Reply to reviewers

On behalf of all co-authors I would like to thank the referees for the extensive revision of our manuscript. In the following, you could find the point-by-point responses to every individual comment, in which you will find your comments in regular style and our responses in italic.

In general, as could be seen in detail within the *track-changes* file, the most of the suggestions have been adopted and the manuscript have substantial improvements in the abstract, the introduction and some parts of the materials and methods sections, including two appendices.

All the figures have been checked with the colorblindness simulator (https://www.color-blindness.com/coblis-color-blindness-simulator/) and slight changes in colors and dot shapes have been done to ensure that readers with color vision deficiencies can also interpret our findings.

Reply to Reviewer #1

Dear authors,

Thanks for your contribution on a hot topic regarding hydrology in mountainous and more generally sloping areas, which can indeed be of interest in many perspectives (risk management, agriculture/forestry, water supply).

<u>Summary of the preprint</u>

The goal of this study is to determine, amongst the parameters (rainfall, groundwater water head, and two soil water content) of an already calibrated 1D physical model, which are the main drivers (in the form of the most important antecedent conditions) of the soil water content. Indeed, this variable has been identified as critical in the occurence of landslides.

To do so, the field data is enriched with synthetic data produced by the 1D physical model, fed by stochastically-generated rainfall events. This augmented data pool is analysed through a machine learning method, combining Random Forest to assess the variable importance of each antecedent parameter conditions with regard to the soil water content output, and K-means clustering to classify the results.

Provided that the output is normalized (ratio between the soil water content and the rainfall event depth), the analysis shows a quite balanced importance of the antecedents (with a slight predominance of the soil water content condition). The triplet {rainfall, water content of the first meter of soil, groundwater head} is the most predictive of the soil water content evolution.

Finally, the classification reveals seasonal behaviors consistent with previous studies and the field observations.

Thank you for so clearly summarizing the contents of our manuscript, which make us believe that, overall, the aim of the paper was clear enough and supported by the results.

General comments

- The article is well written, at the exception of some phrasing and vocabulary issues. Especially, I reckon that "soil cover" is misused as it is a synonym of "land use" and refers to vegetation or anything that covers the soil. In the article, it is used as a synonym of "soil" or more precisely "topsoil". Please correct this, as the land cover is not a variable nor a parameter in your model. Also, "slope response" is a bit spurious phrasing as what you're investigating is the soil water content response in a sloping context, and not that of the slope itself. I would suggest to use soil instead of slope and maybe rethink the title accordingly to avoid confusion (e.g. Understanding hydrologic controls of sloping soils response to precipitation...).

Thank you for your positive evaluation of our work. Indeed, the term "soil cover" is sometimes used in landslide community, to define the mantle of soil (or regolith), that often covers a more compact and stable bedrock formation. We understand that the use of such term is misleading in the broader context of hillslope hydrology, so we will modify it to "soil mantle" throughout the text. We also accept your suggestion to modify the title to "Understanding hydrologic controls of sloping soil response to precipitation through Machine Learning analysis applied to synthetic data".

- The structure is good and most of the sections (or sub-sections) are clear (the description of the site and the field surveys, the clustering analysis etc.).

Thank you again for your positive evaluation of the structure of the manuscript.

- Nevertheless, the aim of the study should be more precisely explained at the beginning. Is it oriented toward a seasonal analysis (what it seems, regarding the Results section and some mentions beforehand) or a crisis-warning model (which is also mentioned both in the Introduction and Conclusion, and would be more consistent with the time resolution of the data)? In my opinion, it can't do both. Depending on the objective, the design choices for the synthetic data (time resolution, separation criterion) would be quite different.

Indeed, the main aim of the study is to understand how the seasonal slope conditions, related to climate forcing, may affect the capability of the soil of retaining rainwater infiltration for a time long enough to potentially determine critical conditions as a consequence of rainfall events (e.g., the triggering of landslides). The time resolution of the data, as well as the criterion adopted to separate events within the rainfall record, are indeed tailored to this aim. In the revised manuscript, we will make it clearer in the final part of the Introduction.

- A figure (a workflow for instance) could be of help to summarize the method.

Thank you for suggesting it. We will consider adding a flowchart in the revised manuscript.

- Some information on the hardware used and the computation time would be welcome, especially by comparing it with another approach (sensitivity analysis).

In the revised manuscript, we will add some considerations about the computational effort. However, we do not think it is worth also adding a sensitivity analysis, which is out of the scope of our study, for the following reasons.

First, we analyzed the dataset mimicking what could be done if, rather than synthetically generated data, one was handling real field monitoring data. In fact, we were mostly looking for a way to identify the major cause-effect relationships between (measurable) inputs and outputs before (possibly, but not necessarily) building a model for the interpretation of such relationships, rather than evaluating the sensitivity of an (already available) model output to variations in the input (although the Random Forest analysis also allows quantifying the information content of each considered input variable).

Second, the sensitivity analysis is usually carried out to evaluate the effects of input (and parameter) uncertainty on model predictions. In this study, the model chain (already calibrated and validated previously: Greco et al., 2013; Comegna et al., 2016; Greco et al., 2018) is used as a tool to generate a (richer) synthetic dataset (this is a common problem in landslide studies, as field monitoring data records, even when they are relatively long, usually contain very few data representative of potentially critical situations). The model is assumed to represent "the reality", and adding a sensitivity analysis may result misleading, as it would move the focus to the performance of the model (which, in general, could also not exist).

Third, the adopted Random Forest analysis, which allows highlighting the most informative combination of measurable variables to predict the output, is somehow a sensitivity analysis as well, as it gives some indications about the relative importance of the input variables on the possibility of predicting the output, without introducing any mathematical model structure, but simply relying on the application of logical operators (IF-THEN-ELSE) between the variables.

In the revised manuscript, we will add paragraphs in the Introduction and in the Materials and Methods sections, to better explain the choice of Random Forest instead of a sensitivity analysis.

Specific comments

- The method is based on a physical model for simulate the variable relationships on one hand (calibration), and produce the synthetic data on another hand (datapool augmentation). Could you not reach the same goal (or at least compare your results) with a sensitivity analysis of the 1D model with respect to the initial conditions (as you're focusing on the antecedents). In that regard, some supplementary elements supporting the choice of the method would be welcome.

As already mentioned in our reply to one of the general comments, we are studying the relationships between the data as if they were measurements collected in the field, so without the recourse to any mathematical model (in our case, the model was just a tool to enrich the available dataset, so to make it significant for statistical analyses). In the revised manuscript, we will add *paragraphs in the Introduction and in the Materials and Methods sections, to better explain the choice of Random Forest instead of a sensitivity analysis.*

- Concerning the synthetic rainfall generation, I assume that the separation criterion has a huge impact on the outcome of your methodology, especially if you aim at identifying the dominant parameter during extreme events (when the risk is the highest). Is there not a contradiction in chosing a 24h separation criterion (i.e. the time for the topsoil to drain almost completely) when the goal of the study is to assess the importance of the prior state of the soil? Could you elaborate a bit on the choice of this separation criterion (is it only based on the physics of the phenomenon, or is it constrained by the number of events you can bring into the machine-learning and still be computationaly-reasonable)?

The separation of events within the continuous rainfall record aims at linking the occurrence (or non-occurrence) of critical conditions to a rainfall event, so that they can be considered as a direct consequence of that rainfall event. This is commonly made when empirical predictive tools (e.g., rainfall thresholds: Segoni et al., 2018a, b; Guzzetti et al., 2020; Piciullo et al., 2020) are implemented as part of early warning systems, e.g., against rainfall-induced landslides or debris flows, and the definition of the separation criterion is usually made empirically, looking at the performance of the predictor with different choices of the separation criterion.

From a physical viewpoint, especially when one is interested in the separation between the role of antecedent conditions, i.e., related to previous precipitation (and drainage/evapotranspiration) history, from the direct effects of the last precipitation event, it is quite complex to define a suitable separation criterion, especially if dealing with slow processes activated by precipitations, such as the infiltration through the unsaturated soil layer. In fact, to completely separate what depends on "previous" precipitation from what is linked to the last rainfall event, one should wait for the infiltration process initiated by previous precipitations to be finished, and, in a soil layer of few meters thickness, it may take several days. Extending so much the dry time interval between two separate events, especially during rainy seasons, would imply the aggregation of several events in a single one, thus leading to long rainy periods, rather than events, thus preventing the desired separation of antecedent conditions from direct effects of events. So, as we have defined the "response" of the soil layer as its attitude to retain infiltrated rainwater after the end of a rain event, looking at the moisture of the topsoil layer seemed a good trade-off: topsoil moisture controls the infiltration at the ground surface, hence when gravitational drainage from the topsoil is already over (the field capacity has been reached), the infiltration of a new rainfall input through the ground surface would not depend (or it would only little depend) on the remnants of the infiltration process caused by previous precipitation. In this respect, we tested a separation dry interval of 24 hours, commonly used when the available rainfall data are at daily resolution (Berti et al., 2012; Leonarduzzi et al., 2017; Peres et al., 2018), and anyway in line with the empirical choices that are commonly made in the early warning community (Segoni et al., 2018a).

As we mentioned in the paper, the choice of 24 hours for the separation dry interval leads to about 50 rainfall events per year (i.e., 53061 rainfall events in 1000 years). The adopted Machine

Learning techniques for the analysis of the dataset (K-means clustering and Random Forest algorithm) can handle larger datasets, thus the adoption of a shorter separation time interval, which would lead to a larger number of separated rainfall events, could be feasible from the computational effort point of view. However, we chose 24 hours for the previously explained reasons.

- Did you try to assess the sensitivity of your method to the value of this separation criterion?

No. We did not test different separation criteria.

- On a related matter, by producing very diverse events in terms of duration and intensity, are you not risking to blur (to average) the relative importance of each parameter that might differ depending of the type of event? Said otherwise, if the goal is to find the sensitivity of your model to the prior conditions when an extreme event occurs, why not choose only extreme events for the analysis? Or for seasonality, why not splitting the rainfall chronicles beforehand and therefore acquire more specific relative importance of the antecedents?

As already mentioned in our reply to previous comments, the goal of the study, which was not clearly described in the Introduction, is not to find the sensitivity of a model, but to find the most important cause-effect relationships between data, which could be a useful information to build a model. However, although extreme rainfall events are more likely leading to critical conditions in terms of increase of water storage in the soil, considering antecedent conditions may help not only to explain why extreme rainfall events sometimes do not lead to critical conditions, but also why sometimes ordinary (or not so extreme) rainfall events do cause critical conditions. About a-priori considering seasonality, our dataset clearly shows that, owing to climate variability, seasons are often anticipated or delayed, and the idea is that monitoring suitable variables may allow recognizing the actual establishment of "seasonal" conditions.

- I understood that the previous studies were focusing on analysing and predicting the seasonal changes. Now, in this study, is this time-scale still relevant? I thought that the hour-time resolution was aiming at refining a model and a monitoring network more crisis-oriented (informing on the risk of a landslide to occur for example).

As already mentioned in our reply to one of the previous comments, the main aim of the study is to understand how the seasonal slope conditions, related to climate forcing, may affect the capability of the soil of retaining rainwater infiltration for a time long enough to potentially determine critical conditions as a consequence of rainfall events (e.g., triggering of landslides). So, data at hourly resolution are required for the assessment of hazard in real time, while the assessment of antecedent conditions requires a longer timescale, and the relevant data might be also acquired at a coarser resolution (indeed, both soil moisture and groundwater level dynamics are much slower than rainfall). Hence, it could be possible to adopt different time resolutions for rainfall data (e.g., *hourly) and for hydrological data (groundwater and soil moisture could be acquired at daily resolution). However, if one manages a monitoring network capable of hourly resolution, then the same dataset can be used for both short timescale predictions (hazard assessment) and long timescale processes (infiltration/drainage/evapotranspiration affecting antecedent conditions).*

- Line 54: please, explicit what you mean by "long timescales".

We will modify the sentence, by writing "timescales of weeks or even months, much longer than the duration of rainfall events, typically ranging between some hours and few days".

- Line 98: Would not it be "at the contact between soil and bedrock" as soil cover is the description of what covers the surface of the soil? (Same remark for all "soil cover" occurrences)

The sentence will be modified to "Recent studies show that the response of the soil mantle to precipitation is affected by the wetness of the interface with the underlying bedrock, which controls the leakage of water from the soil to the fractured limestone".

- Lines 98-104: That refers to epikarst. You should maybe cite supplementary studies and modelling approaches outside your workgroup (Perrin et aL, 2003; Hartmann et al., 2014; Dal Soglio et al., 2020 for instance).

Agreed. We will specify that the uppermost weathered part of the bedrock is indeed the epikarst, and we will add the relevant suggested references.

- Line 117: "identified" appears a bit confusing here, and can be understood only once the Method section has been read. I would suggest "sorted" or "chosen". I also suggest to rephrase the whole sentence, whose syntax seems wrong to me.

In the revised manuscript, the sentence will be rewritten as: "After sorting the rainfall events within the 1000 years hourly timeseries, a dataset is built with the antecedent conditions one hour before the beginning of each rainfall event. It includes the previously listed variables plus the total rainfall event depth and the change in water stored in the soil cover at the end of each rainfall event."

- Line 321: Note that a purely 1D approach does not account for flow accumulation and possible secondary infiltration (runoff that infiltrate during its course downstream). That could overestimate the influence of the groundwater level by underestimating the amount of infiltration.

Obviously, heterogeneities of the soil mantle (either morphological, e.g., slope inclination, soil mantle thickness, or physical, e.g., soil layers with different hydraulic properties) may induce 3D *effects in the flow processes. However, 3D effects are expected to be not particularly significant in the studied slopes, for several reasons. First, owing to the geometry of the slopes (i.e., hundreds of meters long with a soil mantle of few meters), the water potential gradients are such that significant deviations of the flow from the vertical direction (or, more precisely, from the direction orthogonal to ground surface) can occur only when the soil approaches saturation, so that capillarity gradients become small and gravitational gradient prevails (along a steeply inclined slope, in this condition the component of the gradient parallel to the slope becomes significant). In addition, the attainment of soil saturation is very unlikely, owing to the very high porosity (as high as 75%). Furthermore, the high inclination angles, in most slopes larger than 35°, imply that slope failure (landslide) would occur before soil attains saturation. Finally, the very high hydraulic conductivity (as high as 30 mm/h), together with the usually unsaturated soil conditions (soil capillary potential rarely overcomes -0,5 m: Cascini et al., 2014; Comegna et al., 2016; Napolitano et al., 2016), makes overland runoff very small, even during the most intense rainfall events (Greco et al., 2018; Marino et al., 2020). In short, lateral redistribution of infiltration flow can be considered quite small in the studied slopes. In the revised manuscript, we will add more information about the characteristics of the studied slopes and soil (Section 2.1), and we will give some justification of the use of the simplified 1D model in Section 2.2.2.*

- Lines 373-376 and 491-493: How do you support this direct relationship between the water level in the aquifer and the one in the stream? No dedicated parameter appears in the mathematical description of the model. Moreover, a direct proportionality might not be true, especially during extreme events (droughts and floods).

Indeed, in the description of how the epikarst aquifer is schematized in the model used to generate the synthetic data (lines 327-332), we have only written that it is modelled as a "linear reservoir, that releases water "as deep groundwater recharge and spring discharge". This conceptualization of the aquifer behavior implies that the streamflow (supplied by the springs) is proportional to the water level in the perched aquifer. We understand that, written in this way, it is not clear to the reader, so we will extend the explanation in the revised manuscript. Obviously, the assumption of a linear relationship linking aquifer water level and spring outflow is a simplification of the reality, and we agree that deviations from linearity are expected, especially in extreme conditions. However, synthetic groundwater level data are used only to separate "low" levels (clusters 1 and 3 of Figures 8, 9 and 10) from "high" (cluster 2 of Figures 8, 9 and 10) or "very high" levels (cluster 4 of Fig. 10), and the same could be made with stream level data, which is probably easier to be measured in the field, compared to the groundwater level in a temporary aquifer.

- Once again, I think that the Conclusion and the contribution of this study would be more valued if it were compared to another approach (local sensitivity analysis for instance).

We hope that, in the revised manuscript, the goal of our study, which did not result clear in the current version, will become clearer. We are not dealing with the development of a mathematical model of the behavior of the soil mantle of the studied slope, but we are rather analyzing field data *(though synthetic) to understand the major cause-effect relationships between (measurable) variables. This analysis may be carried out in absence of any model, to interpret field data.*

Technical corrections

- In text reference ordering (reference grouped in the same parenthesis) should follow a consistent pattern (chronologically I would suggest, see HESS editors to make sure).

Thank you for catching this inconsistency. In the revised manuscript, in all cases we will follow the chronological order.

- Replace "soil cover" by "soil" or "topsoil" all along the article if you're agreeing with my previous statement on the meaning of these terms.

We will use the word "soil mantle" in place of "soil cover".

- Replace "slope" by "sloping soil" or simply "soil" as needed.

We will follow this Reviewer's suggestion. The title of the revised manuscript will also be changed accordingly.

- Line 136: "is" is missing between "results" and "quite variable".

We will rewrite as "In these three areas, the thickness of the soil covers is quite variable".

- Line 439: "developed" is not appropriate here. Use "performed" or "carried out".

We will replace "developed" with "carried out".

- Line 477: is" is missing between "soil storage" and "less connected". Maybe you shoud rephrase this sentence.

We will rephrase the sentence as: "The importance of h_a *on the response of the soil mantle suggests that, in some conditions, the change in soil storage is affected by the capability of water exchange between the soil mantle and the underlying aquifer, as it will be discussed in the following sections".*

- Figures 5 and 6: the unit for the h_a (groundwater level) is mm. Shouldn't it be m? What is the base level?

Thank you for suggesting. We will express h_a in meters, as this unit is much more convenient for the groundwater level. The groundwater level is referred to the base of the epikarst, which is assumed 14 m below the interface between the soil mantle and the bedrock (Table 1). We will specify this in the revised manuscript.

<u>*References*</u>

Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. Journal of Geophysical Research: Earth Surface, 117(4). https://doi.org/10.1029/2012JF002367

Cascini, L., Sorbino, G., Cuomo, S., & Ferlisi, S. (2014). Seasonal effects of rainfall on the shallow pyroclastic deposits of the Campania region (southern Italy). Landslides, 11(5), 779–792. https://doi.org/10.1007/s10346-013-0395-3

Comegna, L., Damiano, E., Greco, R., Guida, A., Olivares, L., & Picarelli, L. (2016). Field hydrological monitoring of a sloping shallow pyroclastic deposit. Canadian Geotechnical Journal, 53(7), 1125–1137. https://doi.org/10.1139/cgj-2015-0344

Dal Soglio, L.; Danquigny, C.; Mazzilli, N.; Emblanch, C.; Massonnat, G. Taking into Account both Explicit Conduits and the Unsaturated Zone in Karst Reservoir Hybrid Models: Impact on the Outlet Hydrograph. Water 2020, 12, 3221. https://doi.org/10.3390/w12113221

Greco, R., Comegna, L., Damiano, E., Guida, A., Olivares, L., & Picarelli, L. (2013). Hydrological modelling of a slope covered with shallow pyroclastic deposits from field monitoring data. Hydrology and Earth System Sciences, 17(10), 4001–4013. https://doi.org/10.5194/hess-17-4001-2013

Greco, R., Marino, P., Santonastaso, G. F., & Damiano, E. (2018). Interaction between perched epikarst aquifer and unsaturated soil cover in the initiation of shallow landslides in pyroclastic soils. Water (Switzerland), 10(7). https://doi.org/10.3390/w10070948

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., & Melillo, M. (2020, January 1). Geographical landslide early warning systems. Earth-Science Reviews. Elsevier B.V. https://doi.org/10.1016/j.earscirev.2019.102973

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014, September 1). Karst water resources in a changing world: Review of hydrological modeling approaches. Reviews of Geophysics. https://doi.org/10.1002/2013RG000443

Leonarduzzi, E., Molnar, P., & McArdell, B. W. (2017). Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. Water Resources Research, 53(8), 6612–6625. https://doi.org/10.1002/2017WR021044

Marino, P., Comegna, L., Damiano, E., Olivares, L., & Greco, R. (2020). Monitoring the hydrological balance of a landslide-prone slope covered by pyroclastic deposits over limestone fractured bedrock. Water (Switzerland), 12(12). https://doi.org/10.3390/w12123309

Napolitano, E., Fusco, F., Baum, R. L., Godt, J. W., & De Vita, P. (2016). Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). Landslides, 13(5), 967–983. https://doi.org/10.1007/s10346-015-0647-5

Peres, D. J., Cancelliere, A., Greco, R., & Bogaard, T. A. (2018). Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. Natural Hazards and Earth System Sciences, 18(2), 633–646. https://doi.org/10.5194/nhess-18-633-2018

Perrin, J., Jeannin, P. Y., & Zwahlen, F. (2003). Epikarst storage in a karst aquifer: A conceptual model based on isotopic data, Milandre test site, Switzerland. Journal of Hydrology, 279(1–4), 106–124. https://doi.org/10.1016/S0022-1694(03)00171-9

Piciullo, L., Tiranti, D., Pecoraro, G., Cepeda, J. M., & Calvello, M. (2020). Standards for the performance assessment of territorial landslide early warning systems. Landslides, 17(11), 2533–2546. https://doi.org/10.1007/s10346-020-01486-4

Segoni, S., Piciullo, L., & Gariano, S. L. (2018a). A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides. Springer Verlag. https://doi.org/10.1007/s10346-018-0966-4

Segoni, S., Rosi, A., Fanti, R., Gallucci, A., Monni, A., & Casagli, N. (2018b). A regional-scale landslide warning system based on 20 years of operational experience. Water (Switzerland), 10(10). https://doi.org/10.3390/w10101297

Reply to Reviewer #2

This paper aims at understanding the responses of slopes to precipitation. To do so the authors rely on mathematical and machine learning models that have been developed by using synthetic data inspired by ground measurements.

This is an interesting topic and the authors have performed a significant number of experiments to shed light on the responses of slopes (although I don't think these experiments account for the effects of slope) to precipitation in mountainous regions.

The organization of the paper is clear however, the English language needs to be improved.

Thank you for your positive evaluation of the manuscript. We agree that it is more appropriate to write "soil response", and we will change it throughout the revised paper, as well as in the title, which will be modified in "Understanding hydrologic controls of sloping soil response to precipitation through Machine Learning analysis applied to synthetic data". We will double-check the English language, so to fix all grammar and spelling mistakes.

Below are my comments:

The title of the paper is "slope responses to precipitation" but actually only a 1-D simplified model has been performed therefore, no slopes have been modelled. For such a study, I would expect a 2-D or even a 3-D model with the slope and lateral flow. This is a big limitation of the study as accounting for slopes will change the results presented in this paper.

As already mentioned in the reply to the previous comment, the title of the paper will be modified to: "Understanding hydrologic controls of sloping soil response to precipitation through Machine Learning analysis applied to synthetic data". However, we would like to make some remarks about the supposed big limitation due to the choice of a simplified 1D model to mimic the field data. Obviously, heterogeneities of the soil mantle (either morphological, e.g., slope inclination, soil mantle thickness, or physical, e.g., soil layers with different hydraulic properties) may induce 3D effects in the flow processes. However, 3D effects are expected to be not particularly significant in the unsaturated soil mantle of the studied slopes, for several reasons. First, owing to the geometry of the slopes (i.e., hundreds of meters long with a soil mantle of few meters), the water potential gradients are such that significant deviations of the flow from the vertical direction (or, more precisely, from the direction orthogonal to ground surface) can occur only when the soil approaches saturation, so that capillarity gradients become small and gravitational gradient prevails (along a steeply inclined slope, in this condition the component of the gradient parallel to the slope becomes significant). In addition, the attainment of soil saturation is very unlikely in the studied soil, owing to its very high porosity (as high as 75%). Furthermore, the high inclination angles, in most slopes larger than 35°, imply that slope failure (landslide) would occur before soil attains saturation. Finally, the very high hydraulic conductivity (as high as more than 30 mm/h), together with the usually unsaturated soil conditions (soil capillary potential rarely overcomes -0,5 m: Cascini et al., 2014; Comegna et al., 2016; Napolitano et al., 2016), makes overland runoff very small, even during the most intense rainfall events (Greco et al., 2018; Marino et al., 2020).

In short, lateral redistribution of infiltration flow can be considered quite small in the soil mantle of the studied slopes. In the revised manuscript, we will add more information about the characteristics of the studied slopes and soil (Section 2.1), and we will give some justification of the use of the simplified 1D model in Section 2.2.2.

I am not familiar with the term “soil cover” what does it exactly mean?

Soil cover is used in the landslide research literature to define a mantle of soil, of various origins, covering a more compact and stable bedrock. We understand that this is misleading for the broader readership of hillslope hydrologists, so we will change it to “soil mantle” throughout the entire revised manuscript.

Write “Precipitation” without and s

Thank you for catching this mistake. We will make the correction throughout the entire manuscript, including the title.

The abstract is quite long, shorten it.

Indeed, the abstract is currently more than 2000 characters. Although the guidelines for article preparation do not fix limits, it is far too long. In the revised manuscript, we will completely rewrite it. We have already drafted it, and it looks that there is room to reduce it to about 60% of its current length.

L17-26: from the paper, most of the work was based on synthetic data.

In the revision of the abstract, we will clarify that the study deals with the synthetic dataset.

One of the objectives of this paper was to find out the variables to be measured. This objective has not been addressed. It is also obvious that soil moisture, groundwater levels, and rainfall are the variables that should be monitored.

Field monitoring for the assessment of rainfall-induced geohazards usually involves only rainfall measurements (Peruccacci et al., 2017). Only recently the importance of soil moisture measurements for the prediction of shallow landslides and debris flows is being recognized (Lazzari et al., 2018; Mirus et al., 2018; Marino et al., 2020). About groundwater level, it is usually considered an informative variable only for deep-seated landslides, as in that case the slip surface of the landslide can be below the groundwater table. Differently, to the best of our knowledge, it has been never recognized useful also for shallow landslide prediction. Hence, in the context of our study, monitoring it is far from being obvious.

We understand that, in the Introduction, the focus on geohazards, and specifically on shallow landslides and debris flows was not clearly stated (see also the following comment from this Reviewer). We will modify the final part of the Introduction to better describe the context and the goal of the study.

The paper did not link landslides to slopes. The authors should clearly state that their work aims to identify the parameters controlling the responses of slopes to precipitation with implications for landslides since the landslides have not been specifically addressed.

We agree. We will rewrite parts of the Introduction to state more clearly that the response of the soil mantle to precipitation is studied for rainfall-induced geohazard assessment and, more specifically, for shallow landslide and debris flow hazard assessment. Also, in the discussion of the results we will add some text to explicitly state the implications that the different conditions may have for slope stability assessment.

L56: "has been early identified" reword.

We will rephrase the entire sentence between lines 55 and 58, which will become: "While the importance of soil moisture conditions on slope runoff and drainage has been recognized long since (Ponce & Hawkins, 1996; Tromp-Van Meerveld & McDonnell, 2006a, 2006b), …"

L83-84: "but where particularly destructive rainfall triggered landslides occurred." Reword

We will rephrase lines 83-85 as: "This research focuses on a case study in an area frequently hit by destructive rainfall-triggered shallow landslides".

L94 avoid starting a sentence with "not only". "Not only" then what?

We will rephrase lines 95-99 as: "Recent studies show that the response of the soil mantle to precipitation is affected not only by rainfall characteristics and antecedent soil moisture, but also by the wetness of the soil-bedrock interface, which controls the leakage of water into the underlying fractured limestone (Marino et al., 2020a, b).

L112: change to 1000-year without the s or remove the hyphen.

Thank you for the suggestion. We will use 1000 years throughout the entire manuscript.

L149: remove a priori.

Agree. "A priori" is not necessary and will be removed.

L220: change to November 11th 2021

Thank you. We will delete the word "the".

L260: is the synthetic rainfall close to reality? Is the rainfall consistent with the climatology of the area. Could you show the comparisons between real rainfall data and the ones you have created? Are the wet and dry interval consistent with reality? Could you provide some comparisons with the real-world data?

Thank you for raising these issues. In the submitted manuscript, we decided to describe very briefly the stochastic NRSP model used for synthetic rainfall generation, giving some references to let the interested readers get more information. We understand that we gave too little information, given that the synthetic rainfall series plays an important role in our methodology. In the following, we give detailed information to the Reviewer, so that he can judge about how the generated synthetic rainfall resembles the real experimental record. In the revised manuscript, we will try to find a trade off between the sake of brevity (the synthetic rainfall generation is here only a tool, but it is not the core of the study) and the need for more information. Possibly we will put some of the information in an appendix.

The NRSP stochastic model of rainfall (Neyman and Scott, 1958; Rodriguez-Iturbe et al., 1987a, b; Cowpertwait et al., 1996) describes the process of point rainfall as a superposition of randomly arriving rain clusters, each containing several rain cells with constant intensity. The hyetograph within a cluster is obtained by the superposition of the intensity of the various cells belonging to the cluster. It has been calibrated based on 17 years experimental data (2000-2016) of rainfall depth at 10 min resolution, recorded by the rain gauge managed by Civil Protection in Cervinara. The calibration has been carried out by minimizing, for rainfall aggregated at various durations, the difference between the following quantities, estimated by the model and calculate from the experimental data: mean, variance, lag 1 autocorrelation, probability of dry interval, probability of transition from dry-to-dry interval, probability of transition from wet-to-wet interval. The calibration procedure is based on the one proposed by Coptwertwait et al. (1996), and it is described in detail in Peres and Cancelliere (2014). To account for the seasonality of rainfall, these quantities have been calculated month by month in the experimental record (Fig. R1), suggesting that the calibration of the NRSP model should be carried out separately for seven homogeneous periods (September, October, November, December-March, April, May-June, July-August).

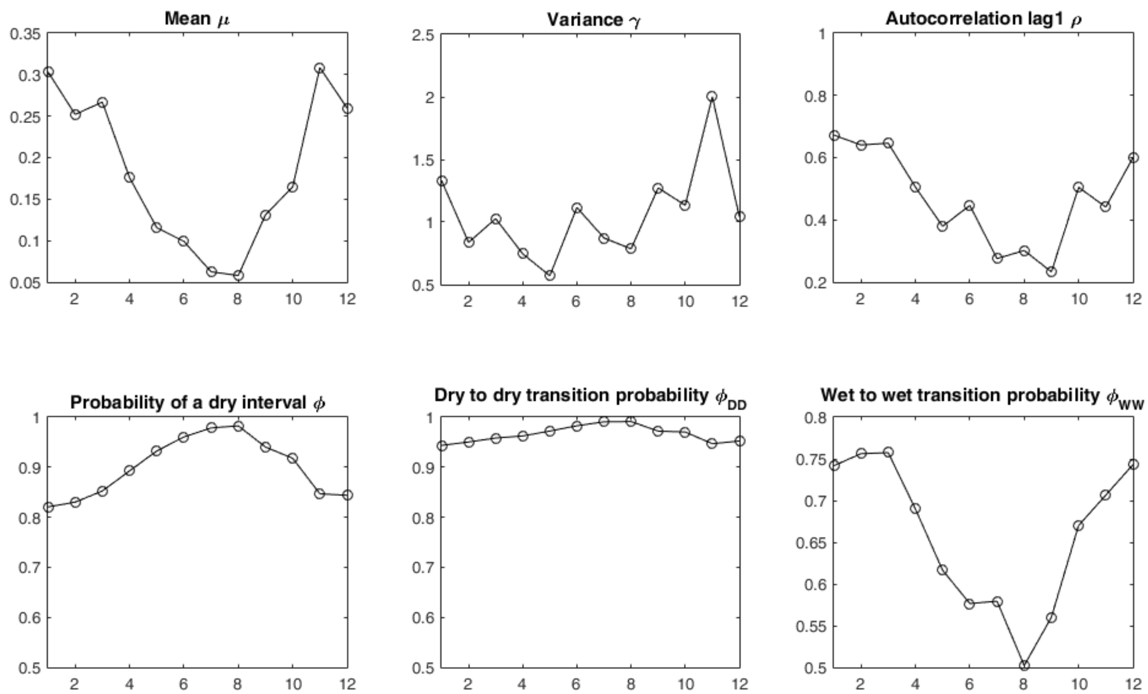


Figure R1. Monthly plot of hourly rainfall characteristics calculated based on the experimental data of the rain gauge of Cervinara.

Table R1 gives the obtained parameters of the NRSP stochastic model, where λ [h^{-1}] represents the parameter of a Poisson process describing the arrival of clusters; ν [-] is the mean number of cells in a cluster, also described by a Poisson process; β [h^{-1}] is the parameter of an exponential probability distribution describing the arrival times of each cell in a cluster, expressed as the number of time intervals of 10 minutes starting from the beginning of a cluster; η [h^{-1}] is the parameter of an exponential probability distribution describing the duration of rain cells; ξ [$h^b mm^{-b}$] is the parameter of a Weibull probability distribution describing the rain intensity of cells, with cumulative probability function $F(x;\xi,b) = 1 - exp(-\xi x^b)$, in which x is cell rain intensity and the parameter $b = 0.8$ has been set a priori (Cowpertwait et al., 1996).

Table R1. Parameters of the NSRP model.

Parameter	Sept	Oct	Nov	Dec-Mar	Apr	May-Jun	Lug.Aug
λ [h^{-1}]	0.015	0.00524	0.00257	0.0238	0.00809	0.00386	0.00900
ν [–]	2.68	36.4	57.1	2.60	38.7	21.6	1.40
β[h^{-1}]	0.265	0.156	0.0167	0.813	0.123	0.116	24.5
η[h^{-1}]	1.41	57.3	1.43	0.280	15.5	8.59	1.23
ξ[h^b mm^{-b}]	0.330	0.047	0.450	0.967	0.186	0.158	0.268

The adherence of the rainfall generated with the stochastic model to the experimental rainfall data has been tested by evaluating rainfall characteristics different from those used for the calibration. For instance, Figure R2 shows the comparison of the rainfall depth cumulated over one year for the experimental data and the synthetic data generated with the calibrated NRSP model.

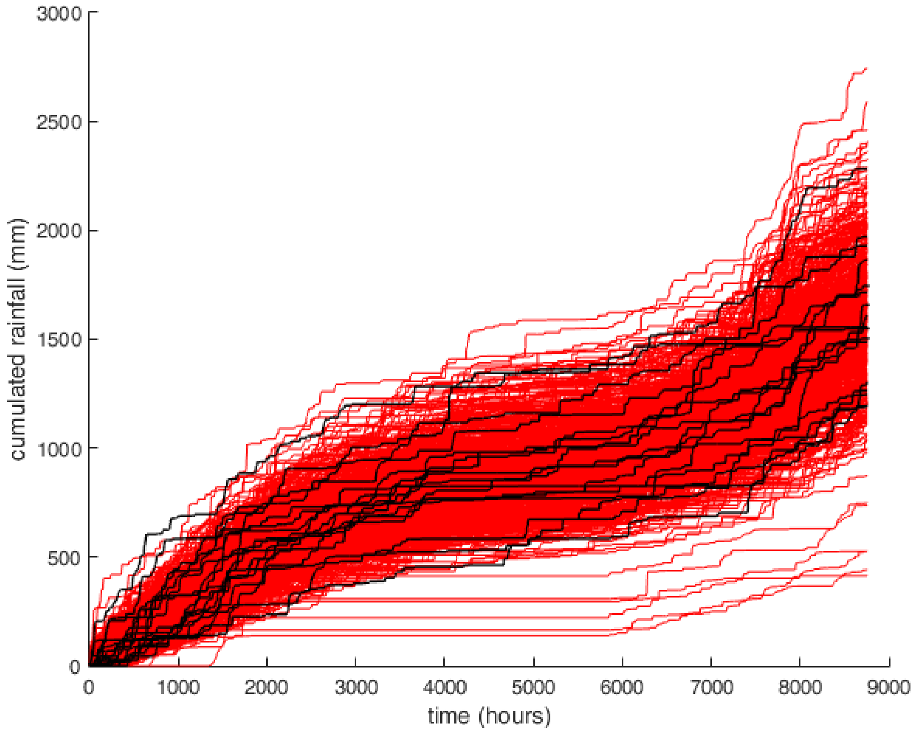


Figure R2. Comparison of observed (black) and simulated (red) cumulated rainfall plots in a year.

In figure R3, the boxplot of the maximum hourly rainfall in one year, observed in the experimental dataset of 17 years, is compared with the same boxplot referred to 20 series of 17 years randomly extracted from the generated 1000 years synthetic rainfall series. Several synthetic 17 years intervals show a distribution of the maximum hourly rainfall close to the observed one.

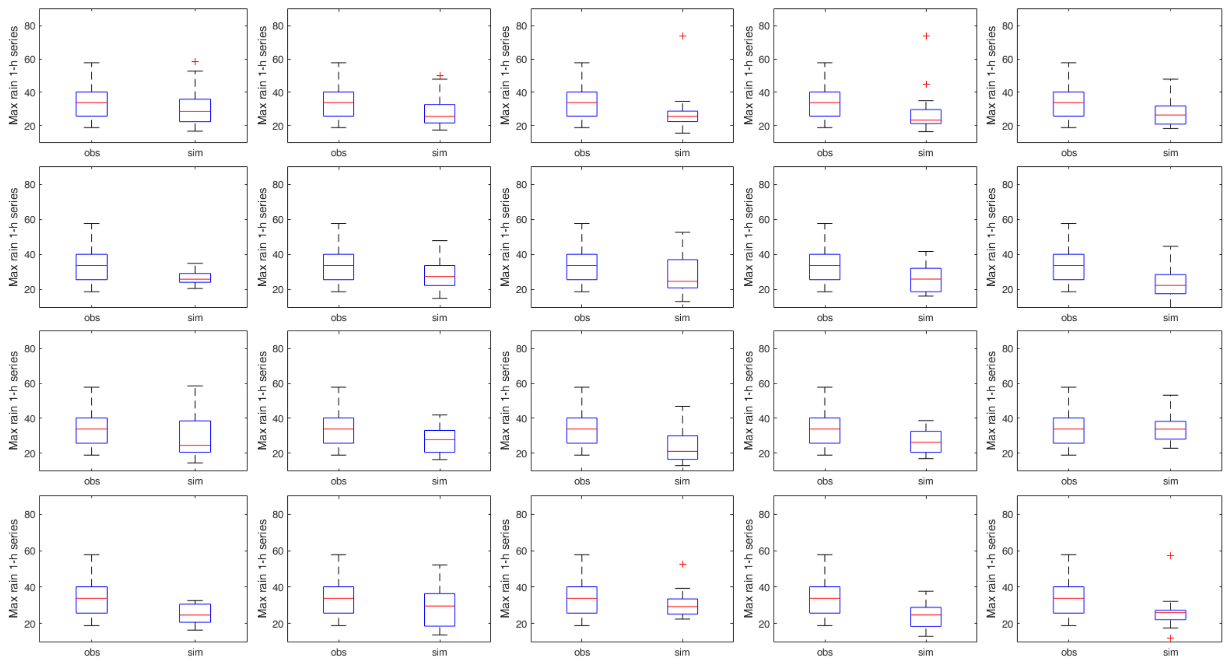


Figure R3. Comparison of observed and simulated distributions (boxplots) of the maximum hourly precipitation in a year, for series of the same length. Each panel shows the distribution for the 17 observed years (boxplot is always the same), and 17 randomly picked simulated years.

Regarding the required comparison between synthetic and observed wet and dry intervals, figure R4 shows the scatterplot of duration and total rain depth of the events, sorted with a separation "dry" interval of 24 hours with less than 2 mm rainfall from the experimental dataset (blue dots) and the synthetic dataset (grey dots). The plots show how the synthetic data contain the observed ones, and that the shape of the dot clouds looks quite similar.

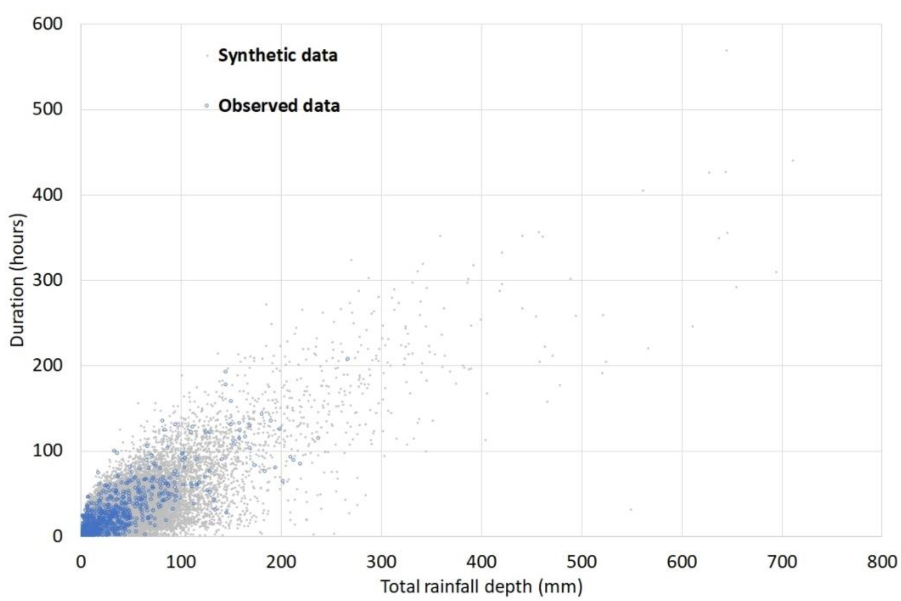


Figure R4. Scatterplot of rainfall event duration vs. total rainfall event depth. The events have been sorted within the rainfall datasets by considering a separation "dry" interval of 24 hours with less than 2 mm rainfall. Blue dots represent events extracted from the 17 years experimental rainfall dataset; grey dots represent events extracted from the 1000 years synthetic rainfall dataset.

Figure R5 shows the frequency distributions of the durations of dry intervals belonging to the 17 years rainfall dataset, and the same distribution for the dry intervals extracted from the 1000 years synthetic dataset: the two distributions look nearly identical.

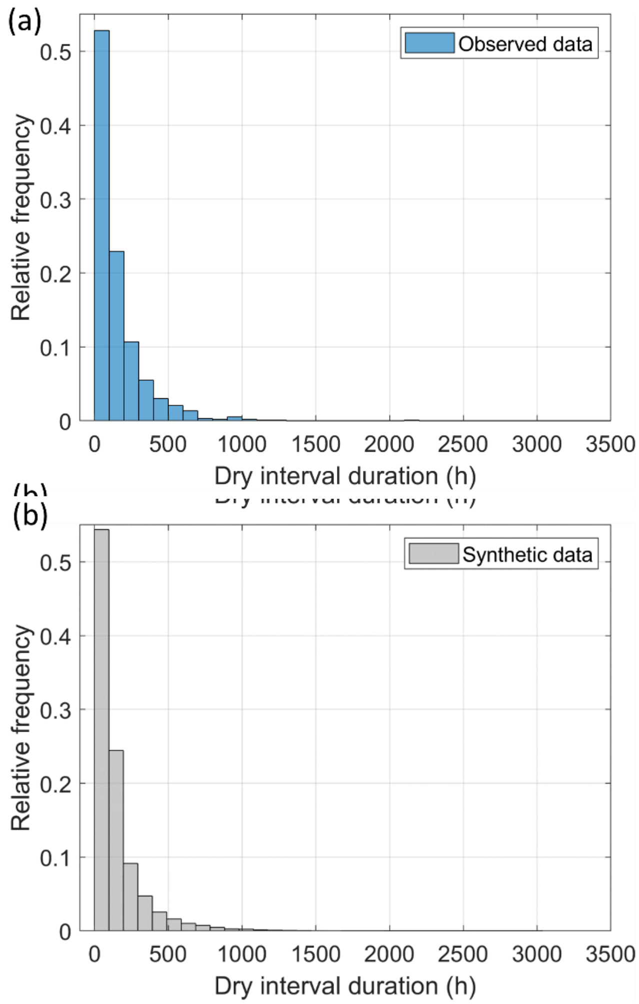


Figure R5. Frequency distributions of dry interval durations for events extracted from the 17 years experimental rainfall dataset (a) and events extracted from the 1000 years synthetic rainfall dataset (b). The events have been sorted within the rainfall datasets by considering a separation "dry" interval of 24 hours with less than 2 mm rainfall.

Did you use a specific code for the hydrologic modeling, if yes please provide reference to the code, if not add the detail about the code you use.

We solved the integration of the 1D Richards' equation, with the conditions assigned at the upper boundary and at the interface with the bedrock, coupled with the continuity equation of the linear reservoir schematizing the perched aquifer in the epikarst, with a self-made finite difference code written in Matlab environment. We believe that adding too many details about the numerical solver in the paper is not necessary, as it is a standard numerical technique, and it would further increase the length of an already long paper.

Even if these test cases are synthetic, you should compare the results to some observations to check if it fits within the boundaries of the variables. For example, soil moisture and groundwater levels can be compared to ground measurements.

The comparison with field data is indeed what we meant to do with Figure 5 (and Figure 6). From those figures, you can directly compare the few measured values of soil moisture of the upper 100 cm of the soil profile with the synthetic data. About the synthetic groundwater level data, they are compared with stream water level data. The reasons for this choice are several:

- So far, we have measurements of stream water level, while only recently we have installed two piezometers in the epikarst.

- The streams are supplied by groundwater coming from the fractured bedrock with very little contribution of overland runoff (less than 1% of the rainfall) only during the most intense rainstorms (it is revealed by the timing of the observed hydrographs in response to rainfall as well as by measurements of electric conductivity of stream water: Marino et al., 2020), so there might be a close relationship linking stream water level and groundwater level.

- Installing piezometers in the fractured limestone is a complex operation, owing to the mechanical resistance of the rock, which obliges to the use of powerful drilling machines; we have recently installed two piezometers (July 2020), but one of them could penetrate the limestone only for less than a couple of meters, as the machine that could be carried in that steep part of the slope (a light one) was not able to drill more depth; the second piezometer, which is at the foot of the slope, in a much less steep terrain, penetrates 16 meters below the ground, but there we have found a different kind of soil mantle (not only pyroclastic soil, but also some meters of alluvial deposits), in total more than 10 meters thick; as we had no clue of the degree of interconnection of the fractured system in the limestone, we decided to extend the pervious part of the piezometer (the filter) to almost the entire penetration depth in the limestone (1,5 meters for the first piezometer, 5 meters for the second one), as a shorter filter at the base of the piezometer (as it is usually done) would increase the risk of not intercepting any connected fracture; in this way, there is more chance for water to enter the piezometer, but, as it may enter at any height along the filter and then pond at the base of the piezometer, we cannot convert the water depth that we measure in the piezometer into a groundwater level; during the 2020/2021 hydrologic year we did not measure any water in the piezometers (2020 was a quite dry year), but in December 2021, after a quite rainy autumn (more than 900 mm between September and December), for the first time water appeared in both the piezometers, confirming that the temporary aquifer actually develops in the epikarst during rainy periods; until summer 2022, the piezometric measurements were made irregularly with a

freatimeter, but in autumn 2022 we have installed an automatic sensor inside the piezometer on the steep terrain (the first one), and this winter we have observed a slight increase of groundwater level once the cumulated rainfall from September exceeded 800 mm.

- Stream water seems to appear and disappear consistently with the groundwater fluctuations, although, so far, we have too few data to demonstrate it; however, measuring stream water level is much easier than groundwater level in the studied context, and it could be an effective surrogate of groundwater level.

- The use that we do with the synthetic groundwater level data (that could be done with field data, either of groundwater or of stream water level) is just to discriminate between "high" level and "low" level, as a proxy to identify active subsurface drainage conditions.

The colored dots of Figures 5 and 6 also show that the seasonality of the synthetic variables is consistent with that of the observed variables.

To understand the effects of each variable on the hydrologic processes, a sensitivity analysis could be performed. Why did the authors choose machine learning technique? I understand that the clustering needs a machine learning technique but to clearly highlight the importance of a variable to a given hydrologic processes, a sensitivity analysis could be performed.

This comment, as well as similar comments made by the other Reviewers, clearly indicates that, in the Introduction, we failed to describe the aims of the study. In the revised manuscript, we will add paragraphs in the Introduction and in the Materials and Methods sections, to better explain the choice of Machine Learning (and specifically Random Forest) instead of a sensitivity analysis. In fact, we believe that adding a sensitivity analysis, which is out of the scope of our study, would be misleading, for the following reasons.

First, we analyzed the dataset mimicking what could be done if, rather than synthetically generated data, one was handling real field monitoring data. In fact, we were mostly looking for a way to identify the major cause-effect relationships between (measurable) inputs and outputs before (possibly, but not necessarily) building a model for the interpretation of such relationships, rather than evaluating the sensitivity of an (already available) model output to variations in the input (although the Random Forest analysis also allows quantifying the information content of each considered input variable).

Second, the sensitivity analysis is usually carried out to evaluate the effects of input (and parameter) uncertainty on model predictions. In this study, the model chain (already calibrated and validated previously: Greco et al., 2013; Comegna et al., 2016; Greco et al., 2018) is used as a tool to generate a (richer) synthetic dataset (this is a common problem in landslide studies, as field monitoring data records, even when they are relatively long, usually contain very few data representative of potentially critical situations). The model is assumed to represent "the reality", and adding a sensitivity analysis may result misleading, as it would move the focus to the performance of the model (which, in general, could also not exist).

Third, the adopted Random Forest analysis, which allows highlighting the most informative combination of measurable variables to predict the output, is somehow a sensitivity analysis as well, as it gives some indications about the relative importance of the input variables on the possibility of predicting the output, without introducing any mathematical model structure, but simply relying on the application of logical operators (IF-THEN-ELSE) between the variables.

L655: what does it mean "monitoring antecedent conditions"

The first paragraphs of the Conclusions section describe what we mean with "antecedent conditions": the values of mean soil moisture of the uppermost 100 cm of the soil mantle (θ_{100}) and the water level in the perched aquifer stored in the epikarst (h_a) before the onset of each rainfall event. In the revised version of the manuscript, we will specify again what are the variables that the results, obtained with the analysis of the synthetic dataset, suggest being useful, if monitored in the field, to predict the soil attitude to retain infiltrating rainwater.

Figure 7: change to a and b instead of left and right. Same for Figure 9

We will fix it in the revised manuscript.

<u>*References*</u>

Cascini, L., Sorbino, G., Cuomo, S., & Ferlisi, S. (2014). Seasonal effects of rainfall on the shallow pyroclastic deposits of the Campania region (southern Italy). Landslides, 11(5), 779–792. https://doi.org/10.1007/s10346-013-0395-3

Comegna, L., Damiano, E., Greco, R., Guida, A., Olivares, L., & Picarelli, L. (2016). Field hydrological monitoring of a sloping shallow pyroclastic deposit. Canadian Geotechnical Journal, 53(7), 1125–1137. https://doi.org/10.1139/cgj-2015-0344

Cowpertwait PSP, O'Connell PE, Metcalfe AV, Mawdsley JA (1996) Stochastic point process modelling of rainfall. I Single-site fitting and validation. J Hydrol. https://doi.org/10.1016/S0022-1694(96)80004-7

Greco, R., Comegna, L., Damiano, E., Guida, A., Olivares, L., & Picarelli, L. (2013). Hydrological modelling of a slope covered with shallow pyroclastic deposits from field monitoring data. Hydrology and Earth System Sciences, 17(10), 4001–4013. https://doi.org/10.5194/hess-17-4001-2013

Greco, R., Marino, P., Santonastaso, G. F., & Damiano, E. (2018). Interaction between perched epikarst aquifer and unsaturated soil cover in the initiation of shallow landslides in pyroclastic soils. Water (Switzerland), 10(7). https://doi.org/10.3390/w10070948

Lazzari, M., Piccarreta, M., & Manfreda, S. (2018). The role of antecedent soil moisture conditions on rainfall-triggered shallow landslides. Natural Hazards and Earth System Science, (December), 1–11.

Marino, P., Comegna, L., Damiano, E., Olivares, L., & Greco, R. (2020). Monitoring the hydrological balance of a landslide-prone slope covered by pyroclastic deposits over limestone fractured bedrock. Water (Switzerland), 12(12). https://doi.org/10.3390/w12123309

Marino P, Peres DJ, Cancelliere A, et al (2020) Soil moisture information can improve shallow landslide forecasting using the hydrometeorological threshold approach. Landslides 17:. doi: 10.1007/s10346-020-01420-8

Mirus BB, Becker RE, Baum RL, Smith JB (2018) Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. Landslides. doi: 10.1007/s10346-018-0995-z

Napolitano, E., Fusco, F., Baum, R. L., Godt, J. W., & De Vita, P. (2016). Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). Landslides, 13(5), 967–983. https://doi.org/10.1007/s10346-015-0647-5

Neyman, J., & Scott, E. L. (1958). Statistical Approach to Problems of Cosmology. Journal of the Royal Statistical Society: Series B (Methodological), 20(1), 1–29. https://doi.org/10.1111/j.2517-6161.1958.tb00272.x

Peres DJ, Cancelliere A (2014) Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. Hydrol Earth Syst Sci 18:4913–4931. https://doi.org/10.5194/hess-18-4913-2014

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., & Guzzetti, F. (2017). Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology, 290, 39–57. https://doi.org/10.1016/j.geomorph.2017.03.031

Rodriguez-Iturbe I, Febres De Power B, Valdes JB (1987) Rectangular pulses point process models for rainfall: analysis of empirical data. J Geophys Res. https://doi.org/10.1029/JD092iD08p09645

Reply to Reviewer #3

General comments

The manuscript describes a field study of a slope, enriched by simulation results from a 1D soil column model. There is relevance in the author's work and presented data, however, unfortunately, I struggle to find the novelty and contribution to current research in this work. All in all, it seems more like an analysis of the behaviour of a simple 1D-model (without considering the model's parameters), and it does not become clear what added value this research offers. I would not count monitoring antecedent conditions as new insight (conclusions, line 655)?

It is clear from the comments of this Reviewer (as well as from those of the other Reviewers), that the way we wrote the Introduction, especially the description of the goal of the study, resulted unclear to the reader. The main aim of the study is to understand how the seasonal slope conditions, related to climate forcing, may affect the capability of the soil of retaining rainwater infiltration for a time long enough to potentially determine critical conditions as a consequence of rainfall events (e.g., the triggering of landslides).

To this aim, to enrich the available field dataset, in which very few times the soil mantle approached critical conditions (as it often occurs in studies dealing with geohazard assessment, as the hazardous conditions are rare per definition), we generated a synthetic dataset with already existing models, calibrated and validated in previous studies based on experimental data collected in the field (Greco et al., 2013; Comegna et al., 2016; Greco et al., 2018), but here the model is just a tool to generate the dataset. Once generated, the dataset represents "the reality", and we analyzed it mimicking what could be done if, rather than synthetically generated data, one was handling real field monitoring data. In fact, we were mostly looking for a way to identify the major cause-effect relationships between (measurable) inputs and outputs before (possibly, but not necessarily) building a model for the interpretation of such relationships, rather than evaluating the sensitivity of an (already available) model output to variations in the input (although the Random Forest analysis also allows quantifying the information content of each considered input variable, but without introducing any mathematical model structure, as it is based on the application of logical rules (IF-THEN-ELSE) to classify the input variables).

This is the reason why we did not test the effects of uncertainty/variability of model parameters: it was simply out of the scope of our study. The obtained results show, for the assessment of the hazard of rainfall-induced landslides and debris flows, the potential value of supplementing the monitoring of rainfall (which is the only monitored variable in nearly all the real world applications) with the monitoring of soil moisture before the rainfall (this is something that is recently being recognized by many researchers, but rarely adopted in operational hazard assessment systems), and the monitoring of the water level in the shallow aquifer developing in the uppermost part of the underlying bedrock. This is quite a novel result, to our knowledge never proposed to predict the response to precipitation of a shallow unsaturated soil mantle. Groundwater level is usually considered an informative variable only for deep-seated landslides, as in that case the (deep) slip surface of the landslide can be below the groundwater table. In the studied geomorphological context, monitoring these antecedent conditions is indeed a novelty.

In the revised manuscript, besides rewriting the Introduction, we will more clearly underline this novel aspect in the discussion of the results, as well as in the Conclusions.

The 1D model comes necessarily with many simplifciations (which of course also has its advantages). But the shortcomints are not adressed (and the advantages - fast runtime etc. - not really exploited). E.g. heterogeneity of hydraulic conductivities on a slope, differences in lower aquifer pressure based on position (top, bottom, local gradient) in the slope, parameter uncertainty in general, differences in layer thicknesses across the slope, etc.

We thank this Reviewer for raising this issue. Obviously, heterogeneities of the soil mantle (either morphological, e.g., slope inclination, soil mantle thickness, or physical, e.g., soil layers with different hydraulic properties) may induce 3D effects in the flow processes. However, 3D effects are expected to be not particularly significant in the soil mantle of the studied slopes, for several reasons. First, owing to the geometry of the slopes (i.e., hundreds of meters long with a soil mantle of few meters), the water potential gradients are such that significant deviations of the flow from the vertical direction (or, more precisely, from the direction orthogonal to ground surface) can occur only when the soil approaches saturation, so that capillarity gradients become small and gravitational gradient prevails (along a steeply inclined slope, in this condition the component of the gradient parallel to the slope becomes significant). In addition, the attainment of soil saturation is very unlikely, owing to the very high porosity (as high as 75%). Furthermore, the high inclination angles, in most slopes larger than 35°, imply that slope failure (landslide) would occur before soil attains saturation. Finally, the very high hydraulic conductivity (as high as 30 mm/h), together with the usually unsaturated soil conditions (soil capillary potential rarely overcomes -0,5 m: Cascini et al., 2014; Comegna et al., 2016; Napolitano et al., 2016), makes overland runoff very small, even during the most intense rainfall events (Greco et al., 2018; Marino et al., 2020). In short, lateral redistribution of infiltration flow can be considered quite small in the soil mantle of the studied slopes. In the revised manuscript, we will add more information about the characteristics of the studied slopes and soil (Section 2.1), and we will give some justification of the use of the simplified 1D model in Section 2.2.2.

About the variability of the groundwater table depth, this is also obviously true (and indeed, observations made in two piezometers, recently installed at two different altitudes along the slope, confirm that the groundwater table depth may be quite different). However, the use that we make of the groundwater level information is to discriminate "low" levels (clusters 1 and 3 of Figures 8, 9 and 10) from "high" levels (cluster 2 of Figures 8, 9 and 10) or "very high" levels (cluster 4 of Fig. 10). Depending on the availability of monitoring instruments, this could be made with a single piezometer, as well as with several piezometers (but, although with different levels, if the groundwater level in a piezometer is high, it will be likely high also in the others, unless they are so far from each other that they are monitoring disconnected groundwater systems). This aspect will be better clarified in the discussion of the results of the revised manuscript.

Given such a relatively simple model, it should be possible to run uncertainty analysis or sensitivity analysis – and that possibility should be exploited.

Then, it is unclear why a Random Forest algorithm is used to emulate the physically-based model outputs – why not simply use an ensemble of the physically-based model, and do some uncertainty and sensitivity analysis on that? Also, some basic considerations when using RF have been ignored (hyperparameter search, sound cross validation strategy, discussion of size of dataset etc).

Therefore, I recommend to address the analysis of the physically-based model more extensively before venturing into a ML analysis of its results.

This comments again clearly indicate that, in the Introduction, we failed to describe the aims of the study. In the revised manuscript, we will add paragraphs in the Introduction and in the Materials and Methods sections, to better explain the choice of Machine Learning (and specifically Random Forest) instead of a sensitivity analysis. In fact, we believe that adding a sensitivity analysis, which is out of the scope of our study, would be misleading, for the following reasons.

First, we analyzed the dataset mimicking what could be done if, rather than synthetically generated data, one was handling real field monitoring data. In fact, we were mostly looking for a way to identify the major cause-effect relationships between (measurable) inputs and outputs before (possibly, but not necessarily) building a model for the interpretation of such relationships, rather than evaluating the sensitivity of an (already available) model output to variations in the input (although the Random Forest analysis also allows quantifying the information content of each considered input variable).

Second, the sensitivity analysis is usually carried out to evaluate the effects of input (and parameter) uncertainty on model predictions. In this study, the model chain (already calibrated and validated previously: Greco et al., 2013; Comegna et al., 2016; Greco et al., 2018) is used as a tool to generate a (richer) synthetic dataset (this is a common problem in landslide studies, as field monitoring data records, even when they are relatively long, usually contain very few data representative of potentially critical situations). The model is assumed to represent "the reality", and adding a sensitivity analysis may result misleading, as it would move the focus to the performance of the model (which, in general, could also not exist).

Third, the adopted Random Forest analysis, which allows highlighting the most informative combination of measurable variables to predict the output, is somehow a sensitivity analysis as well, as it gives some indications about the relative importance of the input variables on the possibility of predicting the output, without introducing any mathematical model structure, but simply relying on the application of logical operators (IF-THEN-ELSE) between the variables.

<u>Specific comments</u>

Section 1. Introduction

Line 94 ff: Remember that you are describing location-specific aspects – "fractured limestone", depth of soil above bedrock – i.e. this is not generally applicable.

Thank you for catching this. Indeed, this paragraph refers specifically to the presented case study (i.e., slopes with a shallow pyroclastic soil mantle covering a fractured limestone bedrock). Therefore, it should be moved to the final part of the Introduction, where the characteristics of the case study are briefly anticipated.

Line 105ff: This paragraph with your objectives could be formulated more clearly. E.g. the first sentence – please reformulate and state more clearly what your objectives are.

Apart from some sentences, which should be rewritten to improve the language and style, this paragraph must be totally rewritten, as it misled the Reviewers (and it would mislead all the readers of the paper). In fact, as already pointed out in the replies to previous comments from this Reviewer, the focus of the study is not on the physically based model (which is instead, by our mistake, mentioned firstly in the paragraph of the objectives of the study), but on the interpretation of field monitoring data. Coupled with the NRSP stochastic model of rainfall, the model was just a tool to generate a rich synthetic dataset, which was then analyzed as if it were obtained by field measurements. This approach was chosen because field data series always contain few data representative of potentially critical conditions (in other words, landslides and debris flows, as well as other rainfall-induced geohazards are rare phenomena), so to have enough data to carry out statistically significant analyses. The final part of the Introduction will be completely reformulated to state the goal of the study more clearly.

Section 2. Material and methods

Line 140, line 146: Make clear that you describe/summarize the methods applied in *your* study (e.g. by adding "see section 2.2")

In the revised manuscript we will make more clear that this initial part of section 2 anticipates what is then described in the following subsections.

Figure 1: I suggest a visualizing a DEM as background in one of the images – maybe the smaller inset? However, it remains unclear whether the inset is necessary at all, or one map would do just fine. Indicate the location of the monitoring station. Moreover, the outline of the inset seems incorrect, as well as the red star indicating the main scarp does not match the location indicated in the inset.

We will use a DEM instead of a photo in the smaller inset, so to give the reader information about the morphology of the studied slope. We will also be more precise in selecting the zoomed rectangle, as well as we will move the red star in the correct position (thank you for catching this mistake), which should be the main scarp of the landslide.

Figure 2: This is based on data from Damiano et al. 2012?

Indeed Figure 2 is adapted from Damiano et al. (2012), and this should have been mentioned in the caption. However, as the layered nature of the soil cover is an information that is never exploited in this study, we are considering summarizing its description in the revised manuscript, likely eliminating the figure.

Line 196: By “pyroclastic ashes” you here refer to the entire soil profile? Or only to the layer “Volcanic ashes” in Fig. 2? The terminology used around here should be made more clear.

We agree that we should be consistent, using the word “pyroclastic” throughout the entire manuscript.

Figure 3: Please improve this figure. E.g. north up, show its outline in Figure 1.

In the revised manuscript, we will change the photo, using one with the standard orientation (North upward).

Section 2.1.1 partly lacks details – what has been measured exactly? How long? What temporal resolution? (part of it comes later in section 3.2, and should be noted here)

Thank you for the suggestion. In the We will add more information about the monitoring campaign, and we will also move to Section 2.1.1 the information given between lines 487 and 492 (Section 3.2), leaving there only a small mention.

Section 2.2.1: Reference for the NSRP model is lacking. Also, some kind of comparison (various statistics?) of the synthetic time series with the observed time series would be appreciated.

Thank you for raising these issues. In the submitted manuscript, we decided to describe very briefly the stochastic NRSP model used for synthetic rainfall generation, giving some references to let the interested readers get more information. We understand that we gave too little information, given that the synthetic rainfall series plays an important role in our methodology. In the following, we give detailed information to this Reviewer, so that he can judge about how the generated synthetic rainfall resembles the real experimental record. In the revised manuscript, we will try to find a trade off between the sake of brevity (the synthetic rainfall generation is here only a tool, but it is not the core of the study) and the need for more information. Possibly we will put some of the information in an appendix.

The NRSP stochastic model of rainfall (Neyman and Scott, 1958; Rodriguez-Iturbe et al., 1987a, b; Cowpertwait et al., 1996) describes the process of point rainfall as a superposition of randomly arriving rain clusters, each containing several rain cells with constant intensity. The hyetograph within a cluster is obtained by the superposition of the intensity of the various cells belonging to the cluster. It has been calibrated based on 17 years experimental data (2000-2016) of rainfall *depth at 10 min resolution, recorded by the rain gauge managed by Civil Protection in Cervinara. The calibration has been carried out by minimizing, for rainfall aggregated at various durations, the difference between the following quantities, estimated by the model and calculate from the experimental data: mean, variance, lag 1 autocorrelation, probability of dry interval, probability of transition from dry-to-dry interval, probability of transition from wet-to-wet interval. The calibration procedure is based on the one proposed by Coptwertwait et al. (1996), and it is described in detail in Peres and Cancelliere (2014). To account for the seasonality of rainfall, these quantities have been calculated month by month in the experimental record (Fig. R1), suggesting that the calibration of the NRSP model should be carried out separately for seven homogeneous periods (September, October, November, December-March, April, May-June, July-August).*

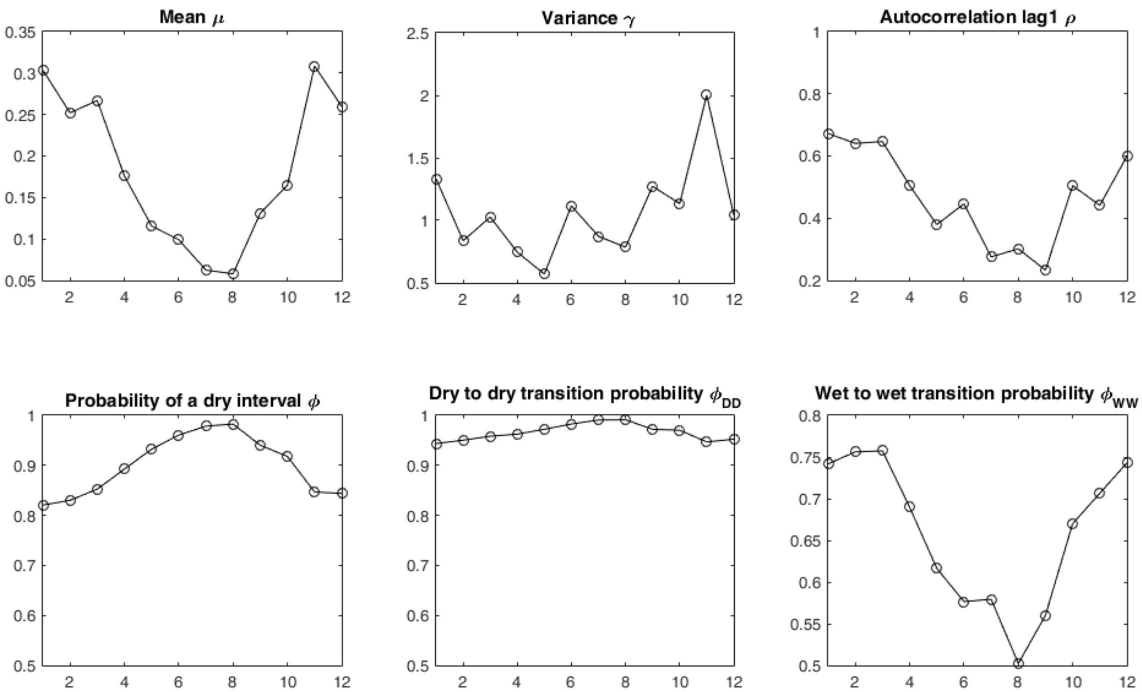


Figure R1. Monthly plot of hourly rainfall characteristics calculated based on the experimental data of the rain gauge of Cervinara.

Table R1 gives the obtained parameters of the NRSP stochastic model, where λ [h^{-1}] represents the parameter of a Poisson process describing the arrival of clusters; ν [-] is the mean number of cells in a cluster, also described by a Poisson process; β [h^{-1}] is the parameter of an exponential probability distribution describing the arrival times of each cell in a cluster, expressed as the number of time intervals of 10 minutes starting from the beginning of a cluster; η [h^{-1}] is the parameter of an exponential probability distribution describing the duration of rain cells; ξ [$h^{b}mm^{-b}$] is the parameter of a Weibull probability distribution describing the rain intensity of cells, with cumulative probability function $F(x;\xi,b) = 1 - exp(-\xi x^{b})$, in which x is cell rain intensity and the parameter $b = 0.8$ has been set a priori (Cowpertwait et al., 1996).

Table R1. Parameters of the NSRP model.

Parameter	Sept	Oct	Nov	Dec-Mar	Apr	May-Jun	Lug.Aug
λ [h^{-1}]	0.015	0.00524	0.00257	0.0238	0.00809	0.00386	0.00900
ν [–]	2.68	36.4	57.1	2.60	38.7	21.6	1.40
β[h^{-1}]	0.265	0.156	0.0167	0.813	0.123	0.116	24.5
η[h^{-1}]	1.41	57.3	1.43	0.280	15.5	8.59	1.23
ξ[h^b mm^{-b}]	0.330	0.047	0.450	0.967	0.186	0.158	0.268

The adherence of the rainfall generated with the stochastic model to the experimental rainfall data has been tested by evaluating rainfall characteristics different from those used for the calibration. For instance, Figure R2 shows the comparison of the rainfall depth cumulated over one year for the experimental data and the synthetic data generated with the calibrated NRSP model.

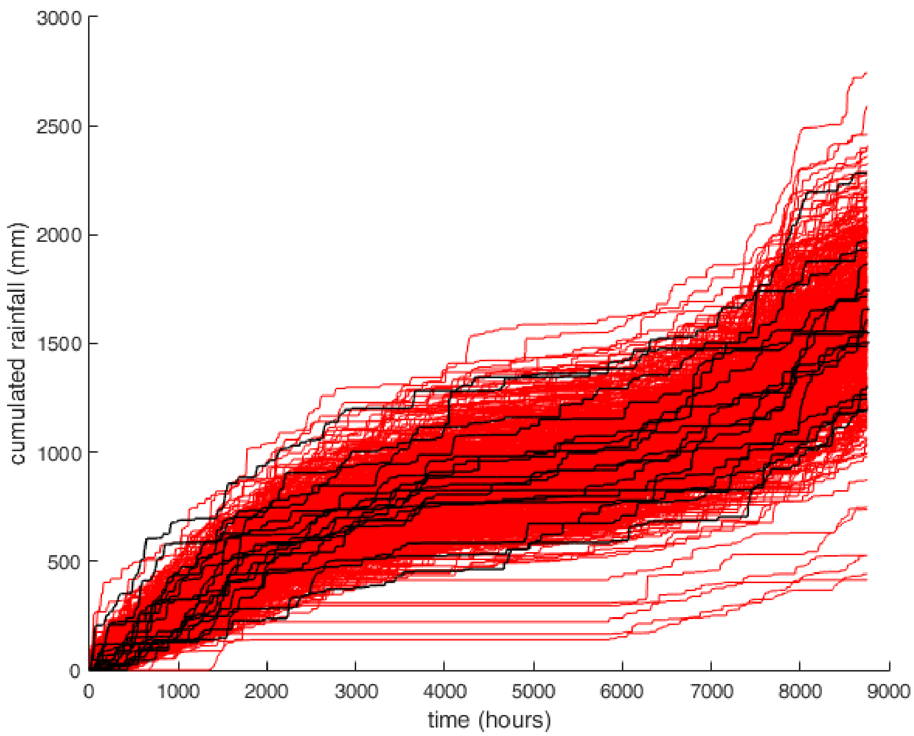


Figure R2. Comparison of observed (black) and simulated (red) cumulated rainfall plots in a year.

In figure R3, the boxplot of the maximum hourly rainfall in one year, observed in the experimental dataset of 17 years, is compared with the same boxplot referred to 20 series of 17 years randomly extracted from the generated 1000 years synthetic rainfall series. Several synthetic 17 years intervals show a distribution of the maximum hourly rainfall close to the observed one.

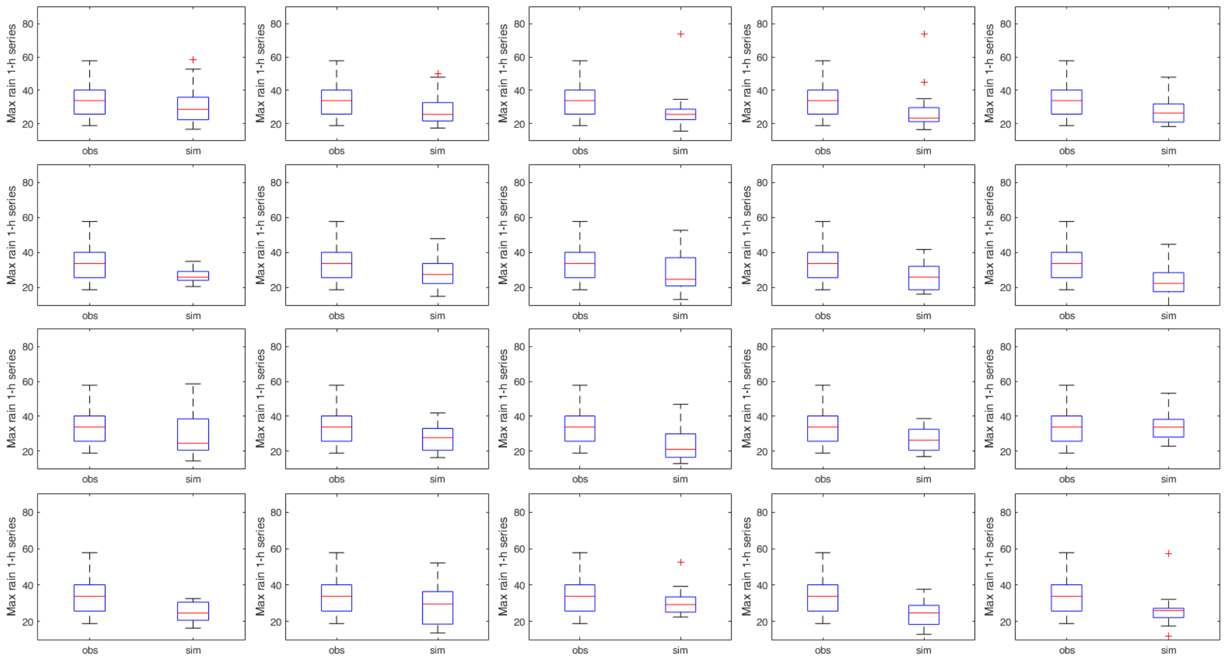


Figure R3. Comparison of observed and simulated distributions (boxplots) of the maximum hourly precipitation in a year, for series of the same length. Each panel shows the distribution for the 17 observed years (boxplot is always the same), and 17 randomly picked simulated years.

Regarding the comparison between synthetic and observed wet and dry intervals, figure R4 shows the scatterplot of duration and total rain depth of the events, sorted with a separation "dry" interval of 24 hours with less than 2 mm rainfall from the experimental dataset (blue dots) and the synthetic dataset (grey dots). The plots show how the synthetic data contain the observed ones, and that the shape of the dot clouds looks quite similar.

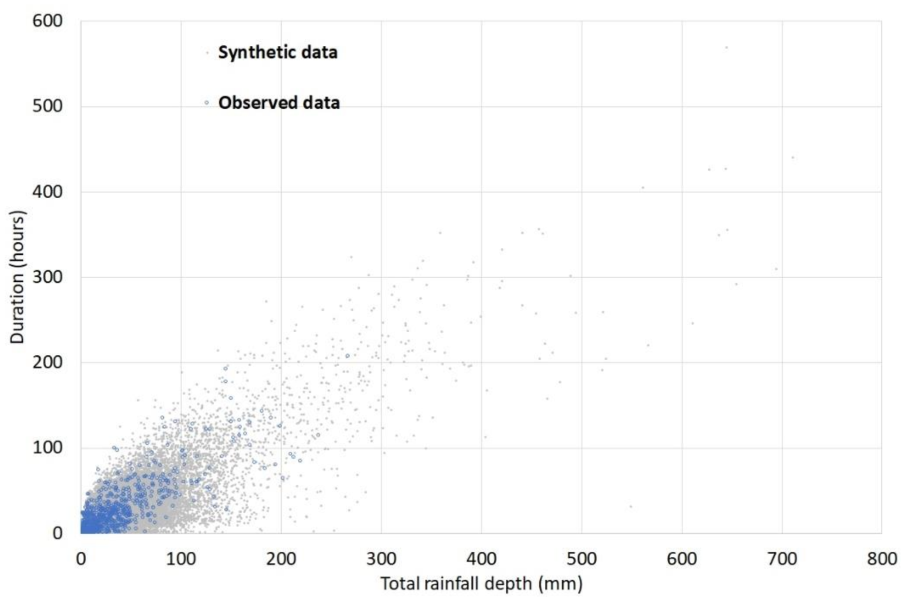


Figure R4. Scatterplot of rainfall event duration vs. total rainfall event depth. The events have been sorted within the rainfall datasets by considering a separation "dry" interval of 24 hours with less than 2 mm rainfall. Blue dots represent events extracted from the 17 years experimental rainfall dataset; grey dots represent events extracted from the 1000 years synthetic rainfall dataset.

Figure R5 shows the frequency distributions of the durations of dry intervals belonging to the 17 years rainfall dataset, and the same distribution for the dry intervals extracted from the 1000 years synthetic dataset: the two distributions look nearly identical.

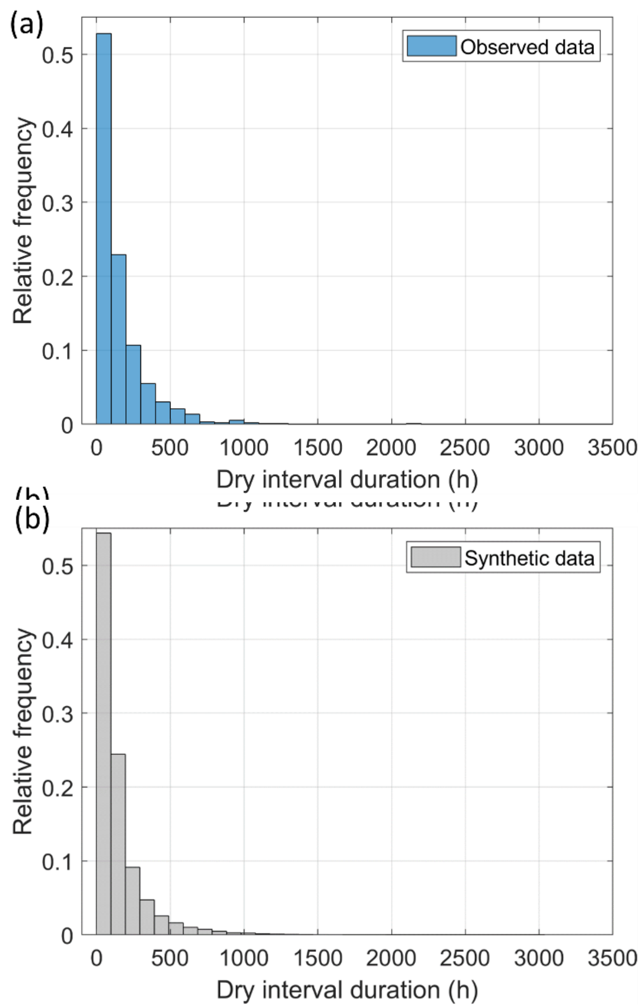


Figure R5. Frequency distributions of dry interval durations for events extracted from the 17 years experimental rainfall dataset (a) and events extracted from the 1000 years synthetic rainfall dataset (b). The events have been sorted within the rainfall datasets by considering a separation "dry" interval of 24 hours with less than 2 mm rainfall.

Line 284ff: For your definition of separate rainfall events – can you argue for the threshold of 24h with less than 2mm rainfall to separate events? This seems very little – does the slope really drain in such a short time? I.e. have effects of preceding events really disappeared after such a short time? I do not follow your argument that the volumetric water content at 10cm depth is sufficient to conclude that. Also, you show model results in Figure 4, right? That is not directly apparent to the reader.

The separation of events within the continuous rainfall record aims at linking the occurrence (or non-occurrence) of critical conditions to a rainfall event, so that they can be considered as a direct consequence of that rainfall event. This is commonly made when empirical predictive tools (e.g., rainfall thresholds: Segoni et al., 2018a, b; Guzzetti et al., 2020; Piciullo et al., 2020) are implemented as part of early warning systems, e.g., against rainfall-induced landslides or debris flows, and the definition of the separation criterion is usually made empirically, looking at the performance of the predictor with different choices of the separation criterion.

From a physical viewpoint, especially when one is interested in the separation between the role of antecedent conditions, i.e., related to previous precipitation (and drainage/evapotranspiration) history, from the direct effects of the last precipitation event, it is quite complex to define a suitable separation criterion, especially if dealing with slow processes activated by precipitations, such as the infiltration through the unsaturated soil layer. In fact, to completely separate what depends on "previous" precipitation from what is linked to the last rainfall event, one should wait for the infiltration process initiated by previous precipitations to be finished, and, in a soil layer of few meters thickness, it may take several days. Extending so much the dry time interval between two separate events, especially during rainy seasons, would imply the aggregation of several events in a single one, thus leading to long rainy periods, rather than events, thus preventing the desired separation of antecedent conditions from direct effects of events. So, as we have defined the "response" of the soil layer as its attitude to retain infiltrated rainwater after the end of a rain event, looking at the moisture of the topsoil layer seemed a good trade-off: topsoil moisture controls the infiltration at the ground surface, hence when gravitational drainage from the topsoil is already over (the field capacity has been reached), the infiltration of a new rainfall input through the ground surface would not depend (or it would only little depend) on the remnants of the infiltration process caused by previous precipitation. In this respect, we tested a separation dry interval of 24 hours, commonly used when the available rainfall data are at daily resolution (Berti et al., 2012; Leonarduzzi et al., 2017; Peres et al., 2018), and anyway in line with the empirical choices that are commonly made in the early warning community (Segoni et al., 2018a).

We will clarify that Figure 4 shows synthetic data.

Section 2.2.2: I miss a discussion of the limitations of the 1D model. Assuming that some lateral in the soil layer (and deeper aquifer) exists, this results in different groundwater level across different parts of the slope / for different gradients. Etc… You set up a single model (with a single parameter set) – ignoring the heterogeneity of soil thickness, hydraulic conductivites, etc. one would find along the slope? Given such a simple 1D model, I would at least recommend to perform some kind of sensitivity analysis / parameter uncertainty estimation / ensemble model run.

Please, see our answer to a previous similar comment about the limitations of 1D model. Again, it clearly arises that the Reviewer was misled by our unclear description of the goal of the study, *which is not about evaluating the performance of a model, but about how to extract information about cause-effect relationship from a dataset of hydrological variables describing the response to precipitation of the pyroclastic soil mantle of the studied slope. We analyzed the data as if they came from field monitoring, although, to get a richer dataset, we generated a synthetic dataset. The model, already developed, calibrated, and validated in previous studies (Greco et al., 2018), was here used just as a tool to generate the synthetic dataset, by coupling it with the NRSP stochastic model of rainfall. Hence, a sensitivity analysis of the model output to parameters is out of the scope of this study.*

Section 2.2.3/Line 362ff: Referring back to my comment to line 332ff: You mention that you extract variables before the onset of each rainfall event, as the "would be measurable in the field". E.g. those are aquifer water level – which, I assume, is largely different on the top of the slope from the bottom of the slope.

As we already replied to a previous comment from this Reviewer, the groundwater table depth is indeed variable throughout the slope (observations made in two piezometers, recently installed at two different altitudes along the slope, confirm that the groundwater table depth may be quite different). However, the use that we make of the groundwater level information is to discriminate "low" levels (clusters 1 and 3 of Figures 8, 9 and 10) from "high" levels (cluster 2 of Figures 8, 9 and 10) or "very high" levels (cluster 4 of Fig. 10). Depending on the availability of monitoring instruments, this could be made with a single piezometer, as well as with several piezometers (but, although with different levels, if the groundwater level in a piezometer is high, it will be likely high also in the others, unless they are so far from each other that they are monitoring disconnected groundwater systems). This aspect will be better clarified in the discussion of the results of the revised manuscript.

Line 370: An actual quantification of soil water content based on satellite observations is hard (rather than a relative value), especially on such small scales – this limitation should be mentioned.

We will briefly mention the limitations of satellite products compared to field observations.

Section 2.3.1

Unfortunately, it is hard to understand how you obtain your dataset with triplets of variables. Why did you chose three of the four to predict? You are predicting simulated change in soil water storage, right? What is the time interval? Or is it just aggregated values per rainfall event? How many datapoints? And what are your variable inputs? Only the four mentioned variables? Why did you not also run the 1D physically based model with various parameterizations – I assume the outputs would have looked quite different. Also, why did you chose RF, and how did you decide for number of trees, splits etc (hyperparameters)?

Again, this Reviewer has been misled by the confusion in the description of the goal of this study. So, we have already replied to the point about evaluating the effects of parameters uncertainty/variability on model output, which is out of the scope of this study.

Some of the requested information is already given in the manuscript: we evaluate the change of soil storage between after and before any rainfall event; the number of data is given (around 53000); the choice of the variables is described as the outcome of Random Forest analysis. We did not test other choices of the variables to be monitored, as we wanted to stick to what can be easily obtained by means of currently available instruments (i.e., satellite products, field soil moisture measurement networks, piezometers, stream water stage sensors).

About the choice of the RF, as we already pointed out in a previous answer, it was made to mimic what could be done if, rather than synthetically generated data, one was handling real field monitoring data. In fact, we were mostly looking for a way to identify the major cause-effect relationships between (measurable) inputs and outputs before (possibly, but not necessarily) building a model for the interpretation of such relationships, rather than evaluating the sensitivity of an (already available) model output to variations in the input, and Random Forest allows quantifying the information content of each considered input variable without introducing any mathematical model structure, but just relying on the application of logical operators (IF-THEN-ELSE) between the variables.

It looks clear that, aiming at brevity, we gave too little information about how the RF was implemented. Here we provide detailed information to this Reviewer about the training and validation of the RF model, as well as about the choice of hyperparameters.

To evaluate the performance of the Random Forest model, the cross-validation technique was used. In cross-validation, the dataset is divided into k equal parts, also known as folds. Then, for each fold, the Random Forest model is trained on the remaining k-1 folds of the data and tested on the remaining fold. We chose k=5, so that the process was repeated 5 times, every time using a different fold (20% of the dataset) as the validation set. A performance metric was calculated for each fold, to estimate how well the RF model perform on new data. We used the explained variance score, computed as follows:

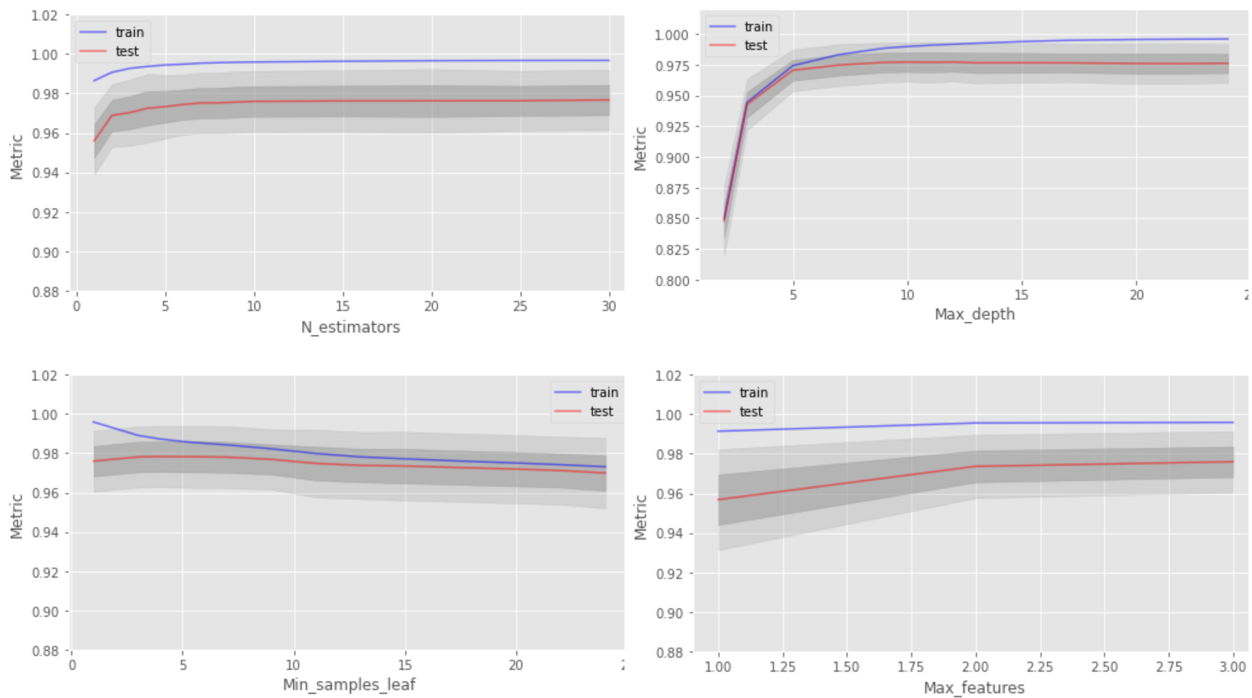
$$metric = 1 - \frac{Var(y - \hat{y})}{var(y)}$$

$Var(y - \hat{y})$ and $Var(y)$ are the variance of prediction errors and actual values respectively. Higher values of explained variance indicate better performance. In addition, the tuning of the hyperparameter of the model was performed based on the cross-validation results to select the optimal set of hyperparameters. In other words, the random forest model was fitted k-times to the data provided by the cross-validation, changing the value of the following hyperparameters once at time:

- *n_estimators: the number of trees to build in the forest;*
- *max_depth: the maximum depth of each decision tree in the forest;*
- *min_samples_leaf: the minimum number of samples required to be at a leaf node;*

- *max_features: the number of features to consider when looking for the best split.*

The following plots show the trend of the hyperparameters vs. the performance metric, the explained variance score. The blue and red lines are the average values of the performance metric computed for the train datasets and the test dataset, respectively, provided by the cross-validation process. The gray and light gray bands represent average values of the metric +/- the standard deviation and two times the standard deviation, respectively.



According to the evidence shown by the previous plots, the search of the optimal parameters was carried out using the following ranges for the hyperparameters:

- *n_estimators was fixed to 10;*
- *max_depth: [5, 10, 15];*
- *min_samples_leaf: [1, 3, 5, 7, 9],*
- *max_features: [2,3].*

The obtained best parameters are max_depth=10; min_samples_leaf=3; max_features=3.

As well known, RF algorithm can work well with high dimensional or multidimensional data, but having a high number of features can lead to overfitting. Therefore, it's important to adjust the hyperparameters (e.g., max_features) to prevent overfitting and create a robust model. Anyway, the synthetic dataset is characterized by only three features (the three variables quantifying antecedent conditions), and a very large number of samples (more than 50000), hence overfitting could be excluded, also owing to the cross-validation method used for model training.

Section 2.3.2

Again, what data exactly do you cluster? The covariates described in the previous section? Also (line 411) – you do not really use “spatial” data here, do you?

The clustering is carried out on the triplets that, based on the results of the RF analysis, seems to be the most suitable to describe the effect of antecedent conditions (prior to the onset of each rainfall event) on the attitude of the soil mantle to retain infiltrating rainwater.

In line 411 we meant that k-means clustering evaluates the distance between the dots in the space of the variables to which the clustering is applied. Nothing to do with distance in the field. We will rephrase the sentence to make it clearer.

Section 3.1

Table 3 vs Table 2: Doesn’t this shown effect of the normalization of delta(S) to H simply show that there is a quite good linear correlation between H and delta(S) – why then not simply use a linear regression model?

It is clear that we gave too much emphasis to the analysis of deltaS, so that the Reviewer was misled. The results just show that, obviously, the more it rains, the more soil storage increases, and if one is interested in evaluating the response of the soil mantle to precipitation, one should look at the ratio between deltaS and H. We will remove Table 2 and just briefly explain the choice of deltaS/H.

Figure 5: The water levels should not be shown on logarithmic scale (that just hides deviations?). Also, make clear that you compare simulated aquifer water levels (ha) with observed stream water levels (hs). Can you try to argue better for your statement “a direct relationship links the water level in the aquifer and the water level in the stream” based on this? As you also mention in section 4, line 632: “substantial agreement between synthetic and experimental data” – this has to be quantified.

We beg to disagree on the first remark: if plotted along a Cartesian axis, many of the dots would collapse very close to the zero (the “low” water levels), thus hiding the existence of a cluster containing a large number of antecedent conditions.

The agreement between synthetic and field data is indeed what we meant to show in Figure 5 and Figure 6. From those figures, you can directly compare the few measured values of soil moisture of the upper 100 cm of the soil profile with the synthetic data. About the synthetic groundwater level data, they are compared with stream water level data. The reasons for this choice are several:

- So far, we have measurements of stream water level, while only recently we have installed two piezometers in the epikarst.

- The streams are supplied by groundwater coming from the fractured bedrock with very little contribution of overland runoff (less than 1% of the rainfall) only during the most intense rainstorms (it is revealed by the timing of the observed hydrographs in response to rainfall as well as by measurements of electric conductivity of stream water: Marino et al., 2020), so there might be a close relationship linking stream water level and groundwater level.

- Installing piezometers in the fractured limestone is a complex operation, owing to the mechanical resistance of the rock, which obliges to the use of powerful drilling machines; we have recently installed two piezometers (July 2020), but one of them could penetrate the limestone only for less than a couple of meters, as the machine that could be carried in that steep part of the slope (a light one) was not able to drill more depth; the second piezometer, which is at the foot of the slope, in a much less steep terrain, penetrates 16 meters below the ground, but there we have found a different kind of soil mantle (not only pyroclastic soil, but also some meters of alluvial deposits), in total more than 10 meters thick; as we had no clue of the degree of interconnection of the fractured system in the limestone, we decided to extend the pervious part of the piezometer (the filter) to almost the entire penetration depth in the limestone (1,5 meters for the first piezometer, 5 meters for the second one), as a shorter filter at the base of the piezometer (as it is usually done) would increase the risk of not intercepting any connected fracture; in this way, there is more chance for water to enter the piezometer, but, as it may enter at any height along the filter and then pond at the base of the piezometer, we cannot convert the water depth that we measure in the piezometer into a groundwater level; during the 2020/2021 hydrologic year we did not measure any water in the piezometers (2020 was a quite dry year), but in December 2021, after a quite rainy autumn (more than 900 mm between September and December), for the first time water appeared in both the piezometers, confirming that the temporary aquifer actually develops in the epikarst during rainy periods; until summer 2022, the piezometric measurements were made irregularly with a freatimeter, but in autumn 2022 we have installed an automatic sensor inside the piezometer on the steep terrain (the first one), and this winter we have observed a slight increase of groundwater level once the cumulated rainfall from September exceeded 800 mm.

- Stream water seems to appear and disappear consistently with the groundwater fluctuations, although, so far, we have too few data to demonstrate it; however, measuring stream water level is much easier than groundwater level in the studied context, and it could be an effective surrogate of groundwater level.

- The use that we do with the synthetic groundwater level data (that could be done with field data, either of groundwater or of stream water level) is just to discriminate between "high" level and "low" level, as a proxy to identify active subsurface drainage conditions.

The colored dots of Figures 5 and 6 also show that the seasonality of the synthetic variables is consistent with that of the observed variables.

<u>Technical corrections</u>

Some general language editing is necessary

We will double-check the English language to remove language, syntax and style mistakes.

Title, and also in the manuscript: “precipitations” cannot be said – maybe replace with “precipitation events” or similar

Based also on similar remarks made by another Reviewer, the title will be changed to “Understanding hydrologic controls of sloping soil response to precipitation through Machine Learning analysis applied to synthetic data”, and the wrong plural “precipitations” will be corrected throughout the entire manuscript.

<u>*References*</u>

Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. Journal of Geophysical Research: Earth Surface, 117(4). https://doi.org/10.1029/2012JF002367

Cascini, L., Sorbino, G., Cuomo, S., & Ferlisi, S. (2014). Seasonal effects of rainfall on the shallow pyroclastic deposits of the Campania region (southern Italy). Landslides, 11(5), 779–792. https://doi.org/10.1007/s10346-013-0395-3

Comegna, L., Damiano, E., Greco, R., Guida, A., Olivares, L., & Picarelli, L. (2016). Field hydrological monitoring of a sloping shallow pyroclastic deposit. Canadian Geotechnical Journal, 53(7), 1125–1137. https://doi.org/10.1139/cgj-2015-0344

Cowpertwait PSP, O’Connell PE, Metcalfe AV, Mawdsley JA (1996) Stochastic point process modelling of rainfall. I Single-site fitting and validation. J Hydrol. https://doi.org/10.1016/S0022-1694(96)80004-7

Damiano, E., Olivares, L., & Picarelli, L. (2012). Steep-slope monitoring in unsaturated pyroclastic soils. Engineering Geology, 137–138, 1–12. https://doi.org/10.1016/j.enggeo.2012.03.002

Greco, R., Comegna, L., Damiano, E., Guida, A., Olivares, L., & Picarelli, L. (2013). Hydrological modelling of a slope covered with shallow pyroclastic deposits from field monitoring data. Hydrology and Earth System Sciences, 17(10), 4001–4013. https://doi.org/10.5194/hess-17-4001-2013

Greco, R., Marino, P., Santonastaso, G. F., & Damiano, E. (2018). Interaction between perched epikarst aquifer and unsaturated soil cover in the initiation of shallow landslides in pyroclastic soils. Water (Switzerland), 10(7). https://doi.org/10.3390/w10070948

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., & Melillo, M. (2020, January 1). Geographical landslide early warning systems. Earth-Science Reviews. Elsevier B.V. https://doi.org/10.1016/j.earscirev.2019.102973

Leonarduzzi, E., Molnar, P., & McArdell, B. W. (2017). Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. Water Resources Research, 53(8), 6612–6625. https://doi.org/10.1002/2017WR021044

Marino, P., Comegna, L., Damiano, E., Olivares, L., & Greco, R. (2020). Monitoring the hydrological balance of a landslide-prone slope covered by pyroclastic deposits over limestone fractured bedrock. Water (Switzerland), 12(12). https://doi.org/10.3390/w12123309

Napolitano, E., Fusco, F., Baum, R. L., Godt, J. W., & De Vita, P. (2016). Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). Landslides, 13(5), 967–983. https://doi.org/10.1007/s10346-015-0647-5

Neyman, J., & Scott, E. L. (1958). Statistical Approach to Problems of Cosmology. Journal of the Royal Statistical Society: Series B (Methodological), 20(1), 1–29. https://doi.org/10.1111/j.2517-6161.1958.tb00272.x

Peres DJ, Cancelliere A (2014) Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. Hydrol Earth Syst Sci 18:4913–4931. https://doi.org/10.5194/hess-18-4913-2014

Peres, D. J., Cancelliere, A., Greco, R., & Bogaard, T. A. (2018). Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. Natural Hazards and Earth System Sciences, 18(2), 633–646. https://doi.org/10.5194/nhess-18-633-2018

Piciullo, L., Tiranti, D., Pecoraro, G., Cepeda, J. M., & Calvello, M. (2020). Standards for the performance assessment of territorial landslide early warning systems. Landslides, 17(11), 2533–2546. https://doi.org/10.1007/s10346-020-01486-4

Rodriguez-Iturbe I, Febres De Power B, Valdes JB (1987) Rectangular pulses point process models for rainfall: analysis of empirical data. J Geophys Res. https://doi.org/10.1029/JD092iD08p09645

Segoni, S., Piciullo, L., & Gariano, S. L. (2018a). A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides. Springer Verlag. https://doi.org/10.1007/s10346-018-0966-4

Segoni, S., Rosi, A., Fanti, R., Gallucci, A., Monni, A., & Casagli, N. (2018b). A regional-scale landslide warning system based on 20 years of operational experience. Water (Switzerland), 10(10). https://doi.org/10.3390/w10101297