Reply to Reviewer #1

On behalf of all co-authors I would like to thank the Reviewer for the time implied in this extensive revision of our manuscript. Responding to every comment individually, you will find your comments in regular style and our response in italic as follows:

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.

<u>General comments</u>

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.

<u>Technical corrections</u>

- 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