This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

EGU General Assembly #EGU23 23 - 28 April 2023

Clustering and Random Forest Analysis for the Identification of Hydrological Controls of Slope Response to Rainfall

Daniel Camilo Roman Quintero, Pasquale Marino, Giovanni Francesco Santonastaso, Roberto Greco

Dipartimento di Ingegneria, Università degli Studi della Campania "Luigi Vanvitelli", Italy

danielcamilo.romanquintero@unicampania.it

https://meetingorganizer.copernicus.org/EGU23/EGU23-11154.html

Università degli Studi della Campania Luigi Vanvitelli

Dipartimento di Ingegneria

Methodology

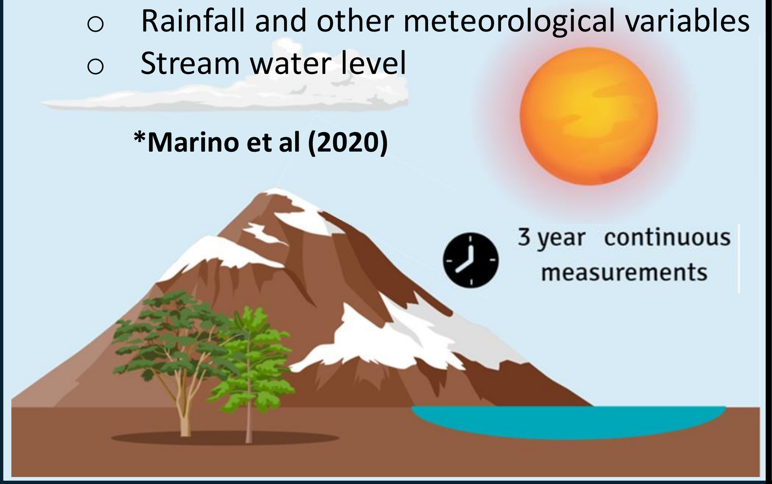
1 Case study

- Cervinara (Campania), Southern Italy
- Landslide prone area covered by loose granular pyroclastic deposits (ashes and pumices) disposed in several layers laying on a fractured limestone bedrock hosting an ephemeral perched aquifer

2 Field monitoring

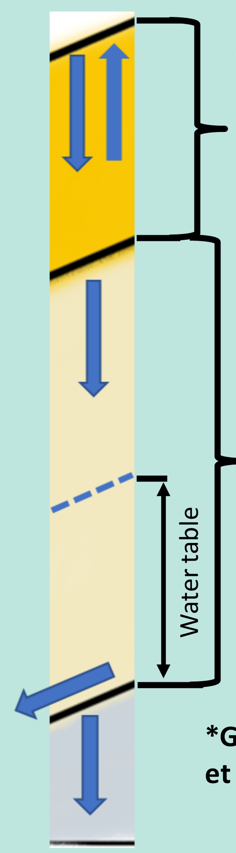
- Carried at the case study since 2017
- Monitored variables
  - Volumetric water content
  - Rainfall and other meteorological variables
  - Stream water level

*Marino et al (2020)



3 Process simulation

*1000-year hourly synthetic rainfall



Soil cover: Water infiltration and evapotranspiration 1D Richards equation

Epikarst: Superficial aquifer – linear reservoir model – Inflow from the **soil cover** and outflow to spring and deeper systems

*Greco et al (2018), Marino et al (2021)

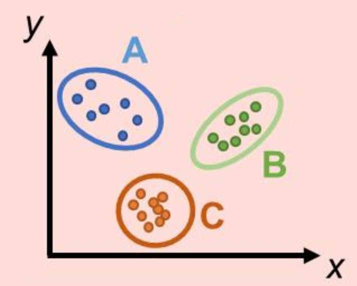
4 Data analysis

Clustering and Machine Learning

*Roman Quintero et al (2023)

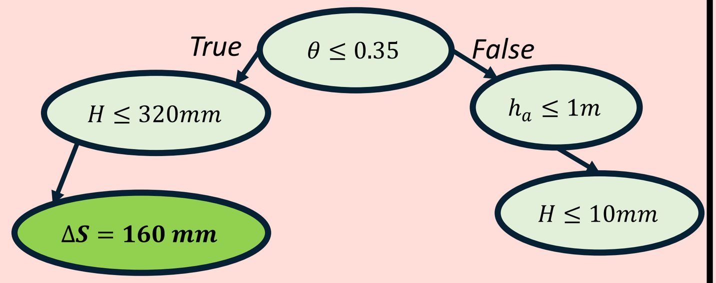
Clustering: k-means clustering

-Optimal number of clusters: Silhouette



Machine learning: Random Forest

-Variable Importance assessment feature



Simulation of field hydrologic processes

Table 1. Hydraulic parameters of the coupled model of the unsaturated soil cover and of the aquifer hosted in the Epikarst (Greco et al. 2018)

Soil cover	Soil cover thickness (m)	2
	Saturated water content (-)	0.75
	Residual water content (-)	0.01
	VG. air entry value index (m^{-1})	6
	VG. shape parameter [n] (-)	1.3
	Saturated hydraulic conductivity (m/s)	$3x10^{-5}$
Epikarst	Epikarst thickness (m)	14
	Effective porosity (-)	0.005
	linear reservoir const (days)	871

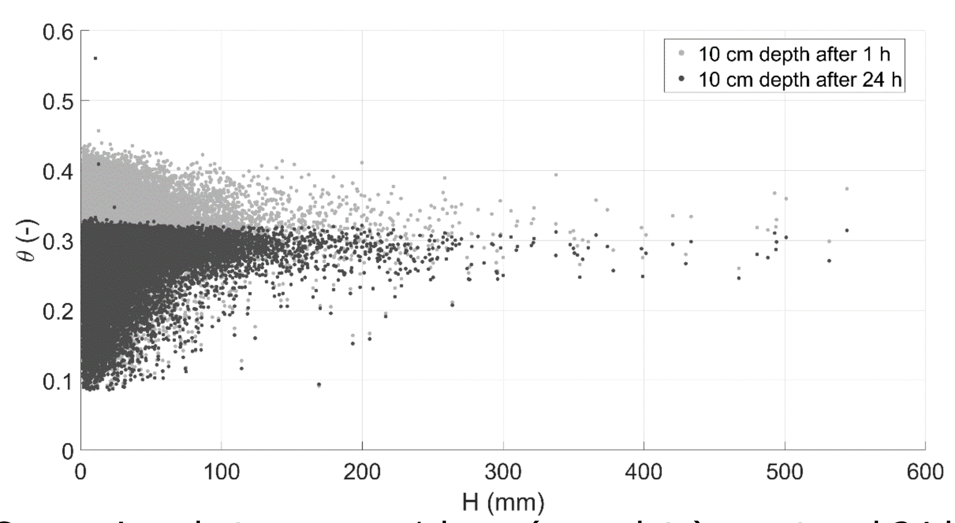


Figure 1. Comparison between θ_{10} 1 hour (grey dots) event and 24 hours (black dots) after the end of any rainfall event according to its total amount H

Sharing is encouraged

The simulations have been done solving a coupled model of Richards 1D equation for the soil cover and a linear reservoir model for the epikarst, fed with a 1000-year synthetic rainfall dataset. The physical parameters are shown in Table 1. Thus, the rainfall events are identified between intervals with more than 2 mm rainfall in 24 h. Raining less than 2 mm in 24h allows the water drain through the uppermost boundary of the soil cover, as seen in Figure 1.

The hydrologic seasonal behavior of the underground antecedent conditions (*i.e.*, prior to the initiation of any rainfall event) is analyzed. Figure 2 shows a comparison between the hydrologic seasonal behavior of the synthetically generated and field monitored data: mean volumetric water content in the soil (θ_{100}), and the epikarst water level (h_a) for the synthetic data, or stream level (h_s) for the field data. The wettest conditions are found typically from December to May, while the driest are seen from August to September, with two intermediate conditions from June to July and October to November.

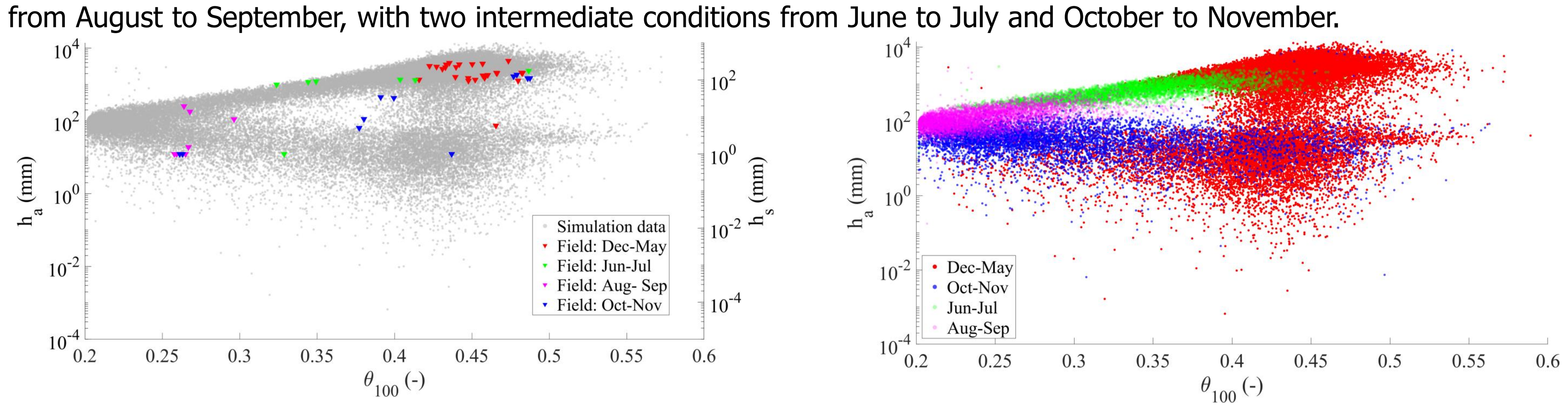


Figure 2. Seasonal behavior of the antecedent underground conditions θ_{100} and h_a (or h_s for field data) for: (a) the field monitored dataset and (b) the synthetic dataset

Slope response to precipitation

Table 2. RMSE and variable importance for H, θ_6, θ_{100} and h_a in the prediction of soil response described as: **(a)** ΔS and **(b)** $\Delta S/H$ evaluated by Random Forest analysis

(a)		Importance			
Dataset	RMSE	H	θ_6	θ_{100}	h_a
$\langle H, \theta_6, h_a \rangle$	5.353	0.963	0.024	-	0.012
$\langle H, \theta_{100}, h_a \rangle$	4.336	0.964	-	0.024	0.010
$\langle H, \theta_6, \theta_{100} \rangle$	4.706	0.962	0.014	0.022	-
$\langle \theta_6, \theta_{100}, h_a \rangle$	24.665	-	0.313	0.340	0.345

(b)		Importance			
Dataset	RMSE	H	θ_6	θ_{100}	h_a
$\langle H, \theta_6, h_a \rangle$	0.213	0.352	0.329	-	0.319
$\langle H, \theta_{100}, h_a \rangle$	**0.197**	**0.293**	**-**	**0.405**	**0.302**
$\langle H, \theta_6, \theta_{100} \rangle$	0.203	0.340	0.261	0.399	-
$\langle \theta_6, \theta_{100}, h_a \rangle$	0.210	-	0.292	0.414	0.293

The data analysis, using the Random Forest, is done based on hydrological variables feasible to be measured in the field: cumulative rainfall event depth (H), mean soil volumetric water content at 6 cm and 1 m depth (θ_6 and θ_{100}) and the ground water level (h_a), both before the rainfall initiation.

The slope response is assessed according to the change in the water stored in the soil cover after a rainfall event ΔS.

The variable importance feature of Random Forest is used here to analyze the best way to assess the slope response and choose the best triplets to be related in order to identify the hydrological controls to slope behavior. Table 2 summarizes the results of the variable importance test.

The Random Forest modelling has been performed assuming the hyperparameters that ensure a stable response: a forest size of 100 trees and a maximum branch splits of 20.

Hydrological behavior of slope

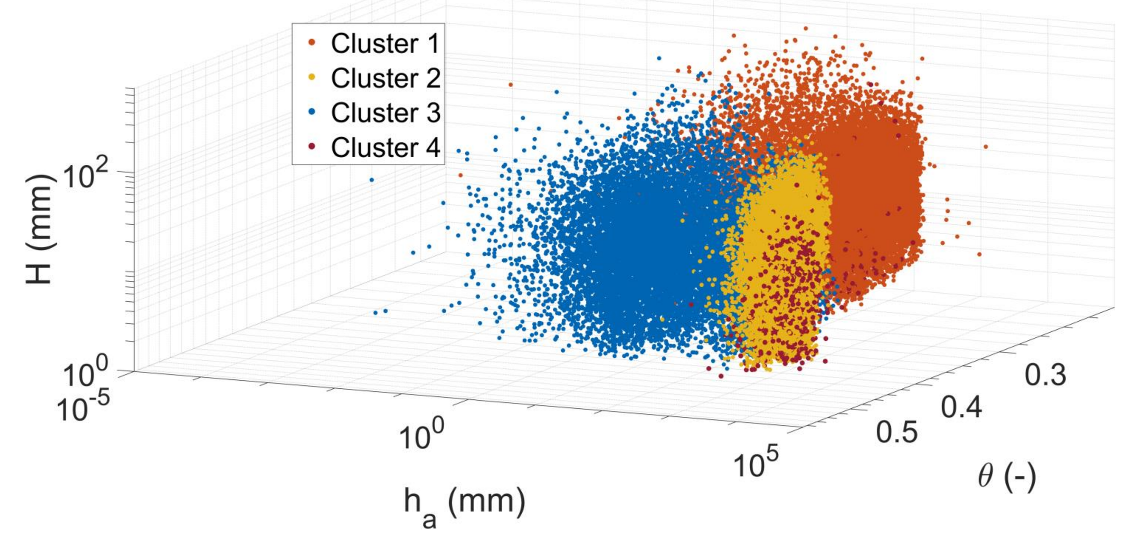


Figure 3. Clusters found within the synthetic data triplets $\langle \theta_{100}, h_a, \Delta S/H \rangle$ represented in the space (θ_{100}, h_a, H)

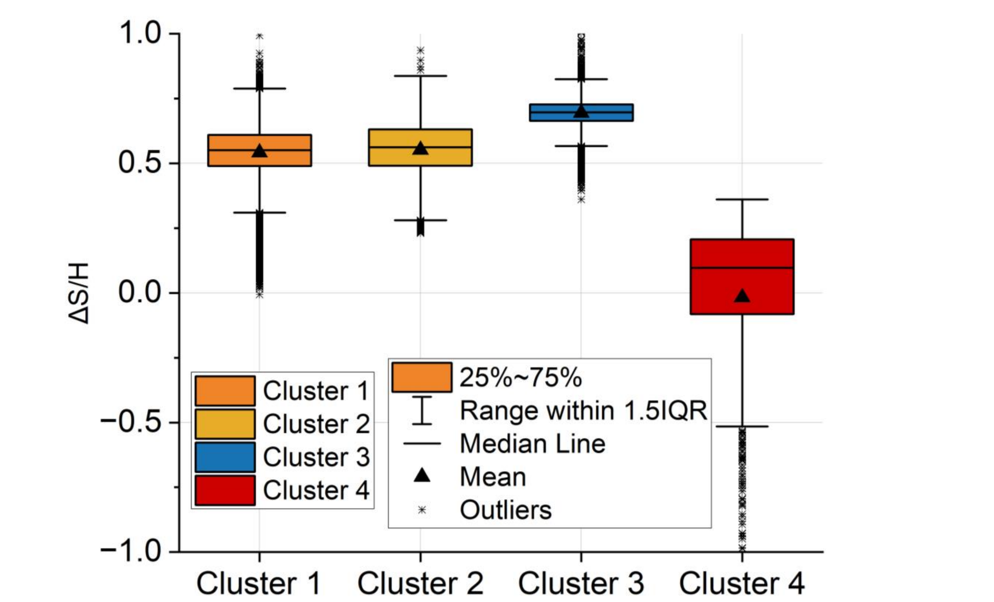


Figure 5. Distribution of the slope response $\Delta S/H$ for the data in each cluster

The slope response is analyzed here as $\Delta S/H$. Its hydrological behavior is explored with the k-means clustering technique among the triplets $\langle \theta_{100}, h_a, \Delta S/H \rangle$ visualized in the log-space (θ_{100}, h_a, H). The optimal number of clusters within the dataset is 4. Figure 3 shows the identified clusters; Figure 4 shows the data distributions of the underground antecedent conditions, for the entire dataset (fig. 4a-b) and for the identified clusters (fig. 4c-h). Finally, Figure 5 shows the distribution of the slope response at each cluster.

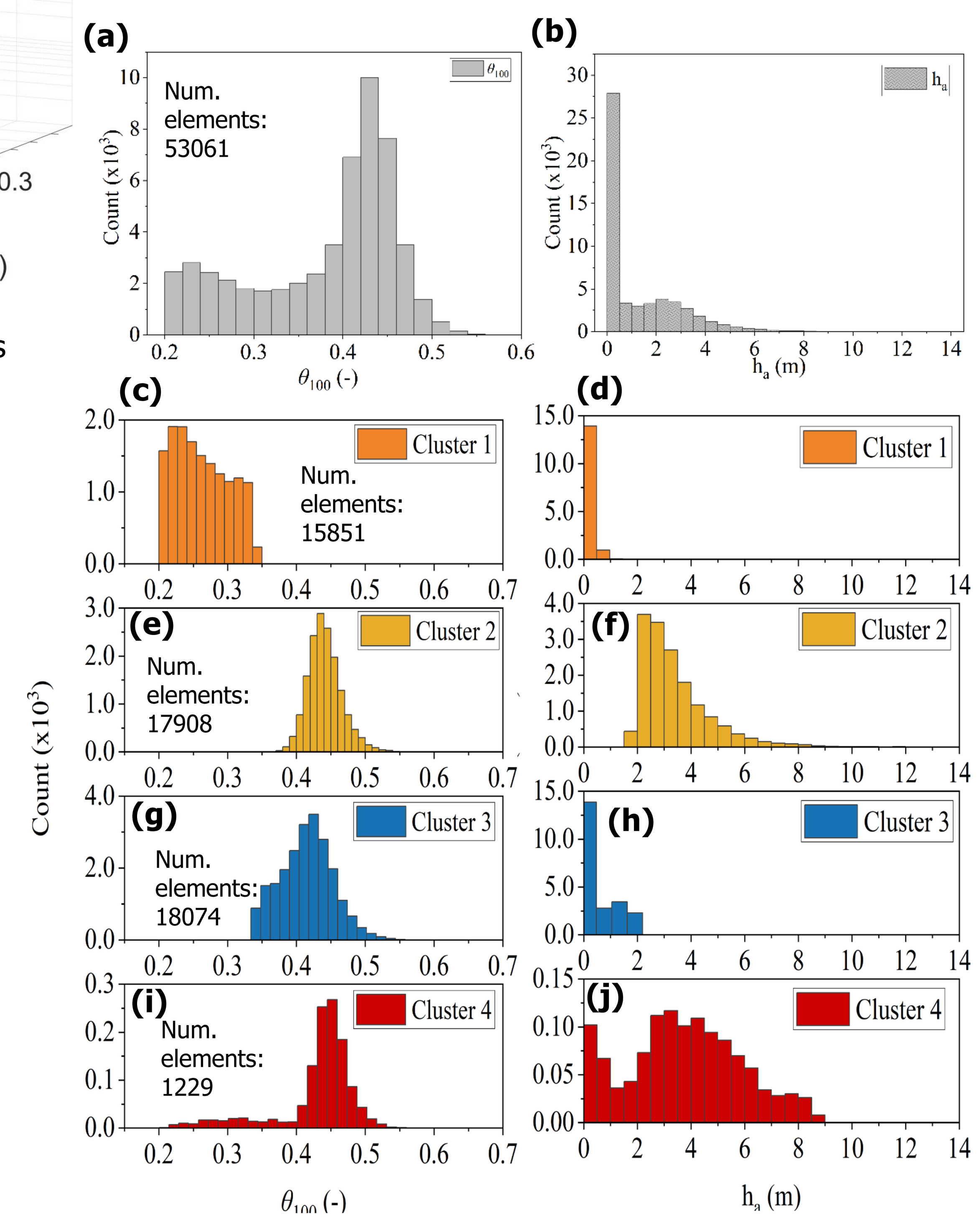


Figure 4. Data distribution of underground antecedent conditions (θ_{100} (left) and h_a (right)) for the complete dataset (a and b), and the identified clusters (c to j) within the triplets $\langle \theta_{100}, h_a, \Delta S/H \rangle$. Modified from: Roman Quintero et al (2023)

Results

The dry antecedent conditions are gathered in cluster 1, described by θ_{100} typically below the field capacity (estimated as $\theta = 0.35$) and low values of h_a. In such cases the slope tends to retain all the rainwater, but evapotranspiration can subtract significant amount of infiltrated water, showing a summer-like behavior.

Inversely, wet soil conditions are found in clusters 2 and 3, with θ_{100} typically above the field capacity:

(i) In cluster 2 wet soil is coupled to high h_a, i.e., conditions normally occurring in late winter and spring. The active drainage lets part of the rainwater drain out of the soil cover to the epikarst, so the slope response is comparable to cluster 1.

(ii) In cluster 3 wet soil is coupled to low h_a, gathering scenarios normally observed in late autumn, when most of the rainwater tends to accumulate in the soil cover due to the impeded drainage through soil-epikarst interface.

Cluster 4 gathers conditions in which the slope drains out much of the rainfall. Such response is normally seen with active drainage conditions, in the transition period from spring to summer.

Nomenclature

H	Total rainfall amount
θ_{100}	Mean volumetric water content at 1 m depth
θ_6	Mean volumetric water content at 6 cm depth
h_a	Epikarst water level
h_s	Stream water level
ΔS	Change in water stored in the soil cover at the end of rainfall

References

1. Greco R, Marino P, Santonastaso GF, Damiano E. Interaction between perched epikarst aquifer and unsaturated soil cover in the initiation of shallow landslides in pyroclastic soils. Water 2018, 10, 948.
2. 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 12(12): 3309.
3. Marino P, Santonastaso GF, Fan X, Greco R (2021). Prediction of shallow landslides in pyroclastic-covered slopes by coupled modeling of unsaturated and saturated groundwater flow. Landslides 18(1): 31-41.
4. Roman Quintero DC, Marino P, Santonastaso GF, Greco R (2023). Understanding hydrologic controls of slope response to precipitations through Machine Learning analysis applied to synthetic data. EGUsphere: 1-41. DOI: 10.5194/EGUSPHERE-2022-1078