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).*

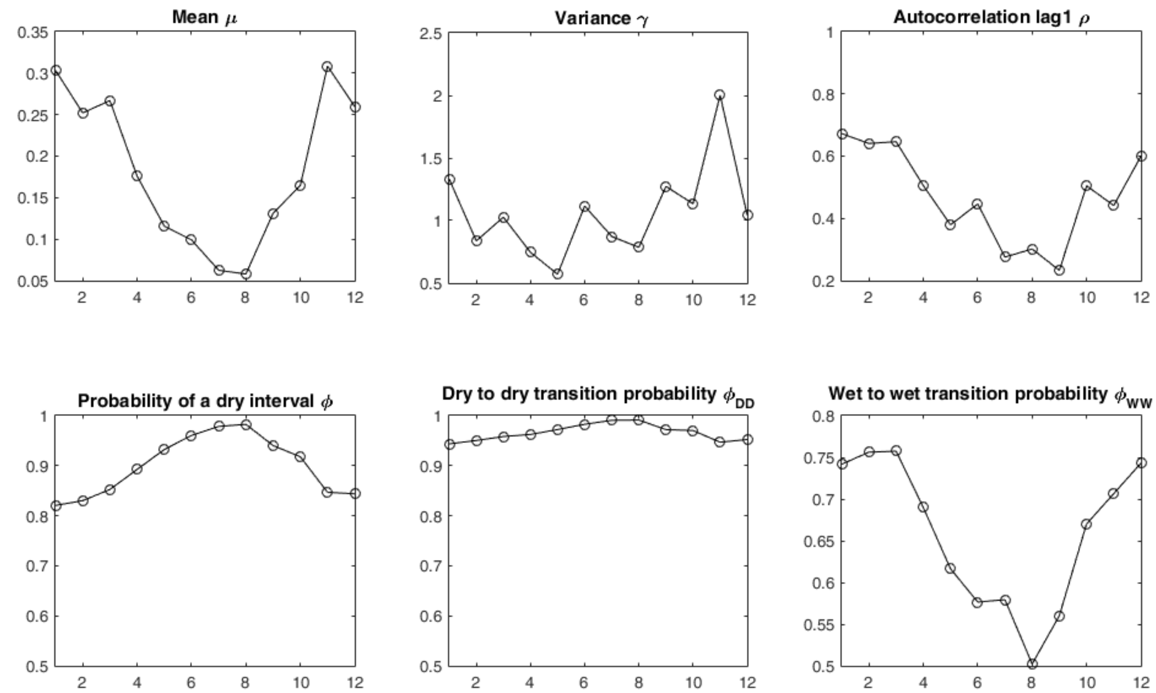

Figure R1. Monthly plot of hourly rainfall characteristics calculated based on the experimental data of the rain gauge of Cervinara.

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.

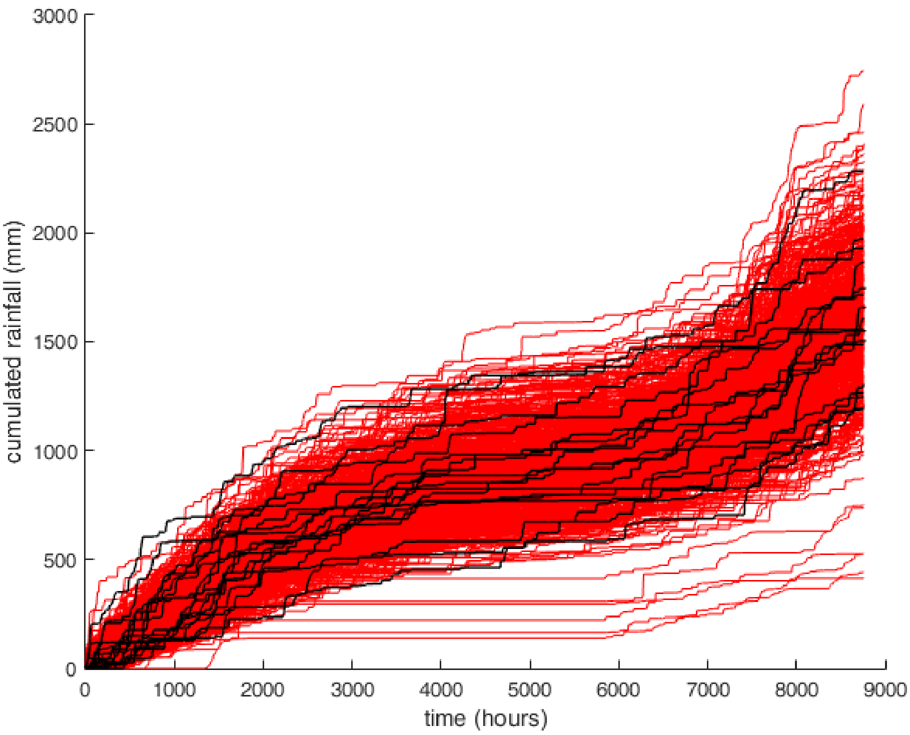

Figure R2. Comparison of observed (black) and simulated (red) cumulated rainfall plots in a year.

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.

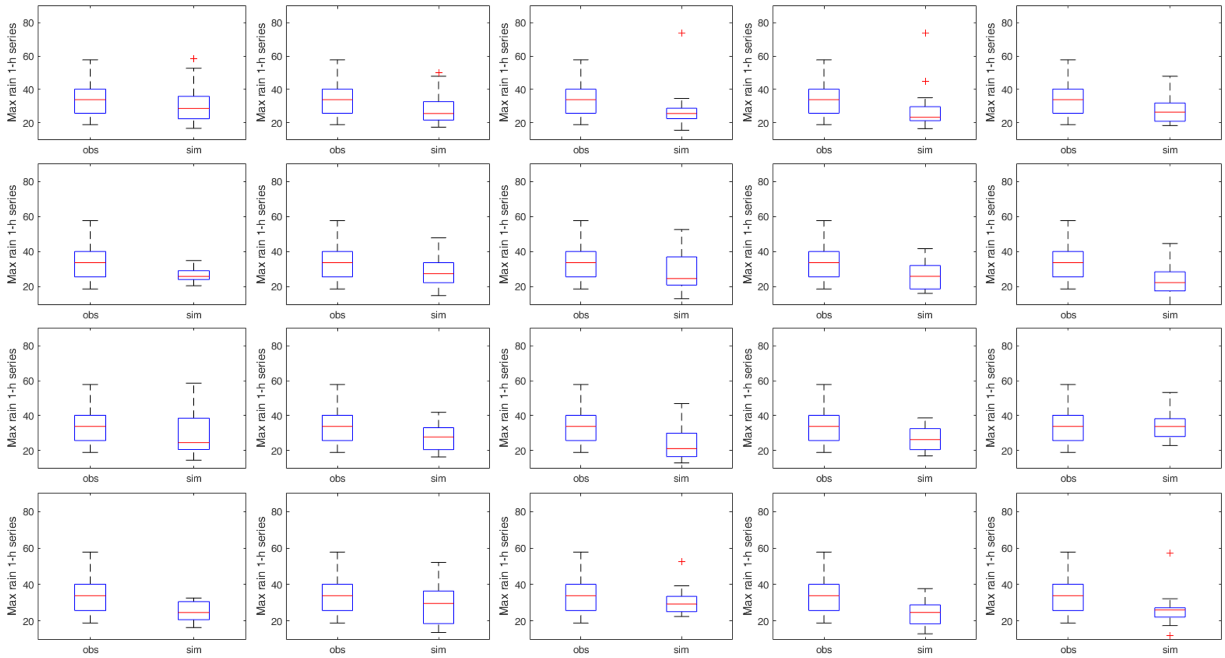

Figure R3. Comparison of observed and simulated distributions (boxplots) of the maximum hourly precipitation in a year, for series of the same length. Each panel shows the distribution for the 17 observed years (boxplot is always the same), and 17 randomly picked simulated years.

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.

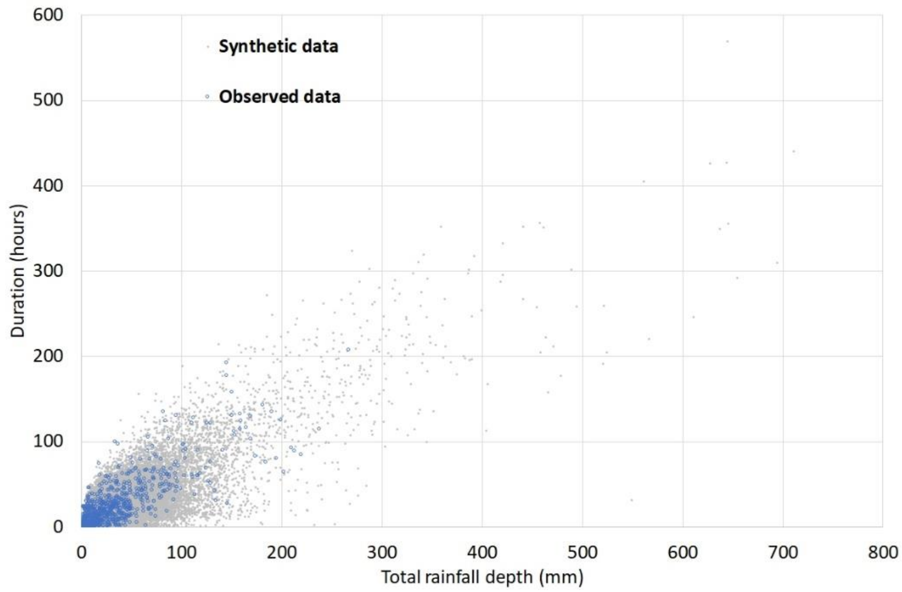

Figure R4. Scatterplot of rainfall event duration vs. total rainfall event depth. The events have been sorted within the rainfall datasets by considering a separation "dry" interval of 24 hours with less than 2 mm rainfall. Blue dots represent events extracted from the 17 years experimental rainfall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

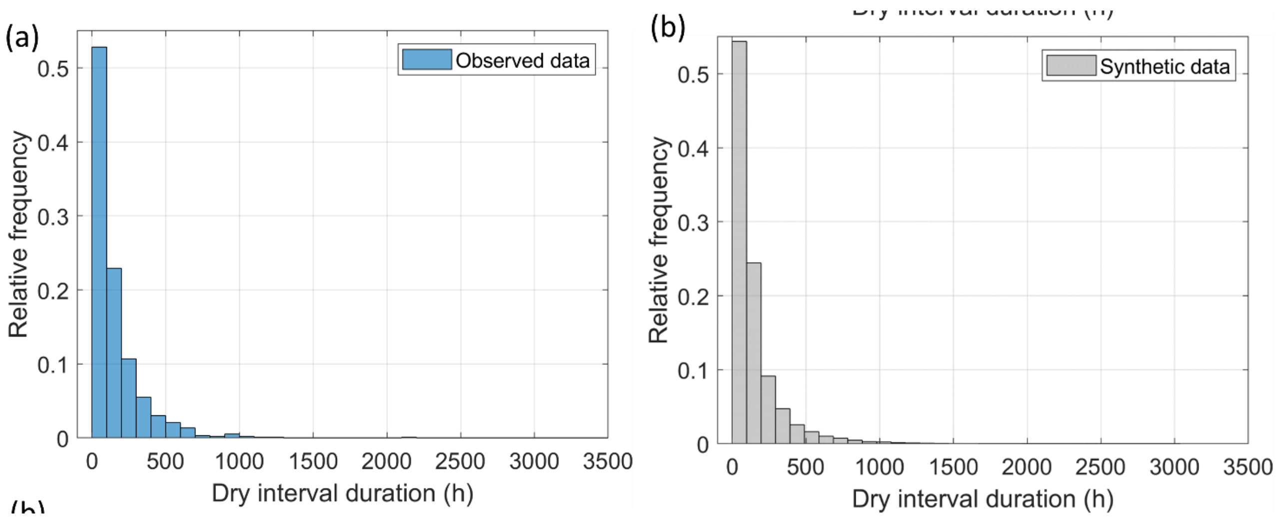

Figure R5. Frequency distributions of dry interval durations for events extracted from the 17 years experimental rainfall dataset (a) and events extracted from the 1000 years synthetic rainfall dataset (b). The events have been sorted within the rainfall datasets by considering a separation "dry" interval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