceed 47.27°.

Based on this, one might conclude that the CDF_{RSA} method effectively applies a higher cutoff slope angle than the 40° threshold used in ST_{RSA}. However, this raises a significant inconsistency: if the CDF_{RSA} method imposes a higher slope angle threshold, the number of cells with non-zero probability should logically be lower, as steeper slope areas generally occupy smaller portions of the terrain. Why does the CDF_{RSA} method result in a significantly higher number of non-zero probability pixels? In Figure 2, it can be seen that lower slope portions are identified as source areas, which contradicts the statement above. This apparent discrepancy (due to the current presentation of the results) requires further explanation and clarification to ensure the consistency and reliability of the reported results.

The ROC curves in Figure 9 show "susceptibility maps "performance, including the runout analysis. I wonder why the authors did not perform validation for identifying the source areas based on mapped rockfall source areas but relied instead on pixel comparisons of the areas, which do not allow a quantitative insight into the capabilities of the three approaches to identify potential source areas. In my understanding, this would be a starting point for the quantitative comparison.

I could not discover any point in assessing the uncertainties in all parts of the analysis except for some buffers around points. If data-driven methods were applied in LAND-SE, I would like to know which part of the data was used and how the model was validated. Also, did the authors use all rockfall source locations for other supervised approaches to estimate the thresholds or only a part of them? The quality of the data is not clearly introduced. Are the data used representative of such an analysis? I miss any discussion or reference about the uncertainties of the field mapping inventory. Is the inventory of deposits used to evaluate the runout models representative?

Discussion and conclusions

The discussion is quite undetailed, repeating the steps of the analysis in large parts and referring to the result figures without really discussing them. The conclusions contain statements that, in my understanding, are not founded on the study's findings, as many relevant aspects that could allow conclusions of the general applicability of the proposed approaches were not appropriately reported in the manuscript.

Specific comments:

Page 2, line 61: *"Deterministic methods identify rockfall source or detachment locations using models based on mechanical principles… ".* Through the text, it seems that the deterministic methods are set in a way that is equivalent to physics-based methods. However, physics-based approaches may also be probabilistic if stochastic elements are included to represent uncertainties or process randomness.

Page 3, line 66: *"Most of the approaches… "*It is unclear which of the previously mentioned approaches are being referred to here.

Page 3, line 84: *"…advanced heuristic methods… ".* Please specify which methods and provide a few references.

Page 3, line 85: *the full sentence: "A heuristic method depends on the site characteristics and its application requires validation… "*This is not specific to heuristic methods as all methods depend to a certain degree on site characteristics and need validation and special adaptation (e.g., data engineering in machine learning). Do you want to express that heuristic methods rely on expert knowledge and rules of thumb tailored to site-specific characteristics? Consider rephrasing.

Page 3, lines 89-90: *"Hybrid methods combine statistical and experimental methods, such as neural networks or machine learning decision analysis, to reduce the amount of data required and improving the accuracy of the results ".* Akward formulation: what are experimental methods? Artificial neural networks belong to machine learning, which builds a wider category that covers shallow and deep learning, among others. They are not inherently a hybrid of statistical and "experimental"/heuristic methods. Further, machine learning can be used in decision analysis, but I have never heard the term "machine learning decision analysis "before (but maybe this is only my perspective). I suggest to rephrase the sentence.

Page 4, lines 128: *"The map was reclassified into 7 geotechnical classes (Sarro et al., 2020; Rossi et al., 2020) ".* Here the reference combines two works using different classes. Namely, Sarro et al. (2020) used 7, while Rossi et al. (2020) used only 5.

Page 5, line 129: You could extend Table 1 to describe the lithologies included in the geotechnical classes in more detail.

Page 5, lines 133-134: I do not understand the unsupervised version of the ECDF-based analysis. To my knowledge, an ECDF requires a set of observations to construct the empirical distribution.

Page 5, line 134: *"..location of source areas (i.e., normally mapped in the field; see Rossi et al., 2020 for details) ".* Why not include them in Figure 1? These data represent a crucial dataset for the first step of the analysis, yet the reader is not provided with information about their quantity or spatial distribution.

Page 5, line 135: *"…needs additional geo-environmental information (see Rossi et al., 2020 for details) "*Because the analysis is used here, specify the used data details without referring to other publications. Keep the paper self-contained.

Page 5, lines 136 ff: Consider rephrasing the paragraph. Can you establish a link between the described data types and the records mentioned below (lines 140-144)?

Page 6, lines 155-156: Above, in line 130ff, the method requires only slope and source areas for the supervised version. However, here it is stated that slope and geology are used without observational data. Could you please clarify this discrepancy?

Page 6, lines 158-159: incomplete sentence.

Page 6, lines 169-173: I wonder why such an analogy to Gran Canaria is necessary here, given that so much previous work has already been conducted on El Hierro. While drawing analogies can be helpful, they should perhaps not play such a dominant role in the decision-making process. If thresholds are defined based on fieldwork conducted directly on El Hierro, this approach would seem more robust and scientifically sound than relying heavily on analogy.

Page 6, lines 175ff: Very clear so far. This represents the 'supervised' ECDF. However, how was the 'unsupervised' ECDF performed mentioned in *line 134*?

Page 7, lines 195-197: this belongs into the result section.

Page 7, line 201: Please, check the reference Rossi et al. (2022) for consistency before mapping was referred to Rossi et al. (2020).

Page 7, line 202: *"The model uses **as** input morphometric parameters… "*Specify which.

Page 8, line 244: *"…qualitative interpretation…"*- The methods employed are quantitative; I'm curious what prevents a quantitative interpretation of the results?

Page 8/9, lines 253-255: *„The rockfall deposits mapping can be affected by uncertainty and to be reliable should be statistically representative of different geo-environmental setting**s** controlling rockfall occurrence and evolution".* So are they representative? Consider including a paragraph in the discussion section to address this point in detail.

Page 9, lines 259-261: „Susceptibility maps" involving the runout show areas affected by the rockfalls, not necessarily their occurrence. This is one of the points where the used terminology is confusing.

Page 9, line 267: *„To validate the models…"* Specify which models (runout?).

Page 9, line 270: *„…ROC plots (Rossi et al., 2010, 2022; Rossi and Reichenbach, 2016)".* ROC was not originally developed in the cited works. Use 'e.g.,' if the reference is provided as an example.

Page 9, line 271: *„…observed susceptibility value…"*. See general comments.

Page 10, line 299-300: *„The comparison of source areas identified with the three methods was performed using spatial overlay in raster format and frequency-based criteria".* This is a comparison of spatial extent, which does not provide any information about model quality but highlights only the differences between the models. For a more robust quality assessment, would it not be better to evaluate the modeled 'source area maps' against the mapped source areas, for example, using metrics such as success rate curves or ROC?

Page 10, line 302: *„No pixels were identified as source area only by STRSA being always associated either with CDF_{RSA} or $PROB_{RSA}$".* Does this imply that the source areas identified by ST-RSA are always contained within the extent predicted by the other models? This underscores the importance of verifying the quality of source area identification using metrics such as ROC. It seems unlikely that a model predicting a susceptible area three times larger would achieve a better balance between true-positive and false-positive rates. If it does, this would suggest that a significant portion of the mapped source areas is located on slopes below the 40° threshold, raising questions about the quality of the data or the validity of the expert-based analysis that defined the empirical threshold.

Page 10, line 306: *„…largest mismatch for ST_{RSA} and $PROB_{RSA}$".* I find this outcome reasonably expected and intuitive. The PROB-RSA method combines categorical parameters, such as lithology, with high-resolution slope data and multivariate, yet linear, models. Including categorical parameters inherently limits the method's ability to capture fine-scale variations, reducing its discriminatory power compared to the ST-RSA approach. The ST-RSA method relies solely on high-resolution slope input and a threshold value that is not directly tied to observational data. The latter may introduce significant uncertainties and add diffusion to the data-driven allocation of source areas. Such points could be addressed in the discussion part.

Page 11, line 335: *„…simplified geotechnical classes…"* It is difficult to follow from the perspective of this study. Are there five classes? How do the simplified geotechnical classes from the previous study relate to this one? This comparison was never introduced. If it is highly relevant, consider including the simplified classes in Table 1, as this information is too specific to rely on a cross-reference alone. Is the outcome from line 336 was not possible to reveal with the 7 used geotechnical classes of this study?

Page 11, lines 338-339: *„These percentages can be explained by the geological and morphological setting".* This is interesting; *p*lease add more details on these settings.

Page 12, line 359: Please include a few explanatory sentences for Table 3 to clarify what each column represents. This will help avoid potential misunderstandings and ensure the table is easily interpretable.

Page 12, lines 363-368: The statement in this passage appears overly optimistic and potentially misleading. Highly biased maps can still be produced when using a biased inventory, even if the differences between maps seem smaller. This is because the ECDF method does not address or validate the quality of the input data. While ECDF helps normalize or standardize the data distribution, making outputs from different datasets appear more comparable, the accuracy and reliability of the resulting maps depend heavily on the quality and representativeness of the input data. If the inventory or underlying assumptions are biased, such as through under-sampling certain areas or over-representing others, the ECDF will only propagate these biases. As a result, the maps may appear statistically consistent but remain systematically inaccurate. This highlights a significant limitation of the study, as no rigorous uncertainty assessment is provided to substantiate the claim. Addressing this gap is crucial to ensure the robustness and credibility of the conclusions.

Page 12, line 374: *„Identical AUC_{ROC} values are obtained for unsupervised and supervised ECDFs, when the same source area identification method is used".* At this point, it is unclear. Does not this contradict the statement that supervised classification is superior?

Page 12, line 375: *„ROC analysis is sensitive to methodological choices and…"* It is unclear in what sense „sensitive" is being used here. Furthermore, it was stated in a few lines above that ROC analysis can be applied regardless of the classification approach adopted.

Page 12, 387ff: *„Where additional data, such as geological or geomorphological information, are accessible, investing time in the mapping of source areas enables the application of probabilistic methods that yield more robust results".* I'm curious why the geomorphological information, if available, is also not utilized in the heuristic approach (this could be a more complex approach than the simple threshold method). Incorporating this data could be an intermediate step between a simplistic slope threshold and a more quantitative statistical analysis, which typically requires sufficient observational data.