https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Spatial dynamic rainfall thresholds for landslide in the Aburrá Valley, Colombia

Oscar Sanchez[1,2] and Edier Aristizábal[1]

[1]Departamento de Geociencias y Medio Ambiente, Universidad Nacional de Colombia,Facultad de Minas, Medellín, Antioquia, Colombia
[2] Early Warning System of the City of Medellín and the Aburrá Valley - SIATA, Área Metropolitana del Valle de Aburrá - AMVA, Medellín, Colombia

Correspondence: Oscar Iván Sánchez Patiño (oisanchezp@unal.edu.co) and Edier Aristizábal Giraldo (evaristizabalg@unal.edu.co)

Abstract. Rainfall-triggered landslides are frequent in the Aburrá Valley, a densely populated Andean valley in Colombia, where warnings must tell where as well as when slopes become critical. We derived rainfall thresholds for each slope unit from 492 dated landslides (2011-2025), four terrain variables, rainfall and the El Niño–Southern Oscillation (ENSO) phase. Non-landslide observations were drawn under three designs that share the landslide's slope unit, its date, or neither, each resampled 100 times. For each design we compared logistic regression, generalized additive models and random forests. A one-day window best described the trigger; for antecedent rainfall, 30 days was chosen among several equivalent windows. Median areas under the receiver operating characteristic (ROC) curve ranged from 0.955 to 0.988. The algorithms differed little within a design, whereas the design changed which predictors carried the signal. Three cut-offs on the ROC curve of the additive model define four alert levels, whose rainfall thresholds change with land cover, slope, soil, ENSO phase and antecedent rainfall. In the larger of two landslide events of 2025 kept out of training, 78 % of the evaluated slope units with landslides reached High or Very high before the report, while 11 % of the valley did.

1 Introduction

Rainfall-triggered landslides cause losses every year in mountain regions around the world, and the toll is heaviest where steep terrain, intense rainfall and dense settlement coincide (Petley, 2012; Froude and Petley, 2018). Colombia is among the most affected countries, and most of its fatalities and damage occur in the tropical Andes (Aristizábal and Sánchez, 2020; Gómez et al., 2023). Of the tools that anticipate these events, rainfall thresholds are the simplest to derive, apply and maintain over large areas, which is why most regional landslide early warning systems in operation use them (Guzzetti et al., 2020). They have been derived empirically from the rainfall recorded before known failures (Caine, 1980; Chleborad et al., 2006; Gariano et al., 2015), from physically based models of slope hydrology and stability (Bogaard and Greco, 2016; Marín et al., 2019), and, more recently, from data-driven and machine-learning models that learn the relation between rainfall, terrain and landslide occurrence from the record (Merghadi et al., 2020; Steger et al., 2022; Moreno et al., 2024).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Physically based models capture the mechanics of failure but need hydrological and geotechnical properties that are rarely available over a whole region, and machine-learning models predict well but need far more data than most inventories contain (Merghadi et al., 2020). Data-driven statistical models sit between the two: with a landslide inventory, a few terrain variables and the rainfall record, they predict well enough to be useful and remain interpretable (Reichenbach et al., 2018; Steger et al., 2024). Most applications, however, still answer two questions separately. Susceptibility maps tell where a landslide may occur, given the terrain, but not when (Reichenbach et al., 2018), and empirical thresholds tell when, given the rainfall, but not where (Guzzetti et al., 2020). The few studies that answer both at once are recent (Schmaltz et al., 2017; Steger et al., 2022; Moreno et al., 2024; Steger et al., 2024, 2026), and they are the ones that can say which slopes are critical on a given day, which is what a warning needs.

The thresholds in use have three further limitations. First, a single threshold for a whole region ignores the terrain within it, so the same rainfall counts as equally dangerous on a forested ridge and on a built-up hillslope; spatially varying thresholds, in which the terrain changes the rainfall needed, address this problem (Crosta, 1998; Steger et al., 2024). Second, intensity-duration thresholds, the most widely used (Caine, 1980; Guzzetti et al., 2020), need the start and end of the triggering storm, which are known only when the inventory records the hour of failure. Thresholds based on accumulated antecedent rainfall absorb part of that uncertainty, are easy to compute from a gauge network and separate a short-term trigger from a long-term preparatory component, as in the margin-of-stability concept (Crozier, 1986; Glade et al., 2005). Third, empirical thresholds rarely use the rainy days on which nothing happened, so how often they would raise false alarms is unknown (Postance et al., 2018; Leonarduzzi and Molnar, 2020).

In Colombia the two questions have also been treated apart. Rainfall thresholds have been derived for several cities and regions (Gómez, 1990; Arango Gartner, 2000; Mayorga Márquez, 2003; Moreno et al., 2006; Gómez et al., 2023) and susceptibility has been mapped for planning (Aristizábal et al., 2019), but few studies couple terrain and rainfall in one model. The Aburrá Valley is an inter-Andean valley of complex geology and geomorphology (Aristizábal, 2004; Aristizábal and Korup, 2025). Its climate is shaped by the Intertropical Convergence Zone, the low-level jets that carry moisture into the Andes and the El Niño–Southern Oscillation (ENSO) (Poveda et al., 2011; Bedoya-Soto et al., 2019; Urrea et al., 2019), and its hillslopes have been occupied rapidly over the past decades (Nieto et al., 2025). Landslides are therefore frequent. Empirical thresholds have been proposed for the valley (Paz and Torres, 1989; Aristizábal et al., 2011), physically based thresholds have been tested in one of its sub-basins (Marín et al., 2019), and susceptibility has been mapped for land-use planning (Aristizábal et al., 2019; Universidad Nacional de Colombia and Área Metropolitana del Valle de Aburrá, 2018), but none of this work joins the where and the when in one tool for early warning.

This study joins the two for the Aburrá Valley, with data-driven models that relate the rainfall of each day to the occurrence or absence of landslides in each slope unit. The models combine the static predisposition of the terrain with the short-term rainfall as the trigger and the long-term rainfall as a proxy for antecedent soil moisture. We first identify the short-term (triggering) and long-term (preparatory) accumulation windows that best separate landslide from non-landslide conditions. We then quantify how the spatio-temporal sampling design of the non-landslide observations conditions the estimated effects and the predictive performance. Finally, we derive spatially varying, dynamic thresholds that can be inspected effect by effect, express them

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

as alert levels for an early warning chain and check them against documented events. The study asks whether, in a tropical mountain valley with a dated inventory and a rain-gauge network, a model that learns from both terrain and rainfall can tell which slopes will be critical on a given day, while raising no more false alarms than a warning system can afford.

2 Data

We first describe the study area and the mapping unit, and then the landslide inventory, which we compiled for this study and which covers 2011 to 2025 (Section 2.1.1). The predictors follow: four static variables of the terrain (slope, curvature, land cover and soil type) and three dynamic variables (short-term rainfall, long-term rainfall and the ENSO phase). All of them were aggregated to slope units.

2.1 Study area and mapping unit

The Aburrá Valley is a narrow valley in the northern part of the Central Cordillera of the Colombian Andes, between $6°00′$ and $6°30′$ N and between $75°15′$ and $75°45′$ W (Fig. 1). We use the political-administrative boundary of the Aburrá Valley Metropolitan Area, which covers ∼1,152 km^2 and is drained from south to north by the Medellín (Aburrá) River. The valley is about 65 km long and deeply incised. Its floor falls from ∼1,800 m a.s.l. in the south to ∼1,000 m a.s.l. in the north, and its flanks rise to high erosional surfaces at ∼2,750 to 3,100 m a.s.l. (Aristizábal, 2004). The bedrock is mostly Paleozoic metamorphic rock (gneisses, schists and amphibolites) intruded by Cretaceous granitoids such as the Antioqueño Batholith. Thick residual soils and slope deposits mantle the hillslopes, and alluvial deposits fill the valley floor (Restrepo and Toussaint, 1988).

Rainfall follows a bimodal cycle driven by the migration of the Intertropical Convergence Zone, with wet seasons in March-May and September-November and drier periods in December-February and June-August (Mejía et al., 1999; Bedoya-Soto et al., 2019). Its amount changes within the valley: in the wettest months it reaches about 250-300 mm in the south and around San Antonio de Prado (Medellín), against 150-200 mm in the north and east (Sánchez Patiño, 2026b). The metropolitan area joins the district of Medellín and nine other municipalities, with ∼4.1 million inhabitants (DANE, 2018). The urban area occupies the valley floor and the lower slopes, and since the second half of the 20th century it has spread onto progressively steeper hillslopes (Nieto et al., 2025).

We used slope units as the mapping unit. A slope unit is a terrain unit bounded by drainage and divide lines, which corresponds to a single hillslope, a set of adjacent hillslopes or a small catchment (Carrara et al., 1991; Alvioli et al., 2016). Grid cells bear no relation to the terrain that fails, so a single hillside can be split into cells with contradictory predictions, and drainage basins are usually too large to isolate the unstable part of a slope; slope units lie between the two (Reichenbach et al., 2018; Chang et al., 2023; Ba et al., 2018). We delineated them from a 5 m DEM with the `r.slopeunits` module (Alvioli et al., 2016, 2020). Of the five parameter settings we compared visually, we kept the one that best balanced detail and computing time: an initial flow-accumulation threshold of 75,000 m^2, a minimum area of 5,000 m^2, a minimum circular variance of aspect of 0.5 and a reduction factor of 10. Units with a mean slope of $10°$ or less were then removed. This takes out the alluvial

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

plain of the Aburrá River, where the reported instabilities relate mostly to river erosion, and remnants of old high plains at the valley edges. Rainfall-triggered landslides are practically absent on such gentle ground. The remaining 61,073 slope units are the unit of analysis (Fig. 1c).

Slope units also suit the inventory. Many records have a position error larger than a grid cell (Section 2.1.1), so an imprecise coordinate that would fall in the wrong cell is much more likely to fall in the right slope unit. Each predictor was aggregated to the slope unit as the mean of its cells for continuous variables and as the most frequent class for categorical ones.

2.1.1 Landslide inventory

The models need landslides with a known date and a reliable location, because each record is matched to the rainfall of a given day and slope unit. We compiled a new inventory from two databases. The first is the catalog of the GEOHAZARDS research group of the Universidad Nacional de Colombia, Medellín (geohazards.com.co), which collects landslides that caused damage, casualties or economic losses, as reported by newspapers, magazines and risk management agencies (Aristizábal et al., 2025). It gives the date of each event but not the time, and locations accurate to about 100 m. The second is the database of the Departamento Administrativo de Gestión del Riesgo de Desastres of Medellín (DAGRD; medellin.gov.co), with 1,318 landslides reported through calls to the 123 emergency line between 2010 and 2024. Each of its records has a date, a time and a location, but the database covers only the district of Medellín.

We screened the records one by one. We discarded those with a doubtful date or an implausible location, such as coordinates outside the valley, on the river bed or on flat terrain, and kept those whose description was precise enough to place them on the terrain and whose date could be confirmed. For part of the events we checked press reports and other online sources to confirm that a landslide had been reported at that place on that date.

For training we kept only the events that can be attributed to rainfall. A record was used if the gauge nearest to it recorded more than 10 mm in the 24 hours before the reported time and more than 50 mm over the previous 30 days, with at least 90 % of the hourly records available in each window. These values are only a filter that confirms it was raining around the time of the event, and they do not define a critical condition. Records attributed to pipe or sewer failures, excavations or collapsed structures were excluded. Because of this filter, no training landslide has less than 10 mm in the previous 24 hours, so the non-landslide observations were restricted to the same rainy conditions (Section 3.2). The models therefore compare landslide and non-landslide conditions within rainy periods.

Two limitations remain. The sources use different names for the same processes, and their descriptions are often too short to tell a shallow slide from a rotational or a small debris flow, so the records could not be classified by type of movement, although different types respond to different controls (Steger et al., 2026). In addition, both databases register events that caused damage or that someone reported, so slopes near houses, roads and infrastructure are over-represented and remote slopes are under-represented (Steger et al., 2022).

The screening left 492 records between April 2011 and June 2025 (Fig. 1c). The 57 records from 1 April 2025 onwards were set aside for the independent check of Section 3.6. Of the remaining 435, 365 met the rainfall condition with a complete gauge record, and these are the landslides used for training. Most records are in Medellín (419, or 85 %), followed by Itagüí (27),

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Girardota (13), Bello (11), Caldas and Copacabana (7 each), La Estrella (5), and Barbosa, Envigado and Sabaneta (1 each). This follows the urban footprint, but also the sources: only Medellín has a systematic record of emergency calls, so the other municipalities appear almost only through events that reached the press or a disaster report.

The records follow the bimodal rainfall cycle. Half of them fall between April and June (249, or 51 %), with 110 in May alone, and a second peak falls between September and November (158, or 32 %), with 71 in November. The drier months are almost empty: February has one record, July three, and the whole December–February period 23.

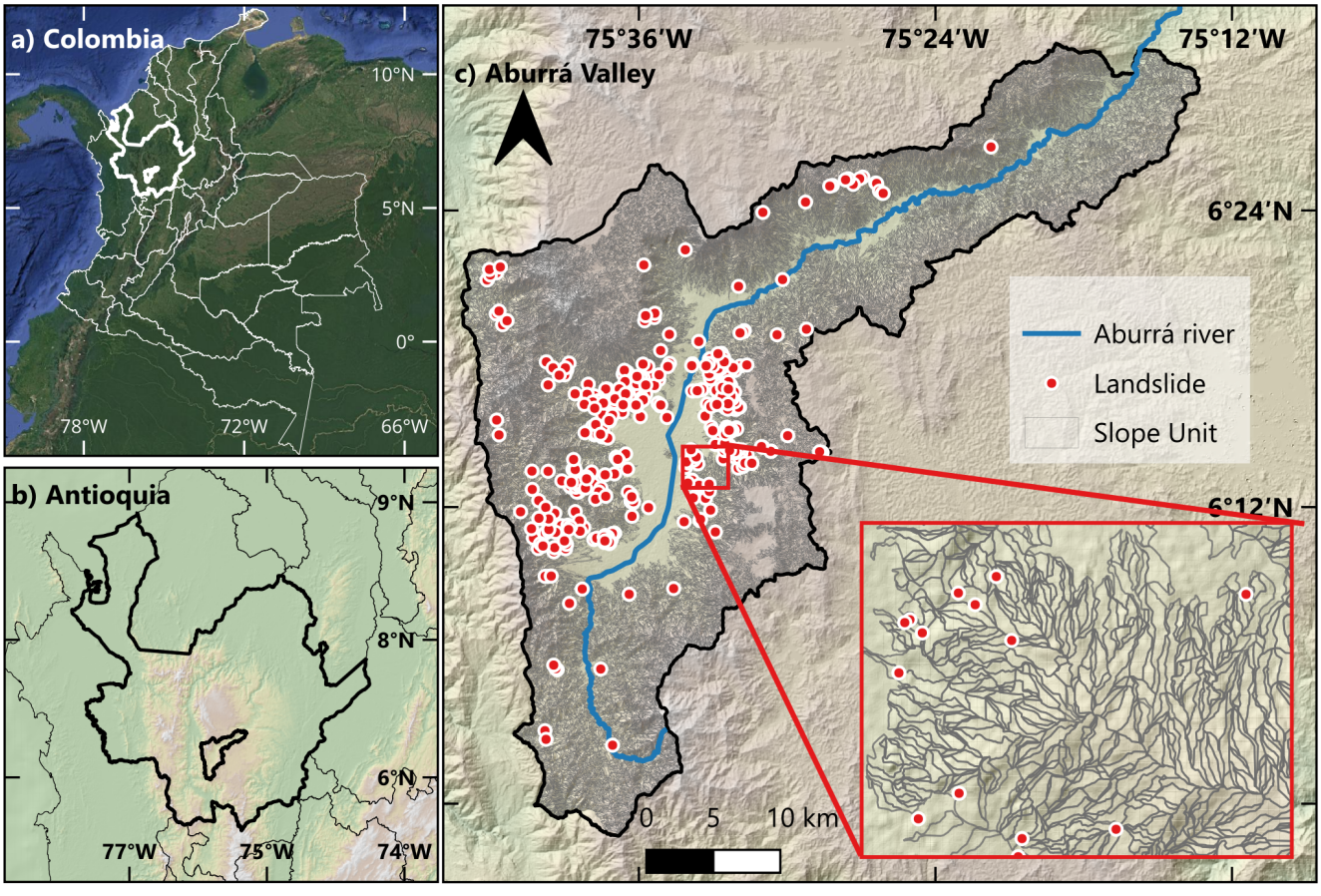


Figure 1. Study area and landslide inventory. (a) Location of the department of Antioquia in Colombia. (b) Location of the Aburrá Valley in Antioquia. (c) The Aburrá Valley, with the boundary of the metropolitan area, the Aburrá River, the slope units (gray) and the 492 landslide records compiled for this study (red dots). Inside the boundary, the areas without slope units are flat terrain excluded from the analysis (mean slope of 10° or less). The red box marks the area enlarged in the inset, where individual slope units can be seen.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Static predisposing variables

How prone a hillslope is to shallow failure depends on several predisposing factors, such as topography, subsurface material, vegetation, hydro-climatic setting and human intervention (Crozier, 1986; Glade et al., 2005; Bogaard and Greco, 2016; Schmaltz et al., 2017). We kept the set of static predictors small, because with the number of landslides available more predictors would add little predictive skill and would make the estimated effects less stable and harder to interpret (Reichenbach et al., 2018). Each slope unit is described by four variables: slope angle, terrain curvature, land cover and soil type. The first two were derived from the 5 m DEM used to delineate the slope units (Section 2.1), and the last two come from thematic maps. Table 1 summarizes their source, unit and aggregation rule, and Fig. 2 shows their spatial distribution.

2.2.1 Slope angle:

The slope is the steepest inclination of the terrain surface with respect to the horizontal, in degrees (Zevenbergen and Thorne, 1987; Moore et al., 1991). We computed it for every cell of the DEM and assigned to each slope unit the mean of its cells. For the map, the units were grouped into the slope classes of the Servicio Geológico Colombiano (Carvajal Perico, 2012). Since units with a mean slope of $10°$ or less had been removed (Section 2.1), five classes remain (Fig. 2a). Most units fall in two of them: $42.1\ \%$ are steep (20-30°) and $29.7\ \%$ very steep (30-45°), whereas extremely steep terrain ($> 45°$) is rare ($0.4\ \%$).

2.2.2 Terrain curvature:

The curvature is the rate of change of the slope. It describes the local shape of the surface and therefore whether surface and subsurface flow converge or diverge (Moore et al., 1991). We used the standard curvature, which combines the profile curvature, measured along the steepest slope, which controls the acceleration of flow, with the planform curvature, measured across it, which controls its convergence. Both were computed by fitting a quadratic surface to a moving 3×3 window of the DEM (Zevenbergen and Thorne, 1987). Negative values indicate concave surfaces (hollows), positive values convex ones (noses), and values near zero planar slopes. Each slope unit took the mean curvature of its cells. For the map, the units were grouped with a threshold of $\pm 0.001\ \mathrm{m}^{-1}$ (Fig. 2b): planar units dominate ($68.9\ \%$), followed by concave ($20.0\ \%$) and convex units ($11.1\ \%$). Slope and curvature enter the models as continuous variables; the classes are used only to make the maps readable.

2.2.3 Land use and land cover:

The land cover map was produced by the remote sensing team of SIATA (siata.gov.co) with the CORINE Land Cover methodology adapted for Colombia (IDEAM, 2010). It was made from a mosaic of Sentinel-2 images (Drusch et al., 2012) taken between 1 December 2024 and 31 January 2025, with a pixel size of 10 m. It has six classes: continuous urban fabric, discontinuous urban fabric, agricultural areas, forests, bare and degraded land, and inland waters. The discontinuous urban fabric mixes scattered housing, green areas and secondary roads, so it is best read as a peri-urban class. Each slope unit took the class covering the largest part of its area (Fig. 2c). Forests dominate $44.3\ \%$ of the units, followed by agricultural areas ($29.5\ \%$), discontinuous urban fabric ($18.8\ \%$), continuous urban fabric ($6.6\ \%$) and bare and degraded land ($0.8\ \%$).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

The map is a single snapshot of 2024–2025, whereas the built-up area grows and vegetation is cleared from one year to the next. The model therefore assumes that this map also describes the earlier part of the record. We accept this simplification here, but an operational version should update the map every year.

2.2.4 Soil type:

The geology of the valley is highly variable, and a lithological map would split the slope units into many classes with few landslides each. We therefore used the 1:10,000 soil map of the Departamento Administrativo de Gestión del Riesgo de Desastres de Antioquia (DAGRAN; dagran.antioquia.gov.co). It records the soil type and its geological origin for the first residual horizon below the organic layer, where most rainfall-triggered landslides in the valley develop. We merged classes with little spatial extent or without landslides in the inventory, to keep the number of categories manageable and because the model cannot learn a class for which it has no data. Four classes remain: silty clay, sandy clay, sandy silt, and blocks, gravel, sand and mud (Fig. 2d). As for land cover, each slope unit took the class covering the largest part of its area. Sandy silt is the most common class (38.7 %), followed by sandy clay (28.3 %), blocks, gravel, sand and mud (19.2 %) and silty clay (13.9 %).

2.3 Dynamic variables

Rainfall data come from the rain-gauge network of the Early Warning System of Medellín and the Aburrá Valley (SIATA; siata.gov.co), which has 142 stations, has operated since 2012 and records rainfall every minute (Ruiz-Carrascal et al., 2026).

2.3.1 Short- and long-term antecedent rainfall:

We described the rainfall before each observation with two accumulations. The short-term antecedent rainfall (STR) is the rainfall of the k periods of 24 hours that end at the observation time, which for a landslide is the hour of its report. The long-term antecedent rainfall (LTR) is the rainfall of the j periods of 24 hours that precede the STR window (Fig. 3). For example, an STR_7 covers the seven days up to the observation, and an LTR_{90} combined with it covers the 90 days that end seven days earlier. For a slope unit i and an observation time t,

$$\mathrm{STR}_{i,t}^{(k)} = \sum_{d=0}^{k-1} R_{i,t-d}, \qquad k \in \{1,\dots,7\}, \tag{1}$$

$$\mathrm{LTR}_{i,t}^{(k,j)} = \sum_{d=k}^{k+j-1} R_{i,t-d}, \qquad j \in \{8,\dots,90\}, \tag{2}$$

where $R_{i,t-d}$ is the rainfall recorded at the gauge nearest to slope unit i during the d-th period of 24 hours before t, and $d = 0$ is the 24 hours that end at t. A window was used only if the gauge had records for at least 90 % of its hours. The two windows are contiguous and do not overlap, so together they cover the antecedent period without gaps. Both variables were computed in the same way for landslide and non-landslide observations.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

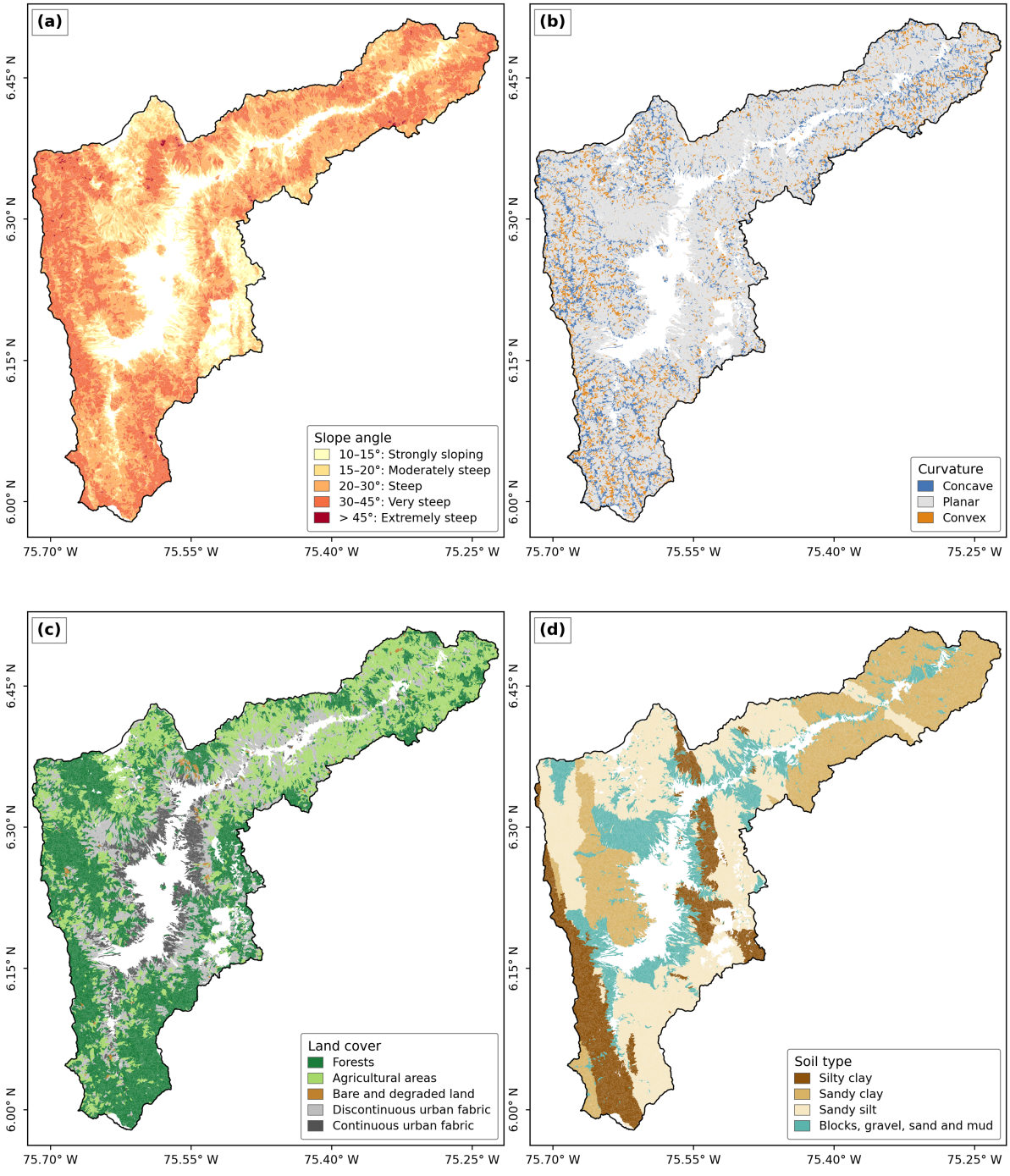


Figure 2. Static predisposing variables of the 61,073 slope units of the Aburrá Valley: (a) mean slope angle, in the classes of the Servicio Geológico Colombiano (Carvajal Perico, 2012); (b) mean curvature, grouped into concave, planar and convex forms with a threshold of ± 0.001 m^{-1}; (c) land use and land cover; and (d) soil type. Continuous variables are the mean of the cells of each slope unit, and categorical variables the class covering the largest part of its area. White areas are the flat terrain excluded from the analysis (mean slope of 10° or less).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

The two variables play different roles. The short-term rainfall is the trigger: it raises pore-water pressure and reduces suction within hours to days. The long-term rainfall is a preparatory condition: it sets how wet the soil is when the trigger arrives, so the same short-term rainfall can be harmless on a dry hillslope and critical on a wet one (Bogaard and Greco, 2016; Steger et al., 2024; Moreno et al., 2024). The lengths k and j were not fixed in advance. We tested all $7 \times 83 = 581$ combinations and selected the pair that best separated landslide from non-landslide conditions (Section 3.3).

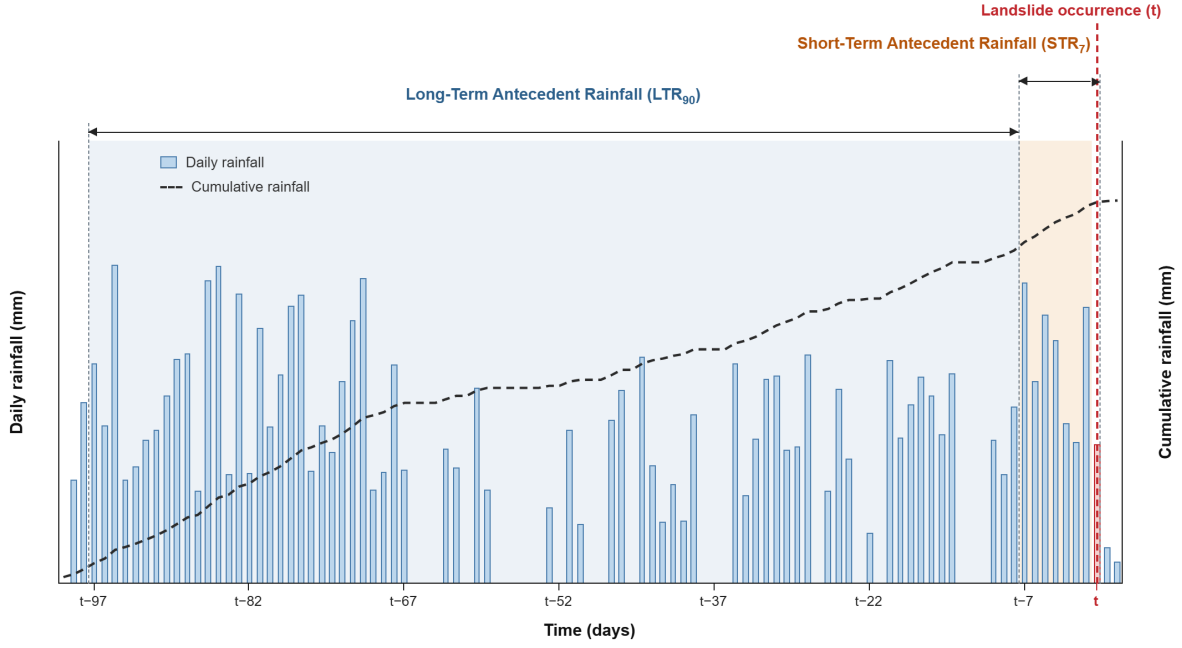


Figure 3. Definition of the two antecedent rainfall variables on a schematic daily rainfall series. The short-term antecedent rainfall (STR, orange) accumulates over the days up to and including the day of the landslide t (red dashed line), here 7 days, and the long-term antecedent rainfall (LTR, blue) over the days that precede the STR window, here 90 days. The two windows are contiguous and do not overlap.

2.3.2 El Niño-Southern Oscillation (ENSO):

ENSO modulates rainfall in the Aburrá Valley at the interannual scale, and we included its phase as the third dynamic variable. ENSO is a fluctuation of the sea surface temperature of the tropical Pacific that alters the atmospheric circulation. In Colombia, El Niño (warm) phases reduce rainfall, mainly in the first rainy season, and La Niña (cold) phases increase it, mainly in the second (Mejía et al., 1999; Pabon et al., 2001; Poveda et al., 2011). This response is largely explained by the Chocó low-level jet, which strengthens during La Niña and carries more Pacific moisture into the Andes (Poveda, 2004). In the Aburrá Valley, La Niña changes the peak rainfall of the wet seasons little, but it raises the rainfall of the dry periods that precede them, and with it the soil moisture before the rains begin (Aristizábal et al., 2026).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

We described the ENSO state with the Oceanic Niño Index (ONI) of the Climate Prediction Center of NOAA (NOAA Climate Prediction Center, 2025), as in Aristizábal et al. (2026). The ONI is the three-month running mean of the ERSSTv5 sea surface temperature anomalies (Huang et al., 2017) in the Niño 3.4 region (5° N–5° S, 170°–120° W), computed against centered 30-year base periods that are updated every five years. Following the NOAA convention, each month was classified as El Niño when ONI $\geq +0.5\,°$C, La Niña when ONI $\leq -0.5\,°$C, and neutral otherwise, and every day took the phase of its month.

The study period contains only a few independent ENSO phases, so the effective sample size behind this variable is much smaller than the number of observations. We include it to test whether the state of the Pacific changes the rainfall needed to trigger a landslide, and we interpret its effect with caution (Section 5.3).

Predictor	Type	Unit	SU aggregation	Temporal scale	Source
Slope angle	continuous	°	mean	static	5 m DEM
Curvature	continuous	m^{-1}	mean	static	5 m DEM
Land cover	categorical	-	dominant class	static	SIATA, Sentinel-2 (10 m)
Soil type	categorical	-	dominant class	static	DAGRAN soil map (1:10,000)
STR (short-term antecedent rainfall)	continuous	mm	nearest gauge	24-h windows	SIATA rain gauges
LTR (long-term antecedent rainfall)	continuous	mm	nearest gauge	24-h windows	SIATA rain gauges
ENSO phase (ONI)	categorical	-	-	monthly	NOAA CPC

Table 1. Static and dynamic predictors used in this study. Continuous terrain variables are the mean of the cells of each slope unit, and categorical ones the class covering the largest part of its area. Each slope unit took the hourly rainfall of the gauge nearest to its centroid, summed over 24-hour windows, so STR and LTR vary in space and time, whereas the ENSO phase is the same for the whole valley and varies only in time. STR and LTR are defined in Eqs. 1 and 2, and the length of their windows was selected as described in Section 3.3.

3 Methods

3.1 Modeling framework

Figure 4 summarizes the framework. We built three training sets that differ in how the non-landslide observations are drawn (Section 3.2), selected the rainfall windows (Section 3.3) and fitted three classification models (Section 3.4). We then evaluated the models (Section 3.5), derived the thresholds from the ROC curve and tested them on two landslide events of 2025 (Section 3.6). The design follows the margin of stability (Crozier, 1986; Glade et al., 2005; Steger et al., 2024): the static conditions set how much stress a hillslope can take, the long-term rainfall reduces that margin by keeping the soil wet, and the short-term rainfall is the stress that finally exceeds it.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

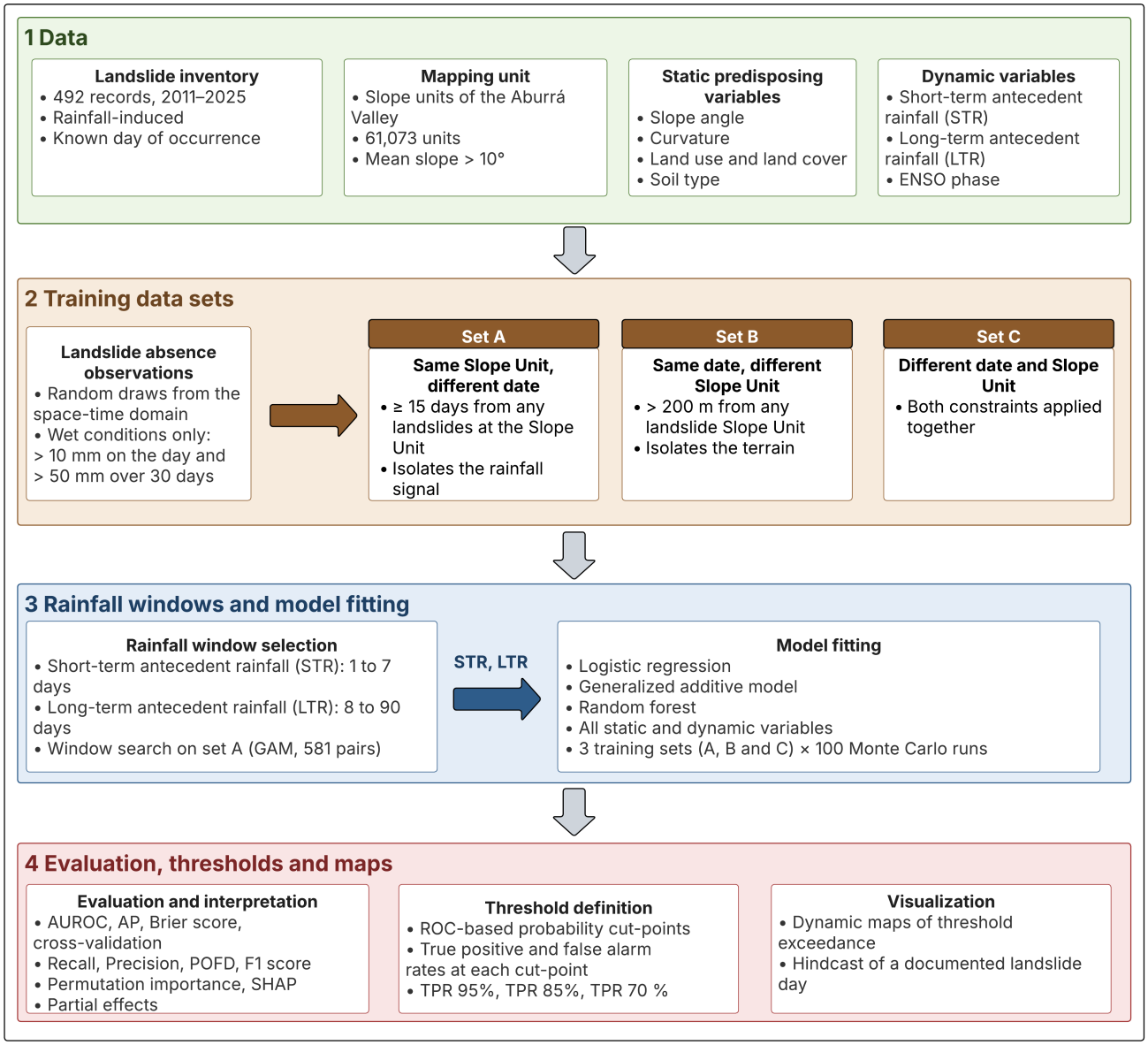


Figure 4. Overview of the modeling framework, organized into data (green), construction of the training data sets (brown), selection of the rainfall windows and model fitting (blue), and evaluation, threshold definition and visualization (red).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.2 Masking and spatio-temporal sampling designs

Non-landslide observations were drawn only from the slope units kept in the analysis, so none comes from the flat terrain removed in Section 2.1; otherwise the models could separate the two classes by telling steep from flat ground. They were also drawn only under rainy conditions, with the filter applied to the inventory: more than 10 mm in the 24 hours before the observation time and more than 50 mm over the previous 30 days. Without this filter most absences would fall on dry days, and the models would only learn to tell rainy days from dry ones. Sampling used only the record before 1 April 2025, which was kept for the independent check (Section 3.6).

We built three training sets, A, B and C, each drawing the absences under a different rule (Table 2). Each set isolates a different part of the signal, so together they show how much of the separation comes from the rainfall, from the terrain, or from both.

- In **set A**, absences are taken in the slope units that had a landslide, on random dates of the whole record that are at least 15 days from any landslide date and meet the rainfall filter. The static variables are then identical in both classes, so the models use only the rainfall and the ENSO phase. Because the contrast is purely meteorological, set A was used to select the rainfall windows (Section 3.3).
- In **set B**, absences are taken on the landslide dates in other slope units, more than 200 m from any unit with a landslide, that meet the rainfall filter at their own location. The local filter matters because rainfall in the valley can be heavy in one sector and light in another on the same day. Both classes then received comparable rain, so the separation rests on the terrain.
- In **set C**, absences differ from the landslides in both slope unit and date, under the same distance and time rules. Terrain and rainfall vary together, so this is the only set in which both can be estimated jointly, and the thresholds are built on it. It is the design used by Steger et al. (2024) and Moreno et al. (2024).

The three sets share four rules: two non-landslide observations per landslide, a ratio far from the real one that keeps the two classes comparable; at least 15 days between two observations of the same slope unit; at least 200 m between an absence and any known landslide; and an 80/20 split into training and test data, stratified by class and grouped by slope unit, so that no slope unit appears in both parts. Each set was resampled 100 times. In every run the absences are drawn again and the models refitted, and we report the median and the interquartile range over the runs, because their spread measures how much a result depends on which absences were drawn.

None of the absences has been checked in the field. They are slope units and dates without a reported landslide, and with an inventory of damage reports some of them are landslides that were never recorded. This lowers the estimated probabilities and raises the thresholds, but it affects the ranking of the observations, and hence the AUROC, much less (Section 5.2).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Set	Slope unit of the absences	Date of the absences	Held fixed	Contrast the set can identify
A	Same as the landslide	Different, more than 15 days from any landslide date, and rainy	Terrain	Rainfall only. Static variables are identical in both classes and are left out of the model
B	Different, more than 200 m away, and rainy	Same as the landslide	Date and rainfall	Terrain only. Both classes received comparable rain on the same day
C	Different, more than 200 m away	Different, more than 15 days from any landslide date, and rainy	Nothing	Terrain and rainfall jointly

Table 2. Sampling designs for the non-landslide observations. The three sets share the terrain mask, the two-to-one ratio of non-landslide to landslide observations and the 80/20 split into training and test data, and each was resampled 100 times.

3.3 Selection of the short- and long-term rainfall windows

We tested all combinations of a short-term window of 1 to 7 days and a long-term window of 8 to 90 days ($7 \times 83 = 581$). For each combination we fitted the additive model on the 100 runs of set A, with the smoothing parameter fixed at $\lambda = 10$ to keep the cost of more than 58,000 fits manageable, and summarized the test AUROC by its median and standard deviation over the runs. Combinations whose median was within one standard error of the best one (the standard deviation divided by $\sqrt{100}$) were treated as equivalent (Hastie et al., 2009), and the choice among them was made on process grounds (Section 5.1). We used set A because its terrain is fixed by construction, so the winning window reflects rainfall alone, and the additive model because it captures non-linear rainfall effects at an acceptable cost. The selected pair was used for all models and sets.

3.4 Classification models

We fitted three models that relax the statistical assumptions step by step: a logistic regression (parametric), a generalized additive model (semi-parametric) and a random forest (non-parametric), the three families most used in the landslide literature (Reichenbach et al., 2018; Moreno et al., 2024; Merghadi et al., 2020). Comparing them shows what flexibility adds to the predictions and what it costs in interpretability. All three model the probability $\pi_{i,t}$ of a landslide in slope unit i at time t from the short- and long-term rainfall ($\mathrm{STR}_{i,t}$ and $\mathrm{LTR}_{i,t}$), the mean slope s_i, the mean curvature κ_i, the soil type c_i, the land cover u_i and the ENSO phase e_t.

The **logistic regression** assumes that every predictor acts linearly on the logit,

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$\log\left(\frac{\pi_{i,t}}{1-\pi_{i,t}}\right) = \beta_0 + \beta_1 \mathrm{STR}_{i,t} + \beta_2 \mathrm{LTR}_{i,t} + \beta_3 s_i + \beta_4 \kappa_i + \sum_l \gamma_l \, d_{l,i,t}, \tag{3}$$

where d_l are the indicator variables of the categorical predictors. The reference levels are forest for land cover and sandy silt for soil type, the most frequent classes in the training data, and the neutral phase for ENSO. The continuous predictors were standardized, so $\exp(\beta_j)$ is the odds ratio for an increase of one standard deviation, and the coefficients were estimated with an L_2 penalty ($C = 1$).

The **generalized additive model** replaces the linear terms of the continuous predictors with smooth functions (Hastie and Tibshirani, 1986; Wood, 2025),

$$\log\left(\frac{\pi_{i,t}}{1-\pi_{i,t}}\right) = \beta_0 + f_1(\mathrm{STR}_{i,t}) + f_2(\mathrm{LTR}_{i,t}) + f_3(s_i) + f_4(\kappa_i) + \gamma_{c_i} + \gamma_{u_i} + \gamma_{e_t}, \tag{4}$$

where each f_j is a penalized B-spline with six basis functions (Eilers and Marx, 1996) and the γ terms are the effects of the soil, land cover and ENSO classes. A common smoothing parameter λ controls how much the curves can bend. It was chosen among five values between 1 and 10^3, equally spaced on a logarithmic scale, by the AUROC on a 70/30 split of the training part, so the test data never enter the choice. The model was fitted with `pyGAM`.

The **random forest** (Breiman, 2001) averages the class proportions of 500 classification trees, each grown on a bootstrap sample with the Gini criterion and $\lfloor\sqrt{p}\rfloor$ predictors drawn at each split ($p = 16$ with the static variables expanded into indicators, and $p = 5$ in set A). Trees had no depth limit, but every leaf had to hold at least five observations. No model used class weights, because they would change the meaning of the predicted probabilities and of the thresholds read from them. The logistic regression and the random forest were fitted with `scikit-learn`, and every random step used a fixed seed.

3.5 Model evaluation and interpretation

All the metrics were computed on the test part of each run and summarized by the median and the interquartile range over the 100 runs. Three of them do not depend on a probability cut-off. The area under the ROC curve (AUROC) is the probability that a random landslide receives a higher score than a random non-landslide observation, so 0.5 means no skill and 1 a perfect separation. The average precision (AP), the area under the precision-recall curve, measures how many alerts would be real; with one landslide for every two absences, its no-skill value is 1/3. The Brier score, the mean squared difference between the predicted probability and the observed outcome, measures calibration, which matters because the thresholds are cut-offs on those probabilities. At a reference cut-off of 0.5 we also report the recall, TP/(TP+FN), the precision, TP/(TP+FP), the false positive rate (FPR), FP/(FP+TN), and the F_1 score, the harmonic mean of precision and recall, where TP, FN, FP and TN are the counts of true positives, false negatives, false positives and true negatives. These depend on the cut-off and only describe the models; the operational cut-offs are derived in Section 3.6.

The Monte Carlo split keeps each slope unit on one side, but every month, terrain class and area of the valley appears on both sides. Since the landslides cluster in a few storms, in the rainy seasons and in the urban classes, a model could score well

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

partly because the test data resemble the training data. We therefore added a cross-validation with five ways of forming the folds, similar to the random, temporal and spatial partitions of Steger et al. (2024): (i) random, with five folds stratified by class and grouped by slope unit, as in the Monte Carlo split; (ii) month, with one calendar month of all the years per fold, so all the observations of a date stay together; (iii) soil type and (iv) land cover, with one class per fold; and (v) rainfall zones, with one of the five homogeneous rainfall zones of the valley per fold (North, East, South, West and San Antonio de Prado, SAP). The zones were delineated from the hourly records of 70 SIATA gauges (2013–2024) by Ward clustering with a spatial contiguity constraint, without using the landslide inventory, and each one is a contiguous part of the valley (Sánchez Patiño, 2026b). Each fold was held out in turn and the three models were refitted on the others with the settings above, including the choice of λ, whereas the rainfall windows were kept fixed. A held-out soil or land cover class is new to the model, which is a form of blocking in predictor space (Roberts et al., 2017); we predicted it with every class seen in training and averaged the probabilities, weighted by the frequency of each class. In set A, which has no static predictors, these two schemes only group the slope units. The cross-validation was run on 20 of the 100 data sets of each training set. Its main measure is the out-of-fold AUROC, computed once from all the held-out predictions of a data set, because it requires ranking observations from different folds against each other, as an alert level does across the valley and the season. The AUROC of each fold shows where the performance falls; it is undefined for a fold with a single class and unstable with fewer than 20 landslides or 20 absences. The cost of each scheme is its paired difference from the random scheme on the same data set.

To see what the models learned, we computed the permutation importance of each predictor (Breiman, 2001), the loss of test AUROC when the predictor is shuffled, averaged over $K = 50$ shuffles,

$$I_j = \mathrm{AUROC} - \frac{1}{K}\sum_{k=1}^{K} \mathrm{AUROC}_j^{(k)}, \tag{5}$$

where $\mathrm{AUROC}_j^{(k)}$ is the value after the k-th shuffle of predictor j. It needs only the predictions, so it applies in the same way to the three models. For the random forest we added SHAP values (Lundberg and Lee, 2017; Lundberg et al., 2020), which split each prediction into the contributions of the predictors, in units of probability, and were computed with the exact algorithm for tree ensembles. The shape of each effect is shown by the smooth terms of the additive model with their 95 % confidence bands, by the odds ratios of the logistic regression, and by the SHAP dependence plots of the random forest. The additive terms are on the logit scale, where a difference of 0.7 roughly doubles the odds. These relationships are shown for a representative run of set C, the one whose additive-model AUROC is closest to the median of the 100 runs, following Steger et al. (2024).

3.6 Threshold definition and dynamic maps

We use the word threshold in two linked senses. The first is a probability cut-off p^* above which an alert is raised. The second is the rainfall that reaches it: for a given slope unit, the combinations of short- and long-term rainfall whose predicted probability exceeds p^*. Because the probability depends on the terrain, this rainfall threshold changes from one slope unit to the next, which is what we call a spatial dynamic threshold (Steger et al., 2024).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

The cut-offs were derived from the additive model of set C. Set C is the only one that estimates terrain and rainfall jointly, and the additive model gives smooth response surfaces whose every term can be inspected. In each run we read from the test ROC curve the first cut-off that reached a target true positive rate (TPR) of 95, 85 and 70 %, and we used the median of each cut-off over the 100 runs. The three cut-offs define the four alert levels of Table 3; the most permissive one is the analogue of the 5 % non-exceedance threshold of empirical rainfall thresholds. Choosing a cut-off weighs missed events against false alarms, so we identify each one by its TPR. Its probability value depends on the two-to-one sampling ratio and should not be transferred to a sample with a different composition without recalibration (Steger et al., 2024). To express the cut-offs in rainfall, we predicted the probability over a grid of short- and long-term rainfall for slope units of contrasting terrain and drew the cut-offs as contour lines, each of which reads as a classical rainfall threshold.

Alert level	Cut-off exceeded	Meaning
Low	none	Below the most permissive cut-off. Fewer than 5 % of the observed landslides occurred under these conditions
Medium	TPR 95 %	Conditions under which landslides have occurred, but which are also common on days without them
High	TPR 85 %	Fewer false alarms, and still most of the observed landslides
Very high	TPR 70 %	Exceptional conditions: false alarms are rare, and seven of every ten observed landslides occurred at or above this level

Table 3. Alert levels of the dynamic maps, defined by the three cut-offs read from the ROC curve. The most permissive cut-off (TPR 95 %) gives the lowest level of alert and the most restrictive one (TPR 70 %) the highest.

The dynamic maps apply the additive model of the representative run, hour by hour, to every slope unit of the valley. A slope unit is classified only when the rainfall at its nearest gauge passes the filter of Section 3.2, because the model was trained under rainy conditions; otherwise it stays at Low. It is not evaluated, and is shown in gray, when its gauge has less than 90 % of the hourly records in a rainfall window or when one of its static predictors is missing or belongs to a class absent from the training data. We mapped two events of the 2025 rainy season with reported emergencies, from a few hours before the rain until after it, and assembled the sequences into animations, available as a video supplement (Sánchez Patiño, 2026a). The maps use the recorded rainfall, so they are a hindcast and not a forecast.

The two events served as an independent check of the maps, counted by slope unit. A slope unit with a landslide is a hit at a given level if it reached that level at any hour between 24 hours before the first report and its own report, and a miss otherwise. A slope unit without landslides that reached the level before the last report is a false alarm. For each level we computed the hit rate, the false positive rate among the evaluated slope units without landslides, and the share of all the slope units of the valley that were alerted. The lead time of a hit is the number of hours between its first hour at the level and the report; since landslides are reported after they occur, it is an upper bound of the warning time. A good result puts most landslides in the upper levels while those levels cover a small part of the valley. The inventory holds only reported landslides, so some false alarms may be landslides that were never reported.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

4 Results

4.1 Sampling designs and rainfall window selection

After the filters of Section 2.1.1, 365 landslides remained for training. Each one was matched with two non-landslide observations wherever the sampling rules allowed it, so every run holds between 1,086 and 1,097 observations, about one third of them landslides.

Figure 5 shows the search for the rainfall windows on set A. Every combination reaches a median AUROC above 0.93, and the short-term window dominates: the best median falls from 0.973 at one day to 0.959 at two days and to 0.941–0.946 at five to seven days. In a paired comparison over the same runs, the best one-day combination beats the best two-day combination in 95 % of the runs. The one-day window is also the most stable, with a standard deviation between runs of about 0.010, against about 0.016 at five and six days. Along the one-day row (Fig. 5c), the median AUROC varies only between 0.965 and 0.973 from 8 to 90 days, less than the standard deviation of any single combination, and the interquartile range of the runs covers the whole curve.

The best combination was a one-day window with a 17-day long-term window, with a median AUROC of 0.973. By the one-standard-error rule, 14 combinations are equivalent to it. All of them use the one-day window, and their long-term windows form two groups, from 15 to 21 days and from 29 to 35 days. We adopted the one-day and 30-day windows, from the second group. In a paired comparison, their AUROC differs from the optimum by a median of -0.0003 and equals or beats it in 47 % of the runs. The choice among equivalent windows is discussed in Section 5.1.

4.2 Model performance

Table 4 and Fig. 6 summarize the performance of the three models. The median AUROC ranges from 0.955 to 0.988 and the median AP from 0.915 to 0.979. All three models score lowest in set B, higher in set A and highest in set C, and the AP and the Brier score rank the sets in the same order. The sets pose different classification problems (Section 3.2), so these values do not rank the designs by quality.

Within a set the models differ little. Their medians fall between 0.972 and 0.979 in set A, 0.955 and 0.965 in set B, and 0.985 and 0.988 in set C, while the interquartile range of any single model spans 0.007 to 0.021. The three models rise and fall together from one run to the next, and the runs show no trend (Fig. 7), so we also compared them within each run (Fig. 8). The additive model beats the logistic regression in 83, 84 and 89 % of the runs of sets A, B and C, by a median of 0.001 to 0.004 in AUROC, and the random forest beats the additive model in 78, 86 and 72 % of the runs, by a median of 0.005, 0.008 and 0.002. At the fixed cut-off of 0.5 the models place their probabilities differently: in set C the additive model reaches a recall of 0.918 against 0.897 for the random forest, with a false positive rate of 0.034 against 0.027.

Figure 9 shows the out-of-fold AUROC of the cross-validation, and Table A1 in Appendix A gives its medians. The random scheme reproduces the Monte Carlo results, with medians at most 0.003 below those of Table 4. In paired terms, holding out a month lowered the AUROC by 0.008 or less in every set and model, and holding out a soil type by 0.004 or less in sets A and C and by 0.007 to 0.012 in set B. In set A no scheme lowered it by more than 0.008. Holding out a rainfall zone lowered it by

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

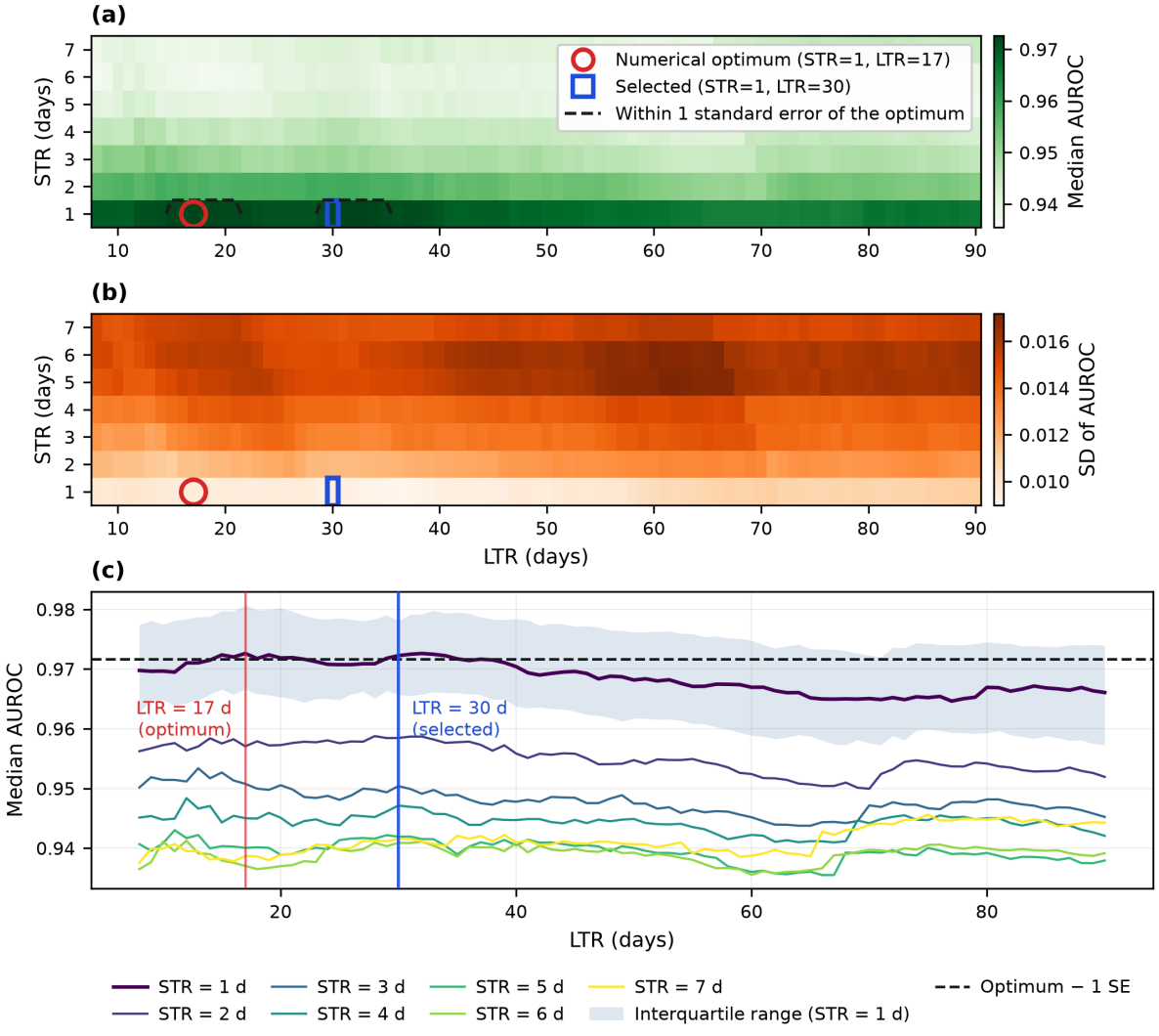


Figure 5. Selection of the rainfall accumulation windows. (a) Median AUROC over the 100 Monte Carlo runs of training set A, for every combination of a short-term window (STR, 1–7 days) and a long-term window (LTR, 8–90 days). The dashed line encloses the combinations whose median lies within one standard error of the best; the circle marks the numerical optimum and the square the combination adopted in this study. (b) Standard deviation of the AUROC between runs, which sets the scale of the sampling noise. (c) Median AUROC as a function of the long-term window, one line per short-term window, with the interquartile range of the 100 runs shaded for STR = 1 day.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Set	Model	AUROC	AP	Brier	F_1	Recall	Precision	FPR
A	LR	0.972 [0.965–0.978]	0.940 [0.922–0.954]	0.062 [0.053–0.070]	0.863	0.849	0.890	0.055
	GAM	0.973 [0.967–0.979]	0.941 [0.923–0.958]	0.060 [0.051–0.068]	0.873	0.863	0.890	0.055
	RF	0.979 [0.975–0.983]	0.957 [0.943–0.967]	0.049 [0.043–0.056]	0.908	0.918	0.903	0.048
B	LR	0.955 [0.943–0.963]	0.915 [0.896–0.935]	0.073 [0.066–0.082]	0.851	0.836	0.858	0.065
	GAM	0.958 [0.948–0.968]	0.926 [0.900–0.942]	0.071 [0.062–0.078]	0.859	0.849	0.868	0.062
	RF	0.965 [0.957–0.975]	0.941 [0.931–0.955]	0.069 [0.061–0.074]	0.863	0.836	0.901	0.045
C	LR	0.985 [0.978–0.989]	0.972 [0.961–0.981]	0.043 [0.036–0.051]	0.915	0.904	0.928	0.034
	GAM	0.986 [0.981–0.990]	0.975 [0.964–0.983]	0.040 [0.034–0.048]	0.921	0.918	0.929	0.034
	RF	0.988 [0.985–0.992]	0.979 [0.974–0.985]	0.042 [0.037–0.047]	0.917	0.897	0.942	0.027

Table 4. Predictive performance of the three models under the three training sets. Values are medians over the 100 Monte Carlo runs, with the interquartile range in brackets for the three metrics that do not depend on the probability cut-off. Lower is better for the Brier score and for FPR; higher is better for the rest.

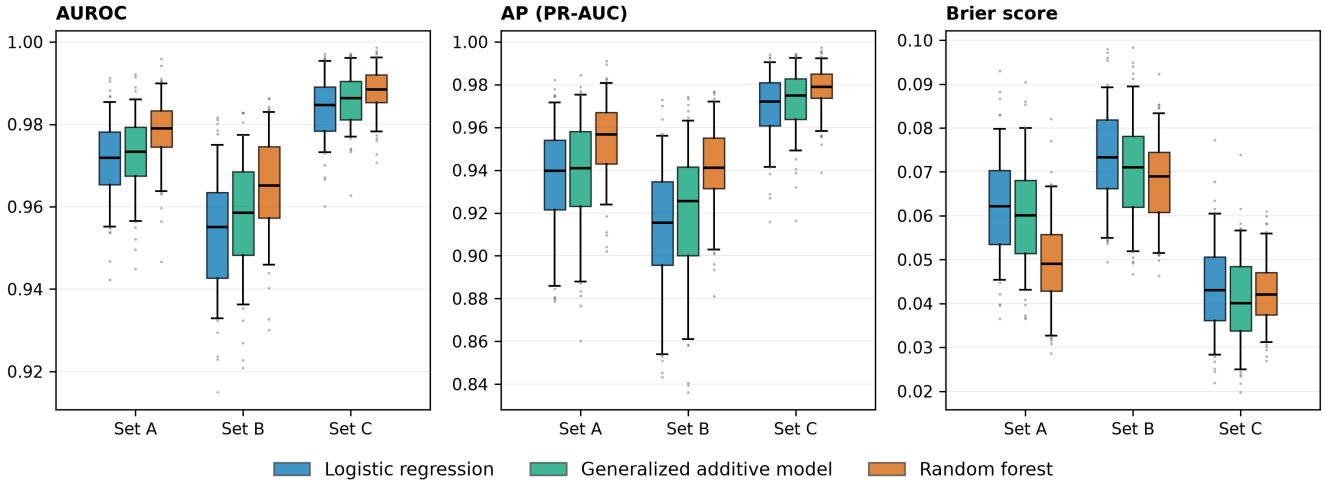


Figure 6. Predictive performance of the three models under the three training sets, over the Monte Carlo runs. (a) AUROC, (b) area under the precision-recall curve (AP), (c) Brier score. Boxes span the interquartile range, whiskers the 5th and 95th percentiles, and the line inside each box is the median. Higher is better in (a) and (b), lower is better in (c).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

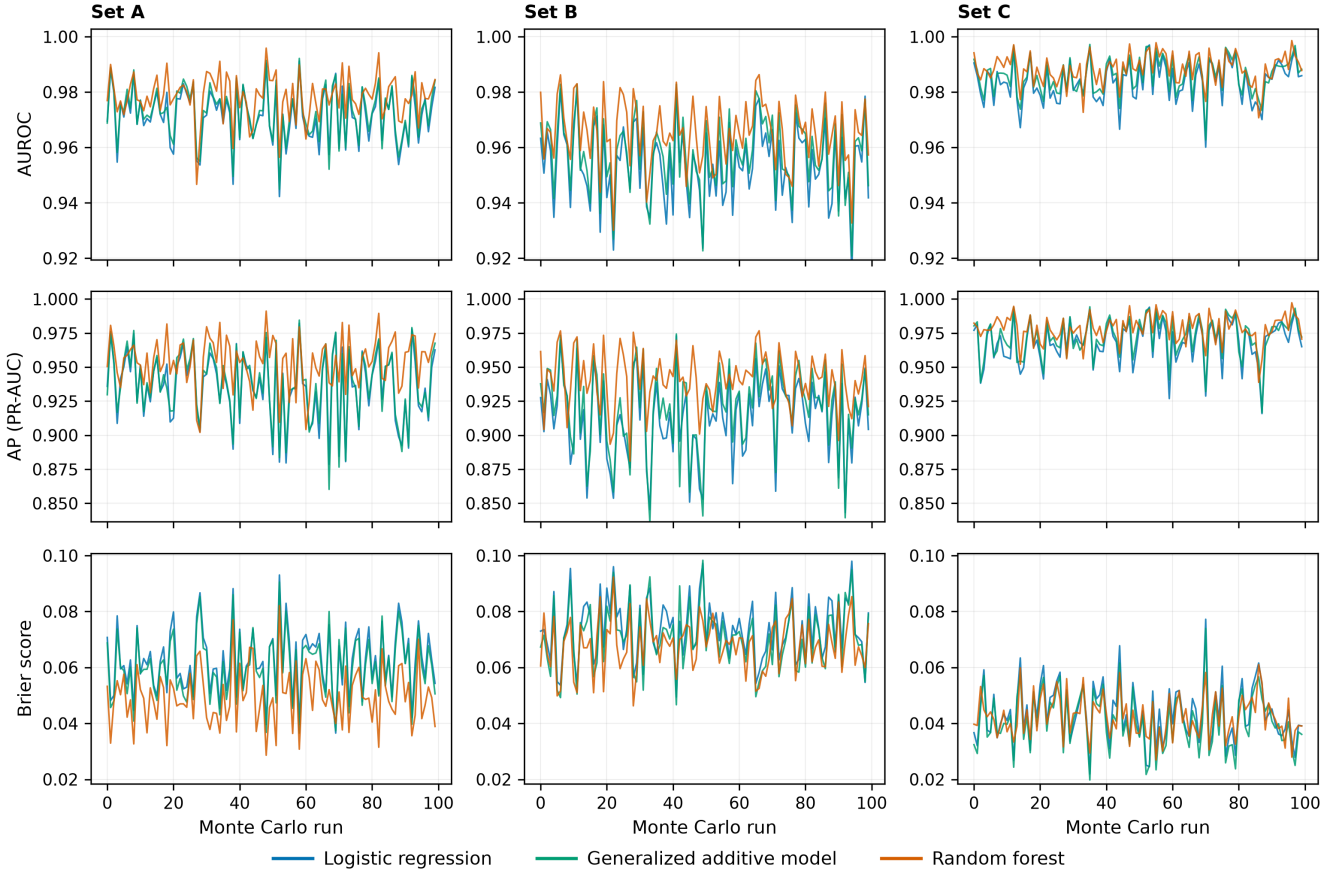


Figure 7. The same three metrics run by run. Rows are metrics and columns are training sets.

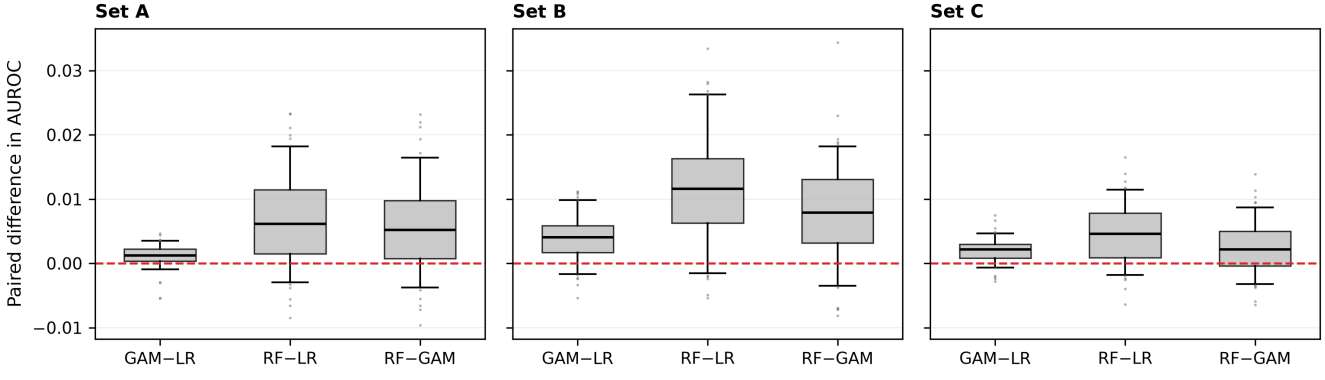


Figure 8. Paired difference in AUROC between models, computed within each Monte Carlo run so that the three models are compared on exactly the same data. The dashed line marks no difference; a box lying entirely above it means that the first model of the pair is consistently better, however small the difference. LR: Logistic Regression, GAM: Generalized Additive Model, RF: Random Forest.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

0.016 to 0.018 in set C and by 0.032 to 0.038 in set B. Holding out a land cover class lowered it by 0.073 to 0.091 in set C, to medians between 0.889 and 0.912, and by 0.197 to 0.256 in set B, to medians between 0.698 and 0.768; under this scheme the random forest lost the least and the logistic regression the most.

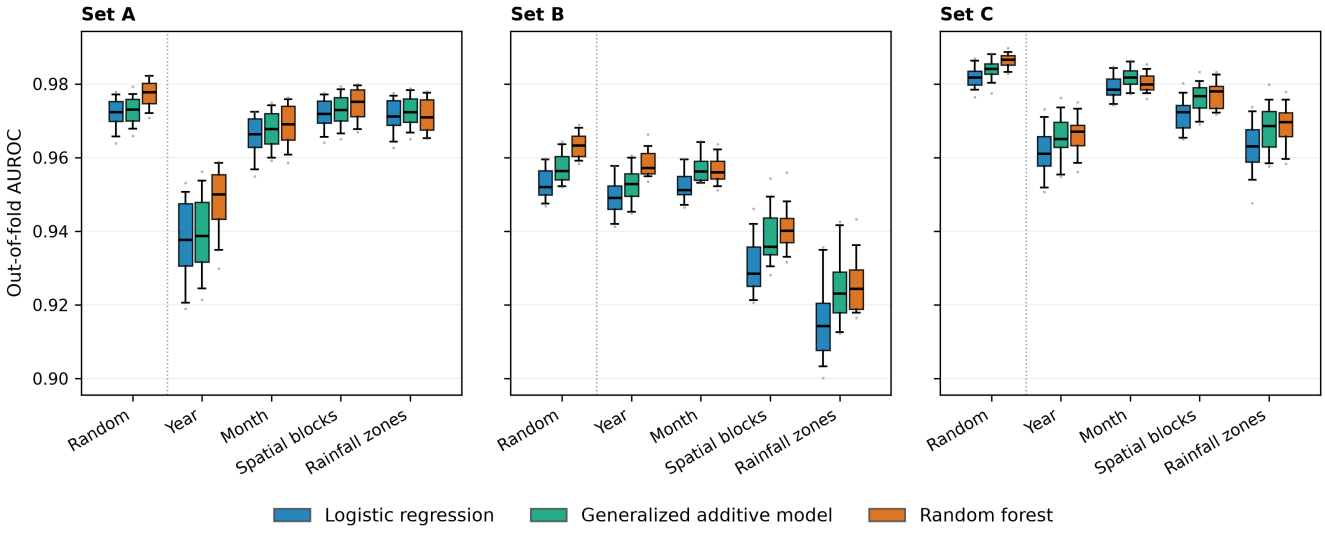


Figure 9. Out-of-fold AUROC of the three models under the five cross-validation schemes, one panel per training set. Each box summarizes the 20 Monte Carlo data sets: the box spans the interquartile range, the whiskers the 5th and 95th percentiles, and the line inside the box is the median. The dotted line separates the random scheme, which is the reference, from the four schemes that hold out a month, a soil type, a land cover class or a rainfall zone. The vertical axis is the same in the three panels.

For the additive model of set C, Fig. A1 gives the AUROC of each held-out fold, against 0.984 for the random scheme. The lowest months were November (0.967, with 61 landslides) and May (0.969, with 74). Within the four land cover classes with enough landslides the AUROC stayed between 0.958 and 0.973, against an out-of-fold value of 0.905 for that scheme, and within the soil types between 0.975 and 0.992. Among the rainfall zones it ranged from 0.964 in the West to 0.988 in the East, except for San Antonio de Prado (0.919), which holds only 8 landslides.

4.3 Variable importance and modeled relationships

Figure 10 shows the permutation importance of the predictors. In set A the short-term rainfall dominates: permuting it costs 0.294 to 0.295 of AUROC, depending on the model, against 0.045 to 0.049 for the long-term rainfall and 0.008 to 0.015 for the ENSO phase, so without it the models fall from about 0.97 to about 0.68. In set B the land cover and the short-term rainfall are about equally important (0.145 to 0.184 and 0.161 to 0.178); the land cover comes first for the logistic regression and the additive model and second for the random forest. In set C the short-term rainfall is first again (0.143 to 0.167) and the land cover second (0.052 to 0.072). The long-term rainfall and the ENSO phase fall to almost zero in set B and keep a small weight in set C (0.012 to 0.014 and 0.006 to 0.008). The slope angle reaches at most 0.007, the soil type 0.006 and the curvature 0.001.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

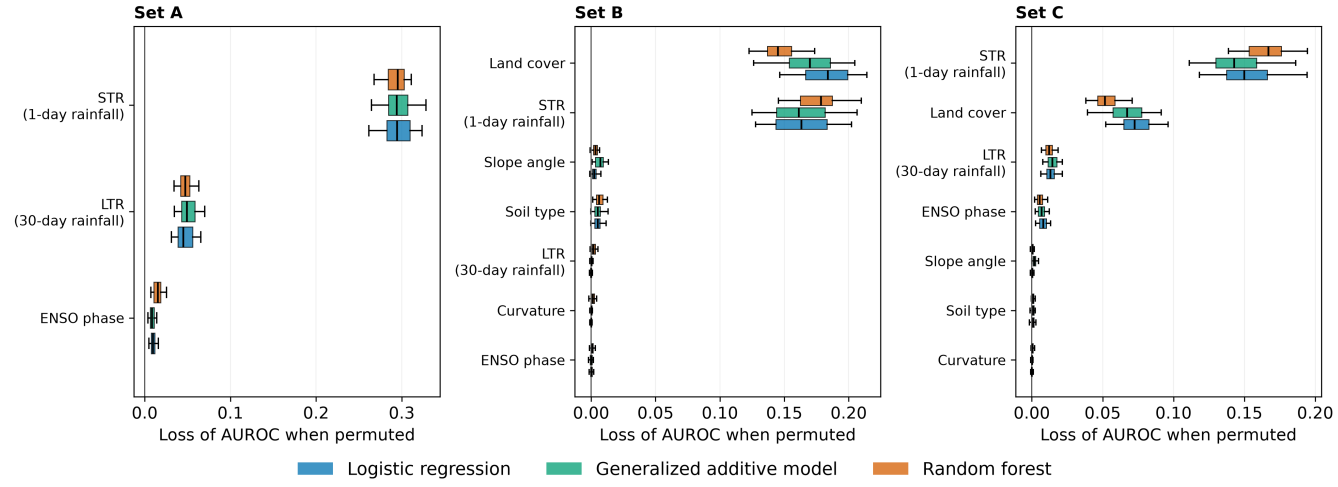


Figure 10. Permutation importance of the predictors under the three training sets, over the Monte Carlo runs. Boxes show the loss of AUROC when the predictor is shuffled in the test part (Eq. 5); they span the interquartile range and the whiskers the 5th and 95th percentiles. Note that each panel has its own horizontal scale.

The partial effects of the additive model show the shape of each relationship (Fig. 11). The short-term rainfall raises the log-odds almost linearly over its whole range, although the observations become sparse above about 100 mm (Fig. 11a). The long-term rainfall also raises them, at a decreasing rate, with an amplitude of about 5.9 in log-odds against about 26 for the short-term rainfall (Fig. 11b); in set A, which has no static predictors, its amplitude reaches about 7.8 (Fig. A2b). The slope effect rises by about 1.2 in log-odds from 10° to a maximum near 30° and falls on steeper ground (Fig. 11c), whereas the curvature is almost flat, with a wide band (Fig. 11d). The land cover shows the largest contrasts, from $+2.7$ for continuous urban fabric to -1.9 for forest and -1.2 for agricultural areas (Fig. 11e). The soil effects are small, and only sandy silt, on the low side, has an interval that excludes zero (Fig. 11f). The ENSO phase has a clear effect, with La Niña at $+0.93$ and El Niño at -0.65, and neither interval crosses zero (Fig. 11g).

The random forest ranks the predictors in the same way (Fig. 12): short-term rainfall first, followed by continuous urban fabric, long-term rainfall, forest and La Niña. High short-term rainfall, continuous urban fabric and La Niña push its predictions up, and forest and agricultural land push them down. The contribution of the short-term rainfall is close to a step (Fig. 13a): in units of probability, its mean is -0.19 between 10 and 20 mm and -0.15 between 20 and 30 mm, jumps to $+0.19$ between 30 and 40 mm, and then rises slowly, from $+0.27$ between 40 and 50 mm to $+0.39$ above 100 mm. The long-term rainfall spans 0.28, from -0.17 to $+0.11$ (Fig. 13b), against 0.73 for the short-term rainfall. In the logistic regression (Fig. A3), the continuous urban fabric multiplies the odds of a landslide by 48 and the discontinuous urban fabric by 5 relative to forest, while agricultural land stays close to forest (0.93). One standard deviation multiplies the odds by 18 for the short-term rainfall and by 2.6 for the long-term rainfall. Relative to the neutral phase, La Niña multiplies them by 3.2 and El Niño by 0.71, while the slope angle and the curvature stay within 15 % of one.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

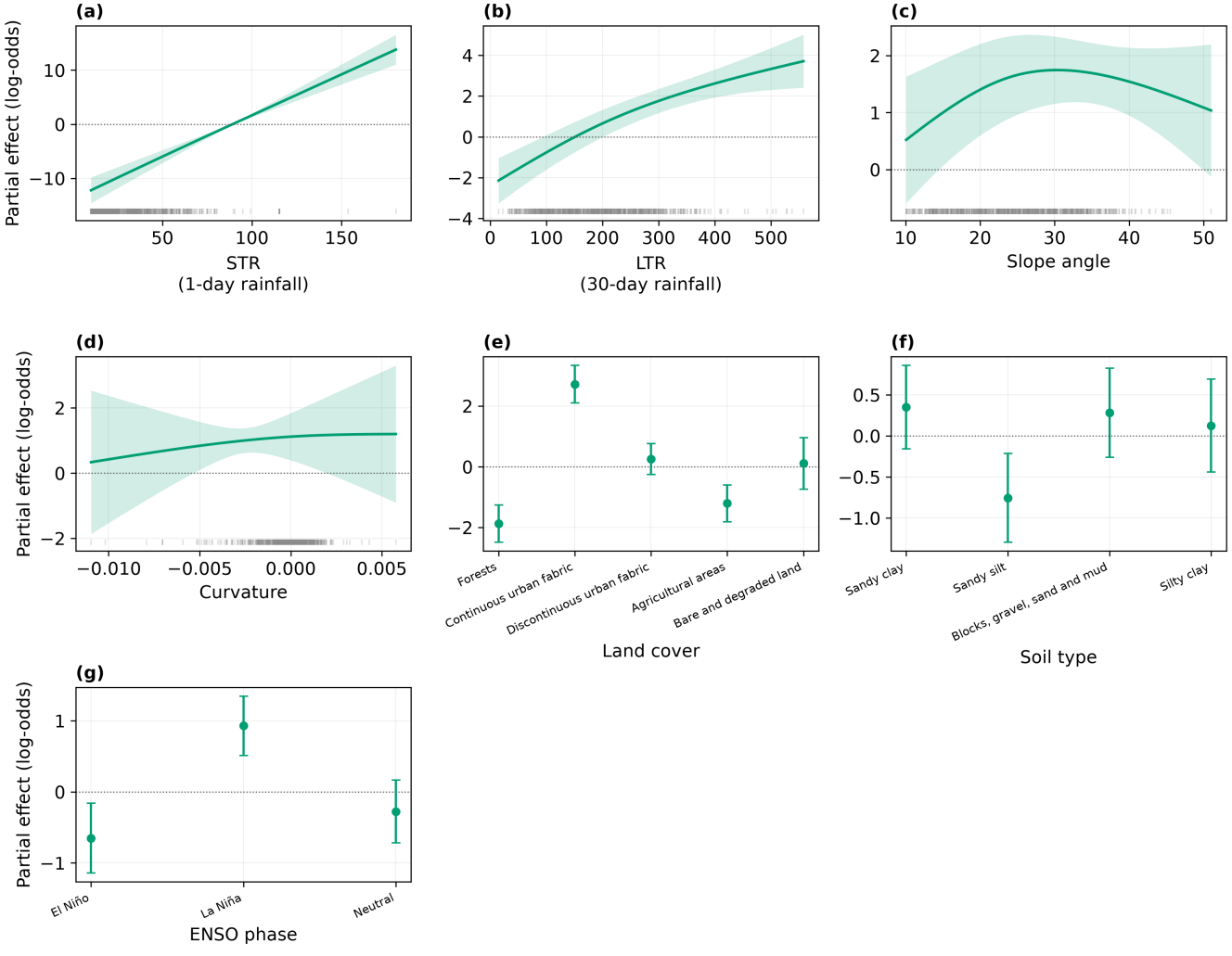


Figure 11. Partial effects of the generalized additive model fitted on the representative run of training set C, on the logit scale and centered on the average conditions. Shaded bands are 95 % confidence intervals and the ticks along the horizontal axis mark the observed values. For the categorical predictors the points are the estimated contribution of each class with its interval.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

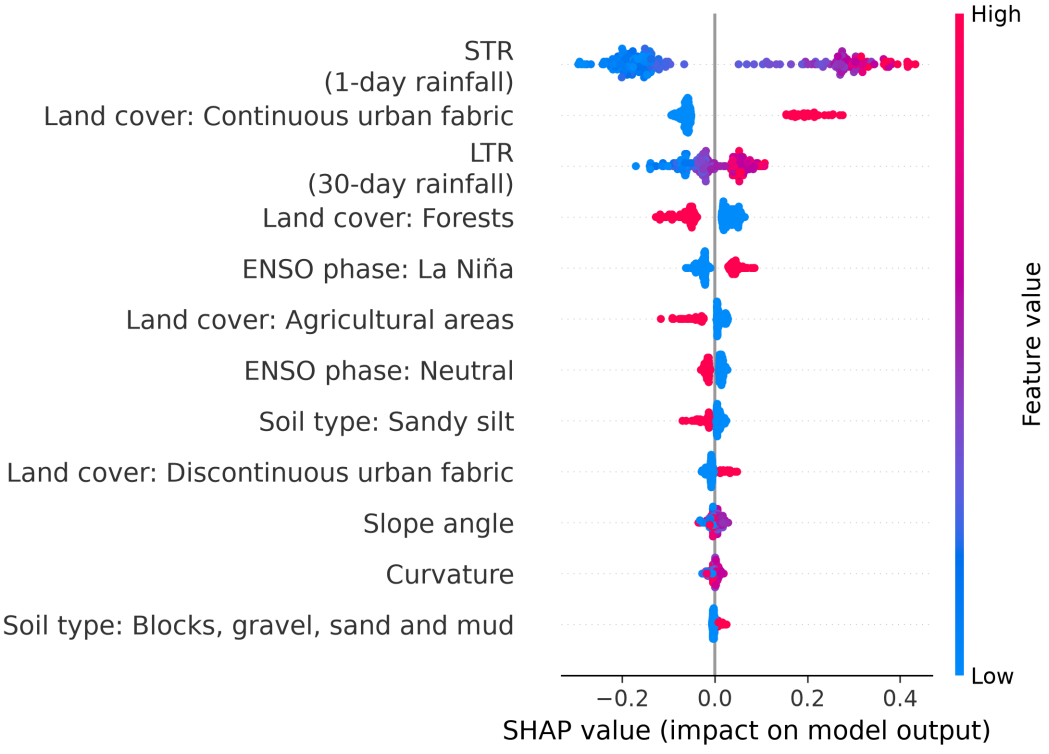


Figure 12. SHAP values of the random forest fitted on the representative run of training set C. Predictors are ordered by the mean of $|\phi_j|$ over the test part, each point is one observation, and its color indicates whether the value of that predictor is high or low for that observation. A point on the right means that the predictor pushed that particular prediction towards a landslide, and a point on the left the opposite.

4.4 Threshold definition and dynamic maps

Figure 14 shows the ROC curve of the additive model of set C. Its median AUROC is 0.986, and 90 % of the runs fall between 0.977 and 0.996. The cut-offs searched for true positive rates of 95, 85 and 70 % reached 96, 88 and 78 %, at probabilities of 0.288, 0.677 and 0.838 and with false positive rates of 5.5, 2.0 and 0.7 %. They exceed the targets because each test set holds about 73 landslides, so the rate moves in steps of about 1.4 percentage points, and the search keeps the first cut-off that reaches the target.

Figure 15 expresses the cut-offs in rainfall for two land covers under El Niño and La Niña, with the slope (32°), the curvature (median) and the soil (sandy silt) held fixed. At the median long-term rainfall of 187 mm under El Niño, a slope unit of continuous urban fabric reaches Medium with about 24 mm of daily rainfall and Very high with 40 mm, whereas a forested unit needs about 53 and 70 mm. These values hold only for the fixed combination of predictors. The panels rest on uneven data: (a) and (c) on 26 and 143 landslides, (b) and (d) on 5 and 17, and the two urban panels hold only 5 and 7 non-landslide observations. In the urban panels most landslides lie well above the Very high line.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

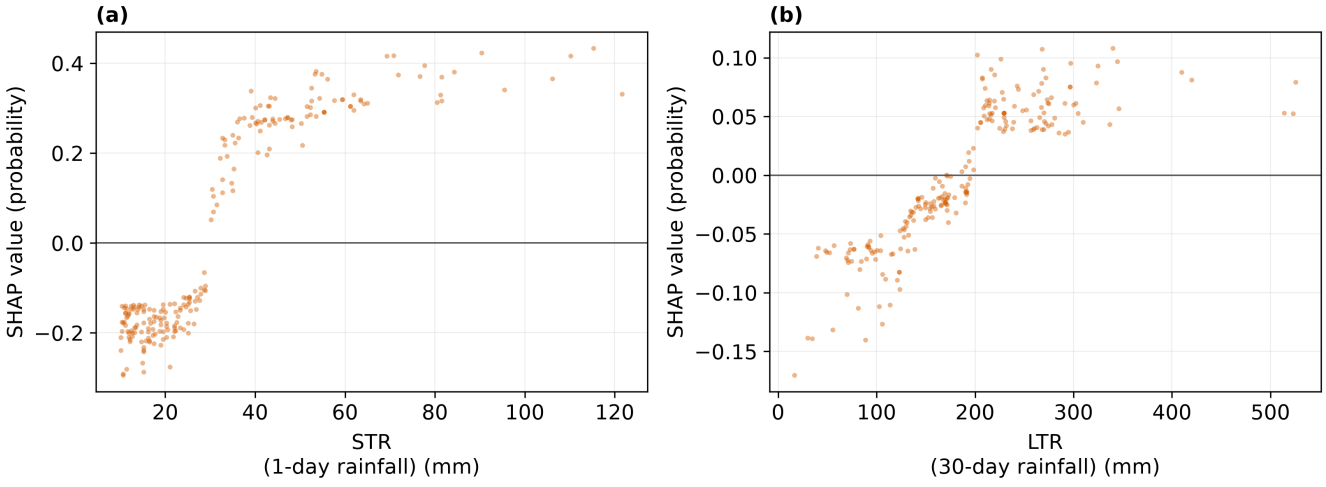


Figure 13. Contribution of the two rainfall variables to each individual prediction of the test part, in units of probability, plotted against their own value, for the random forest fitted on the representative run of training set C. (a) Short-term rainfall; (b) long-term rainfall. Note the different vertical scale of the two panels. The vertical spread at a fixed rainfall amount reflects interactions with the remaining predictors, which the additive model cannot represent.

4.4.1 Dynamic maps

Both events occurred after the 1 April cut-off. The first fell on 27–28 April on the Altavista sector, on the western slopes of Medellín, and the second on 4–5 May on the Olivares sector of Itagüí. Both produced shallow failures, mostly flows and translational slides, that turned into debris flows and destroyed houses and access roads; the Olivares event also caused one fatality. The maps used 140 gauges with records covering each event and the 31 days before it. About a quarter of the slope units could not be evaluated, 27 % in Altavista and 26 % in Olivares, mostly because the nearest gauge had gaps in its record. Table 5 summarizes the check at the three levels. Hourly animations of both events, with the hyetograph of the nearest gauge, are available as a video supplement (Sánchez Patiño, 2026a).

In Altavista the rain fell between midnight on 27 April and the early hours of 28 April, mostly on the western side of the valley between San Antonio de Prado and Altavista, with 24-hour totals of up to 80 mm (Fig. 16). The 380 landslides were reported at 03:00 on 28 April and fall in 178 slope units, 47 of which could not be evaluated. The first units with landslides reached High at midnight, and by 02:00, 98 of the 102 units that reached it before the report were already there (Fig. 17). Of the 131 evaluated units with landslides, 78 % reached High or above before the report and 58 % reached Very high; counted by landslide, 197 of the 283 landslides in evaluated units (70 %) fell in units at High or above. Over the 26 hours examined, 10.8 % of the valley reached High or above and 5.6 % Very high. The units with landslides reached High between zero and three hours before the report, two hours in the median.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

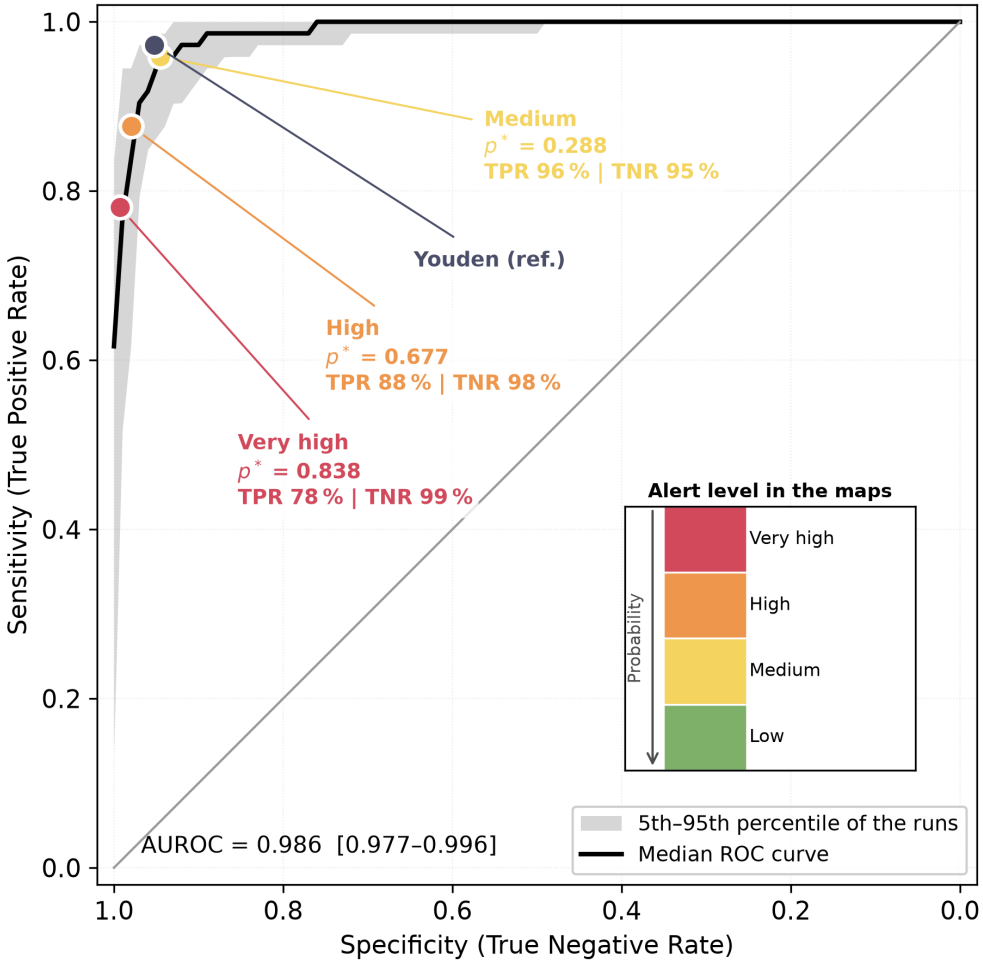


Figure 14. ROC curve of the additive model fitted on training set C, and the probability cut-offs derived from it. The horizontal axis is specificity, decreasing to the right. The black line is the median over the 100 Monte Carlo runs and the gray band spans the 5th to 95th percentile of those runs. The three colored points are the cut-offs that define the alert levels, searched for target sensitivities of 95, 85 and 70 %. The gray point is the Youden-optimal cut-off ($p^* = 0.307$, sensitivity 97 %, specificity 95 %), shown only as a reference. The inset recalls how the cut-offs divide the probability range into the four alert levels used in the dynamic maps.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

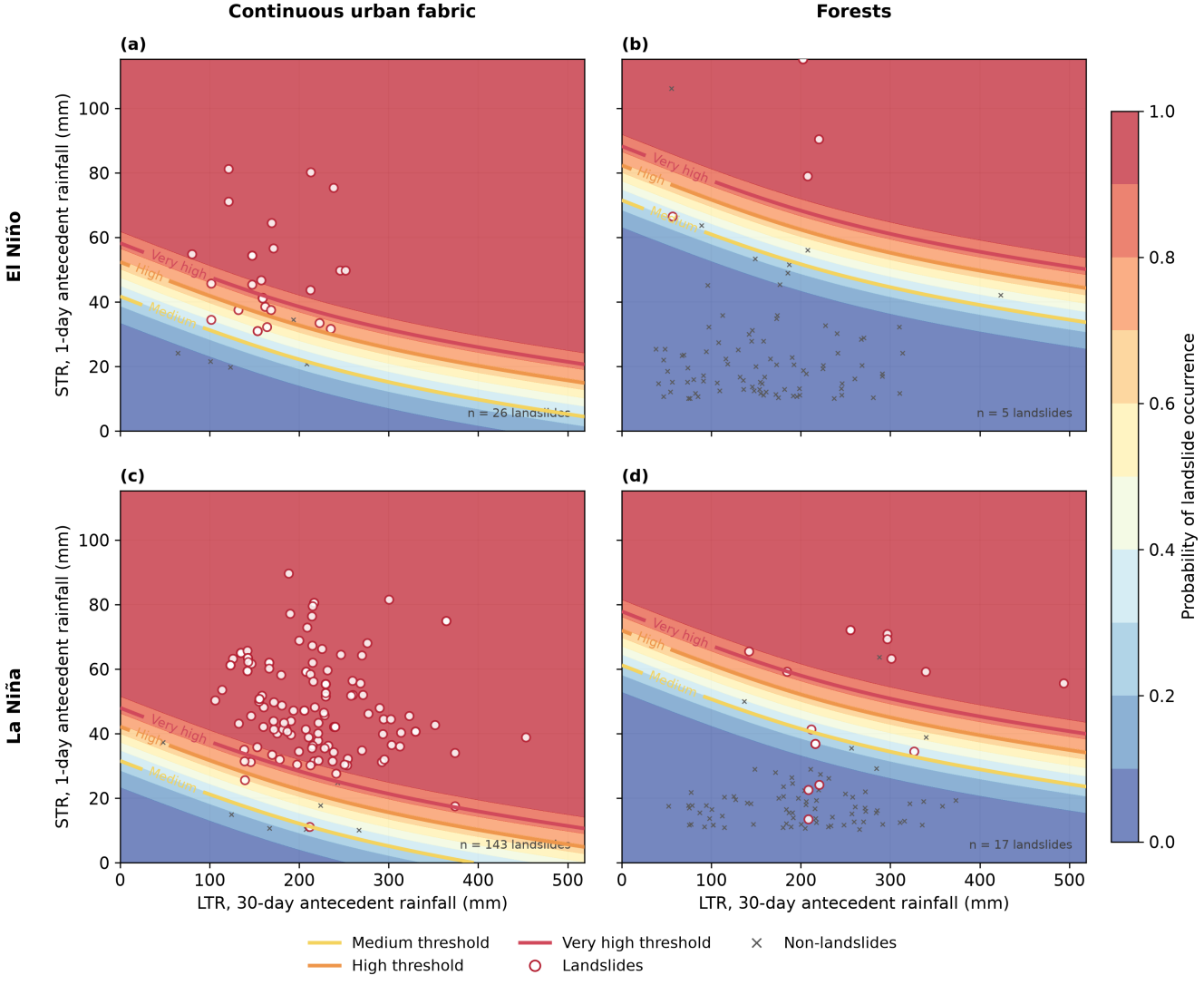


Figure 15. Rainfall thresholds for two land cover classes under the two ENSO phases. All the remaining predictors are held constant across the four panels at a slope of 32°, the median curvature of the study area and sandy silt soil, so the difference between columns is attributable to land cover alone and the difference between rows to the ENSO phase alone. Colors show the probability of landslide occurrence predicted by the additive model over a grid of long-term (horizontal axis) and short-term (vertical axis) antecedent rainfall. The colored lines are the three cut-offs of Fig. 14 and read as classical rainfall thresholds. Points are the landslide and non-landslide observations of the representative run that belong to the corresponding land cover and ENSO phase.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

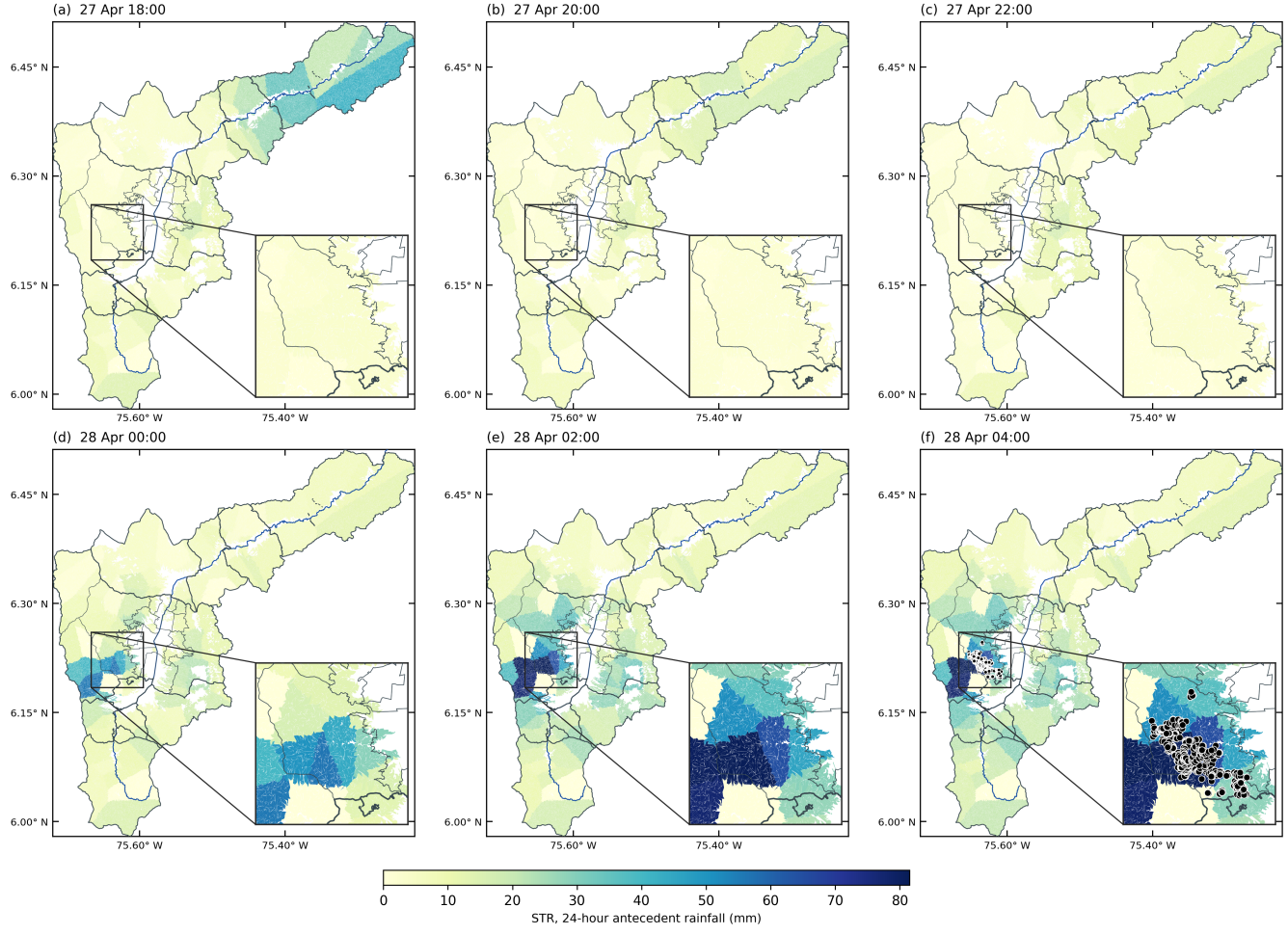


Figure 16. Short-term rainfall accumulated over the previous 24 hours in each slope unit during the Altavista event of 27–28 April 2025. Each panel shows the whole Aburrá Valley and, in the inset, the affected area. Black dots are the landslides reported. White areas are the slope units left out of the analysis for having a mean slope below 10°, and light gray units have no valid rainfall record.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Event	Level reached	Hits	Misses	Hit rate	FPR	Valley alerted (%)	Lead time (h)
Altavista	Medium or above	118	13	0.90	0.30	22.4	3 [1–3]
	High or above	102	29	0.78	0.15	10.8	2 [0–3]
	Very high	76	55	0.58	0.08	5.6	2 [1–3]
Olivares	Medium or above	5	0	1.00	0.64	47.2	8 [4–9]
	High or above	5	0	1.00	0.34	25.1	8 [4–9]
	Very high	5	0	1.00	0.15	11.4	8 [4–9]

Table 5. Independent check of the dynamic maps on the two events of 2025. A slope unit is alerted at a level if it reached that level at any hour between 24 hours before the first report and its own report. Hits and misses are slope units with landslides. FPR is the false positive rate, the fraction of the evaluated slope units without landslides that were alerted, and the valley alerted is the fraction of all 61,073 slope units. Lead time is the median of the hours between the first hour at the level and the report, with the range in brackets; it is an upper bound of the real warning time. Slope units with landslides that could not be evaluated are not counted: 47 of 178 in Altavista and 15 of 20 in Olivares.

In Olivares the rain fell during the night of 4 May and the early hours of 5 May on the western side of Itagüí, the southwest of Medellín and, further south, Caldas, again with 24-hour totals above 80 mm (Fig. 18). Thirty-two landslides were reported at 04:00 (25) and 08:00 (7) on 5 May in 20 slope units, but only 5 of these units, with 8 of the landslides, could be evaluated. All five reached Very high, one at 23:00 on 4 May and four at midnight, between four and nine hours before their report (Fig. 19). The alert also covered a larger part of the valley, including the eastern slopes of Medellín, La Estrella and Caldas: over the 29 hours examined, 25.1 % of the slope units reached High or above and 11.4 % Very high.

Across the three levels (Table 5), moving from Very high down to Medium in Altavista raises the hit rate from 58 to 90 % but multiplies the alerted area by four, from 5.6 to 22.4 % of the valley. In Olivares the evaluated units are hit at every level, while the area at Medium or above reaches 47 % of the valley. The two inventories were compiled with different effort, so their sizes and hit rates are not directly comparable.

5 Discussion

5.1 Selection of the rainfall time windows

The data settled the short-term window but not the long-term one. One day beat every longer short-term window by a margin larger than the noise between runs, and it was also the most stable choice (Fig. 5a,b). The length of the long-term window, by contrast, made almost no difference: from 8 to 90 days the median AUROC changed less than the standard deviation of any single combination (Fig. 5c). Under the one-standard-error rule (Hastie et al., 2009), 14 combinations were equivalent, in two groups of 15–21 and 29–35 days that the paired comparison could not separate. Taking the 17-day optimum would have meant choosing on sampling noise, so we chose on the role of the window in the model.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

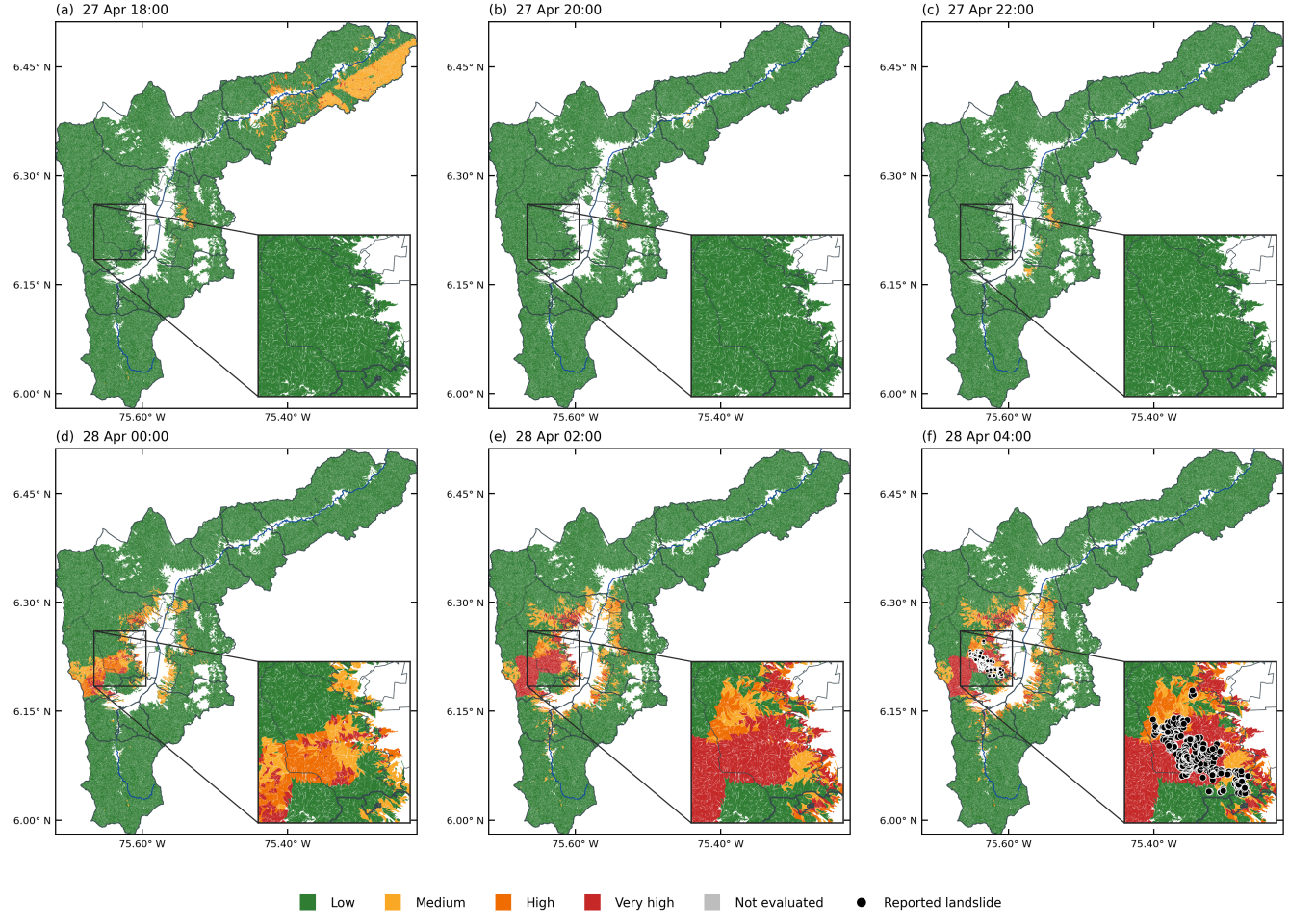


Figure 17. Hourly evolution of the alert level during the Altavista event of 27–28 April 2025, at the same hours as Fig. 16. Levels follow the probability cut-offs of Fig. 14. Gray slope units could not be evaluated. Black dots are the landslides reported up to each hour.

The one-day rainfall is the trigger, the rain that raises pore pressure on the day of failure. The long-term rainfall stands for the antecedent soil moisture, which sets how much of that rain a slope can take before it fails. A monthly total spans several rain episodes and dry spells, which suits a moisture proxy. Earlier work in the valley points to the same length: Aristizábal et al. (2011) found that minimum thresholds appear only when the antecedent rainfall is accumulated over 30 days or more, with a critical minimum near 60 mm in 30 days, and identified the antecedent rainfall as the main conditioning factor. Studies in the Italian Alps that selected their windows with a similar search found 14 to 28 days (Steger et al., 2022; Moreno et al., 2024; Steger et al., 2024). The best length depends on the region (Crosta, 1998), but it is of the order of weeks. The 30-day window lies inside one of the equivalent groups, so it costs no measurable performance, and it can be compared directly with the thresholds already proposed for the valley.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

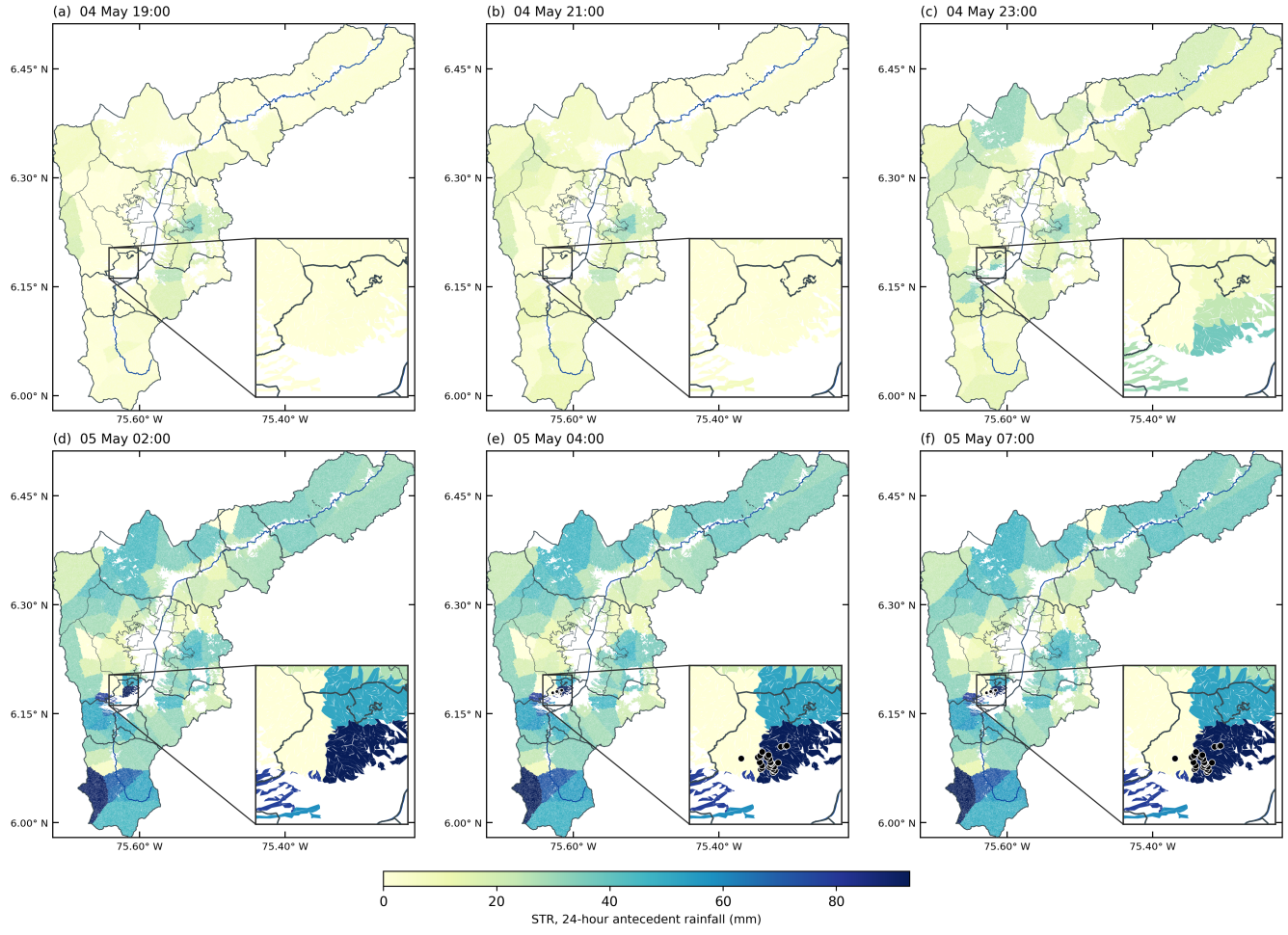


Figure 18. Short-term rainfall accumulated over the previous 24 hours in each slope unit during the Olivares event of 4–5 May 2025, on the same color scale as Fig. 16.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

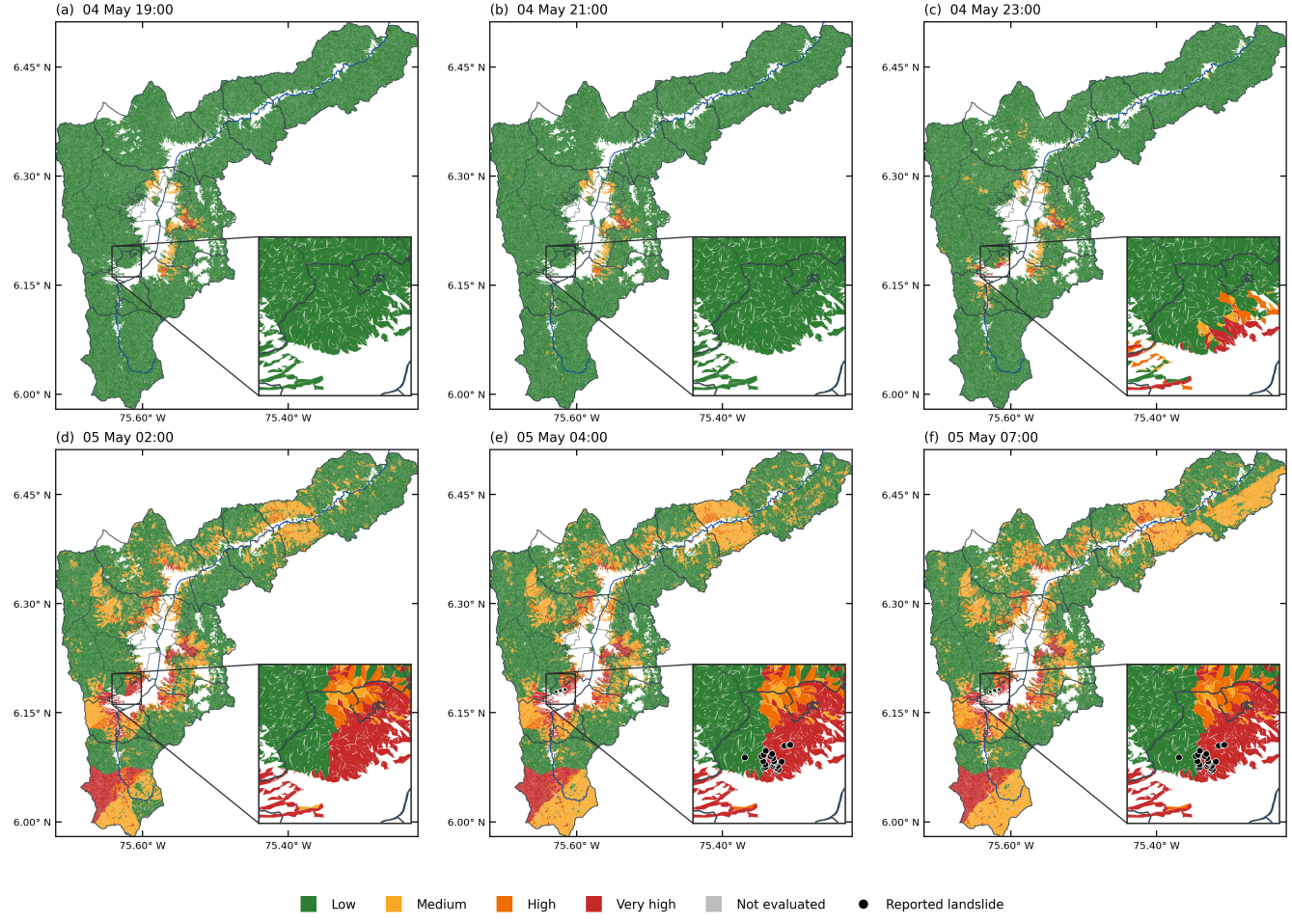


Figure 19. Hourly evolution of the alert level during the Olivares event of 4–5 May 2025, at the same hours as Fig. 18. Gray slope units could not be evaluated.

5.2 Sampling designs and classification models

Within a training set the three models performed almost alike. Their medians differ by less than the spread of any one model between runs (Table 4), and the random forest, although ahead in most paired runs, led by only a few thousandths of AUROC (Fig. 8). The sampling design had a larger effect on what the models learned. Sets B and C use the same predictors, yet the land cover matched the short-term rainfall in set B and fell well behind it in set C (Fig. 10). Comparisons of data-driven landslide models usually focus on the algorithm, but with this inventory the way the absences were drawn mattered more.

All three designs share one weakness: some absences are probably landslides that were never reported. As noted in Section 3.2, this biases the probabilities more than their ranking. It also makes the false positive rates, in the test data and in the 2025 events, upper bounds.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

We built the thresholds with the additive model, which trailed the random forest by only 0.002 in median AUROC in set C, because its behavior can be inspected. The logistic regression forces each effect to be linear on the logit, which the slope effect contradicts (Fig. 11c). The random forest can fit any shape, but it gives no smooth response surface over the rainfall plane from which a threshold can be read and checked. The additive model captures the non-linear slope effect, and each of its terms can be plotted and judged against what is known about the process (Fig. 11). Other space-time landslide studies chose it for the same reason (Steger et al., 2022; Moreno et al., 2024; Steger et al., 2024).

Set C was chosen because its question is the operational one. Sets A and B hold the terrain or the date fixed to isolate part of the signal, which suits diagnosis, but a warning map has to rank any slope unit on any day with terrain and rainfall together, and only set C poses that problem. Its probabilities also have the lowest Brier score (Table 4), which matters because the alert levels are fixed cut-offs on those probabilities.

The cross-validation shows how far these scores hold outside the conditions seen in training (Fig. 9). Holding out whole months cost little, so the Monte Carlo scores do not depend on testing the models on storms they had already seen. Holding out a rainfall zone cost about 0.02 in set C: a model trained on the rest of the valley still ranked the conditions of a zone with another rainfall regime. The lowest zone, San Antonio de Prado, holds only eight landslides, too few to point to a local weakness of the model.

Land cover is the exception. Within a held-out class, the additive model of set C still separated landslides from absences (Fig. A1b), so its response to rainfall carried over to a class it had not seen. What it could not do was place that class relative to the others, because the effect of a class can only be learned from observations of that class. The loss was largest in set B, where land cover weighs as much as the short-term rainfall. The land cover effects of the thresholds are therefore only as reliable as the number of landslides in each class, and the thresholds should not be applied to a class absent from training, which is why the maps leave such slope units unclassified. The rainfall windows were fixed in all folds, so these losses leave out the cost of choosing them, although the clear margin of the one-day window and the flat response to the long-term one make a different choice unlikely.

5.3 Model flexibility, interpretability and plausibility

The three models agree on the leading predictors in every set (Fig. 10), so the reading below does not depend on the model. The short-term rainfall comes first in sets A and C. Without it, the models of set A fall from an AUROC of about 0.97 to about 0.68, and in set C it stays two to three times ahead of the land cover. Only in set B, where both classes share the date, does the land cover match it. What separates the day a slope failed from other wet days is mainly the rain of the last 24 hours, while the terrain mostly tells where. This points to a fast process, consistent with the shallow flows and translational slides that most likely dominate the inventory, whose failure surfaces lie within the first meters of residual soil and respond to changes in pore pressure within hours.

Low permutation importance does not always mean irrelevance, because the measure splits the credit between predictors that carry the same information. In the valley, land cover and terrain go together: the continuous urban fabric occupies the lower and gentler hillslopes and the forest the steeper and higher ones, so once a model knows the land cover, shuffling the

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

slope costs little. The slope still carries information: the additive model fits a clear maximum near 30° (Fig. 11c), the shape expected on residual soils, where gentle ground rarely fails and steeper ground holds too little soil to fail. The curvature, with a flat effect and a wide band (Fig. 11d), adds almost nothing, and the soil adds little.

The permutation importance of the ENSO phase is small, at most 0.015, although landslide reports in the valley are more frequent in La Niña years (Aristizábal et al., 2026). Two reasons can explain the difference. ENSO acts on landslides through the rain it brings, and here that rain is measured at the nearest gauge, so the phase has little left to explain. Aristizábal et al. (2026) compared years without daily local rainfall, so there the phase carried the signal that the gauges carry here. In addition, the rainfall filter places all observations, landslides and absences, on wet days, which removes part of the contrast between phases. The partial effect is nevertheless clear: La Niña raises the probability in the three models, and in the additive model neither its interval nor that of El Niño crosses zero (Figs. 11g, 12 and A3). This effect may be partly an artifact of the sampling. Of the landslides used for training, 98 % fall between 2020 and 2024, a period dominated by one long La Niña and one El Niño, whereas the absences of sets A and C are drawn from the whole record. The La Niña effect may therefore partly mark the years in which most of the inventory was recorded, and with so few changes of phase the two cannot be separated. For warning, ENSO is useful to anticipate which seasons will be more hazardous, but on a given wet day the gauges are the better guide.

The random forest shows a feature of the short-term effect that the additive model cannot. The additive model fits an almost linear rise on the logit scale (Fig. 11a), whereas the forest, free to fit any shape, gives a step (Fig. 13a): its contribution is negative and nearly flat below 30 mm, jumps by about 0.34 near 30 mm and grows slowly above 40 mm, a threshold that the forest found without being told to look for one. Between the 40–50 mm interval and the few days above 100 mm, the contribution of the forest grows by only about 0.12, whereas the additive model keeps raising the log-odds where the observations become sparse. That tail is an extrapolation and should not be read physically. It does not change the maps, because in the additive model a day above 100 mm already reaches the Very high level in every panel of Fig. 15.

For the long-term rainfall the two models agree. In the random forest its contribution spans 0.28 in probability, against 0.73 for the short-term rainfall, and it rises up to about 200 mm before it flattens (Fig. 13b). The additive model also fits a rise that slows as the antecedent rainfall grows (Fig. 11b). Two models with different assumptions find the same shape, which supports reading the long-term rainfall as a preparatory condition whose added effect decreases at high antecedent rainfall.

5.4 Scope and limitations of the dynamic thresholds

The alert levels follow the approach of Steger et al. (2024): the ROC curve ties each cut-off to the two quantities a warning system manages, the share of landslides it captures and the share of false alarms it accepts (Fig. 14). That study combined the TPR95 and TPR25 cut-offs with an optimal one, whereas here all three cut-offs are set by the share of landslides captured, so each level carries its own trade-off. This graded reading needs explicitly sampled non-landslide observations, which classical empirical thresholds lack (Postance et al., 2018; Leonarduzzi and Molnar, 2020). The TPR95 cut-off of Steger et al. (2024) came with a false positive rate of 44 %, against 5.5 % here. The two studies differ in sampling design, rainfall filter and windows, so the comparison is only indicative, but it suggests that in this valley wet days with and without landslides are easier to tell apart.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

The resulting rainfall thresholds change between slope units and between days: in Fig. 15 the lines shift with the land cover and the ENSO phase, and in the model they also depend on the slope and the soil. They fall as the antecedent rainfall grows, as the margin-of-stability concept predicts (Crozier, 1986; Glade et al., 2005): after a wet month a slope is closer to failure and needs less rain on the day. Their dependence on the terrain is what they add to the empirical thresholds already published for the valley. The model, however, also draws thresholds for combinations of predictors that the inventory barely samples. The forest panels rest on 5 and 17 landslides, so their lines are interpolations with wide uncertainty, and the urban panels hold only 5 and 7 absences, so the model has seen few wet days without landslides on urban slopes. In such combinations the thresholds should be used with caution.

The inventory limits the thresholds in two further ways. First, it does not record the type of movement. The damage reports and the two 2025 events suggest that most failures are shallow flows and translational slides, which agrees with the dominance of the one-day rainfall, but deeper or slower movements in the record cannot be ruled out, so the thresholds apply to the mix of processes that gets reported. Second, it comes from damage reports and emergency calls, which concentrate on built-up slopes (Section 2.1.1). The large effect of the continuous urban fabric is therefore partly physical and partly a product of where landslides are reported.

The two events of 2025 support the operational use of the maps, within what two events can show. In Altavista, 78 % of the evaluated slope units with landslides reached High or above before the report while 10.8 % of the valley did, and in Olivares the five slope units that could be evaluated reached Very high four to nine hours before it (Table 5). Landslides are reported after they occur, so these lead times are upper bounds; in Altavista the alert rose with the triggering rain, and the data cannot show whether it came before the failures. The cost of the alerts shows in Olivares, where the rain covered a large part of the valley and a quarter of it reached High or above, although some of those false alarms may be landslides that were never reported. In both events, 26 to 27 % of the slope units could not be evaluated, mostly because the nearest gauge had gaps in its record, and in Olivares this left out 15 of the 20 slope units with landslides, so an operational version needs a rule for such gaps. Both events came after the training period, so this is an independent check, but measuring operational skill needs more events and rainy seasons, with the same temporal hold-out repeated as new events are recorded.

The model is cheap to run and can be inspected effect by effect, and its levels are defined in terms that a risk management agency can use. These properties make data-driven models candidates for regional early warning (Steger et al., 2024). Three steps would bring this one closer to operation. The first is a better inventory, with systematic mapping beyond damage reports, the type of movement and a record long enough to cover both ENSO phases several times. The second is a predictor closer to the process, such as measured or modeled soil moisture, which the antecedent rainfall only approximates. The third is to move from hindcast to forecast by driving the model with radar nowcasts or weather forecasts, after correcting the forecast rainfall against the gauges used in training (Steger et al., 2024); the levels could then move towards impact-based warnings (Steger et al., 2026).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

6 Conclusions

We derived rainfall thresholds for landslides in the Aburrá Valley with data-driven models that relate the rainfall of each day to the occurrence of landslides in each slope unit. The models combined four terrain variables with the short-term rainfall as the trigger, the long-term rainfall as a proxy for antecedent soil moisture, and the ENSO phase. They were trained on 365 dated landslides and on non-landslide observations drawn under three sampling designs, each resampled 100 times.

A one-day window described the trigger better than any longer window, by a margin larger than the noise between runs. The data did not resolve the antecedent window, whose response was flat from 8 to 90 days, and we adopted 30 days, which lies among the equivalent windows and agrees with earlier evidence from the valley. Landslides in the valley thus respond to the rain of the day, and the rain of the previous month sets how much of it is needed.

The three models classified well, with median AUROC between 0.955 and 0.988, and within a design they differed by less than the spread between runs. The sampling design had a larger effect: with the same predictors, the land cover matched the short-term rainfall in one design and fell well behind it in another. Where and when the absences are drawn should therefore be reported and tested as part of any data-driven threshold. Fixing the slope unit isolates the rainfall and suits the selection of the windows, fixing the date shifts the weight towards the terrain, and freeing both gives the design on which an operational model should be trained.

We built the thresholds with the additive model, which trailed the random forest by a few thousandths of AUROC but can be inspected term by term. The short-term rainfall was the strongest predictor in the designs that vary the date. The slope has a clear effect, peaking near 30°, that its association with land cover hides from permutation importance. The effect of the continuous urban fabric is partly physical and partly a product of an inventory built from damage reports, and the La Niña effect cannot yet be separated from the concentration of the inventory in 2020–2024.

The thresholds were read from the ROC curve as four alert levels, each with a known share of landslides captured and of false alarms. The rainfall needed to reach a level changes with the terrain of each slope unit, the antecedent rainfall and the ENSO phase. The model kept its skill when a month, a soil type or a rainfall zone was left out of training, but its thresholds should not be applied to land cover classes that are poorly represented in the inventory or absent from it. In the Altavista event of 2025, after the training period, 78 % of the evaluated slope units with landslides reached High or above while 10.8 % of the valley did. This independent check covers only two events, and before the thresholds enter the warning chain they need more events, a rule for gauges with gaps in their record and an inventory that goes beyond damage reports. The approach needs only a dated inventory and a rain-gauge network, data that many tropical mountain regions already have.

Appendix A: Additional results

Table A1 gives the out-of-fold AUROC of the cross-validation, and Fig. A1 its values fold by fold for the additive model of set C. Figure A2 shows the partial effects of the additive model of set A, and Fig. A3 the odds ratios of the logistic regression for the representative run of set C.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

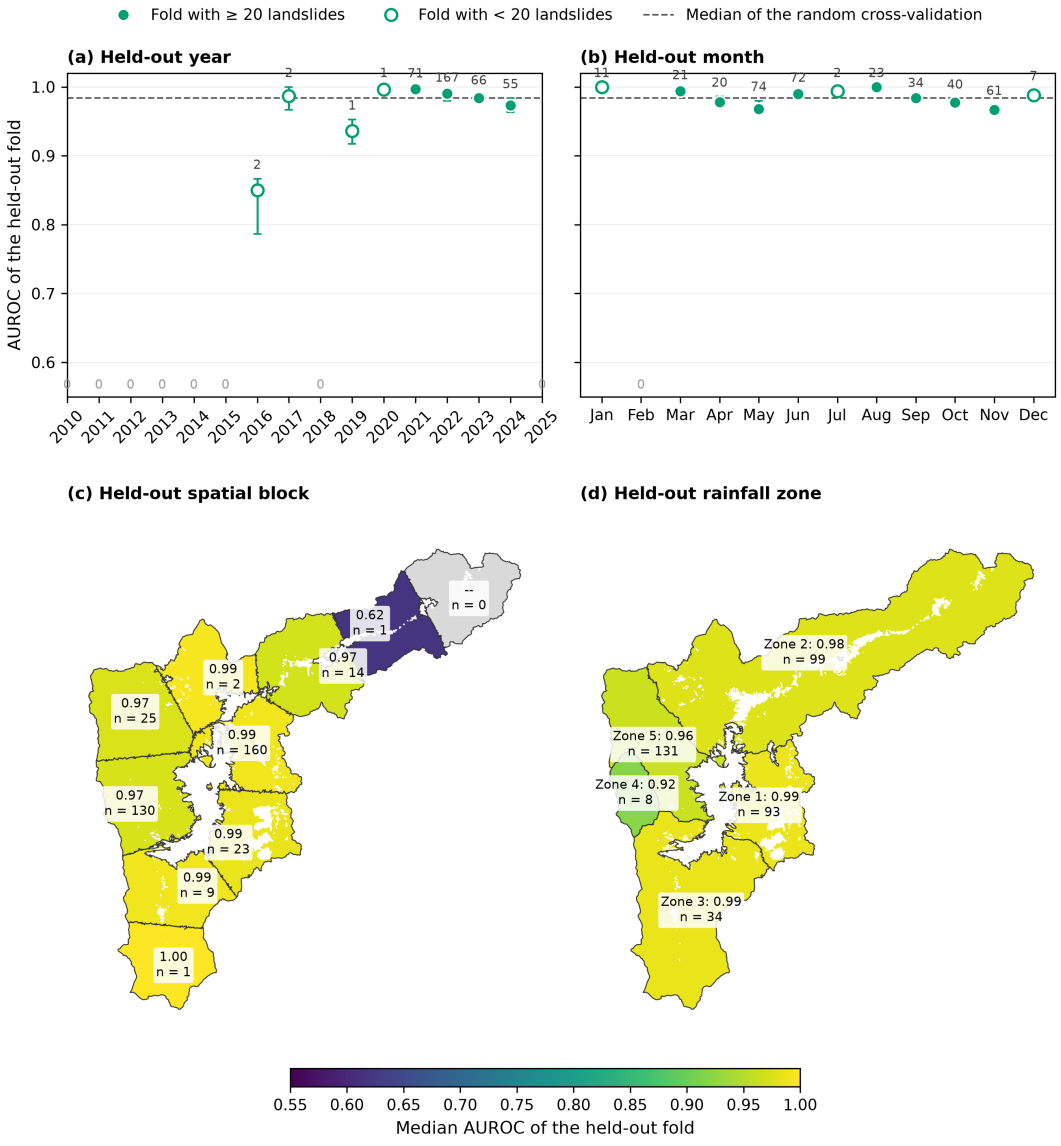


Figure A1. Cross-validation of the generalized additive model of training set C, the model used for the thresholds, fold by fold: (a) calendar months, (b) land cover classes, (c) soil types and (d) homogeneous rainfall zones. In (a) to (c), points are medians over the 20 data sets and bars their interquartile range; the number above each point is the median number of landslides in the fold, and in (b) and (c) the classes are ordered from the most to the fewest landslides. Hollow markers flag folds with fewer than 20 landslides or 20 non-landslide observations, whose AUROC is unstable. The dashed line is the median out-of-fold AUROC of the random scheme. In (d) each zone is colored by its median AUROC and labeled with that value and its number of landslides (n); SAP is San Antonio de Prado (Medellín), and the white areas inside the valley are the flat terrain excluded from the analysis.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

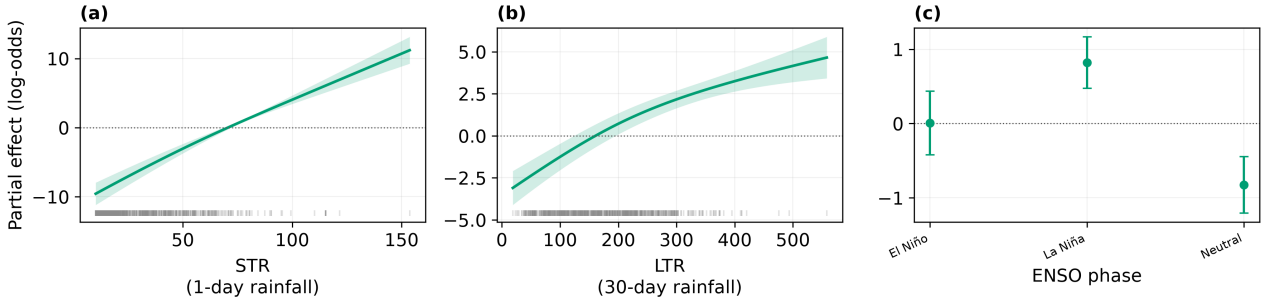


Figure A2. Partial effects of the generalized additive model fitted on training set A, where the only predictors are the two rainfall variables and the ENSO phase.

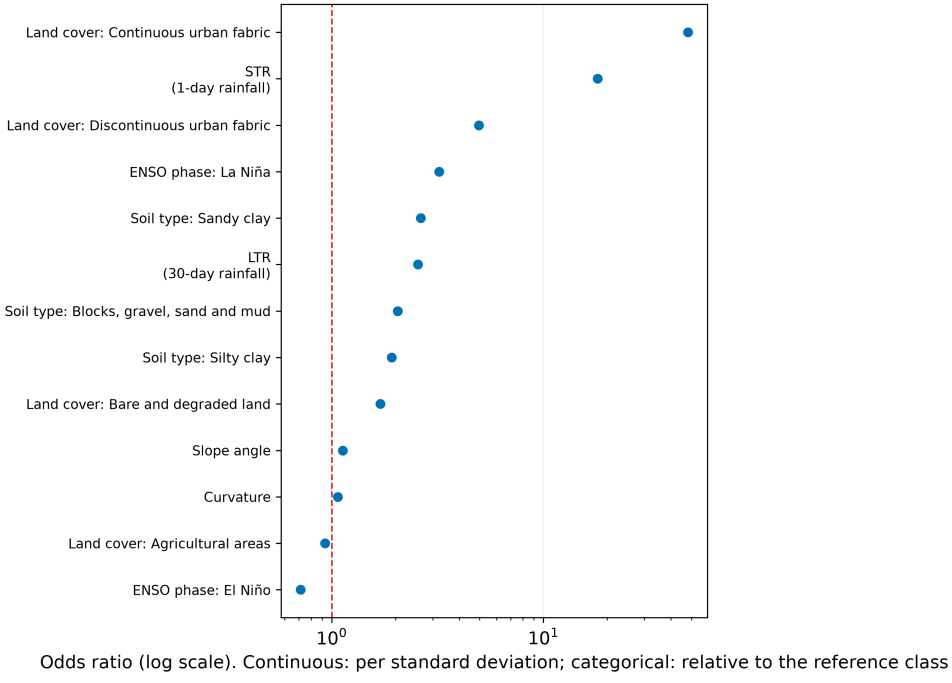


Figure A3. Odds ratios estimated by the logistic regression on the representative run of training set C, on a logarithmic scale. Continuous predictors are expressed per standard deviation and categorical ones relative to their reference class. The dashed line marks an odds ratio of one, that is, no effect.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Set	Model	Random	Month	Soil type	Land cover	Rainfall zones
A	LR	0.972 [0.970–0.975]	0.966 [0.963–0.971]	0.972 [0.970–0.975]	0.970 [0.969–0.975]	0.971 [0.969–0.976]
	GAM	0.973 [0.970–0.976]	0.968 [0.964–0.972]	0.972 [0.971–0.977]	0.971 [0.970–0.976]	0.972 [0.970–0.976]
	RF	0.978 [0.975–0.980]	0.969 [0.965–0.974]	0.976 [0.970–0.978]	0.973 [0.971–0.976]	0.971 [0.968–0.976]
B	LR	0.952 [0.950–0.956]	0.951 [0.950–0.955]	0.946 [0.942–0.950]	0.698 [0.687–0.726]	0.914 [0.908–0.920]
	GAM	0.956 [0.954–0.960]	0.956 [0.954–0.959]	0.950 [0.946–0.954]	0.736 [0.726–0.754]	0.923 [0.918–0.929]
	RF	0.963 [0.960–0.966]	0.956 [0.954–0.959]	0.950 [0.948–0.953]	0.768 [0.760–0.776]	0.924 [0.919–0.930]
C	LR	0.982 [0.980–0.983]	0.978 [0.977–0.981]	0.980 [0.977–0.981]	0.889 [0.879–0.899]	0.963 [0.959–0.968]
	GAM	0.984 [0.983–0.986]	0.982 [0.980–0.984]	0.982 [0.980–0.985]	0.905 [0.895–0.914]	0.969 [0.963–0.973]
	RF	0.987 [0.985–0.988]	0.980 [0.978–0.982]	0.982 [0.980–0.984]	0.912 [0.907–0.920]	0.970 [0.966–0.972]

Table A1. Out-of-fold AUROC of the three models under the five cross-validation schemes. Values are medians over 20 Monte Carlo data sets of each training set, with the interquartile range in brackets. The random scheme uses five folds stratified by class and grouped by slope unit; each fold of the other four schemes is one calendar month, one soil type, one land cover class or one homogeneous rainfall zone. In set A the models do not use the static predictors, so the soil type and land cover schemes only group the slope units.

Video supplement. Hourly animations of the dynamic alert maps for the Altavista and Olivares events, showing the alert level over the whole valley, a zoom of the affected sector with the reported landslides, and the hyetograph of the nearest gauge, are available at https://doi.org/10.5281/zenodo.22243857 (Sánchez Patiño, 2026a).

Data availability. Rainfall data from the SIATA network are available upon reasonable request to the corresponding institutions. Landslide inventories from GEOHAZARDS and DAGRD are subject to institutional access restrictions. The remote-sensing landslide inventory used in this study is described in Aristizábal and Korup (2025).

Code and data availability. The rain-gauge records and the land cover map were provided by the Early Warning System of Medellín and the Aburrá Valley (SIATA), the soil map by the Departamento Administrativo de Gestión del Riesgo de Desastres de Antioquia (DAGRAN), and part of the landslide records by the Departamento Administrativo de Gestión del Riesgo de Desastres de Medellín (DAGRD). These third-party data are owned by those institutions and can be requested from them under their data policies. The remaining landslide records come from the GEOHAZARDS catalog (Aristizábal et al., 2025), and the Oceanic Niño Index is publicly available (NOAA Climate Prediction Center, 2025). The hourly animations of the alert maps are available at https://doi.org/10.5281/zenodo.22243857 (Sánchez Patiño, 2026a). The Python code and the derived data sets (slope units, landslide inventory, training sets and model outputs) are available from the corresponding author on request (oisanchezp@unal.edu.co).

Code availability. The Python code and the derived data sets (slope units, landslide inventory, training sets and model outputs) are available from the corresponding author on request (oisanchezp@unal.edu.co).

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Author contributions. Oscar Sanchez and Edier Aristizábal designed the study. Oscar Sanchez processed the rainfall and landslide data and performed the analyses. All authors contributed to the interpretation of results and to writing and revising the manuscript.

630 *Acknowledgements.* The authors wish to express their gratitude to the SIATA project, led by the Área Metropolitana del Valle de Aburrá (AMVA) and Universidad EAFIT, for providing access to rainfall data from the rain gauge network. We are grateful to the GEOHAZARDS research group at the Universidad Nacional de Colombia, Medellín, for granting access to their landslide inventory for the Aburrá Valley. We also acknowledge the Departamento Administrativo de Gestión del Riesgo de Desastres of Medellín (DAGRD) for providing access to their database of reported landslides within the Medellín district.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F.: Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling, Geoscientific Model Development, 9, 3975–3991, https://doi.org/10.5194/gmd-9-3975-2016, 2016.

Alvioli, M., Guzzetti, F., and Marchesini, I.: Parameter-free delineation of slope units and terrain subdivision of Italy, Geomorphology, 358, 107 124, https://doi.org/10.1016/j.geomorph.2020.107124, 2020.

Arango Gartner, J. D.: Relaciones lluvias–deslizamientos y zonificación geotécnica en la Comuna Dos de la ciudad de Manizales, Trabajo de grado (pregrado), Universidad Nacional de Colombia, Sede Manizales, https://repositorio.unal.edu.co/items/0dcc917f-8726-46ab-87b3-806eb5c3efcf, 2000.

Aristizábal, E.: Geomorphological evolution of the Aburrá Valley, northern Colombian Andes, and implications for landslide occurrence, Master's thesis, Shimane University, Matsue, Japan, https://repositorio.unal.edu.co/handle/unal/85795, 2004.

Aristizábal, E. and Korup, O.: Linking Landslide Patterns to Transient Landscapes in the Northern Colombian Andes, Journal of Geophysical Research: Earth Surface, 130, https://doi.org/10.1029/2024JF008027, 2025.

Aristizábal, E. and Sánchez, O.: Spatial and temporal patterns and the socioeconomic impacts of landslides in the tropical and mountainous Colombian Andes, Disasters, 44, 596–618, https://doi.org/10.1111/disa.12391, 2020.

Aristizábal, E., González, T., Montoya, J. D., Vélez, J. I., Martínez, H., and Guerra, A.: Análisis de umbrales empíricos de lluvia para el pronóstico de movimientos en masa en el Valle de Aburrá, Colombia, Revista EIA, pp. 95–111, 2011.

Aristizábal, E., López, S., Sánchez, O., Vásquez, M., Rincón, F., Ruiz-Vásquez, D., Restrepo, S., and Valencia, J. S.: Evaluación de la amenaza por movimientos en masa detonados por lluvias para una región de los Andes colombianos estimando la probabilidad espacial, temporal, y magnitud, Boletín de Geología, 41, 85–105, 2019.

Aristizábal, E. V., Correa, D., García Cadavid, C., Valencia Londoño, A. M., and Zabaleta Benavides, E. A.: Geohazard: base de datos por movimientos en masa y avenidas torrenciales en Colombia y el departamento de Antioquia, Ciencia e Ingeniería Neogranadina, 35, 131–147, https://doi.org/10.18359/rcin.7786, 2025.

Aristizábal, E. V., Sanchez, O. I., and Korup, O.: Landslide timing and rainfall regimes in a rapidly urbanizing tropical mountain valley in Colombia, Natural Hazards, 122, 306, 2026.

Ba, Q., Chen, Y., Deng, S., Yang, J., and Li, H.: A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment, Earth Science Informatics, 11, 373–388, https://doi.org/10.1007/s12145-018-0335-9, 2018.

Bedoya-Soto, J. M., Aristizábal, E., Carmona, A. M., and Poveda, G.: Seasonal shift of the diurnal cycle of rainfall over medellin's valley, central andes of Colombia (1998–2005), Frontiers in Earth Science, 7, https://doi.org/10.3389/feart.2019.00092, 2019.

Bogaard, T. A. and Greco, R.: Landslide hydrology: from hydrology to pore pressure, Wiley Interdisciplinary Reviews: Water, 3, 439–459, https://doi.org/10.1002/wat2.1126, 2016.

Breiman, L.: Random forests, Machine learning, 45, 5–32, 2001.

Caine, N.: The rainfall intensity–duration control of shallow landslides and debris flows, Geografiska Annaler. Series A, Physical Geography, 62, 23–27, https://doi.org/10.2307/520449, 1980.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., and Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes and Landforms, 16, 427–445, https://doi.org/10.1002/esp.3290160505, 1991.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.

Carvajal Perico, J. H.: Propuesta de estandarización de la cartografía geomorfológica en Colombia, Colección Guías y Manuales, Servicio Geológico Colombiano, Bogotá, ISBN 978-958-99528-2-5, 2012.

Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S. R., Huang, J., and Zhou, C.: Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors, Journal of Rock Mechanics and Geotechnical Engineering, 15, 1127–1143, https://doi.org/10.1016/j.jrmge.2022.07.009, 2023.

Chleborad, A. F., Baum, R. L., and Godt, J. W.: Rainfall thresholds for forecasting landslides in the Seattle, Washington, area—exceedance and probability, Open-File Report 2006-1064, U.S. Geological Survey, https://doi.org/10.3133/ofr20061064, 2006.

Crosta, G.: Regionalization of rainfall thresholds: an aid to landslide hazard evaluation, Environmental Geology, 35, 131–145, 1998.

Crozier, M. J.: Landslides: Causes, Consequences and Environment, Croom Helm, London, 1986.

DANE: Resultados Censo Nacional de Población y Vivienda 2018, Tech. rep., Departamento Administrativo Nacional de Estadística, Bogotá, 2018.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Eilers, P. H. and Marx, B. D.: Flexible smoothing with B-splines and penalties, Statistical science, 11, 89–121, 1996.

Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Natural Hazards and Earth System Sciences, 18, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, https://doi.org/10.1016/j.geomorph.2014.10.019, 2015.

Glade, T., Anderson, M. G., and Crozier, M. J.: Landslide hazard and risk, John Wiley & Sons, 2005.

Gómez, D., Aristizábal, E., and García, E.: Antecedent rainfall influence on landslides in the Colombian Andes, https://www.researchgate.net/publication/369552890, https://www.researchgate.net/publication/369552890, preprint, 2023.

Gómez, D., García, E. F., and Aristizábal, E.: Spatial and temporal landslide distributions using global and open landslide databases, Natural Hazards, 117, 25–55, https://doi.org/10.1007/s11069-023-05848-8, 2023.

Gómez, J.: Relación entre infiltración y movimientos en masa en Bucaramanga, Anales de Ingeniería Civil, 1990.

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M.: Geographical landslide early warning systems, Earth-Science Reviews, 200, 102 973, https://doi.org/10.1016/j.earscirev.2019.102973, 2020.

Hastie, T. and Tibshirani, R.: Generalized additive models, Statistical science, 1, 297–310, 1986.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H.: The elements of statistical learning: data mining, inference, and prediction, vol. 2, Springer, 2009.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons, Journal of Climate, 30, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

IDEAM: Leyenda nacional de coberturas de la tierra. Metodología CORINE Land Cover adaptada para Colombia, escala 1:100.000, Instituto de Hidrología, Meteorología y Estudios Ambientales, Bogotá, 2010.

Leonarduzzi, E. and Molnar, P.: Deriving rainfall thresholds for landsliding at the regional scale: daily and hourly resolutions, normalisation, and antecedent rainfall, Natural Hazards and Earth System Sciences, 20, 2905–2919, 2020.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.



Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 4765–4774, 2017.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nature machine intelligence, 2, 56–67, 2020.

Marín, R. J., García, E., and Aristizábal, E.: Umbrales de lluvia para deslizamientos superficiales basados en modelos físicos: aplicación en una subcuenca del Valle de Aburrá (Colombia), Dyna, 86, 312–322, 2019.

Mayorga Márquez, R.: Determinación de umbrales de lluvia detonante de deslizamientos en Colombia, Meteorología Colombiana, https://www.osti.gov/etdeweb/biblio/20626492, 2003.

Mejía, J. F., Mesa, O., Poveda, G., Vélez, J. I., Hoyos, C. D., Mantilla, R., Barco, J., Cuartas, A., Montoya, M. I., and Botero, B. A.: Distribución espacial y ciclos anual y semianual de la precipitación en Colombia, DYNA, 127, 7–26, 1999.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., and Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, Earth-Science Reviews, 207, 103 225, https://doi.org/10.1016/J.EARSCIREV.2020.103225, 2020.

Moore, I. D., Grayson, R. B., and Ladson, A. R.: Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications, Hydrological Processes, 5, 3–30, https://doi.org/10.1002/hyp.3360050103, 1991.

Moreno, H. A., Vélez, M. V., Montoya, J. D., and Rhenals, R.: La lluvia y los deslizamientos de tierra en Antioquia: análisis de su ocurrencia en las escalas interanual, intraanual y diaria, Revista EIA, 5, 59 – 69, 2006.

Moreno, M., Lombardo, L., Crespi, A., Zellner, P. J., Mair, V., Pittore, M., van Westen, C., and Steger, S.: Space-time data-driven modeling of precipitation-induced shallow landslides in South Tyrol, Italy, Science of the Total Environment, 912, 169 166, 2024.

Nieto, S., Aristizábal, E., and Ozturk, U.: Coupled evolution of a city and landslides, Environmental Research Letters, 20, https://doi.org/10.1088/1748-9326/adc4d8, 2025.

NOAA Climate Prediction Center: Cold & Warm Episodes by Season (Oceanic Niño Index, ONI v5), https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php, 2025.

Pabon, J., Eslava, J., and Gomez, R.: Climatic large-scale characteristics of the tropical Americas (in spanish), Meteorología Colombiana, 4, 39–46, 2001.

Paz, C. and Torres, A.: Precipitación y su influencia sobre algunos deslizamientos ocurridos en las laderas del Valle de Aburrá, Ph.D. thesis, Universidad Nacional de Colombia, 1989.

Petley, D.: Global patterns of loss of life from landslides, Geology, 40, 927–930, https://doi.org/10.1130/G33217.1, 2012.

Postance, B., Hillier, J., Dijkstra, T., and Dixon, N.: Comparing threshold definition techniques for rainfall-induced landslides: A national assessment using radar rainfall, Earth Surface Processes and Landforms, 43, 553–560, 2018.

Poveda, G.: La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna, Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales, 28, 201–221, https://doi.org/10.18257/raccefyn.28(107).2004.1991, 2004.

Poveda, G., Alvarez, D. M., and Rueda, Ó. A.: Hydro-climatic variability over the Andes of Colombia associated with ENSO: a review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots, Climate Dynamics, 36, 2233–2249, 2011.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.

Restrepo, J. J. and Toussaint, J. F.: Terranes and continental accretion in the Colombian Andes, Episodes, 11, 189–193, https://doi.org/10.18814/epiiugs/1988/v11i3/006, 1988.

https://doi.org/10.5194/egusphere-2026-6736
Preprint. Discussion started: 18 November 2026
© Author(s) 2026. CC BY 4.0 License.



Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, 2017.

Ruiz-Carrascal, D., Acevedo, A. B., Zuluaga, M. D., Ramírez, G., Ruiz, D. F., Avellaneda-Jiménez, D. S., Carmona, A. M., Guzmán, C., Marulanda-Tobón, A., Puerta-Yepes, M. E., et al.: The Early Warning System of the City of Medellín and the Aburrá Valley- SIATA, Colombia: a+ 20-year prolific journey, Climate Services, 42, 100 653, 2026.

Sánchez Patiño, O. I.: Appendix_B_Thesis_Master_Oscar_Sanchez, https://doi.org/10.5281/zenodo.22243857, video/audio, version v1, 2026a.

Sánchez Patiño, O. I.: Definition of spatio-temporal rainfall thresholds for landslides in the Aburrá Valley, M.S. thesis, Universidad Nacional de Colombia, Medellín, Colombia, in review, 2026b.

Schmaltz, E. M., Steger, S., and Glade, T.: The influence of forest cover on landslide occurrence explored with spatio-temporal information, Geomorphology, 290, 250–264, 2017.

Steger, S., Moreno, M., Crespi, A., Zellner, P. J., Gariano, S. L., Brunetti, M. T., Melillo, M., Peruccacci, S., Marra, F., Kohrs, R., et al.: Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models, Natural Hazards and Earth System Sciences Discussions, https://doi.org/10.5194/nhess-2022-271, preprint, 2022.

Steger, S., Moreno, M., Crespi, A., Gariano, S. L., Brunetti, M. T., Melillo, M., Peruccacci, S., Marra, F., De Vugt, L., Zieher, T., et al.: Adopting the margin of stability for space–time landslide prediction–A data-driven approach for generating spatial dynamic thresholds, Geoscience Frontiers, 15, 101 822, 2024.

Steger, S., Spiekermann, R., Moreno, M., Lehner, S., Enigl, K., Crespi, A., and Schlögl, M.: Dynamic spatial modelling of mass movement impacts for large areas: a data-driven framework for impact-based early warning, Natural Hazards and Earth System Sciences, 26, 3637–3665, 2026.

Universidad Nacional de Colombia and Área Metropolitana del Valle de Aburrá: Estudios básicos de amenaza por movimientos en masa, inundaciones y avenidas torrenciales en los municipios de Caldas, La Estrella, Envigado, Itagüí, Bello, Copacabana y Barbosa, para la incorporación de la gestión del riesgo en la planificación territorial, Tech. rep., Universidad Nacional de Colombia and Área Metropolitana del Valle de Aburrá, Medellín, 2018.

Urrea, V., Ochoa, A., and Mesa, O.: Seasonality of Rainfall in Colombia, Water Resources Research, 55, 4149–4162, https://doi.org/10.1029/2018WR023316, 2019.

Wood, S. N.: Generalized additive models, Annual Review of Statistics and Its Application, 12, 497–526, 2025.

Zevenbergen, L. W. and Thorne, C. R.: Quantitative Analysis of Land Surface Topography, Earth Surface Processes and Landforms, 12, 47–56, https://doi.org/10.1002/esp.3290120107, 1987.