Answer Reviewer 1

General comments

This is a well-structured manuscript that provides a clear methodology to address some of the uncertainties in rockfall susceptibility map production. Overall, I found it easy to read and follow, and it provides a valuable contribution to the more general issue of producing hazard maps at larger (regional) scales. I have a few questions, primarily around the technical implementation of some of the methods and how to interpret some of the later figures, however, none of these are major issues and I look forward to seeing the published revised version.

Thank you very much for the positive feedback and for the observations and suggestions that will improve the comprehension and readability of the manuscript. We have revised the text considering all your comments and suggestions. Below, we present a point-by-point response to address all your suggestions.

We hope that the revised version of the manuscript will be compliant with the quality standards of the journal. We sincerely appreciate the revision and the opportunity to provide a new version of the article.

Specific comments

1. Locations of past rockfall events vs locations of future rockfall events

In this work there is the inherent assumption that past rockfall areas not only inform future rockfall areas but are also still "on-the-list". After a rockfall has occurred, should the probability of a future event at the same location decrease? At some point you presumably will run out of rocks to fall?

In the introduction, we have added the following text to explain better the meaning of susceptibility.

The landslide susceptibility measures the degree to which a terrain can be affected by future slope movements. In other words, it is an estimate of "where" landslides are likely to occur. In mathematical language, can be defined as the probability of spatial occurrence of slope failures, given a set of geo-environmental conditions and a set of past landslide locations (Reichenbach et al., 2018).

Reference: Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-science reviews, 180, 60-91.

2. The term "probability" without any conditional information

It is slightly misleading to have probability == 1 for a whole set of pixels without some sort of dependency. A probability of 1 is a certain event. These maps (e.g., figure 4) – are essentially saying there are huge areas where a rockfall will definitely pass through. Does this mean that the next event to occur will definitely pass through all of these places? Or that all of these places will eventually have a rockfall event? (if so, what time frame). I presume the values are just to simplify calculations – maybe some context or an example statement of what p == 1 at a specific pixel means might help?

The value == 1 could be an implication of the numerical procedure (i.e., ECDF are generated using talus in the supervision with the value of 1 assigned to the highest trajectory count; in the channel upwards, those value can be higher and then ECDF assign a value 1) and a simplification due to the used colour scale.

Pixels classified as 1, doesn't mean that the next event will definitely pass through them but these places will eventually have a rockfall event in a future that is not defined (not included in the definition of susceptibility). By assigning a probability value of 1, we aim to convey that these specific locations present the highest likelihood of being affected by rockfalls. Regarding the time frame, this analysis exclusively considered susceptibility; determining the timing of rockfall events would require further analysis of triggering factors such as rainfall or seismicity, or alternatively, conducting more detailed studies. We acknowledge this comment, as it presents an important point for future research to enhance the rockfall assessment on El Hierro.

In response to the reviewer's suggestion, we have added the following clarification:

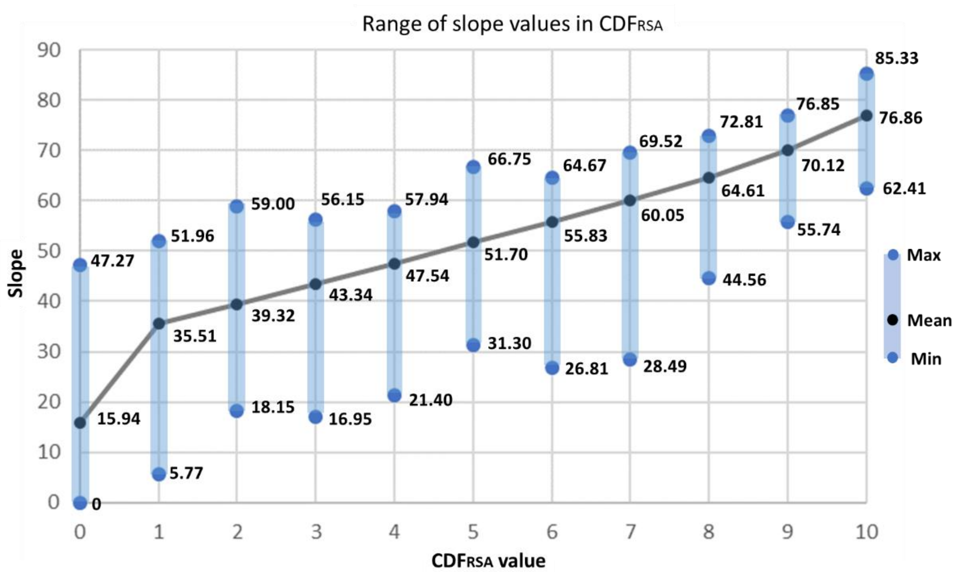
"The resulting map presents values ranging from 0 to 1, representing a probabilistic estimate of the likelihood of each pixel being affected by a rockfall event. Consequently, pixels equal to 1 indicate areas where the susceptibility zonation evaluate a higher probability of rockfall occurrence."

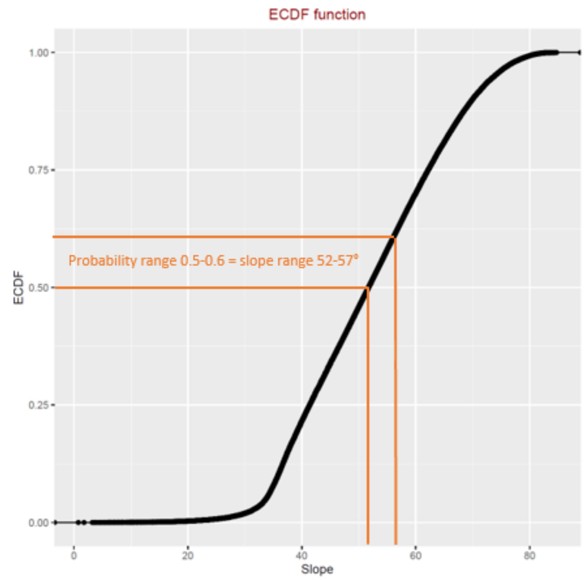
3. Slopes of past rockfall events informing CDFRSA

If I understood this method correctly, any area with a slope LESS than the minimum observed slope of an existing rockfall source area has a probability of 0, and similarly, any place with a slope GREATER than the maximum observed slope has a probability of 1. It seems unlikely that the most extreme events have already been observed (especially given the small areas of the island where the rockfall data come from). It would be more realistic (and fair to the method) to have assumed observed values were (e.g.) 5 and 95 % values of the true slope distribution (actual % should be dependent on the number of data you're using to build the ecdf). Some idea of the range of these slope values would be beneficial in the text too (e.g., are we looking at 42 – 42.5 degree slopes or 30 to 57 degrees).

Many thanks for your feedback. To clarify, we have added the ECDF graphs derived for the source area identification, which illustrates the distribution of slope angle values and the corresponding the probabilities. In the figure we show graphically the slope angle range associated to the probability ranges 0.5-0.6. Due to space constraints and the already high number of tables and figures in the manuscript, the authors have decided not to include it; however, we will include a text indicating the range of values.

"The slope values corresponding to a classification of 1 range from 62° to 82°, with a mean slope of 77°. In contrast, the slope values associated with a classification of 0 do not exceed 47.27°, exhibiting a mean slope of 16°."

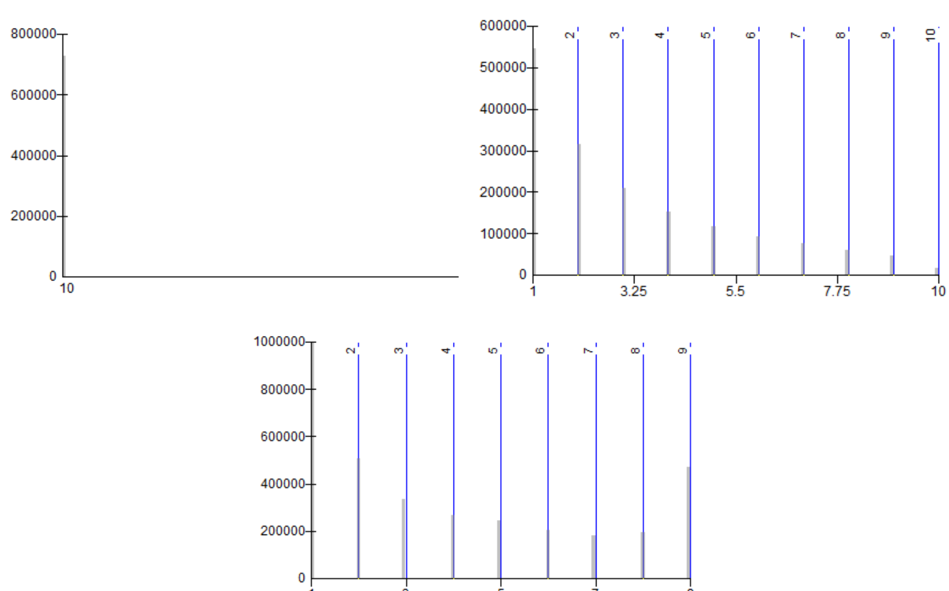




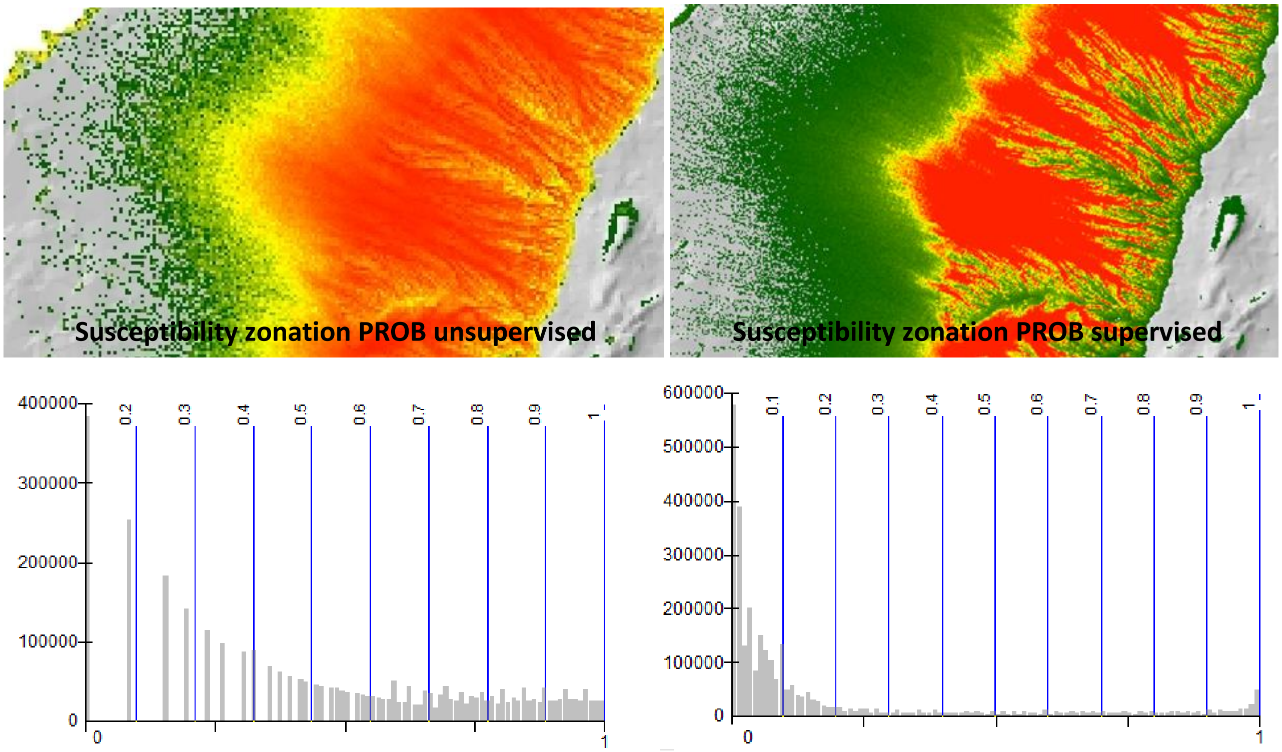
4. How many rocks == a rockfall? And an idea (histogram?) of how many pixels these rockfalls are hitting on their way down (e.g., are most simulated trajectories only passing through 1 (start) or 2 pixels or is it closer to 100?).

We appreciate the reviewer's comment. The information requested by the reviewer would be relevant and interesting, both form a hazard evaluation perspective and to deal with the issue of rockfall boulder fragmentation. However, the rockfall runout simulation software we have used, which mainly focuses on regional rather than hillslope scale analyses, does not produce in output such information. Histograms of each source area are included below.

In the case of ST_{RSA}, since the assigned values are binary (0 for non-source areas and 1 for source areas), we multiplied the values by 10, resulting in 10 simulations launched from each identified source area. For CDF_{RSA} and $PROB_{RSA}$, where values range from 0 to 1, representing the probabilistic likelihood of a pixel functioning as a source area, the number of simulations was scaled between 0 and 10 based on the probability. This approach appropriately weights the probability of rockfall occurrence in the susceptibility zonation.



Due to the concurrent simulation launches from all source areas, it is challenging to discretize the rockfall trajectories originating from individual pixels. Nonetheless, we will consider conducting a similar analysis to that presented in Table 1, either as a new table or integrated into a figure.



5. Data resolution

I might have missed it but I couldn't see where the information was on pixel size(s), whether the resolution of data for the different maps were identical to start or if interpolation was necessary, whether there was any concern or consideration of the method sensitivity to raster size (e.g., is it 10 x 10 m or 200m x 200m) and how do these compare to rockfall source areas?

Thank you for your comment. For the susceptibility analysis in the island of El Hierro, a 5 x 5 m pixel size was employed, providing sufficient sensitivity for the results. In the following sections, we differentiate between the layers required to identify the source areas based on the different approaches and the layers used in the simulations. It has been indicated in the text.

6. The external validation data

Lines 242 to 244 suggest that external data (not used in the rest of the analyses) were used to "validate the models". This is a great idea – but we are not told enough about this external data, the number, the extent, whether they are completely separate from the data used in the rest of the manuscript or are essentially just a subset. I also didn't completely understand the buffer idea – how do these values (5, 50, 100m) align with pixel size?

To explain better the rational of buffer selection we added the following text at the end of 3.3.

Different buffer sizes allow to consider uncertainty due to local conditions and boulders locations. In the proposed approach the location of mapped boulders is used to evaluate the rockfall susceptibility zonation.

Commonly this information is mainly used to evaluate runout models verifying if simulations reach entirely or partially the boulder locations.

7. Supervised is better

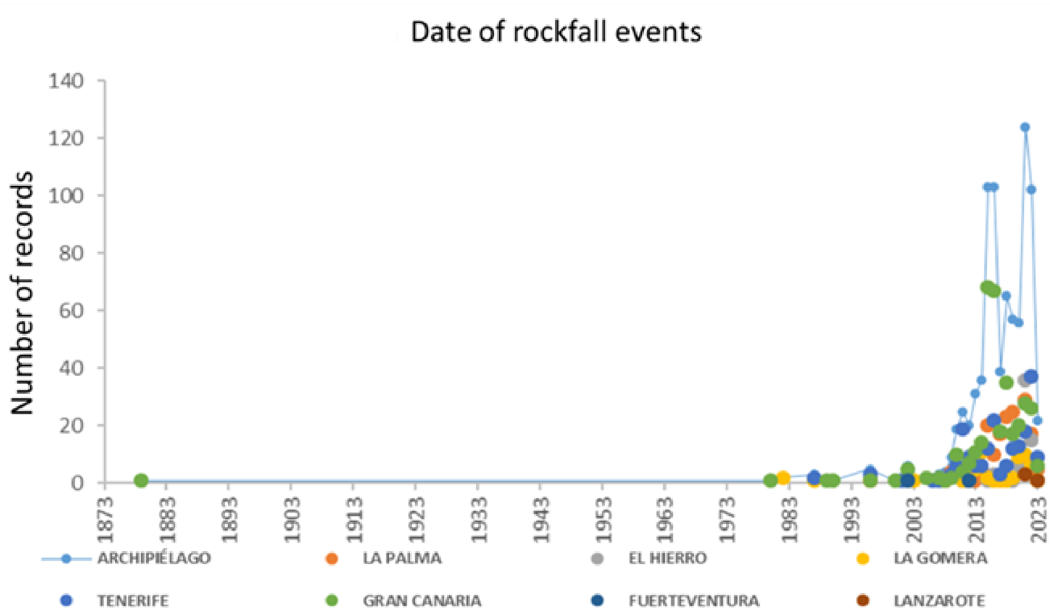
Given the finding that the supervised classification system is better – some discussion around the original rockfall data used is necessary – how biased is the original dataset by (e.g.) proximity to populations? Size of rockfall? Age of rockfall? And was any sensitivity analysis done via subsetting this data during the classification process?

Many thanks for your comment. It is fine to add additional description of data, but we believe we can just add text. It will be also important to stress the possible impact of selecting biased or inaccurate data on modelling of both source areas and runout simulation. To underline the importance of the expert-based work to create unbiased and representative datasets we have added the following at the end of second paragraph of section 3.3. Here you need to rewrite the answer including the mentioning of adding the text below to the manuscript.

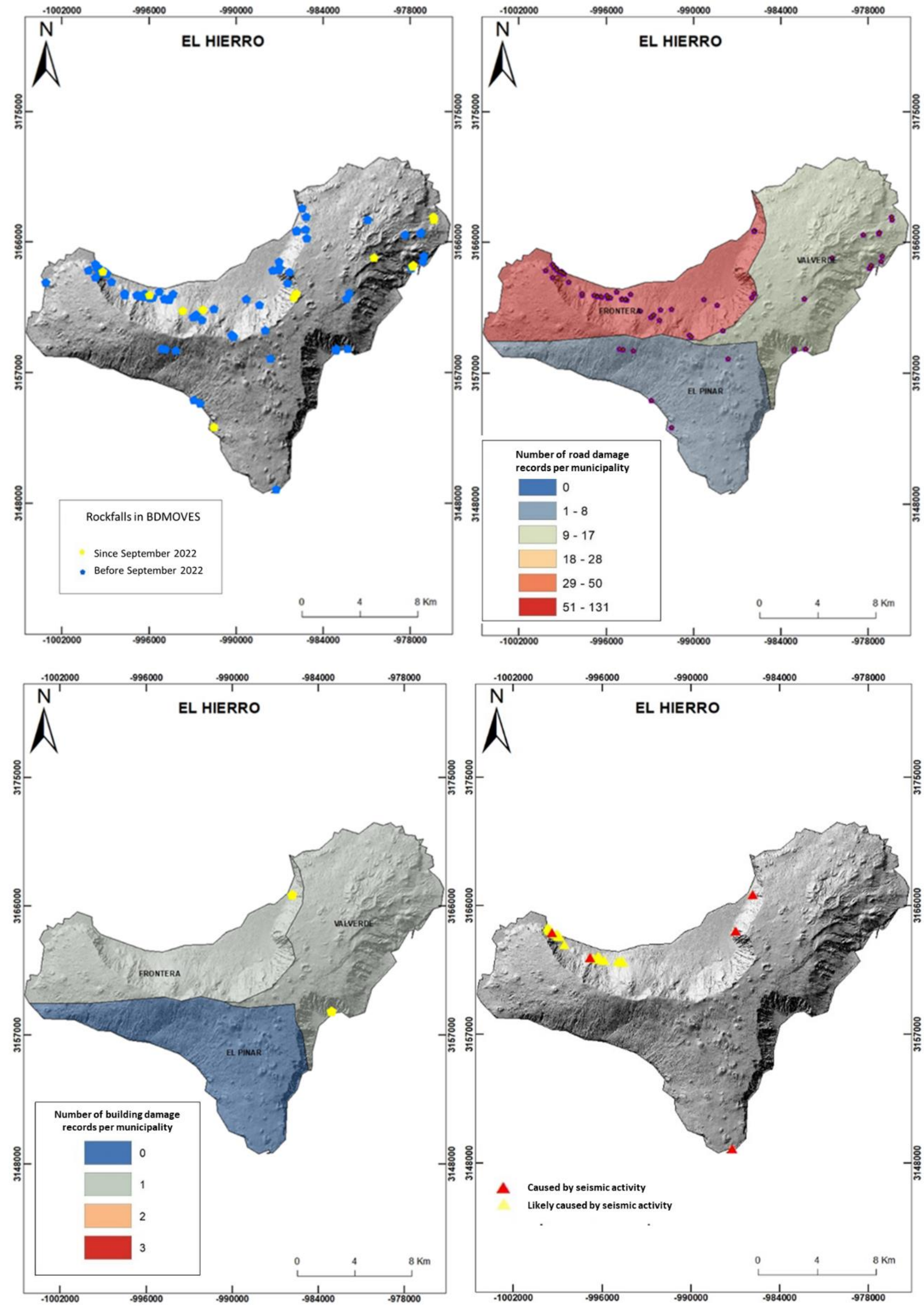
Despite the expert-based rockfall deposits mapping is affected by uncertainty, it is relevant to provide for the analysis a representative map accounting for different geo-environmental setting which control rockfall occurrence and evolution.

For the number of rockfalls used to validate this work, we have the following data: *field investigations conducted from 2012 to 2018 (47 records), aerial images (84 records), and the MOVES database (BDMoves) (78 records).* Some of the specific details requested by the reviewer are outlined below. To address some of the questions raised, they are presented as graphs or maps. Since the analysis of the inventory is not the main objective of the paper, it will not be included in the manuscript.

The following graph shows the number of rockfall records in the Canary Islands from 1879 to the present (2023). They are represented by colored circles corresponding to each island. As can be seen, the available records are mainly concentrated in the last 10 years, with a particular emphasis on the most recent years. This is due to the fact that work on MOVES database began during this period, when the IGME-CSIC, together with IRPI-CNR, collaborated on several international projects, and a citizen observatory was established in the Canary Islands to collect rockfall data through an app.



The following maps show the rockfall inventory in MOVES for the island of El Hierro. In the first map, you can see the events recorded in the database before September 2022 (blue) and since September 2022 (yellow).



Additionally, to address the impact on infrastructure, a map of road damage by municipality is shown. The purple dots indicate the locations of road network damage. Finally, a map of building damage by municipality is presented. The yellow dots indicate the locations of damage to buildings. This information is conditioned by the available data in the database, as in many cases (especially smaller events below 1 m³) there is no information on the damage caused. As can be seen, the main impact is on the road network. Regarding fatalities due to rockfalls, there are records of only two people who unfortunately lost their lives in Pozo de las Calcosas in the mid-20th century.

Regarding the triggering factors, they are mainly due to rainfall; however, in El Hierro, as mentioned in the article, they are also conditioned by the occurrence of rockfalls due to seismic activity. An example is the event presented in the paper that occurred in the Roquillos tunnel during the 2011 seismic-volcanic crisis.

In the text, we have added the following information regarding the number of events included in the rockfall inventory.

"To classify and validate the simulated rockfall runout, we have prepared a map (Figure 1) that shows two types of areas: (1) areas affected by rockfalls (red polygons), where we have identified detached boulders through various field investigations conducted from 2012 to 2018 (47 records), aerial images (84 records), and the MOVES database (BDMoves) (78 records). We have point features that we converted into polygons by applying a 50-meter buffer to account for uncertainty in data collection; and (2) areas with no evidence of rockfall activity (green polygons), determined by a heuristic analysis based on field observations and geomorphological and topographical maps. These mainly include flat areas where no evidence of rockfalls has been observed, identified using Geomorphon and topographical data, and verified during the field campaign."

Technical corrections

Line 37 - There are several places where rockfalls should be rockfall, this is one of them, I have not noted them throughout.

Many thanks. We have revised the text and made the corresponding changes.

Line 49 – Sources → source

Done.

Line 86 – EL → El

Done.

Line 130 – evidences → evidence

Done.

Line 188 – I think dependent should be independent here? Unless you're doing some subsetting of the dependent variable?

Thanks for your comment but dependent is correct (Rossi 2020)

Lines 94 – 95 [and repeated on lines 354 – 357] – this looks like it was a valuable exercise – but the results aren't included in the paper?

We appreciate the reviewer's comment, however, we would be grateful if the reviewer could kindly provide further clarification on this point, as we would be pleased to address it.

Lines 203 & 208 – physically → physics

Done

Line 247 – regardless the adopted → regardless of the adopted

Done

Lines 256 – 269 – most of this text is just Table 1 in words and could be removed.

We appreciate the reviewer's comment and suggestion. However, we would prefer to keep this part of the paper as it is to maintain continuity with the rest of the article and to clarify the comparison of the results.

Line 280 – 66,80 → 66.80

Done

Lines 345-347 – this seems completely out of place – suggest removing or adding some context so it is obvious why it is included.

Many thanks for your comment. We have added this information.

To derive rockfall susceptibility maps, the trajectories values can be classified using different systems, including Equal Interval, Natural Break, Quantile, Standard Deviation, Head/Tail Breaks and Landslide Percentage (Alqadhi et al., 2022; Baeza et al., 2016; Cantarino et al., 2019; Tehrani et al., 2022; Wang et al., 2016), in order to make a qualitative interpretation of the results.

Line 350 – helps reducing…. → helps to reduce differences and homogenise zonation

Done

Table 2 – I just can't wrap my head around this table, but I am assuming this is just me!

Thanks for comment. We have prepared the following table to make the information clearer.

COMPARISON OF RSA MAPS		TOTAL (RSA1 ∪ RSA2)		INTERSECTION (RSA1 ∩ RSA2)		ONLY RSA-1		ONLY RSA-2	
RSA-1	RSA-2	Pixels (#)	Area (Km^2)	Pixels (#)	Area (Km^2)	Pixels (#)	Area (Km^2)	Pixels (#)	Area (Km^2)
ST	CDF	1628115	40.70	727536	18.19	67	0.0017	900512	22.51
ST	PROB	3399705	84.99	727490	18.19	19	0.005	2672196	66.80
CDF	PROB	3482657	87.06	1543701	38.59	82971	2.07	1855985	46.40

Figure 2 – Would be good to add the number of red pixels for each of the source area maps (LHS ones) or add some text in the caption (e.g.) "see Table 1 for pixel count".

Done. We have added the text proposed in the capitation.

Figure 3 – I don't quite understand the numbers in the legend compared to the maps – in the map the visible pixels are mainly green and dark blue, but according to the legend, green should be one of the least observed and there should be more peach/pink colours??

Many thanks for your comment. Figure 3 will be reviewed and modified taking into account the reviewer's comment.

Figure 5 – because of the differences in x-axis scale it is virtually impossible to compare these ecdfs. Would it be possible to add the unsupervised line in a different colour to the supervised ones as well so we can see what's going on?

Many thanks for your comment. We prefer to keep the figure the way it is, since the insets layouts is the same for previous and successive figures.

Figure 6 – I thought my printer had broken. Also, it's not "probability of rockfall trajectories" because you can't have -1 probability, it is the difference in probabilities of rockfall trajectories.

Yes, the reviewer is correct. The map represents the difference in susceptibility calculated using different source areas, which explains the presence of a negative value, indicating a higher probability for the second of the two values. We have changed the legend of the figure.

Figure 7 – where do the zeros get binned in these histograms? Would probably be better (given that 0 is a "match") to have 0 as the mid-point of the central bin.

Many thanks for your comment. Zero values are few and are included in the left bin. We believe that this way to bin the data is more efficient to show better the positive or negative differences between maps. Indeed, we have also used the same in Figure 6 to better describe this feature.

Figure 8 – Needs a better explanation when talked about in the text, and what is the total number of "measurements" number of total pixels across the island? Or only those that were included in at least one of the models as $p > 0$? Or other?

Many thanks for your comment. The total number of pixels corresponds only to those included in at least one of the models with a probability greater than 0 ($p > 0$). We have added the following text to clarify the figure:

The 2D hexagonal bin count heat maps (Figure 8), derived for the different pairs of susceptibility maps, confirm these results showing a better alignment along the bisector of the higher frequency counts (i.e., dark reddish hexagons) obtained for supervised susceptibility maps (Figure 4 d, e, f). These plots are divided into hexagonal bins, and each bin is colored based on the count of susceptibility maps value. Dark reddish shades indicate a higher frequency of measurements within the corresponding hexagon, while lighter areas may indicate sparse values.

Figure 9 – This one also needs some extra information as I don't understand it – how did you measure "true positives"? and as the "empirical ROC" is consistent across rows, and all the points appear to be on the line – how is model fit calculated? And if the numbers are probability threshold reference values – should these be in the opposite order and vaguely align with the numbers of the y-axis to estimate fit? Also, they're all black lines as far as I can tell.

Regarding the way the True Positives are calculated, by definition a ROC curve (e.g., Fawcets, 2006) is derived comparing observed/mapped areas affected (i.e., pixels within polygons mapped as "without any significant evidence of potential boulders reaches" taking a value of 0 in the analysis) or not affected by rockfall (i.e., pixels within polygons mapped as "reached by rockfall boulders" taking a value of 1 in the analysis) and probabilistic estimates of rockfall susceptibility (pixels of the modelled susceptibility zonation ranging from 0 to 1). Each point along the ROC curve is derived selecting a probability threshold (between 0 and 1) used to classify the probabilistic prediction (i.e., the values below the threshold take a 0 value, those above a value of 1), then the comparison of the classified probability and the observed values allows to derive a confusion matrix having one value of TN, TP, FN and FP and to derive TPR (True Positive Rate) and FPR (False Positive Rate) values which identify a single point along the ROC curve. To obtain the entire curve, multiple threshold values need to be used, for each of those different values of TN, TP, FN and FP and TPR and FPR are calculated and constitute the point which define the ROC. The above explain why the order of the probability values in the ROC are correct.

Regarding the same values of ROC area obtained for the unsupervised and supervised ECDF for given source area identification methods (i.e., row wise in Figure 9), this can be explained because the two probabilistic transformations (using different ECDFs) are merely a reclassification of the same original information, which the trajectories' count obtained by the models. Overall, the order of the data does not change, and performance of the unsupervised and supervised models must not change. Different is when considering the final values after the reclassification which show clearly the effectiveness of using the supervised ECDF, with this being described and demonstrated in the manuscript. We prefer not to add this detailed explanation, in the text to make the reading smoother. We slightly modify the text of the first paragraph in the section 4.3 as follows.

The graphs show that the model with the best performance is obtained by using the $PROB_{RSA}$ *source areas* (AUC_{ROC}*=0.88), followed by the* CDF_{RSA} (AUC_{ROC}*=0.84), with* ST_{RSA} *performing the worst* (AUC_{ROC}*=0.78).*

References

Fawcett, Tom (2006). "An Introduction to ROC Analysis" (PDF). Pattern Recognition Letters. ***27*** *(8): 861–874.*