Response to Referee #1

Palacio Cordoba, J., Mergili, M., and Aristizábal, E.: Probabilistic landslide susceptibility analysis in tropical mountainous terrain using the physically based r.slope.stability model, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-223, in review, 2019.

We are grateful for the constructive comments on our paper. we are pleased to know that our article is well structured and of interest to the scientific community. We have included the suggestions in the revised manuscript. Please find below our responses in detail. Changes in the manuscript are highlighted in blue letter.

- I find a figure with the SHALSTAB and SHIA_Landslide models to be lacking.
In Figure 6, we show the results obtained with the models SHALSTAB and SHIA_Landslide. The SHALSTAB map was modified to 3 classes. considering a single class for "unstable" and "unconditionally unstable". for the construction of the maps is considered the range found according to the distance to the perfect classification. (L303 - 306).

- The statistical indexes of the table 4 are not defined in the "Data and procedure" section.
The statistical indexes of Table 4 are now defined. (L269-273).

- ROC curves for SHALSTAB and SHIA_Landslide models are missing. This validation technique is the most robust and, in my opinion, the most appropriate to elucidate which method has the best performance.
The ROC curve was added for the SHIA Landslide model as suggested. For the SHALSTAB model the ROC curve could not be realized as this model does not yield a factor of safety or any continuous value for each cell, like the other models evaluated. In SHALSTAB, each cell is classified as unconditionally stable, unconditionally unstable, or potentially unstable – a concept which is, unfortunately, incompatible with the concept of the ROC curve.

-The elaboration of ROC curves of each model needs a more detailed explanation. I am not sure how the A, B, C and D curves presented in figure 5 have been calculated.
The explanation was added and modified in the "Data and procedure" section. (L263 - 275)

- I am surprised that authors do not discuss about some of the most relevant publications about the comparison of stability analysis methods implemented in GIS environments. I encourage authors to read and discuss the results obtained by Cervi et al. (2010) and Zizioli et al. (2013) among others.
We find the discussion by Cervi et al. (2010) on the different models very interesting, especially the discussion on SHALSTAB and considerations of the model (L376-377):
"Related to this aspect, Cervi et al., (2010) expose the limitations of SHALTAB due to the hydrological assumptions (flow in steady-state conditions)."

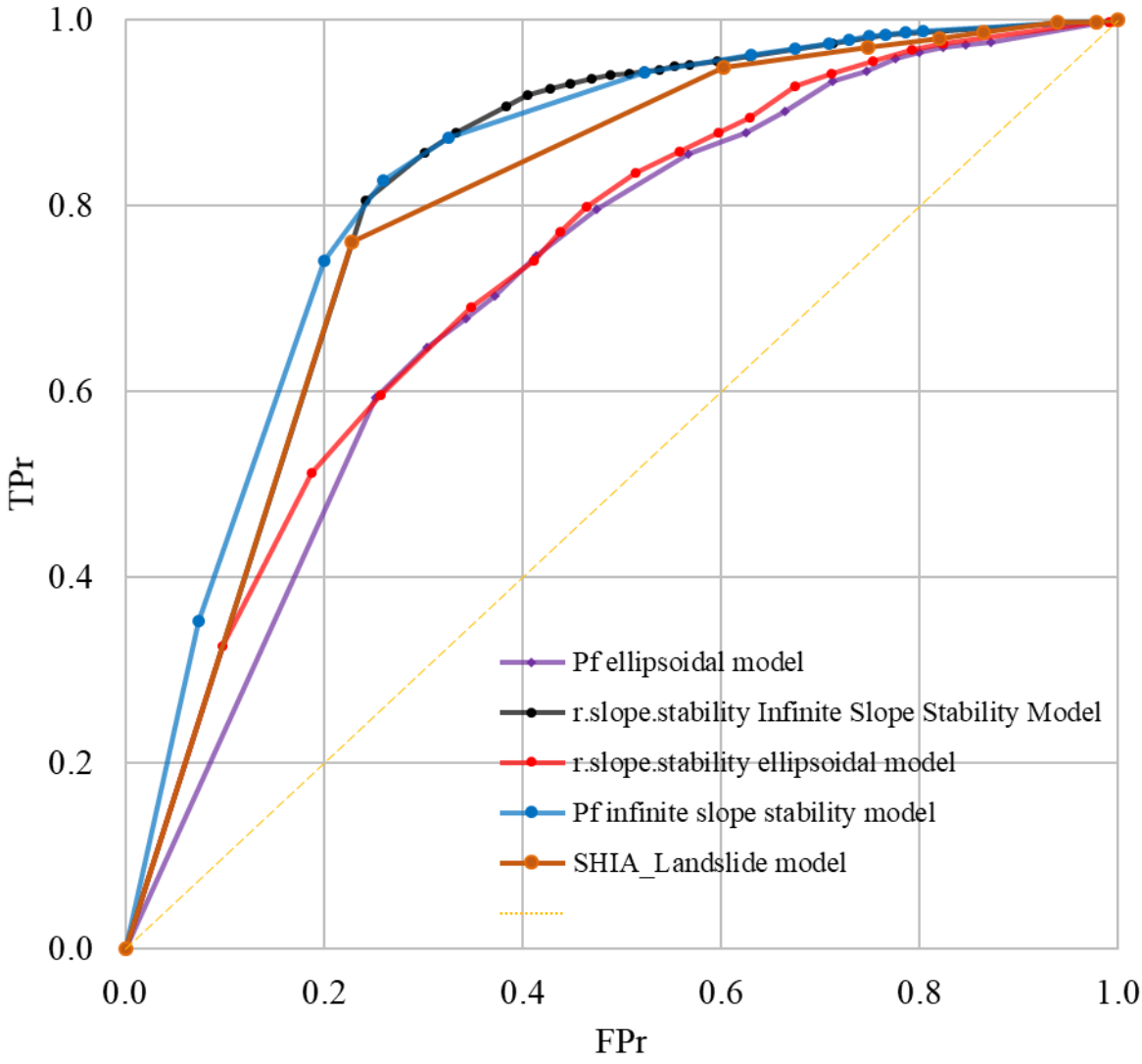
Also the paper of *Zizioli et al. (2013)* was very helpful (L105-106):
"In terms of scale or spatial resolution, physically-based models are suggested to be applied to finer-scale study areas and their results are strongly influenced by the level of detail in the input data (Zizioli et al., 2013)"

OTHER REMARKS:
-Line 46 delete "because of its equatorial location", it is a repetition.
Line 198. Change "Source: Adapted from (Aristizábal et al., 2016; Qiu et al., 2007)" by "Adapted from Aristizábal et al., (2016) and Qiu et al., (2007)".
-Line 316. Change "071" by "0.71".

Thank You, these errors were corrected!

-Figure 5. I think that the figure is confused. Please change the style of the figure to understand it at a glance.

-Change A, B, C and D codes by the name of the models. Use similar colors according to the type of the models.
We have updated Fig. 5 accordingly, also including the ROC curve for the result obtained with the SHIA_Landslide model.



Response to Referee #2

Palacio Cordoba, J., Mergili, M., and Aristizábal, E.: Probabilistic landslide susceptibility analysis in tropical mountainous terrain using the physically based r.slope.stability model, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-223, in review, 2019

We are grateful for the constructive comments on our paper and have included the suggestions in the revised manuscript. Please find below our responses in detail. Changes in the manuscript are highlighted in blue letter.

- in the world? in South America?
South America – corrected.

- I don't understand the relation between the colombian andes and the world. I think the sentence has to be revise because for the reader the relation is not evident. How many landslides or proportion of landslides are triggered by rainfall in Colombian Andes ? I think the percentage mentionned by Petley has to be put above in the text.
The paragraph was rewritten for clarification (L37-40):
"Petley (2008) mentioned that in 2007, 89.6% of the fatalities due to landslides worldwide were related to rainfall-triggered landslides. In tropical environments and complex terrain such as the Colombian Andes, a high percentage of landslides are triggered by heavy or prolonged rainfall (Van Westen and Terlien, 1996; Terlien, 1998)."

- I suggest: Therefore there is a strong societal and economic request to include landslide susceptibility etc.....
Thank You - Corrected according to the suggestion. (L53-54):

- It will be interesting to have information related to the landslide in South America and to know the "position" of Columbia in term of number of landslide, number of fatalities etc.... in this paragraph
There are not many studies with a specific rank, however considering the recommendation of the reviewer additional to the rank of Mergili et al (2015) a new rank according to Petley and Sepulveda (2015) was incorporated (L60-65):
"*... where Colombia is the country with the second highest number of fatal landslides in Latin America and Caribbean, only Brazil shows a slightly higher number of fatal landslides (119) (Sepúlveda and Petley, 2015). Mergili et al. (2015) expose that Colombia, according to EM-DAT, presents a victim/event ratio of 77.34 (3171/41) and is only exceeded by Peru and Ecuador in this respect. However, the EM-DAT database only considers events with high numbers of fatalities: actually, the real number of fatal landslides is much higher (Aristizábal and Gómez, 2007; Mergili et al., 2015).*"

- You have to improve this part because since SAFELAND project, cascini 2008 Corominas et al 2014 and van Westen et al 2007 it is possible to divide the method by qualitative and quantitative method. I think you have to be in the recent scientific paradigm to describe the different method and approach for LSA and LHA. Data driven are separate from Physically in the group Quantitatve methods (see Corominas et al 2014) and the different works of the JTC 1.
The LSA classification was adjusted according to Corominas et al. (2014) (L66-69):
"*... qualitative or by quantitative methods (Corominas et al., 2014). Qualitative methods correspond to knowledge-driven approaches based entirely on the judgment of experts using geomorphological criteria in the field (van Westen et al., 2000). Quantitative approaches are subdivided into data-driven methods and physically-based models. Statistical or data-driven methods ...*"

- it is very nice to cite ourself but Mergili never introduce the concept of FoS, I prefer you put Duncan for instance. I think the text og Mergili et al 2014 a is cited afetr in the text.
The Mergili et al. (2014) and Ahmed et al. (2012) citations were eliminated, we think that the reference to Lam and Fredlund (1993) is enough.

- idem of the comments above.
The Mergili citation was eliminated.

- Mergili is not the first to mention this problem. I think there is in the litterature some key paper mentionning this difference, I suggest to change that.
Thanks for the observation, more references have been included. However the authors think it is difficult to establishe who was the first. The references only try to provide some books and papers where this issue is discussed in detail (L80):
"*... (Bishop, 1954; Carson & Kirkby, 1972; Crozier, 1986; Duncan & Wright, 2005) ...*"

- i think the transition between the two paragraph is not very well structured. the first sentence will be put above in the text. fater you can speak about Columbia.
Thanks for the recommendation, the paragraph was adjusted. (L105-112)

- I have a problem, this a scientific paper and i don't see the scientfic problematic. The introduction is well written, the problematic misses. Where are the scientific questions? I think you have to include a litte paragraph between to explain the problematic and why you want to use R slope stability and what do you expect with it.
Yes, this is an important point which we have not explained in sufficient clarity in the original manuscript. We have extended the explanation accordingly (L105-120):
"*In terms of scale or spatial resolution, physically-based models are suggested to be applied to finer-scale study areas, whereas data-driven approaches are recommended for broader-scale landslide susceptibility analyses (van Westen et al., 2006; Corominas et al., 2014). In fact, the implementation of data-driven methods or physically-based models for the incorporation of landslide hazard mapping into land use planning has been regulated in several countries. In Colombia, according to the Decree 1807/2014, the implementation of deterministic or probabilistic physically-based methods is obligatory in urban and urban expansion areas, whereas statistical and knowledge-driven models are only permitted for rural areas.*
In order to meet this requirement, we have to obtain a more detailed knowledge on the suitability of the available physically-based landslide susceptibility models, with the final goal to use model ensembles in order to obtain a broader, more robust picture of the landslide susceptibility conditions. We think that r.slope.stability could be a candidate to enrich such an ensemble, considering shallow planar and deep-seated ellipsoidal failure surfaces. Consequently, the scientific aims of the present study are (i) to evaluate the suitability of r.slope.stability for physically-based landslide susceptibility mapping in tropical mountainous terrain, and (ii) to identify its fit with other potentially suitable models, helping to learn about strengths, limitations, and uncertainties.
Following these aims, ..."

- Maybe this the point for a problematic, why you want to compare results with another results obtaines with another tools ? I think this one of the key point of your paper and yu have to develop it in the introduction
See the response to the previous comment, particularly (L113-116):
"*In order to meet this requirement, we have to obtain a more detailed knowledge on the suitability of the available physically-based landslide susceptibility models, with the final goal to use model*

ensembles in order to obtain a broader, more robust picture of the landslide susceptibility conditions. We think that r.slope.stability could be a candidate to enrich such an ensemble, …"
and (L118-119):
"*… (ii) to identify its fit with other potentially suitable models, helping to learn about strengths, limitations, and uncertainties.*"

- is it possible to differenciate the type of landslides in the figure ?
This is a good point, such a differentiation would be useful, indeed. However, it is unfortunately not possible, as the landslide database did not provide this information: it is just described in a general way that rotational and translational landslides were found (Hermelin et al., 1992; Garcia, 1995). (L159-160):
"However, there is no distribution map which permits the differentiation between translational and rotational landslides"

- you speak about rotational and translational landslide above, have you got this typology ?for the area?
See above comment.

- i think a desripton of instabilities is a prerequest for this type of study. The reader want to know which type of phenomena wll be modelled and the number etc etc.... I suggest to add a paragraph about labdslides type, number, a a bried description of them with the main predisposing factors
Yes, we agree. The description provided in the first paragraph of this section was rather short in the initial manuscript, so we have complemented it according to the suggestion of the reviewer (L168-174):
"*A total of 699 landslides were identified and mapped in the La Arenosa catchment. According to Velásquez and Mejía (1991), most of the landslides started as shallow translational slides and transformed into debris flows, from very to extremely rapid with high water content (Fig. 1). The landslide bodies were small with respect to the flow length, and the slip surfaces were parallel to the slope surface. Field studies carried out by The majority of the landslides initiated within residual soils in hollows and open slopes with a slope steepness ranging from 35° to 42° (Velásquez and Mejía, 1991). The same authors described that the depth of failure surface was less than 3 m and corresponded to the contact of the residual soil with the saprolite.*"

- the sentence has to rewritten. for instance "The tool is able to take into account planar failure with an infinite slope stability module and another type of failure with a slip surface module . "
Thanks for this recommendation, the formulation was adjusted accordingly. (L184-185):
"The tool is able to take into account planar failure with an infinite slope stability module and a slip surface module"

- I suggest to inverse the two paragraph about the modules
1. planar module
2. slope surface module
in order to be consistent with the sentence above and the order
Thanks for the recommendation, the order was adjusted accordingly.

- I suggest to split the paragraph of the too; in two part
1. description of the oS modules with slope stability model and infinite slope stability model
2. probabilistic model
the paragraph will be clearer
Thank You also for this recommendation, we have split the paragraph accordingly

- I understand the concept and what you mean, but the sentence is not very clear. I suggest to rewritte it.
Yes, we have rewritten this sentence as follows (L213-214):
"*The slip surface model includes a probabilistic component where the dimensions of the ellipsoids are randomly varied for the computation of FoS.*"

-please explain why you use this approach. Justify it please. Is there a strong relationship between slope and thickness of surficial formations and weateherd layers? If this is the case, you can refer to the paper written by Thiery et al., 2019 in engineering geology, in addition to Aritzabal 2013, the authors refer to the relation between slop degerres and different weathered horizons in cristalline rocks.
This is an important point. We had only provided the Aristizábal (2013) reference, because in that work they established the relationship between the slope and the soil weathering profile according to field work in this catchment. However according to the suggestion of the reviewer the references to Thiery et al. (2015) was included as well as a reference to Catani et al. (2010).

- for comparison of what? of another results with another tool ? if this is the case please explain or better put it in the introduction as it mentionned in my antecedent comments
The sentence was rewritten for clarification (L247-248):
"*The r.slope.stability model is applied with the probabilistic approach and deterministic approach for comparison with other models tested in the catchment.*"

- in R slope stability you have different shape of probability fnction. I think it will be good to explain that in the paragraph dedicated to the model.
Yes – Section 3 (L213-214):
"Thereby, rectangular, normal, log-normal, or exponential probability density functions are employed."

- which method ?
The paragraph was partly rewritten. The maximum width and length of the ellipsoid is determined according to what is seen in the field according to the inventory. The number of ellipsoids to be tested (1000) was the one where the results did not vary too much when the number was further increased and where data processing does not require a high-performance computer.

The text in the revised manuscript read as follows (L255-259):
"*... 1000 simulated surfaces touching each cell. Preliminary tests have indicated that this value represents a good compromise between computational efficiency and accuracy of the results. Each ellipsoidal slip surface is defined by the coordinates of the centre and the variable lengths of the three half-axes, the aspect, the inclination and the offset centre above the terrain, ...*"

- why you use this values? they correspond to the size of landslides (min, max?) If you put a pargraph about the slope instabilities and description of them, these values could be justified. i encourage you to write this paragraph as suggested before
The assumed dimensions (and also the depths) of the simulated landslides have been defined according to the landslide inventory (we have added a more detailed description of the landslides to Section 2 – see above comment).

- justify the value. The max depth of trabslational landslide is 2.8 m ? see the comment just before about size and depth justification.
See response to the comment above.

-evaluation or validation? I think you validate the models with the inventory. The evaluation will be made with the different previous maps and the expert judgement. Please revise this pargraph.
We are aware that there is no general agreement on the terminology with regard to "*evaluation*" and "*validation*". We think that, as the reference information (landslide inventory) is also uncertain, the term "*validation*" would be to strong, and rather prefer to keep "*evaluation*".

- put the descriptions in a table. These indices are very well known by the scientific community working on landslides with different spatial susceptibility and hazard approaches
We agree that those are very well-known indices. However, we still think that particularly for this reason, an additional table is not justified. So, we rather prefer to keep the description of these parameters as it is.

- "to evaluate the performance of the tool". Moreover for me the problem of this type of procedure (just statistical comparison) is that the judgement of the expert is neglected. results of this such models have to be kept as 'indicators' . You are the expert of this area, normally you are able (and this is the duty of the scientific) to have a criticism about results and the different location of high, or another susceptibility classes. For instance sometimes the models give good ROC curve and good landslide recognition, but some areas are not taken into account or inverselly you can obtain huge areas with very high susceptibility but in reality (real world) it is impossible (because you don't take into account some paraleters) and with your data you oversimplify the real world.
At the beginning of the paragraph, we have added *"... to evaluate the consistency between the model results.*" We cannot evaluate model performance by the comparison with the results of other models, because they might be as true or wrong as the r.slope.stability results.

We fully agree that it is necessary to somehow combine expert opinion and simulation. The present paper focuses very much on simulation-based approaches of slope stability analysis. Evaluation of the model builds on a landslide inventory, which has been mapped by experts through subjective judgement, based on the available evidence. Consequently, there is – in our opinion – a strong expert-based component in the evaluation of the model results, even though the evaluation itself is a quantitative one. However, expert judgements are also necessary for converting the model results into hazard classes and associated recommendations. This is made clear in the final part of the revised conclusions (L451 -462; see also response to the last comment).

- what is the goal of this comparison/evaluation of r slope stability results? what do you want to explain with this comparison? You have to add some explanations about the goal, the objective of this comparison
In general, using different models for one and the same area and landslide event increases the robustness of the results, as each model is based on different assumptions and different ways of implementation of the equations. Similar results yielded by different models indicate some degree of plausibility, whereas different results help to identify possible shortcomings, limitations, or even bugs. (L113-116):
"*In order to meet this requirement, we have to obtain a more detailed knowledge on the suitability of the available physically-based landslide susceptibility models, with the final goal to use model ensembles in order to obtain a broader, more robust picture of the landslide susceptibility conditions. We think that r.slope.stability could be a candidate to enrich such an ensemble, …*"
and (L118-119):
"*… (ii) to identify its fit with other potentially suitable models, helping to learn about strengths, limitations, and uncertainties.*"

- which type of landslide you try to model with infinite slope model ? all ?

The inventory available to us does not allow for the distinction between shallow translational and deep-seated rotational landslides, but we know that both shallow and deep-seated landslides occurred in the 1990 event (see comment above). Therefore, we have used both models in parallel, in order to cover both types of landslides, expecting that the "truth" is somewhere in between the two results. A major advantage of r.slope.stability is that it includes both the infinite slope stability model and the sliding surface model. A possible future strategy could be to combine both models in one single probabilistic simulation.

- you have two type of landslide? you compare tyoe by type as suggested by Thiery et al., 2007 , Corminas et al., 2014 ? This is important because each landslide type has its own prediposing factors diferent from one to another one. I am dubitative, are the mudslide has the same prediposing factors as the soil slip ? The reader is not able to know because no paragraph about landslides is included in the paper.
We have tried to clarify this issue in Section 2 by describing the landslides in more detail (see comment above).

- why you used this thresholds for the FoS classes ? please justify beacuse in some studies authors divide it in 4 classes (0-1/1-1.2/1.2-1.5/>1.5)
Yes, we are aware about this issue – to our knowledge, there is no generally accepted "standard" classification. We just consider the chosen classification useful for our purpose of comparison, without claiming its general validity.

- I repeat but why calculate two maps ? please justify in the text if you have two type of landslides (translational shallow and deeper landslides with rotational or another slip surface)
Please see above comment.

- and in an expert point of view..... the different probability can be divide following the different thesholds defining by the JTC 1 works, they give somme classes or by old paper written by Kirsten HA. (1983) Significance of the probability of failure in slope engineering. Civ Eng 25:17–27
The paper of Kirsten (1983) was checked in detailed and the text was modified, however the thresholds were not modified considering that they were used just to compare between different models, they are not proposed for further use. According to Kirsten (1983), different thresholds should be used for specific scenarios and according to the local conditions. The following sentence was added (L92-94):
"Although FoS represents a quantitative – and seemingly objective – approach to evaluate slope stability, it has to be used carefully. Authors such a Kirsten (1983) mention that different values of FoS could be obtained from slopes with equal probability of failure. However, several software packages use this concept …"

- i don't understand (once again) you compare all landslide type with each final map. in this sense you compare maybe shallow landslides with map simulated for deeper landslides. This is not correct.
This issue is related to the missing distinction of shallow and deep-seated landslides in the inventory. However, we know that both shallow and deep-seated landslides occurred in the 1990 event (see comment above). Therefore, we have evaluated both models in parallel with the entire inventory, in order to cover both types of landslides, expecting that the "truth" is somewhere in between the two simulation and evaluation results. A major advantage of r.slope.stability is that it includes both the infinite slope stability model and the sliding surface model. A possible future strategy could be to combine both models in one single probabilistic simulation.

- ok and what is the signification really ? it is not very clear for me
The following explanation was added (L362 –369):

"*However, both deterministic analyses of r.slope.stability show the highest false alarm rates. The major reasons for this result – and for the similar patterns yielded with SHIA_Landslide – are most likely an underestimate of the geotechnical stability of the material, the neglect of the effects of vegetation, and/or the overestimate of water saturation. Geotechnical testing is often performed on material which is disturbed in the one or the other way, so that the resulting parameters do not necessarily represent the natural conditions over larger areas. Roots could lead to some degree of stabilization, and it could also be the vegetation retaining sufficient water to avoid full saturation of the soil throughout the catchment.".*

- Never you mention that before an dthe fact that you try to replicate this event !!! Please revise the taxt before to explain this point !!!!
We have mentioned this aspect in the last paragraph of the introduction (L121-123):
"*This model is evaluated using a landslide inventory prepared after a major and destructive rainfall-triggered multi-landslide event in the La Arenosa catchment on 21 September 1990.*"

- yes but this is not the only tool which can used probabailistic analysis TRIGRS now made the same type of calculation, ALICE by the BRGM can made the same calculation. I think you must mention them in your discussion? some references are available on the web.
Yes, this is a good point. We have added the following sentence (L383-385):
"*Besides r.slope.stability, there are several physically-based models which provide a probabilistic module, such as TRIGRS (Baum et al., 2008) and ALICE (Thiery et al., 2017). Much more work has to be done in this direction.*"

- or a ground water level because inmany cases this is not the rainfall introduced in the model but the considered GWL see and mention Fressard et al 2016 please, they explain that
Yes, this is an important aspect. We have replaced "*rainfall event*" by "*hydro-meteorological event*" and added the following text to the description of the study area: (L135-138):
"*Most of the landslides in the catchment are triggered by such events, even though it is often not the rainfall itself, but the response of the groundwater level – also depending on its initial state due to antecedent rainfall conditions – which is decisive for the occurrence of landslides (Fressard et al., 2016).*"

- slope conditions of what ? I think you want to write the slope instabilities or slope stability conditions
Thanks also for this suggestion, we have changed the text to "*slope stability conditions*".

- and the type of failure, but you never mention that. please you have to improve your text to mention this point !!!
It was included in Section 2 (please see response to above comment).

- Why ? maybe during this event only translationnal and shallow landslides occured ?
Thank You for this (and also the next two) comments. This issue was not clearly explained in the original manuscript, so that we have now tried to clarify it (L405-408):
"*… which is confirmed by the reference information: Even though Hermelin et al. (1992) and Garcia (1995) mentioned that both shallow and deep-seated landslides were observed after the studied rainfall event, Velásquez and Mejía (1991) underlined the dominance of shallow landslides.*"

- because you have only translationnal landslide ?!!???
Please see response to the above comment.

- ok but in your area is it only this type of landslides ? if yes why you calculate another type of slope failure ?

Please see response to the above comment.

- YES !!! but in your case ?
We have tried to explain this aspect in a clearer way in the revised manuscript (L415-419):
"*… or – if both mechanisms are relevant – evaluate the model results with different sub-datasets of the inventory. In the present study, due to the lacking distinction of landslide mechanisms in the inventory at the level of individual landslides, we applied both the infinite slope stability model and the slip surface model in order to cover a broad range of mechanisms. The result of the evaluation against the entire landslide inventory was in line with the generally reported dominance of shallow translational landslides.*"

- For more readibility, it will be bettre to put the same number of classes for each map (3 or 4)
The map derived with Shalstab was reduced to to 3 classes, considering a single class for "unstable" and "unconditionally unstable" (Fig. 6D).

- I am desapointed because you don't speak about the real validity of the final map and if the different areas simulated in different susceptibility classes correspond to slope considered by expert as susceptible or not. For you if the landslides are well recognized by the model(s) the maps are ok. This not the case every where. The different maps could be used for future developments (you explain that in your introduction) but now which map you could used to produce regulatory map ? it will be a very interesting discussion and one pargraph about that will improve the paper and give your point of view about that. You can find one interesting exemple Thiery et al. (2017) in GEOSCIENCES about the used of this type of simulation by PBM for future developements. I encourage you to consult the paper.
Thanks for the recommendation – indeed, this is a very important aspect. We see a major role of expert opinion in the conversion of slope failure probabilities to hazard classes and recommendations for action, something which is a highly non-trivial task. We conclude the revised manuscript with the following piece of text, also mentioning another important aspect (L451 -462):

"In Colombia, hazard mapping is mandatory for use land planning in urban areas. Results like the slope failure probability (Fig. 6F and Fig. 6G) are considered suitable for this purpose. However, they only represent one step on a long way:

(i) Much of the landslide risk in the Arenosa catchment is not so much related to the failure of unstable slopes, but rather to the downslope propagation of the mobilized material as debris flows (Velásquez and Mejía, 1991). Therefore, coupling of the slope stability model results with mass flow simulation tools is absolutely necessary (Mergili et al., 2017; Bout et al., 2018).

(ii) Slope failure probabilities – or impact indicator indices (Mergili et al., 2017) – neither take into account the dimension of time nor the fine-scale patterns of the geotechnical characteristics of the catchment. Consequently, those results have to be interpreted in a relative, qualitative rather than an absolute, quantitative sense. The combination of these maps with expert judgement is therefore essential to define suitable thresholds separating the study area into different zones with their individual hazard levels and recommendations for action."