Engineering Geology 371 (2026) 108860

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

On the use of rainfall time series for regional landslide prediction by means of functional regression

Mahnoor Ahmed [a], Vitor Salomão Recacho [b], Giacomo Titti [c,*], Taro Uchida [d], Lisa Borgatti [c], Adriano Barasal Morales [e], Mirko Francioni [a], Luigi Lombardo [e]

[a] Department of Pure and Applied Sciences, University of Urbino Carlo Bo, Urbino, 61029, Italy
[b] Department of Economics, School of Economics, Business and Accounting at Ribeirão Preto, University of São Paulo, Avenida Bandeirantes 3900, Ribeirão Preto, Brazil
[c] Department of Civil, Chemical, Environmental and Materials Engineering, Alma Mater Studiorum University of Bologna, Bologna, 40126, Italy
[d] Institute of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan
[e] University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), PO Box 217, Enschede, AE 7500, Netherlands

ARTICLE INFO

Dataset link: SU based covariates (Original data)

Keywords:
Rainfall-induced landslide
Functional regression
Japan
Early warning system
Space-time prediction

ABSTRACT

Almost fifty years of progress in geoscience have led to the definition of standard approaches for predicting rainfall-induced landslides. These require the approximation of precipitation time series into a scalar cumulative value corresponding to specific time windows. In the meantime, the progress in computer science has led to modeling architectures capable of handling continuous time series for solving similar predictive tasks. The idea behind these tools is that the information contained in a continuous signal is richer than any single summary statistics one can extract from it, thus leading to potentially better performance. This is how we structured this contribution, experimenting on a class of solvers known as functional Generalized Additive Model (FGAM) to predict landslide occurrences in space and time, retaining the combined information of terrain predictors and full rainfall time series. This FGAM is benchmarked against three variations: (i) a functional regression only equipped with the rainfall signal; (ii) a standard GAM with terrain predictors and a scalar representation of the rainfall signal; (iii) a standard GAM only informed with the scalar rain. These benchmarks are tested across a large suite of cross-validations to confirm or reject the assumption on the increased performances. To support such analyses, we tested the four modeling architectures on three separate landslide event inventories triggered by intense precipitation across Japan. The results demonstrate that the main functional regression outperforms the three counterparts. This implies that future development in landslide early warnings should further explore the untapped potential of modern tools inspired to speech/language recognition, a class of models that can use the whole information in complex data cubes, without the need for simplifications such as the those typical of rainfall thresholds.

1. Introduction

Early warnings for rainfall-induced landslides have relied on rainfall thresholds for more than fifty years (Onodera et al., 1974). Rainfall thresholds are defined on the basis of empirical (Glade et al., 2000) or statistical (Satyaningsih et al., 2023) approaches that represent critical rainfall conditions likely to trigger slope failures (Segoni et al., 2018a). These conditions are expressed as a function of rainfall intensity, duration, or accumulated amounts and estimated against the historical occurrence of landslides (Martel et al., 2021). As a result, when the observed or forecasted rainfall exceeds these thresholds, the probability of landslide initiation significantly increases (Crosta et al., 2001). This general approach has been shown throughout the years to produce good predictive performance and a message that can be intuitively received by the population at risk: "if it rains more than a certain amount over a certain time window, this is the level of alert to which we must adhere". This is why rainfall thresholds have been so successful globally, in any mountainous landscape. However, a main limitation affects the standard use of rainfall threshold. It corresponds to the use of the rainfall information separately from the terrain characteristics. In short, rainfall thresholds are estimated independently of the underlying landslide susceptibility, and only in a subsequent moment the two products are often combined together (see, Segoni et al., 2018b; Cao et al., 2022; Huang et al., 2022). This is a technical issue because statistically, the simple multiplication of probabilities (or parameters)

* Corresponding author.
E-mail address: giacomo.titti@unibo.it (G. Titti).

https://doi.org/10.1016/j.enggeo.2026.108860
Received 21 January 2026; Received in revised form 18 April 2026; Accepted 27 May 2026
Available online 3 June 2026
0013-7952/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

to estimate a joint probability is valid only if the variables are mutually independent (Casella and Berger, 2002). In reality, landscape characteristics largely influence the spatio-temporal distribution of rainfall. For instance, the notion of orographic control on precipitation states that topography can act as barriers focusing the rainfall discharge on certain areas rather than others (Bárdossy and Pegram, 2013; Page et al., 2022). Similarly, vegetation is responsible for evapotranspiration and with it, the availability of moisture in the air to be converted back into rain as specific meteorological conditions are met (Guenni and Hutchinson, 1998). The relevance of such landscape properties in explaining the distribution of precipitation over space and time is documented in any weather modeling research where rainfall estimates are interpolated also as a function of such parameters (Chen et al., 2024). The very same parameters appear in the landslide susceptibility literature, with the spatial distribution of occurrence probabilities being unanimously estimated by terrain predictors (Budimir et al., 2015). This is why rainfall thresholds and susceptibility cannot be assumed to be mutually independent, and their posterior combination potentially constitutes a problem from a strict statistical perspective. Therefore, on their own, rainfall thresholds may not provide the best prediction basis for issuing alert levels for landslides, while their standard combination with susceptibility should also be reconsidered for the reasons presented above.

The very same reasoning has recently led scientists to directly integrate rainfall into susceptibility protocols, thus moving away from the static interpretation of the susceptibility, which has traditionally been assumed since the early work of Brabb (1980).

The result has been the definition of a dynamic susceptibility framework (Samia et al., 2019) based on space–time modeling principles (Lombardo et al., 2020). The underlying idea is that both rainfall thresholds and susceptibility modeling share a core element, the binary expression of landslide occurrences into presences and absences. The main difference between the two is that rainfall threshold approaches always handled this binary information in space and time albeit essentially using only rainfall intensity and duration parameters. As for the susceptibility approach, it traditionally handled presence/absence instances in a multivariate framework, though only accounting for the spatial dimension in the data. A space–time modeling approach essentially takes the best of the two, with the modeling target still being expressed into landslide presence/absences, and the multivariate framework typical of susceptibility models being extended to include rainfall (Dahal et al., 2024b). This is why numerous space–time experiments have appeared in recent years, each one targeting different aspects of landslide predictions.

In essence, one can consider the space defined by response and parameters as a large data cube, where the planar level represents the spatial dimension of the landslide data and the height represents the temporal one. As a result, different approaches can be taken to represent these dimensions. In good alignment with the historical literature in landslide susceptibility, space–time examples have investigated various ways of dissecting the landscape, from the most common grid-cells (Steger et al., 2022, 2024), to geomorphologically-oriented partitions such as slope units (Chang et al., 2023; Fang et al., 2024) and catchments (Wang et al., 2022, 2024). In a similar way, the temporal dimension has been tested on a hourly (Moreno et al., 2025), daily (Nocentini et al., 2023), monthly (Wang et al., 2025) and yearly (Dahal et al., 2024a) basis. Notably, all these contributions but one, share the same assumption: the rainfall is aggregated at the scale of the selected temporal unit. This is a procedure that originates from standard rainfall thresholds applications where the effect of the rainfall over time is expressed into its intensity and duration. In a similar manner, most of the space–time landslide prediction models tested so far have aggregated the rainfall into a scalar value, by taking the sum over a specific time window. The same can be seen in the work of Moreno et al. (2024), where a sensitivity analysis explores different aggregation windows for preparatory and triggering rain. Because of this aggregation, no matter how rich or complex the sensitivity analysis may be, the original complexity of the rainfall time series is inevitably simplified. Following the very same observation, few contributions have attempted to retain the complexity of the rainfall signal in space–time landslide predictive models. Fang et al. (2023) represents the first example of this sort, whereby techniques inspired by speech-recognition have been employed to pass entire precipitation signal to a classification model and produce space–time landslide occurrence probabilities across Türkiye. However, their experiment was still tied to the concept of rainfall thresholds, which is why the authors only used rainfall as a predictor, though as continuous time series. Subsequently, the same idea has been tested by Lim et al. (2024) in Vietnam, where the authors also implemented a space–time neural network architecture capable of handling both the continuous rainfall signal as well as the static contribution of landscape characteristics. In both cases, the scalar benchmark was largely superseded by the modeling counterpart processing the rainfall signal as a whole. However, machine learning solutions are often treated as black boxes, even though recent solutions allows for more transparent results (Caleca et al., 2024). Beyond issues of interpretability, deep learning architectures for time-series classification such as Long Short-Term Memory (LSTM) or convolutional networks generally demand extensive labeled datasets to achieve robust generalization across different regions. This prerequisite is seldom satisfied in landslide research, where event inventories with accurate timing are still limited (Qiu et al., 2024, 2025). In addition, these models do not enforce any physically grounded constraints on the temporal organization of the learned features, so the influence of individual days within the rainfall record on the predicted hazard cannot be clearly extracted or given a geophysical meaning. In contrast, statistical functional regression yields an explicit, smooth coefficient function $\beta(t)$ defined over the rainfall time series (see Fig. 3), making it possible to directly identify which portions of the temporal profile exert the strongest hydrological control, which is already a useful insight for designing early warning systems (Zhao et al., 2025). This is why Moreno et al. (2025) opted to test a statistical architecture, known as functional regression (McLean et al., 2014), to model landslides over space and time as a function of static terrain characteristics as well as the dynamic signal of rainfall time series. The authors tested four variations, one where only the rainfall was passed to the model, once in scalar form and once in its functional form. The other two variations essentially stayed the same but adding landscape predictors. Interestingly, the results proved that the use of a scalar and functional rainfall produced significantly better performance only for the benchmark without terrain predictors. When the model featured both rainfall and terrain characteristics, over larger time windows the performance converged to the same level irrespective of the scalar or time series expression of the rainfall signal. This being a stark difference with respect to the other two experiments mentioned above, we started defining the research objective in this work. Our assumption is that the two examples mentioned above where the performance difference was large, both tested very long time series. On the contrary, when the performance difference between a model equipped with a scalar and a functional rainfall was small, the length of the time series was defined over just six days. This is why we planned to test again the functional regression approach introduced by Moreno et al. (2025), but looking over a much longer time window leading to landslide records. To this research plan, we also added few nested experiments, whereby the considered landscape was selected to present a significant variability, and the validation framework was tested over a wide range of data arrangements.

A case study where our group was working on, fit very well this plan. Specifically, as Japan is often hit by major rainstorms, we accessed three landslide inventories, separated in space and time, whose combination yielded a large data cube. By partitioning the respective areas into Slope Units (SUs), we then proceeded to implement: a binomial Generalized Additive Model (GAM; Hastie, 2017) to replicate

an experiment where the rainfall signal is aggregated into a scalar value, once passed (GAM_2) bivariately to the model, and once passed (GAM_1) multivariately alongside terrain predictors. In parallel, we also implemented a binomial functional regression model (FGAM; Greven and Scheipl, 2017), to replicate the experiment where the continuous rainfall time series are passed either ($FGAM_2$) bivariately or ($FGAM_1$) multivariately alongside static landscape properties.

In the remainder of the manuscript, the landslide event inventories of Japan will be detailed in Section 2 together with the spatial and temporal units of choice and related predictors. Section 3 will further explore the functional regression modeling strategy that we employed and the associated diagnostic for benchmarking. Section 4 will then detail the results, whose interpretation will be discussed in Section 5. Ultimately, Section 6 will summarize our views on the future direction and present our concluding remarks.

2. Data overview

2.1. Multitemporal landslide inventory

Landslide inventory data has been retrieved from events documented at the Geospatial Information Authority of Japan (GSI), which surveys and maps disasters resulting from natural hazards, including heavy precipitations, earthquakes, and volcanic activity (https://www.gsi.go.jp/top.html). Specifically, three rainfall-triggered landslide inventories were utilized in this study. The first inventory corresponds to Tropical Storm Talas that brought heavy rainfall to the Kii Peninsula of Japan from 1st to 4th September 2011 (Fig. 1A). Talas resulted in 97 casualties and several thousands of damaged houses upon making landfall. According to the Japan Meteorological Agency, rainfall measurements surpassed previous records of this region at several rain gauge stations at Nara, Mie and Wakayama prefectures, the affected areas of the Kii Peninsula. The three prefectures are connected by the high-relief Kii Mountains whose bedrock is mainly composed by a Cretaceous accretionary complex (Saito and Matsuyama, 2012). Thousands of landslides were mapped via post-aerial photograph interpretation (Uchida et al., 2012) by the National Institute for Land and Infrastructure Management of Japan (NILIM). Marc et al. (2018) redefined the existing inventory using Google Earth and Landsat 5 images to remove landslides mapped before the typhoon, as well as adding new polygons that were not covered in the aerial survey by NILIM (Marc et al., 2018, 2022). After cleaning distorted and amalgamated polygons, the inventory contained 1902 landslide polygons. This area is the largest among the study sites with widespread slope failures including deep rockslides, debris flows and shallow landslides (Saito and Matsuyama, 2012).

The second inventory is present in the Noto Peninsula in Western Japan where heavy rainfall from September 20th to 22nd 2024 triggered several landslides that GSI documented on 23rd and 24th September (Fig. 1B). The inventory was retrieved from GSI as 1540 landslide polygons. The geological makeup of the Peninsula is mainly consisting of pre-Tertiary basement rock characterized by Hida metamorphic rocks along with Tertiary formations of Funatsu granitic rocks (Acharya et al., 2024). In the southern area of the Noto Peninsula, these basement rocks are exposed at the surface, while the central and northern regions are covered by Neogene deposits (Honda et al., 2008).

The third inventory is located in the Miyagi and Fukushima region which represents landslides mapped by Typhoon Hagibis making landfall on 12th October 2019. The resulting 847 landslide polygons which were mapped are available to download from the GSI database (Fig. 1C). The sediment disaster brought by the typhoon was considered to be the largest since 1982 when the statistics started to be recorded with 16 casualties. The affected areas lie in the vicinity of the Kitakami mountains which are overall constituted with the Paleozoic and Mesozoic strata within granite intrusions (Irasawa et al., 2017). The rest of the investigated area is characterized by Kitakami granites of the Early Cretaceous, while the west part mainly by Abukuma granites. Miocene basalt intrusions are also present. The principal landslide type that affected the regions were debris flows and shallow landslides (Irasawa et al., 2020). In this study only debris flows and shallow landslides have been considered.

2.2. Data cube construction

The data matrix we built included observations across all three regions in the form of static and dynamic covariates. The total number of observations in the data matrix, where each observation represents a mapping unit, summed up to 64 250, out of which 49 306 belong to the Kii region, 6056 represent the Noto Peninsula and 8888 are of the Miyagi region. The information across these observations includes 5 static properties and one dynamic functional predictor.

We have selected Slope Units (SUs) as the mapping unit to which the covariate information was aggregated (Alvioli et al., 2016). SUs have been widely used in landslide studies to understand unstable slopes on a regional-scale analysis (Ahmed et al., 2025; Lombardo et al., 2020). In this work, they have been generated via *r.slopeunits* developed by Alvioli et al. (2016). In particular, in the current contribution we integrated *r.slopeunits* with the routine proposed by Ahmed et al. (2025) to numerically optimize the parameter configuration using the function implemented in the SZ-plugin of QGIS (Titti et al., 2025). In total, 18 SU partitions of each study area have been delineated using a matrix of combinations of circular variance of 0.1, 0.4 and 0.7, a minimum area between 10 000 m^2 and 200 000 m^2 and with an initial threshold of 500 000 m^2. The final selection of SU partition for Kii Peninsula is described with minimum area of 20 000 m^2 and 0.3 circular variance, the one for Noto Peninsula is with minimum area of 100 000 m^2 and 0.3 circular variance and the Miyagi region with 200 000 m^2 minimum area and 0.1 circular variance. Fig. 2 shows the summary statistics of the SU partitions, the percentage of unstable SUs in the study areas is 3%, 10% and 4% in Kii, Noto and Miyagi regions, respectively.

The covariates we selected can be grouped in static and dynamic, the former can be assumed time-invariant while the latter time-dependent. All the morphological covariates that we selected, were considered static. They were derived from the GSI DEM, freely accessible through the Basic map information download service of Japan (https://service.gsi.go.jp/kiban/). The GSI DEM is the product of multi resolution aerial survey, photometric survey and topographic maps. The downloaded tiles were converted to GeoTIFF using the QuickDEM4JP plugin (Mohammadreza, 2025) available in QGIS.

We derived the slope steepness (hereafter called Slope), slope aspect and profile curvature from the DEM. The slope steepness is a useful predictor that assesses the gravitational force along the slope for its failure potential (Budimir et al., 2015). Aspect was decomposed into Eastness and Northness converting the native radial information into two linear and ordinal covariates (Ahmed et al., 2023; Titti et al., 2022). Profile curvature was also derived from the DEM to serve as a representation to measure acceleration or deceleration along the slope direction (Saleem et al., 2019). In addition to the morphological covariates, we included the Lithology as a static variable. The map was accessed from the Geological Survey of Japan at a scale of 1:200,000 (https://www.gsj.jp) and was simplified into four major classes which represent sedimentary rocks, accretionary complexes, igneous rocks and metamorphic rocks.

To investigate the triggering factor, we integrated daily precipitation from CHIRPS-GEFS (Harrison et al., 2022). This is a rainfall forecast that provides a daily prediction based on the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System version 12 (GEFS v12) precipitation forecasts, compatible with the Climate Hazards center InfraRed Precipitation with Stations (CHIRPS) to support early warning (Harrison et al., 2022). The rainfall time-series represents the daily cumulative measurement which has been aggregated across 34 days at a spatial resolution of 5 km, currently

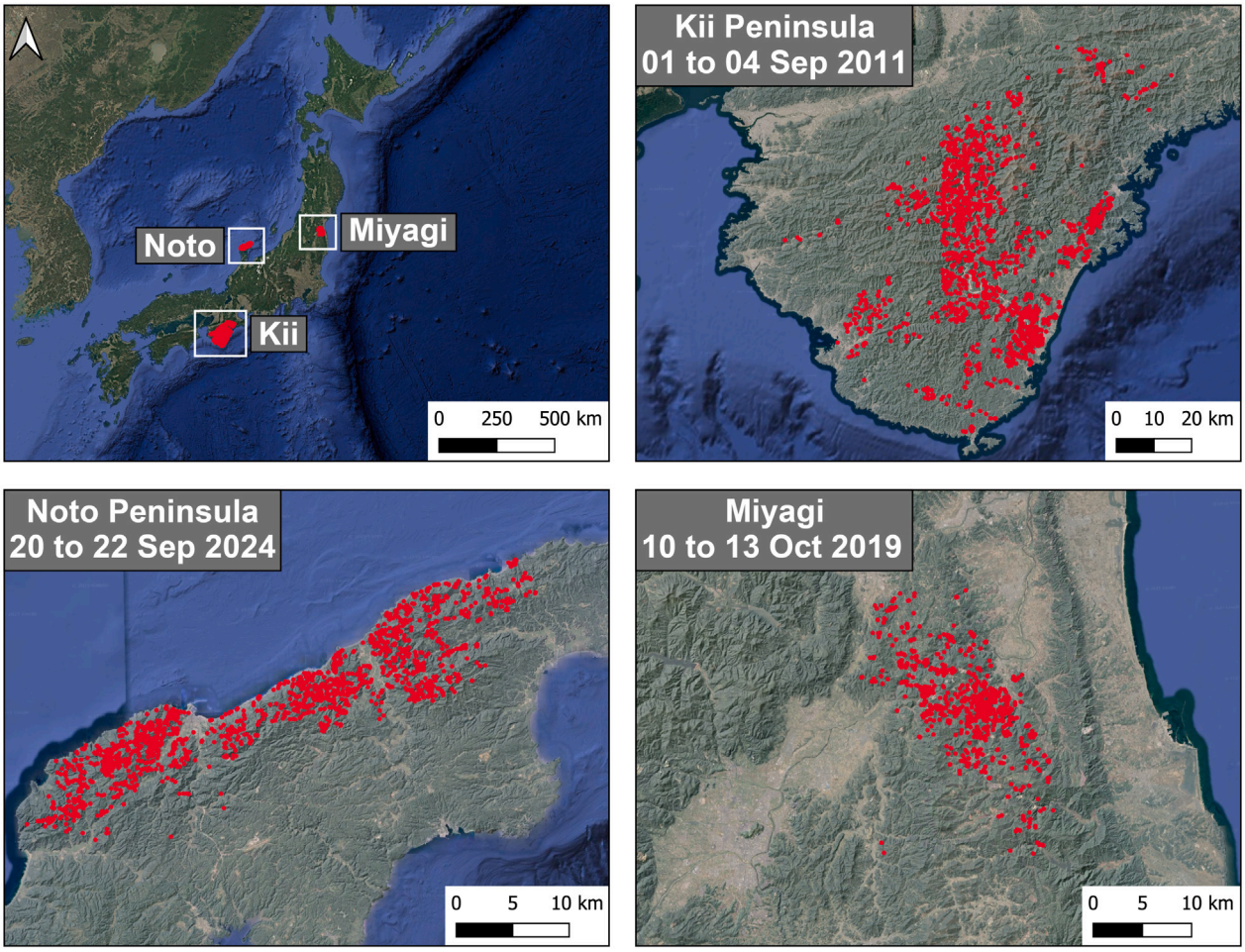


Fig. 1. Overview of the locations across the Japanese landscape where three landslide event inventories have been mapped.

a high-resolution operating forecast dataset. The rainfall events that triggered the landslides in these regions are caused by typhoons or storms, which can last for days. The event day is referred to as the last effective rainy day. The functional predictor is built by using the information of the event day and 33 days prior, thus this period essentially covers the storm-even in its full temporal span while also accounting for antecedent preparatory rainfall.

3. A functional GAM regression strategy

To accurately model landslide phenomena, this study utilized a Functional General Additive Model (FGAM) (Mclean et al., 2014), an extension of the more basic Generalized Linear Models (GLMs). The choice of Functional Generalized Additive Models (FGAM) is justified by their ability to relax the assumptions inherent in linear models. Traditional Generalized Linear Models (GLM), which are commonly employed for modeling landslide susceptibility (Atkinson et al., 1998), exhibit certain limitations attributable to their structural design, notably the assumption of a linear relationship between predictors and the log-odds of landslide occurrence. General Additive Models (GAMs) amend this by allowing the expected value of the response variable to be contingent upon smooth monotonic functions of the linear predictor, with splines serving as these functions in this context. In contrast, the FGAMs approach not only allows for the incorporation of smooth functions such as splines, but more broadly, of functional terms capable of functioning within an n-dimensional framework. In statistics, this property is usually exploited to regress a given response variable against time series. In the context of landslide applications, therefore, the same can be done with respect to rainfall time series, facilitating the representation of the nonlinear temporal dynamics intrinsic to the landslide process.

Notably, both Generalized Additive Models (GAMs) and Functional Generalized Additive Models (FGAMs) can model response variables originating from any distribution of the exponential family. This capability is particularly critical for the present study, as landslide events are encoded as a binary variable, with a response of 1 indicating occurrence and 0 indicating non-occurrence, a data structure often assumed to arise from the Bernoulli case (Bergtold et al., 2010).

In comparison to machine learning methods, the FGAM possesses an additive structure, facilitating the isolation and interpretation of each covariate's respective contribution. This characteristic starkly contrasts with many machine learning approaches frequently operating as black-box models, thus devoid of parameter interpretability.

Beyond these advantages, we stress once more that the principal justification for employing Functional Generalized Additive Models (FGAM) is their capability to incorporate a functional term as a predictor (Kosiorowski, 2014; Mclean et al., 2014; Wood, 2006). This facilitates modeling complex patterns, including spatio-temporal covariates. Such a capacity is particularly significant in contexts where phenomena such as landslides are influenced not solely by individual rainfall events but rather by the cumulative precipitation patterns observed over preceding days or months. The ability of FGAM to model these space–time series effects, such as rainfall, is therefore critical. Technically, this can be conceptualized as a heat map at each temporal instance, resulting in s-dimensional parameters for each covariate period. Ultimately, this leads to a parameter space of s-dimensions

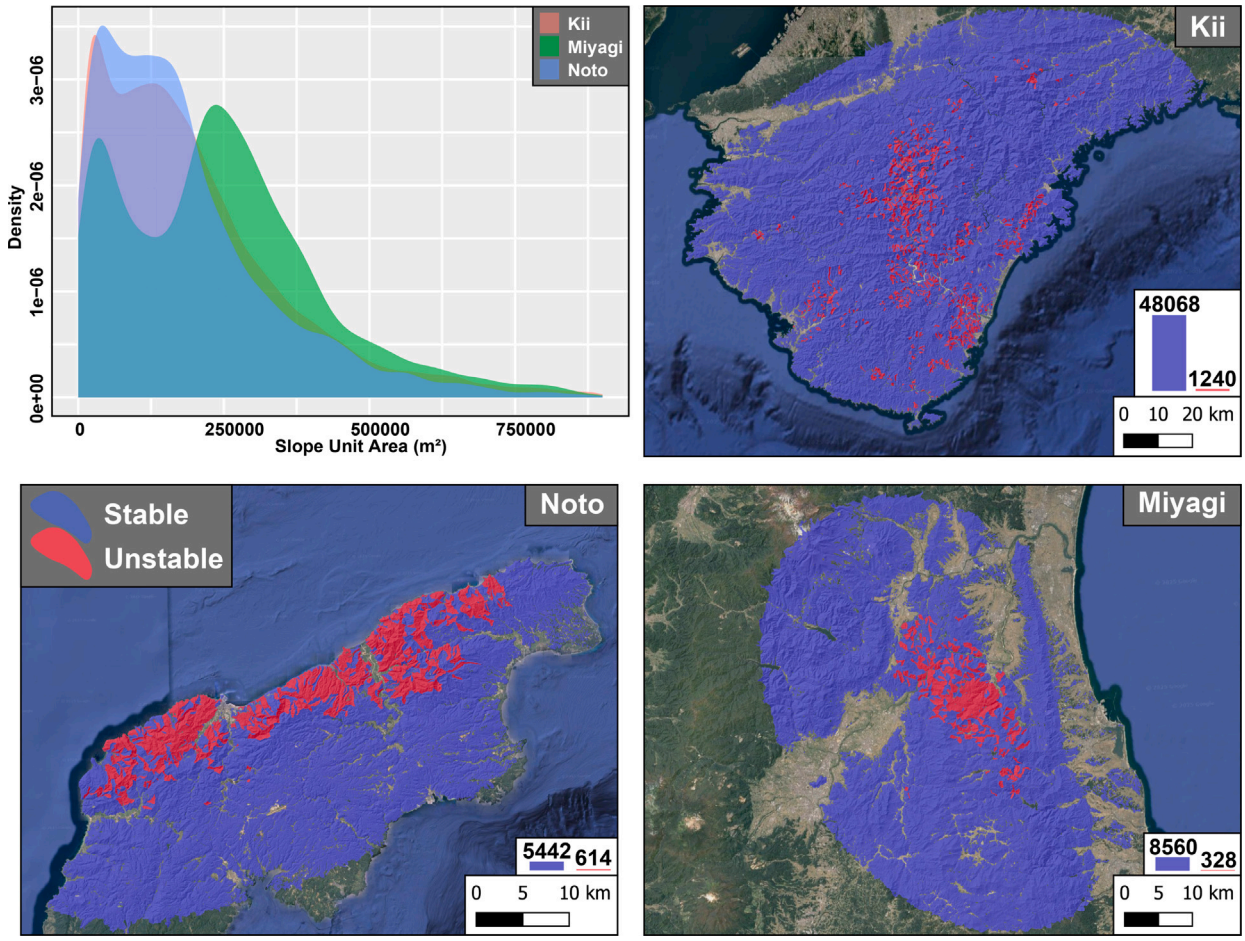


Fig. 2. Summary statistics of the SU partition size and relative proportion of labeled stable/unstable slopes.

multiplied by t-dimensions. A functional term is characterized by the $\mathcal{L}^2$ inner product of a function of coefficients (exhibiting infinite dimensionality), which is typically represented using known basis functions, alongside a function derived from the data. The rationale for expanding the function of coefficients in terms of basis functions lies in its membership within the $\mathcal{L}^2$ space, which inherently possesses infinite dimensions. However, the dimensionality of the function is constrained by the number of observations, rendering the estimation of all terms within the inner product infeasible. By expanding these terms into K basis functions, the challenge of dealing with the inner product is transformed into a manageable problem of K-dimensional order.

As previously stated, the response variable is binary, necessitating the utilization of a Bernoulli distribution to model the presence or absence of landslides. To achieve this, it is necessary to employ a link function, which in this instance, is expressed as follows:

$$\text{logit}(p_{s,t}) = \log\left(\frac{p_{s,t}}{1-p_{s,t}}\right) \tag{1}$$

This is necessary because the smooth terms of Functional Generalized Additive Models (FGAM) exhibit high flexibility, and the sum of the additive components, denoted as $\phi_{s,t}$ in Eq. (2), is not constrained to the interval [0,1]. Therefore, it is essential to transform these components back to this interval to ensure they represent valid probabilities.

Based on the previously presented information, it is possible to introduce the Functional Generalized Additive Model (FGAM) to be employed in this research. This model will incorporate linear covariates, smooth covariates, and functional covariates, as shown below:

$$Y(s,t) \sim Bernoulli(p_{s,t})$$

$$\begin{aligned}
\text{logit}(p_{s,t}) =& \phi_{s,t} \\
\phi_{s,t} =& \alpha + \beta_1 \cdot \text{eastness}_s + \beta_2 \cdot \text{northness}_s + \boldsymbol{\gamma}^\top \text{litho}_s + \\
& f_1(\text{slope}_s) + f_2(\text{profcurvature}_s) \\
& + f_3(\text{precipitation}_{s,t}) \\
f_1(\text{slope}_s) =& \sum_{i=1}^{7} \omega_{1,i} B_{1,i}(\text{slope}_s) \\
f_2(\text{profcurvature}_s) =& \sum_{i=1}^{5} \omega_{2,i} B_{2,i}(\text{profcurvature}_s) \\
f_3(\text{precipitation}_{s,t}) =& \iint \sum_{l=1}^{4} \sum_{m=1}^{3} \omega_{3,l,m} B_{3a,l}(s) \\
& \times B_{3b,m}(t) \text{precipitation}_{s,t}(s,t) ds dt
\end{aligned} \tag{2}$$

where, Y represents the observation of the occurrence of a landslide at time t and location s. The function $\text{logit}(p_{s,t})$ serves as the link function, as defined by Eq. (1). The parameter α denotes the intercept of the model. The parameters β_1 and β_2 represent the coefficients associated with the linear covariates of eastness and northness, respectively. The vector $\boldsymbol{\gamma}$ encompasses coefficients corresponding to various lithological types. The function $f_1(\text{slope}_s)$ is a one-dimensional spline with seven basis functions, utilized for modeling the Slope. Similarly, $f_2(\text{profcurvature}_s)$ is a one-dimensional spline with five basis functions for modeling profile curvature. Additionally, $f_3(\text{precipitation}_{s,t})$ is a two-dimensional spline, each dimension comprising four basis functions, employed to model precipitation dynamics. In the remainder of the manuscript the set of equations corresponding to Eq. (2) will be referred to as FGAM_1.

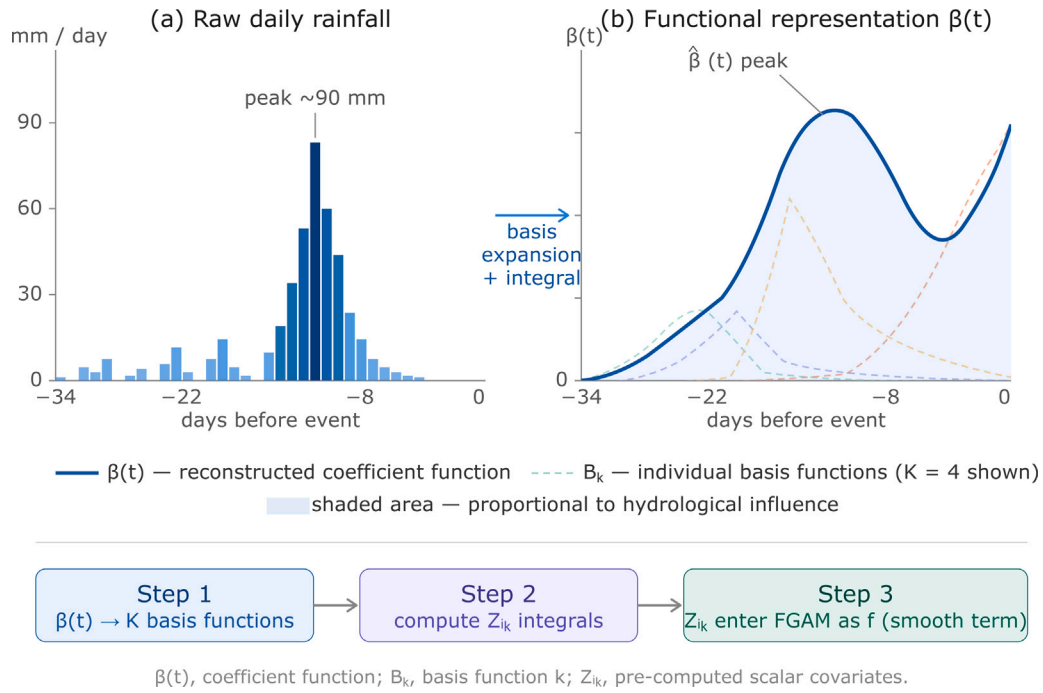


Fig. 3. Schematic illustration of converting a 34-day daily rainfall series into a functional predictor. Panel (a) displays the original discrete measurements. Panel (b) presents the expansion in terms of K B-spline basis functions and the resulting smooth coefficient function $\hat{\beta}(t)$. Taking the integral of the product of the rainfall curve with each basis function produces scalar covariates Z_{ik}, which enter the FGAM as the smooth term f. $K = 4$ basis functions are shown for illustration; the actual K is chosen using the EDF $\ll K$ diagnostic.

Notably, the covariate set described above comes from unreported tests, thorough which we explored a much richer parameter space only to converge to a limited number of covariates for reasons of conciseness and clear interpretation.

In the subsequent section, a comprehensive discussion will be conducted concerning the estimation of the model previously presented.

3.1. Estimation

Similar to Generalized Additive Models (GAMs), Functional Generalized Additive Models (FGAMs) are frequently estimated in a classical way using Restricted Maximum Likelihood (REML) (Mclean et al., 2014). In our analysis, smoothing parameters were estimated via REML using the *refund* package in R, which employs the *mgcv* backend for model estimation. REML was chosen instead of Generalized Cross-Validation (GCV) because GCV is known to favor undersmoothing and may yield multiple local minima, especially when dealing with non-Gaussian responses such as the Bernoulli outcome considered here. In contrast to other techniques such as standard Maximum Likelihood (ML), REML reduces bias in the estimation of smoothing parameters by integrating out the fixed effects prior to estimation, thereby providing more stable and reliable inferences (Wood, 2017). However, FGAMs necessitate the implementation of a penalty due to the functional term being the $\mathcal{L}^2$ inner product, which exhibits infinite dimensionality. To address this issue, two well-established techniques are commonly employed (Goldsmith et al., 2011; Mclean et al., 2014; Reiss and Ogden, 2007). The first involves transforming this inner product into a tensor product and applying Penalized Functional Regression (PFR) to reduce the number of coefficients. The second approach involves utilizing Functional Principal Component Analysis (FPCA), which decomposes the functional data and converts the infinite-dimensional problem into a finite-dimensional one.

The Penalized Functional Regression (PFR) methodology addresses the inherent infinite-dimensional nature of the functional term, denoted as $f(x_i) = \int_{\mathcal{T}} x_i(t)\beta(t)\,dt$. The principal challenge involves the estimation of the unknown coefficient function $\beta(t)$. The PFR method resolves this issue by expressing the coefficient function $\beta(t)$ as a linear combination of a finite set of K basis functions, represented as $b_k(t)$:

$$\beta(t) = \sum_{k=1}^{K} \omega_k b_k(t) \tag{3}$$

where ω_k are the unknown basis coefficients to be estimated. By substituting this basis expansion into the inner product, the functional term is converted from an integral representation into a finite linear sum:

$$f(x_i) = \int_{\mathcal{T}} x_i(t) \left(\sum_{k=1}^{K} \omega_k b_k(t) \right) dt \tag{4}$$

$$= \sum_{k=1}^{K} \omega_k \left(\int_{\mathcal{T}} x_i(t) b_k(t) dt \right) \tag{5}$$

The pre-calculation of the K integral components for each observation i can be conducted in advance. Let us denote $Z_{ik} = \int_{\mathcal{T}} x_i(t) b_k(t)\,dt$. These Z_{ik} values constitute a new set of scalar covariates derived from the functional predictor x_i. Consequently, the functional term is redefined as:

$$f(x_i) = \sum_{k=1}^{K} \omega_k Z_{ik} \tag{6}$$

The transformation described herein is pivotal: it effectively reformulates the originally infinite-dimensional functional problem into a conventional, finite-dimensional parametric representation. Nevertheless, the function $\beta(t)$ retains its smooth functional characteristics. To ensure this smoothness and mitigate the risk of overfitting, a penalty is imposed on the coefficients. This penalty is derived from the "wiggliness" of $\beta(t)$, commonly quantified by its squared second derivative, $\int (\beta''(t))^2 dt$. This results in a quadratic penalty applied to the coefficients:

$$\lambda \int (\beta''(t))^2 dt = \lambda \boldsymbol{\omega}^{\mathrm{T}} \mathbf{S} \boldsymbol{\omega} \tag{7}$$

where $\boldsymbol{\omega}$ is the vector of coefficients $[\omega_1, \ldots, \omega_K]^T$, $\mathbf{S}$ is the known penalty matrix derived from the basis functions, and λ is the smoothing parameter governing the balance between model fit and smoothness. For models containing J smooth components, each component is associated with its own smoothing parameter λ_j.

With the resolution of the infinite-dimensional problem, attention is directed towards the REML estimation method. The REML approach characterizes each unknown smooth function in the model, f_j for $j = , \ldots, J$, where J denotes the total number of smooth terms, as a linear combination of basis functions:

$$f_j(x_j) = \sum_{k=1}^{K_j} \beta_{jk} b_{jk}(x_j) \tag{8}$$

where b_{jk} represent the basis functions and β_{jk} denote the unknown coefficients for the jth smooth term. This formulation is equivalent to the PFR basis expansion in Eq. (6), enabling the use of standard estimation frameworks. The number of basis functions K_j for each smooth term was selected by examining the Effective Degrees of Freedom (EDF) in relation to the maximum degrees of freedom implied by K_j. In particular, a chosen value of K_j was regarded as sufficient when the estimated EDF were clearly smaller than K_j, indicating that the penalty was successfully governing the smoothness and that the model fit was not being unduly restricted by the basis dimension. Smooth effects for the scalar covariates were modeled using cubic regression splines, and tensor product smooths were used for the two-dimensional functional precipitation term, consistent with the default settings of the *refund* and *mgcv* packages. Other basis choices were considered during model construction; however, no formal, systematic comparison was carried out, as differences in fit were minimal and the adopted specification yielded stable and easily interpretable estimates.

For each smooth function, we associate at least one functional that quantifies the degree of smoothness, represented by the expression $\boldsymbol{\beta}_j^T \tilde{\mathbf{S}}_j \boldsymbol{\beta}_j$. Here, $\tilde{\mathbf{S}}_j$ denotes a matrix of known coefficients. The measure of smoothness typically assesses entities such as spline penalties; however, it can also encompass more intricate constructs such as tensor product smooth penalties (Wood, 2006), which is the scenario pertinent to our model. To comprehend the workings of REML, consider a straightforward penalization in a one-dimensional spline, represented by $\int f_j''(x)^2 \, dx$. The model can then be estimated by minimizing the penalized log-likelihood:

$$\min_{\boldsymbol{\beta}} \left\{ -l(\boldsymbol{\beta}) + \sum_{j=1}^{J} \lambda_j \boldsymbol{\beta}_j^T \tilde{\mathbf{S}}_j \boldsymbol{\beta}_j \right\} \tag{9}$$

where l represents the log-likelihood, while λ_j denote the smoothing parameters that control the trade-off between the objectives of achieving an optimal fit and maintaining model smoothness (Wood, 2017).

In the subsequent section, a discussion will be provided on the cross-validation methodology employed in this study for model validation. Additionally, alternative models utilized for benchmarking purposes will be presented.

3.2. Benchmarking

To evaluate the performance of the $FGAM_1$ model developed in this study, we employed cutoff-independent and cutoff-dependent techniques. We examine whether our hypothesis, which incorporates rainfall time series data, provides superior predictive accuracy compared to the conventional landslide model that solely utilizes the absolute accumulation of rainfall prior to the event.

Before understanding the benchmarking process in detail, it is necessary to discuss the models that were used to test the performance. The first model is the GAM model that only uses the cumulative rainfall data antecedent the landslide events (without a temporal dimension in this model). Referred as GAM_2 in Eq. (10), it is a conventional model often used for global forecasting of landslides. The second model is also a GAM model that in addition to cumulative rainfall also includes the same static covariates which are used for the $FGAM_1$ model as shown in Eq. (2). The second GAM is called GAM_1 (see Eq. (10)). The third model is similar to the FGAM shown in Eq. (2), except for the static covariates which are absent. It includes only the precipitation function term. The model is called $FGAM_2$ (see Eq. (10)). The algebraic structure of the benchmark models are encapsulated in the following equation:

$$\begin{aligned} GAM_1 =& \alpha + \beta_1 \cdot \text{eastness}_s + \beta_2 \cdot \text{northness}_s + \gamma \text{litho}_s + \\ & f_1(\text{slope}_s) + f_2(\text{profcurvature}_s) + f_3(\text{precipitation}_s) \\ GAM_2 =& \alpha + f_3(\text{precipitation}_s) \\ FGAM_2 =& \alpha + f_3(\text{precipitation}_{s,t}) \end{aligned} \tag{10}$$

The choice of this model for comparison is due to the GAM_2, which is a commonly utilized model globally for emergency systems. Therefore, if we aim to propose a new model, it must demonstrate superior performance compared to this standard. The GAM_1 and $FGAM_2$ models are employed to evaluate whether the covariates exert an influence on the model.

The initial examination involves determining the relevance of incorporating time-series data. Subsequently, simulations are conducted for $FGAM_1$ and $FGAM_2$ using an incremental portion of the time-series data, commencing from the earliest available date and progressively increasing until the entire 34-day period is utilized. For GAM_1 and GAM_2, given their reliance solely on accumulated precipitation data prior to the landslide event, the simulation begins with the cumulative rainfall on day 0, subsequent day up to day 34.

The initial cutoff-independent procedure implemented is random cross-validation, initiated with ten partitions beginning at ratios of 5%, 10%, 20%, and proceeding incrementally to 90% of the dataset designated for training. The remaining portion is reserved for validation purposes, its made 100 times replicated bootstrap. Consequently, for each model at every specified ratio value, we obtain a distribution characterized by its mean and variance. The subsequent methodology is Leave-one-out cross-validation. Within the dataset, we possess data on landslides precipitated by three distinct storm events. The procedure involves selecting two storm events for the purpose of model training, while the remaining storm event is utilized as test data.

The benchmarking process, which is dependent on cutoff values, is employed to assess the accuracy of each model in the classification of true positive and true negative outcomes of observations. This procedure is executed using a bootstrap method replicated 100 times. For each replication, across varying ratios and models, the true positive rate and true negative rate are calculated.

4. Results

The strength of statistical modeling has always been the clarity it offers when it comes to interpreting the role of covariates in the estimation of the response variable. This is valid also in the case of our space–time functional regression model, which is why we begin this section with an overview of the covariate effects. Fig. 4 presents each $FGAM_1$ model component, with individual effects being already expressed as marginal probabilities (hence after applying the logit transformation to the estimated regression coefficients). The first two panels showcase the contribution of Slope and Profile curvatures as splines, whereas the third panel highlights the categorical effect estimated for the Lithology, followed by the linear effects of Eastness and Northness. The last panel corresponds to the functional effect estimated for rainfall, with the two axes expressing the joint intensity and duration contributions to the marginal probability.

From a geomorphological perspective, all the selected covariates present interpretable effects. The Slope shows an overall near-sigmoidal behavior, whereby up to steepness values of around 10° are associated with low occurrence probabilities. From this range until ~35°, landslide

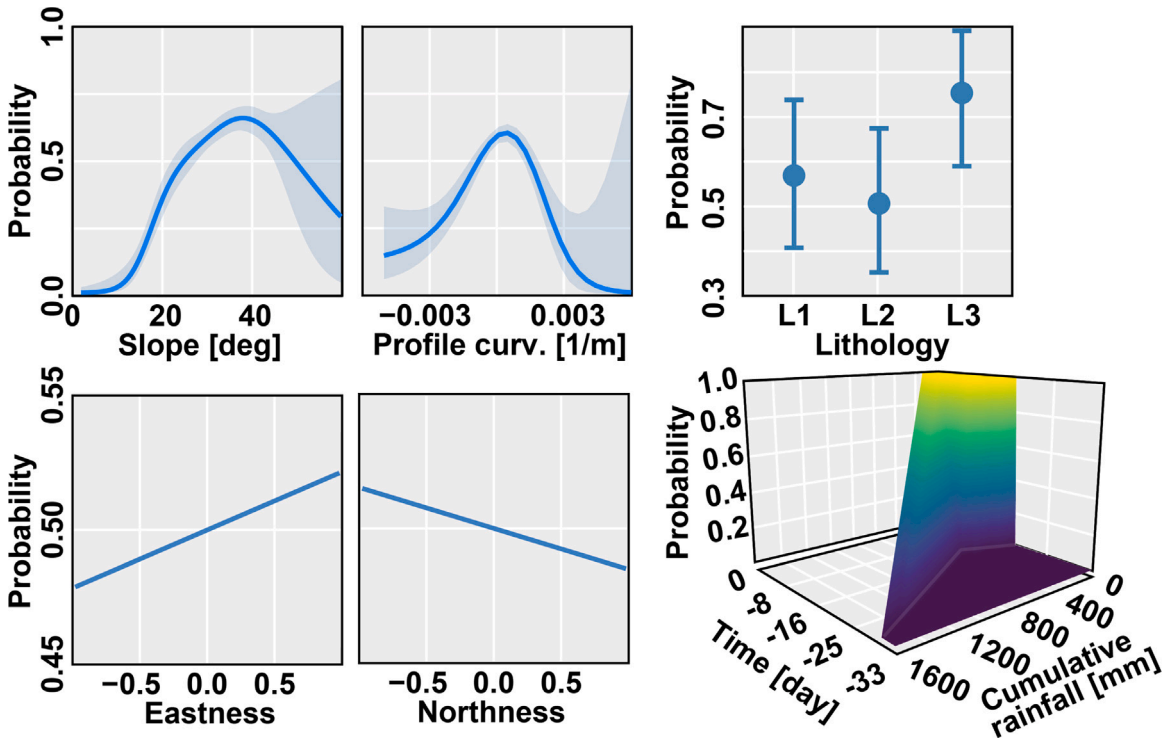


Fig. 4. Covariates' effects expressed in the probability scale of $FGAM_1$ model. The Lithology acronyms should be read as: L1 for Sedimentary Rocks; L2 for Igneous Rocks; L3 for Accretionary Complexes.

occurrence probabilities increase almost linearly with steepness, only to find an inflection right after, until the maximum steepness calculated for all SUs, which corresponds to a probability of approximately 0.3. This is a reasonable behavior that matches the considered landslide type, primarily in the form of shallow landslide and debris flows. In fact, low steepness is well known to represent stable conditions, hence the low probabilities. As for higher values, the range corresponding to the central linear trend appears also geomorphologically sound, and the same is valid for the last slope range where the steepness becomes too large to be even hosting a soil column or thin unconsolidated deposits. We recall here that the absence of soils is usually assumed for near-vertical steepness values. However,it is virtually impossible to produce near-vertical SUs, which is why we interpret the spline effect for the slope consistent with physical assumptions. In the spline estimated for the Profile curvature, the effect reaches a peak at a probability of around 0.55 corresponding to near planar terrains. As for the negative and positive curvatures, respectively corresponding to convex and concave terrains, landslide probabilities sharply decrease to their minimum. This is an effect consistent to the nature of slope instabilities mapped across the three inventories. As for the Lithology control on slope stability, Accretionary Complexes appear to be the only class significantly associated to slope instability, as opposed to Sedimentary and Igneous rocks. Notably, as we used Lithology as a factor variable, the effect of Metamorphic Rocks is not reported as it internally constitutes the reference factor class. The linear contributions of Eastness and Northness promote instabilities in SUs, respectively facing East and South. For these two, a clear interpretation is more difficult to provide, as aspect is well known in the literature to potentially act as a proxy for strata attitude, but also for dry and wet conditions due to sunlight exposure. The last panel of Fig. 4 carries part of the novelty of this manuscript, with the marginal probabilities varying as a continuous function of both rainfall intensity and duration. What stands out the most is that if one looks exclusively at the oldest day before the landslide occurrence date (day −33), no matter the rainfall intensity, the landslide occurrence probability is invariably zero. However, as the rainfall time series progressively approaches the reference day of the landslide occurrence, one can notice a sharp increase in probabilities. Interestingly, the maximum of the effect is reached around one or two days before the landslide date. We recall here that the three inventories we considered were all mapped after three separate rainstorms, each reportedly lasting around three days. In this sense, the probability peak estimated not at the landslide occurrence date but rather a couple of days before seems perfectly reasonable.

Having demonstrated that the covariates' effects cannot be rejected from a physical standpoint, we proceeded with testing our model capabilities against some benchmark runs and across various cross-validation strategies. The first test is summarized in Fig. 5 where we explore the model performance at varying lengths of the considered time window. In doing so, we recall here that our tests featured four architectures, one where the functional regression is the same as before ($FGAM_1$), one where only the continuous rainfall signal is passed to the functional regression model ($FGAM_2$), and two corresponding non-functional GAM variations (GAM_1 and (GAM_2)) where the rainfall is aggregated into a single scalar value. The information conveyed by Fig. 5 is quite straightforward, with GAM_2 being consistently the model with the lowest performance, reaching the AUC peak at 29th day before the landslide occurrence. The same date is also where GAM_1 reaches its maximum performance, with the most interesting aspect being the starting point of the two respective lines. In fact, GAM_2 starts at an approximate AUC of 0.73 whereas GAM_1 starts at around 0.80. This large performance difference is exclusively brought by static terrain predictors. Looking into the functional extensions, an aspect needs to be remarked, as both lines begin one day before the landslide occurrence. The reason for this is because one rainfall time point alone would not constitute a time series, whereas this consideration does not apply to the GAM counterparts, as they naturally deal with single rainfall values anyway. Another interesting observation is that not only the two functional models are essentially outperforming the corresponding models with aggregated rainfall, but they are also much more stable. In Fig. 5, the two GAM architectures show an erratic behavior over time. The performance mostly drops up to ten days before the landslide date, recovering after that up to day −29, only to drop once again at the

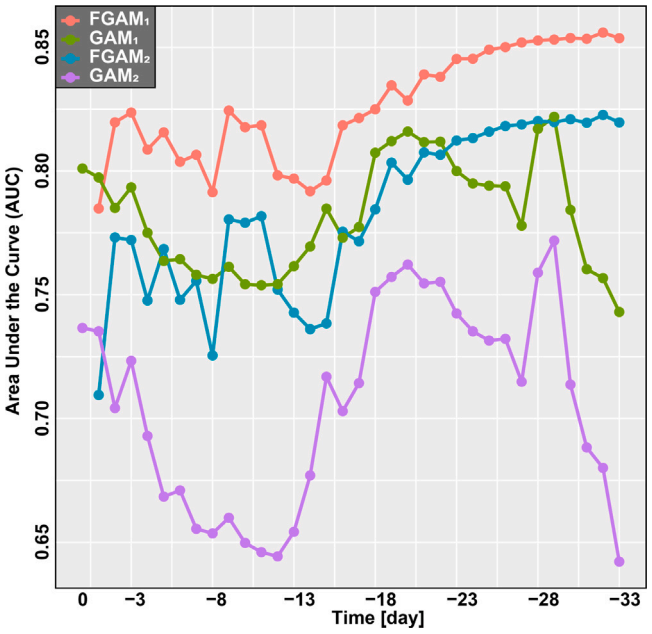


Fig. 5. Overview of the test results at varying time windows for the two functional models $FGAM_1$ and $FGAM_2$, and the two respective standard benchmarks GAM_1 and GAM_2.

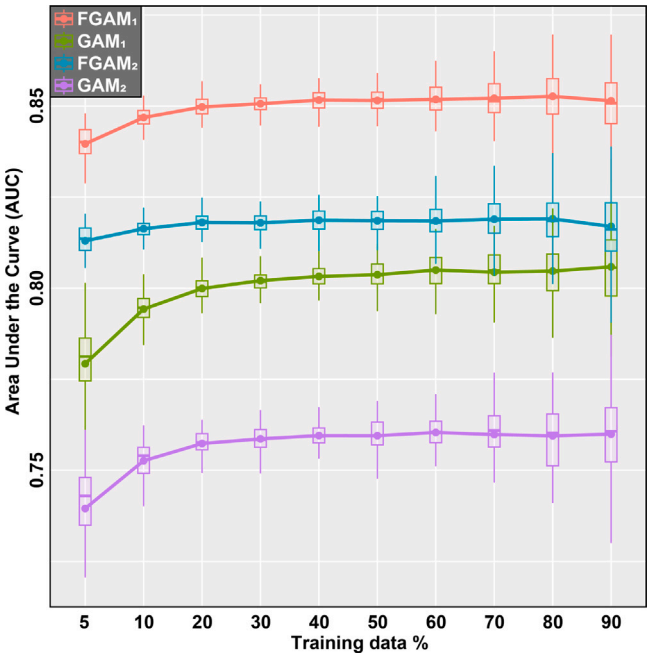


Fig. 6. Benchmark tests for model stability at various proportion of training data. The performance in the AUC axis are computed on the complementary validation subset. The boxplot are drawn by bootstrapping the precedure 100 times.

end of the considered time window. The same does not happen for the two functional models, as they both quickly increase as the considered time window increases, but with a much smoother behavior, especially after day −16. Ultimately, another relevant aspect to mention is that the performance difference between GAM_1 and GAM_2, and the performance difference between $FGAM_1$ and $FGAM_2$ are not the same. The vertical distance between the two functional lines is much smaller than the vertical difference between the two GAMs, and more discussion on the implications of such observation will be presented in Section 5.

Beyond the considerations on the required time window, we moved our tests towards model stability at various data populations. Fig. 6 presents the predictive performance of the four models described above, corresponding to the case where the four models include rainfall information for up to 33 days before the landslide date. These tests are implemented by randomly extracting more and more data from the original data matrix, and testing the prediction skill on the complementary subset. To provide confidence intervals, we bootstrap the same operation 100 times for each data proportion. What becomes evident is a consistent tendency of the FGAM architectures to outperform the GAM ones, although the magnitude of improvement varies across cross-validation settings. Here, performance differences are examined in terms of whether they shift a model across the classification thresholds suggested by Hosmer and Lemeshow (2000). For example, from poor to acceptable or from acceptable to excellent, rather than focusing solely on raw differences in AUC. One can even draw a horizontal line at around 0.81 AUC, below which, no matter how the FGAM models have been perturbed, they consistently produce acceptable to good classification results (Hosmer and Lemeshow, 2000), with AUC values remaining above 0.81 in random cross-validation.

In a similar manner, the two GAM models consistently return performance below 0.81 AUC. Additionally, it becomes evident that the FGAM require much less information to reach a performance asymptote, whereas the GAM model counterparts improve much more as more data are used for training. Ultimately, the vertical difference between the two model pairs is much narrower for the FGAM options as opposed to the standard GAM ones, an additional indication of model robustness.

Notably, the AUC is a cutoff-independent model diagnostic, which is why, while running the same tests presented in Fig. 6, we also measured the confusion matrix obtained for each training data proportion and for each of the 100 bootstrap replicates. The resulting cutoff (set to be the Youden index; Fluss et al., 2005) dependent diagnostics are reported in Fig. 7. There, we present how accurately each model classifies stable SUs (along the four x-axis where TN means True Negative and N negative value) and how accurately it classifies unstable SUs (along the four y-axis where TP is True Positive and P is positive value). As for the palette, each color represents a training data proportion, with repeated symbols corresponding to each of the 100 bootstrap replicates. The resulting patterns should be interpreted with an ideal model clustering the results towards the top right corner. The opposite location, in the bottom left corner would mark bad performance in classifying both stable and unstable SUs. Reading the information delivered in Fig. 7 from this perspective, one can notice that the closest model results to the top right corner are reached by $FGAM_1$. Interestingly, $FGAM_2$ even performs slightly better when it comes to TP/P, indicating that a good landslide prediction can indeed be achieved by using the rainfall signal on its own. However, it loses in comparison along the TN/N axis, demonstrating that static covariates contribute to better discriminating stable slopes. Another element to stress is that the results of $FGAM_2$ are much more clustered together, something that should not be confused with model stability. In fact, this is quite normal as the higher variability shown by the $FGAM_1$ is to be attributed to the variability brought by static covariates. Examining the GAM counterparts, one can immediately notice a drastic shift towards the left corner of each panel, a demonstration that the scalar approximation of the continuous rainfall signal has very negative repercussions on the prediction of stable SUs. This being said, the prediction ability of unstable SUs is affected to a much lesser extent as compared to the stable SU case.

What we have shown so far are results from random cross-validations, a type of test known to return overoptimistic performance (Anderssen et al., 2006). For this reason, we complement our analyses including spatial cross-validation (Brenning, 2012), which here we interpret by testing all possible combinations where the four

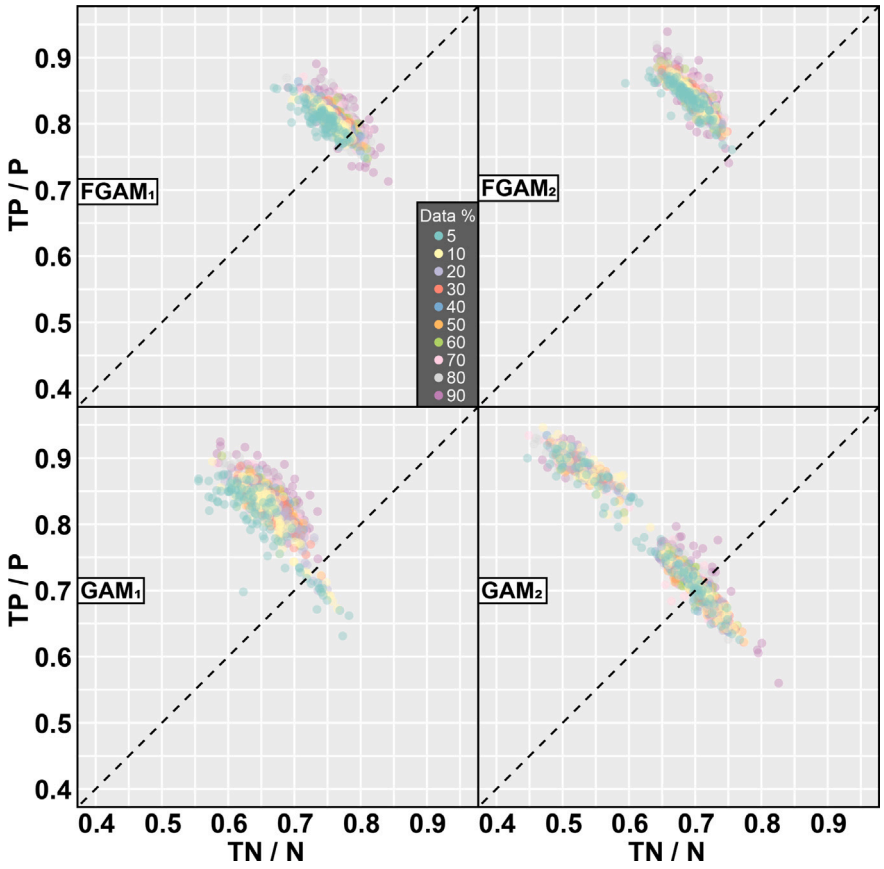


Fig. 7. Overview of the test results for 100 bootstrapped model obtained by training $FGAM_1$, $FGAM_2$, GAM_1 and GAM_2 with various random proportions of the whole data cube. Each model performance is measured in terms of TN/N and TP/P. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of the leave-one-inventory out cross-validations for each of the four model variations: $FGAM_1$, $FGAM_2$, GAM_1, and GAM_2.

Training	Test	Model	AUC	FP	TP	TN	FN	AUC rank
Miyagi & Noto	Kii	$FGAM_1$	0.7488715	17 915	985	30 151	255	Moderate
Kii & Miyagi	Noto	$FGAM_1$	0.6807794	2714	470	2728	144	Acceptable
Noto & Kii	Miyagi	$FGAM_1$	0.6133637	4886	300	3674	28	Acceptable
Miyagi & Noto	Kii	GAM_1	0.7065708	26 494	1147	21 572	93	Moderate
Kii & Miyagi	Noto	GAM_1	0.5775250	3403	505	2039	109	Low
Noto & Kii	Miyagi	GAM_1	0.5259396	5838	303	2722	25	Low
Miyagi & Noto	Kii	$FGAM_2$	0.7462982	23 814	1150	24 252	90	Moderate
Kii & Miyagi	Noto	$FGAM_2$	0.7571729	1961	493	3481	121	Moderate
Noto & Kii	Miyagi	$FGAM_2$	0.7225076	3593	311	4967	17	Moderate
Miyagi & Noto	Kii	GAM_2	0.7098441	29 283	1201	18 783	39	Moderate
Kii & Miyagi	Noto	GAM_2	0.6030044	2379	414	3063	200	Acceptable
Noto & Kii	Miyagi	GAM_2	0.6058821	4469	297	4091	31	Acceptable

model variations we explored are trained over two landslide event inventories and validated over the remaining one. The summary results of this procedure are shown in Table 1. There, one can observe a consistent tendency of the functional regression framework to outperform the basic GAM counterpart, although the magnitude of improvement varies across regions. It is worth noting that the AUC values derived from spatial cross-validation are lower than those obtained via random cross-validation, ranging from 0.61 to 0.75 for $FGAM_1$. This reflects the increased challenge of achieving generalization across geographically distinct landslide inventories. According to Hosmer and Lemeshow (2000), this range corresponds to moderate to acceptable predictive ability; however, it should be interpreted cautiously because of the structural heterogeneity among the three study areas. In the discussion section, we will present additional reflections on the errors associated with each model variation when predicting data in new areas.

5. Discussion

Two aspects have been explored in this contribution. The first corresponds to the use of space–time modeling architectures, while the second refers to the use of continuous rainfall time series as part of these architectures. Neither aspects are really under discussion on whether they are better of their standard alternatives. Space–time modeling architectures allow the incorporation of both static and dynamic predictors in the process of predicting rainfall-induced landslides. This is a more holistic approach as opposed to performing susceptibility and rainfall threshold analyses independently of each other. In a similar manner, one can safely assume that the use of the entire rainfall signal is more informative than a single summary statistic for predictive purposes. These two aspects have already been investigated in recent scientific contributions, which is why in the following

two sections, we will focus on reflecting on strengths (Section 5.1) and weaknesses (Section 5.2) of a functional regression approach to landslide prediction.

Before moving to such considerations, we would like to discuss here a more neutral aspect of this contribution. In fact, one of the most common dilemmas in landslide susceptibility modeling is the choice of a suitable mapping unit (e.g., Erener and Düzgün, 2012). Such choice has been almost dominated by grid-cells since the 1980's (Reichenbach et al., 2018), with more recent periods experiencing a rise in the use of SUs (Caleca et al., 2025). To further distinguish this contribution, we tested a SU partition, as opposed to the experiments presented in Moreno et al. (2025) where the mapping unit was expressed in a squared lattice, 30 m × 30 m in size. Therefore, despite there is no reason to assume that an FGAM should not be feasible when partitioning an area into SU, this contribution still constitutes the first example in the literature where such evidence has been formally presented for landslide prediction purposes.

5.1. Supporting arguments

In general, the use of multivariate regression naturally allows for incorporating a wide set of predictors. This is why we argue that splitting the landslide prediction task into a susceptibility assessment step, separated from the estimation of rainfall threshold,s is not a valid strategy from a basic numerical perspective. A multivariate framework allows for estimating covariates' effects simultaneously, accounting for reciprocal influence on the landslide presence/absence distribution. This is the main reason why the literature on space–time landslide modeling has seen a spike in recent years. However, several options belong to the multivariate spectrum of techniques, both of statistical and machine learning nature. One of the main strengths of statistical models is its clear interpretability, something we also demonstrate here. To stress once more this aspect, we recall that machine learning has recently developed a sub-branch also dedicated to interpretability, offering options to gain insight on predictor contributions. This sub-branch is usually referred to as explainable AI, and to this wide umbrella belong well known techniques such as Local Interpretable Model-agnostic Explanations (LIME; Zafar and Khan, 2021) or SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017). However, despite the visibility these methods have gained in recent years, they still constitute post-hoc techniques. In other words, they are not metrics internal to the model itself, but rather estimates calculated after the model has been trained. The strength of statistical models is that when examining the regression coefficients or, like in Fig. 4, the probabilities associated to each covariate, one is gaining insight on the model itself, in terms of its constitutive parameters. This is why, despite the recent success of machine learning and its explainable features (Caleca et al., 2024), statistical models still offer a good balance between prediction power and interpretation.

To further evaluate the interpretability of our framework, we examined potential concurvity among the smooths in our functional regression model. Concurvity is the nonlinear counterpart of multicollinearity in additive models; when present, it can destabilize parameter estimates and hinder the clear attribution of effects to individual predictors (Wood, 2017). This issue is especially pertinent in our study areas, where orographic controls are well recognized and rainfall often concentrates over steep slopes, implying a possible physical dependence between the precipitation and slope predictors. The maximum concurvity index for the functional precipitation term was 1.0, which could be interpreted as problematic at first glance. However, this value is recognized as being overly conservative for tensor product smooths, since it reflects a theoretical maximum rather than the actual extent of shared structure. The reason is that the diagnostic procedure tests whether the tensor product can be reconstructed from the separate marginal bases; because the tensor product basis, by construction, includes the marginal basis functions as a subspace, the test will almost always indicate near-complete overlap. In effect, it is detecting redundancy that is built into the basis design itself, rather than true confounding present in the data (Wood, 2017). The observed and estimated concurvity values for precipitation were 0.61 and 0.35, respectively, both comfortably below the common 0.8 threshold. For slope and profile curvature, all concurvity indices were also well under 0.8 (observed: 0.40 and 0.04, respectively), indicating that the additive FGAM formulation effectively disentangles the effects of the static terrain predictors from the dynamic rainfall signal. This transparency stands in contrast to many machine learning methods, in which such dependencies are absorbed into model weights and remain opaque to the analyst.

Taking advantage of the statistical models interpretability, we refer to the results shown in the 3D surface on Fig. 4. There, a characteristic of our model and the assumptions we have made become quite clear when focusing on the horizontal plane. From a slope stability perspective, one would interpret that as the time passes, moving towards the landslide occurrence date, the rainfall should be assumed to increase, as this measure well correlates to the expected pore-pressure increase. However, the figure clearly shows that we have cumulated the rainfall backwards (from the day of landslide occurrence to the −33rd day), because the only known factor is the landslide occurrence date, which is why the time series is assumed to start from there and gets progressively summed to the rainfall discharged before. This is a quite analogous procedure to that of standard rainfall threshold approaches (Uwihirwe et al., 2020). Here, we highlight this aspect to exemplify how the very same concepts that have laid the foundation for landslide early warnings over the last 50 (Mirus et al., 2025) years can be directly integrated into a functional regression framework.

Looking specifically at this functional regression idea, we will reflect here on the patterns shown in Fig. 5. There, it becomes indisputable that the use of the whole rainfall signal yield unmatched results in comparison to architectures that only approximate rainfall. Therefore, the question is not whether this is worthy from a pure modeling perspective but rather if it is worth from an operational one. Interestingly, this operational aspect will be discussed here as a strength, but also in the next Section, as a potential weakness. The next step of any short-term landslide prediction model should be its evaluation when issuing alerts for the potential population at risk. In this sense, it is fundamental to know whether a model suitably classifies stable and unstable SUs. This is something we graphically summarized in Fig. 7, where the accuracy of the model is separately plotted in a space where one dimension refers to the presence (TP/P) and the other to the absence (TN/N) data. The idea is that a model should ideally predict 100% of both presence/absence instances, something which in practice is never achieved. Therefore, from an operational perspective, it is important to reflect on the errors a given model produces. A model that consistently produces False Negative (FN) is a model whose alerts will potentially put the life of people at risk (Palau et al., 2023). However, a model that consistently produces False Positive (FP) is a model whose credibility will decrease over time, thus the informed population may progressively tend to underestimate the alert information (Piciullo et al., 2020). This is the reason why Fig. 7 supports once more the use of a functional regression approach. The first thing the plot communicates is that at varying data populations, the two functional regression architectures ($FGAM_1$ and $FGAM_2$) produce very consistent predictions. Conversely, the results we observe from one training data proportion to another for the two standard GAMs (GAM_1 and GAM_2) produce quite strong differences. As a result, one can state that the robustness of model equipped with rainfall time series is much higher than that of a model that uses a scalar rainfall representation. Moving back to the idea of specifically examining the accuracy, the closest results to the top right corner of each panel in Fig. 7 (100% TP/P and 100% TN/N) are those coming from $FGAM_1$. The global behavior across the tests with varying proportions of training data shows that $FGAM_2$ ($\mu = 83.8\%$, $\sigma = 2.45\%$) even produces slightly higher performance

than $FGAM_1$ (μ = 80.7%, σ = 2.06%) when predicting landslides. However, $FGAM_2$ also produces 6.3% more errors on average when it comes to recognizing stable slopes (or the N instances along the x-axis). In fact, the global behavior along the TN/N axis across all tests has μ = 76.2% and σ = 2.04% for $FGAM_1$ and μ = 69.9%, σ = 2.37% for $FGAM_2$. This implies that an early warning exclusively based on rainfall time series, and not informed about the terrain characteristics, would produce significantly more FN. These are locations that $FGAM_2$ would recognize as stable whilst in reality they have experienced landsliding. We stress here that this type of error is the most dangerous and further reflections will be presented later in the text. Extending the same reasoning to GAM_1 and GAM_2, the average TP/P measures across data training replicates are 82.8% (σ = 3.58%) and 76.3 (σ = 8.52%), respectively. This means that the functional approach increases the capacity of recognizing unstable SUs, although all four variations are responsible for excellent prediction skills. Where GAM_1 and GAM_2 underperform compared to their functional counterparts, is seen in the classification of stable SUs. There, we can notice a drastic shift towards the left side of the TN/N axis, with GAM_1 producing a global accuracy with μ = 66.4% (σ = 3.05%) and GAM_1 even going down to a μ = 64.7% (σ = 8.34%). This implies a significant increase in FN when using a scalar simplification of the rainfall signal. In turn, from an operational perspective, this would produce much more unreliable early warnings and potentially expose local populations to danger. The very same considerations can be made when the four model variations are spatially cross-validated. The added value here is that one can interpret these results to be reflecting how transferable these models are. What we can observe in Table 1 are very similar patterns to those mentioned above for the pure random cross-validation strategy. The two functional models largely outperform their respective standard GAMs, although some differences are to be noted even within the same type of architecture. Specifically, $FGAM_2$ classifies slightly better than $FGAM_1$ those locations hosting a landslide, but fails in a relative sense at classifying stable slopes. This can be interpreted with a tendency of predicting landslides "everywhere", which is why the functional model shows smaller errors in the FN. However, this is done at the expenses of distinguishing slopes where landslides do not take place, hence leading $FGAM_2$ to also returning much larger FP than $FGAM_1$.

Because of this, we opted to look further at the results by focusing on diagnostics such as Precision and Recall (Cook and Ramadas, 2020; Loche et al., 2022). We recall here that Precision is calculated as TP/(TP + FP), whereas Recall is obtained as TP/(TP + FN). These two respectively deliver "how accurate the positive predictions are" and "how complete the positive detections are". Their results, plotted in a typical Precision-Recall curve space, are shown in Fig. 8.

There, the final verdict is reached in favor of $FGAM_1$. This model variant produces high Recall values in line with those of the other three, while also returning overwhelmingly higher Precision values.

Therefore, we finally decided to produce maps of the associated predictions, which we show in Fig. 9. The idea is to present the probability patterns produced by $FGAM_1$ and plot the results of the other models as differences. In such a way, we highlight what areas would be largely misclassified with respect to our best solution, especially in light of a potential application for an early warning system. Visualizing the results on a map, one can understand why $FGAM_2$ was close to $FGAM_1$, as the map for Kii appears quite similar. We recall here that Kii is an area much larger than the other two, where instead the probability differences with respect to $FGAM_1$ appear increasingly larger as the model complexity decreases. Another observation is made with patterns arising between $FGAM_1$ and the other two GAM alternatives. In both cases, (third and fourth row of Fig. 9), large portions of the area are covered in green, a color indicating that both GAM_1 and GAM_2 largely overestimated the probability.

To conclude this section, we would like to stress that the rainfall signal used in the model comes from weather forecast. This was a choice we made specifically to test the model in a forward-looking manner. Several contributions in the literature can be found where the rainfall data used for landslide predictive purposes comes from observations (e.g., rain gauges; Rossi et al., 2017). They serve the purpose of setting the reference of the optimal warnings one could build because observed rainfall intensity constitutes the best measure one could assume to have access to Segoni et al. (2018a). Conversely, forecast precipitation intrinsically suffers from biases and uncertainties, thereby affecting how well one can explain landslide occurrences (Intrieri et al., 2013; Cremonini and Tiranti, 2018). Therefore, we conclude this section stressing once more that the reported results could be used to directly design an early warning, as the overall modeling structure and procedure complies with the requirements for alert systems (Guzzetti et al., 2020), of course specific to the three areas we tested, at least in the current model form.

5.2. Opposing arguments

We believe that the functional regression approach we propose here is a step towards better rainfall-induced landslide predictions. However, we also believe that several limitations still apply to this framework, which is what we will reflect on in this section.

For instance, looking once more at Fig. 5, it becomes clear that the use of rainfall time series reaches an asymptote at the −23rd day, with the remaining signal being only responsible for almost negligible improvements. Interestingly, the performance associated to GAM_1 and GAM_2 reach their peaks at the day −29. We present this argument to highlight the need of defining a suitable time window from which to extract the rainfall time series. The difference between a functional and a regular GAM approach is that, for longer time series, the performance simply plateaus, whereas models that use a scalar rainfall representation behave much more erratically. This should be a strength rather than a limitation. However, we decided to mention the time window length here because the current approach leaves a big question unanswered. We recall here that the reason we tested a similar modeling architecture to that presented by Moreno et al. (2025) is because its performance was not better than a standard model with a scalar rainfall, as it was instead obvious in the experiments run by Fang et al. (2023) and Lim et al. (2024). Our assumption was that the rainfall time series was maybe too short, even though it was selected as such because the modeled shallow landslide type. This is an aspect that we have yet to address, which is why we placed the reflections on the time window in this section. In fact, the three landslide events considered in this experiment were triggered by tropical storms in the respective regions, mostly resulting in shallow landslides, often classified as debris flows. As a result, two interlinked limitations arise. The first comes in the form of limited use for type-specific early warnings, whereby one would like to ideally predict shallow and rapid landslides separately from deep and potentially slower ones. As for the second limitation, not having access to a complete landslide classification, hinders our ability to produce geomorphologically-sound interpretations of the rainfall time series as part of a functional regression approach. Ideally, we would need a multi-temporal inventory where metadata also reports various landslide classes. Unfortunately, this is quite a rare information to get access to, which is why we tested our model on inventories that are assumed to be essentially monotypic. Fortunately, in recent years the community is increasingly generating inventories in which landslides are systematically classified (Berti et al., 2025). Aside from the newly made inventories, recent methodological developments even allow to classify landslides according to their mechanism in old inventories, as long as topological information is available (Bhuyan et al., 2024). In this sense, we plan to access or synthetically generate inventories reporting landslide types and explore the functional regression framework. Our assumption, as it is also the case for rainfall thresholds (Frattini et al., 2009; Fusco et al., 2022), is that deeper and larger landslides would require even longer time series. The rationale implies that for the rainfall to percolate into the hillslope materials and reach the potential

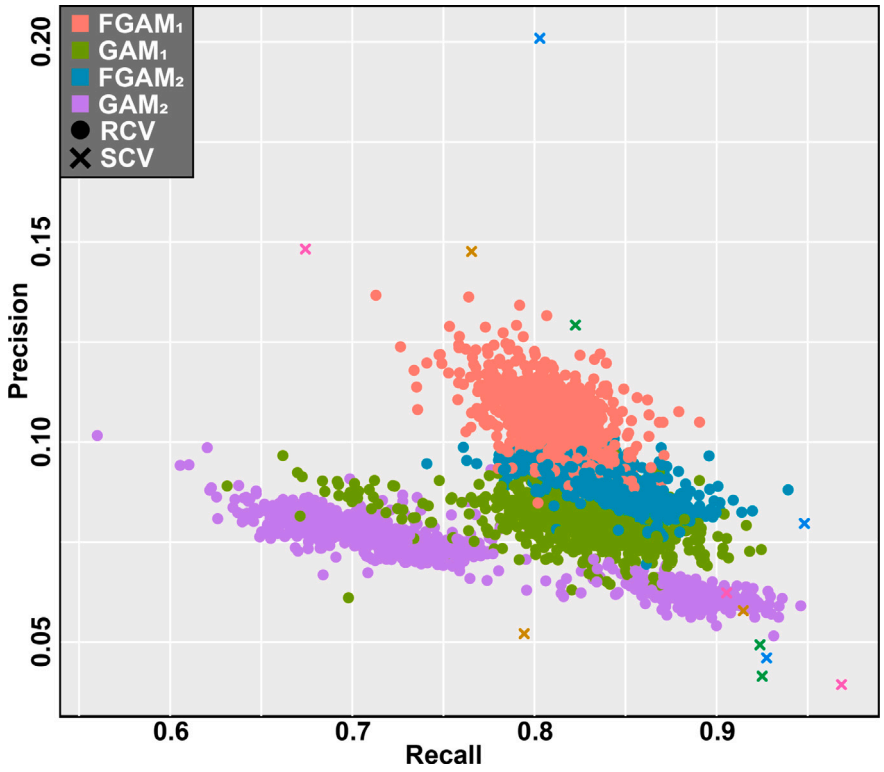


Fig. 8. Random (RCV) and Spatial (SCV) cross-validation results, plotted in a Precision-Recall space.

sliding surface, it would take more time as the area contributing to infiltration increases. Anyway a region-specific sensitivity analysis of the chosen time window length, which could shed light on the hydrological mechanisms triggering landslides in each of the three study areas, remains a topic for future work. It was not considered in this study because the number of events in each region were not large enough to enable robust cross-validation.

Assuming the nature of failure to be mainly shallow, the time-window for triggering such failures should be relatively short. The same foundational idea is behind many existing early warning systems where the antecedent window for cumulative rainfall is usually considered between 3 to 30 days (Segoni et al., 2018b). To better understand the rainfall influence on landsliding and capture its preparatory and triggering effects, we opted for a relatively wide time-window, which is why we extend our time series reaching up to 34-days from the assumed triggering day. In our case, this choice was justified in Fig. 5, where a peak in performance is reached at around −23 days, and before which the AUC seems to reach an asymptote.

A fundamental requirement of the proposed functional regression approach is the availability of high-frequency rainfall data and landslide inventories where the occurrence time of each event is documented with adequate accuracy (Rani et al., 2026). In the three Japanese inventories examined in this study, landslide occurrence dates were recorded on a daily scale, which allowed the 34-day rainfall time series to be easily aligned with each event. However, in typical applications, landslide inventories are usually assembled as one-time snapshots after major events, with the exact timing of individual landslides either missing or recorded only at a monthly or seasonal resolution (Petley, 2012; Kirschbaum et al., 2015). In such data-scarce settings, a spline-based method would encounter substantial difficulties, because the temporal mismatch between the rainfall time series and the recorded landslide date would inject noise into the estimated coefficient function $\hat{\beta}(t)$. This problem is especially pronounced in the crucial days just before the event, where the probability mass is most heavily concentrated. The greater the temporal uncertainty in the inventory, the more the functional signal will be smoothed and weakened, which can lead to an underestimation of the rainfall effect in the vicinity of day 0. In this sense, the continued use of static susceptibility maps and empirical thresholds in data-scarce settings remains appropriate, and the functional regression framework introduced here should be viewed as a tool suited to contexts where both high-quality rainfall records and temporally accurate inventories are available. Future research could investigate how robust the functional spline is to artificially added temporal noise, as a means of determining the minimum timestamp precision needed for reliable functional estimation. However, an even narrower temporal certainty when dating the landslide events should be preferable moving forward in dynamic landslide studies, and even more so, in data-scarce situations. Luckily, recent contributions explored such problem constraining the triggering time to just few days in central Africa (Dewitte et al., 2021; Deijns et al., 2022). Second to the temporal resolution of the failures, the spatial and temporal resolution of rainfall also weighs in. The idea is that a functional GAM should still be able to capture the rainfall influence better than alternatives even in data-scarce situations, as even simpler models such as rainfall thresholds should also be affected by misrepresenting rainfall patterns. We assume so, because the functional aspect would still produce probabilistic estimates but based on a relatively richer information basis as compared to methods that only require scalar approximations. For instance, in areas with sparse rain gauge networks, any model would struggle, not just a functional GAM, a GAM or a standard rainfall threshold method. In case where one would employ satellite rainfall estimates, aside from the problems caused by spatial and temporal resolution, one would face a data where extremes tend to be often smoothed out compared to rainfall records produce by local gauges. However, even though the time series should be smoother, its signal as a whole should still capture certain dynamics which a scalar representation would likely miss out. As a result, one should still assume an FGAM approach to produce better performance compared to standard alternatives.

We recall here that the goodness-of-fit metrics returned outstanding performance for $FGAM_1$, approximately excellent for $FGAM_2$ and

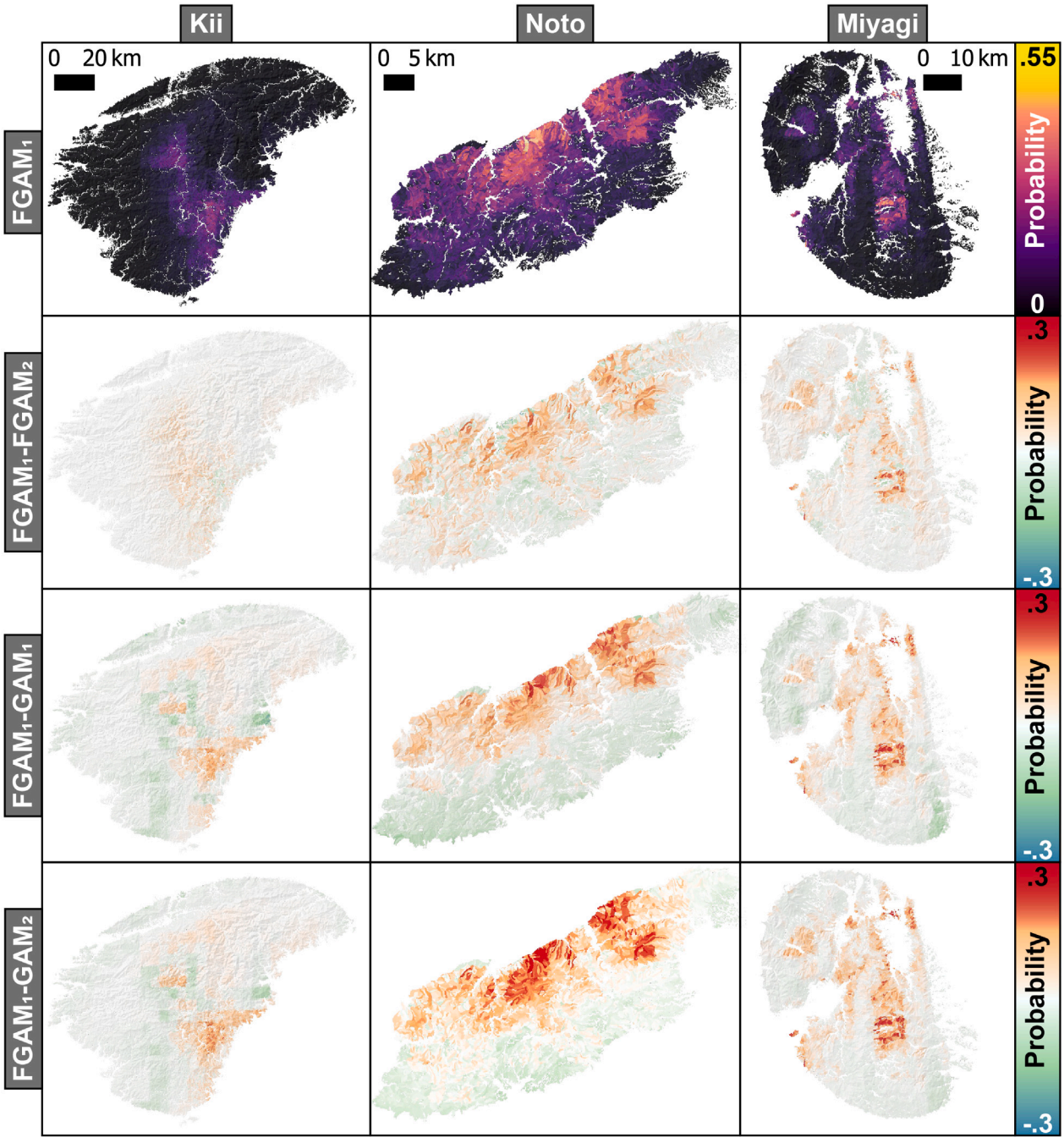


Fig. 9. Landslide occurrence probabilities, plotted for the best model $FGAM_1$ in the first row. The other rows present the relative differences with respect to $FGAM_1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

GAM_1, and only acceptable for GAM_2. Comparing to this benchmark level, none of the cross-validation results produces outstanding performance diagnostics and all appear to be slightly worse both in the case of random and spatial CV routines. While it is difficult to explain the cause of this output, there still can be possible points of discussion. Specifically $FGAM_2$ consistently yields better results for all three test sites, since it only governs the functional rainfall signal. However, once topographic interactions are introduced (in $FGAM_1$), which can significantly differ in the diverse geographic makeup of Japan, the AUC drops. Leave-one-out focuses on testing the model by completely exempting a region, which influences the learning capability of the model within the topographic heterogeneity of the training inventories. Thus, model performance obviously suffers when terrain and/or rainfall patterns differ from one dataset to another.

Mirroring the previous Section, we will conclude the current one reflecting on the operational aspects of our functional regression approach. There is an obvious reason why rainfall thresholds are still the main backbone of many landslide early warning systems worldwide. As we have demonstrated in this contribution, their simplicity places them at a disadvantage when compared to more complex approaches from a pure numerical perspective. However, from an operational perspective one cannot solely base the judgment on numerical metrics, but rather consider the social component that exists in the process of delivering an early warning to a non-scientific audience (Sedhain et al., 2025). The rainfall thresholds and the alerts that arise from them come with a very simple message, easily received by the agencies that issue the warning as well as the population at potential risk, irrespective of the

educational background. This is because rainfall amounts are quite intuitive and straightforward to understand.

The same cannot be said when the message involves probabilities, especially if those come from a functional regression approach.

This is why pure performance may not necessarily be the best criterion to use, as agencies may be more inclined to work with systems that they can control (see, Sedhain et al., 2025).

In a similar manner, one can argue that the given population at risk does not need to know the background information and relative modeling, but rather focuses on the alerts communicated to them. Several studies, not necessarily in the context of landslides, have already argued that a warning message understood by many is also a message that one is more inclined to follow Basher (2006). And, if this increases compliance, chances are that the right behavior under dangerous conditions is also assumed by many, thus reducing the overall risk as compared to a better alert system which people may be more reluctant to receive (Morss et al., 2008). Our discussion on pros and cons of complex modeling architecture therefore concludes with a reflection on how to best translate the significantly better performance demonstrated by our functional regression framework into better risk mitigation actions. This aspect inevitably requires an additional collaborative phase where stakeholders such as civil protection agencies are involved in the process, potentially co-creating risk management protocols for ensuring the right behavior when hazard situations arise (Vollmer et al., 2025).

6. Conclusion

The status of current technological advancement calls for a review of certain assumptions the scientific community has made for decades. Among the possible rooms for improvements for landslide prediction, space–time models have shown great potential. However, many modeling architectures fall in the category of space–time, and the most tested ones so far in the context of landslides take inspiration from traditional rainfall thresholds, whereby the continuous precipitation signal is simplified into a scalar representation by taking the cumulative value for specific time windows. In the same category of space–time model, also falls the functional regression, a model archetype capable of handling predictors in the form of time series. The present contribution has explored this context by testing whether a functional regression strategy can lead to better landslide predictions. Certainly, there are too few tests in the literature exploring this theme, and a systematic experimentation in different landscapes, rainfall patterns and landslide types should take place before any definite conclusion can be made. However, the results we show here demonstrate our assumption across any of the tests we run, which is why in light of such an observation we recommend the community at large to critically review the standard use of scalar rainfall aggregates. Currently, we have only experimented on a landslide occurrence probability framework (susceptibility). However, similar considerations can be made for predicting landslide sizes in space and time (intensity) or the combination of both terms leading to a full hazard modeling approach based on functional regressions. This aspect is where future tests should also be directed towards, hopefully confirm that the richer rainfall information contained in continuous time series can improve landslide predictions across most of the hazard and risk assessment existing protocols.

CRediT authorship contribution statement

Mahnoor Ahmed: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Vitor Salomão Recacho:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Giacomo Titti:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Taro Uchida:** Writing – review & editing, Supervision, Data curation. **Lisa Borgatti:** Writing – review & editing, Supervision. **Adriano Barasal Morales:** Writing – review & editing. **Mirko Francioni:** Writing – review & editing, Supervision. **Luigi Lombardo:** Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Research Link Provided

SU based covariates (Original data) (Github)

References

Acharya, A., Malla, A.B., Guragain, S., Dahal, R.K., Hasegawa, S., 2024. Engineering geological challenges following the 2024 noto earthquake, Japan. Asian J. Eng. Geol. 1 (1 and 2), 7–16.

Ahmed, M., Tanyas, H., Huser, R., Dahal, A., Titti, G., Borgatti, L., Francioni, M., Lombardo, L., 2023. Dynamic rainfall-induced landslide susceptibility: a step towards a unified forecasting system. Int. J. Appl. Earth Obs. Geoinf. 125, 103593.

Ahmed, M., Titti, G., Trevisani, S., Borgatti, L., Francioni, M., 2025. Is higher resolution always better? A comparison of open-access DEMs for optimized slope unit delineation and regional landslide prediction. Nat. Hazards Earth Syst. Sci. 25 (7), 2519–2539.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., Guzzetti, F., 2016. Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. Geosci. Model. Dev. 9 (11), 3975–3991.

Anderssen, E., Dyrstad, K., Westad, F., Martens, H., 2006. Reducing over-optimism in variable selection by cross-model validation. Chemometr. Intell. Lab. Syst. 84 (1–2), 69–74.

Atkinson, P., Jiskoot, H., Massari, R., Murray, T., 1998. Generalized linear modelling in geomorphology. Earth Surf. Process. Landf.: J. Br. Geomorphol. Group 23 (13), 1185–1195.

Bárdossy, A., Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water Resour. Res. 49 (8), 4545–4565.

Basher, R., 2006. Global early warning systems for natural hazards: systematic and people-centred. Philos. Trans. R. Soc. A 364 (1845), 2167–2182.

Bergtold, J.S., Spanos, A., Onukwugha, E., 2010. Bernoulli regression models: Revisiting the specification of statistical models with binary dependent variables. J. Choice Model. 3 (2), 1–28.

Berti, M., Pizziolo, M., Scaroni, M., Generali, M., Critelli, V., Mulas, M., Tondo, M., Lelli, F., Fabbiani, C., Ronchetti, F., et al., 2025. RER2023: the landslide inventory dataset of the May 2023 Emilia–Romagna meteorological event. Earth Syst. Sci. Data 17 (3), 1055–1074.

Bhuyan, K., Rana, K., Ferrer, J.V., Cotton, F., Ozturk, U., Catani, F., Malik, N., 2024. Landslide topology uncovers failure movements. Nat. Commun. 15 (1), 2633.

Brabb, E.E., 1980. Slope-Stability Studies in the San Francisco Bay Region, California. In: Professional Paper, Number 1239, U.S. Geological Survey, Washington, D.C., U.S. Geological Survey Professional Paper 1239.

Brenning, A., 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In: 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 5372–5375.

Budimir, M., Atkinson, P., Lewis, H., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides 12 (3), 419–436.

Caleca, F., Confuorto, P., Raspini, F., Segoni, S., Tofani, V., Casagli, N., Moretti, S., 2024. Shifting from traditional landslide occurrence modeling to scenario estimation with a "glass-box" machine learning. Sci. Total Environ. 950, 175277.

Caleca, F., Lombardo, L., Steger, S., Tanyas, H., Raspini, F., Dahal, A., Nefros, C., Mărgărint, M.C., Drouin, V., Jemec-Auflič, M., et al., 2025. Pan-European landslide risk assessment: From theory to practice. Rev. Geophys. 63 (1), e2023RG000825.

Cao, Y.-m., Guo, W., Wu, Y.-m., Li, L.-p., Zhang, Y.-x., Lan, H.-x., 2022. An hourly shallow landslide warning model developed by combining automatic landslide spatial susceptibility and temporal rainfall threshold predictions. J. Mt. Sci. 19 (12), 3370–3387.

Casella, G., Berger, R.L., 2002. Statistical Inference, second ed. Duxbury, Pacific Grove, CA.

Chang, Z., Huang, F., Huang, J., Jiang, S.-H., Liu, Y., Meena, S.R., Catani, F., 2023. An updating of landslide susceptibility prediction from the perspective of space and time. Geosci. Front. 14 (5), 101619.

Chen, C., He, Q., Li, Y., 2024. Downscaling and merging multiple satellite precipitation products and gauge observations using random forest with the incorporation of spatial autocorrelation. J. Hydrol. 632, 130919.

Cook, J., Ramadas, V., 2020. When to consult precision-recall curves. Stata J. 20 (1), 131–148.

Cremonini, R., Tiranti, D., 2018. The weather radar observations applied to shallow landslides prediction: a case study from north-western Italy. Front. Earth Sci. 6, 134.

Crosta, G.B., Frattini, P., et al., 2001. Rainfall thresholds for triggering soil slips and debris flow. In: Proc. of the 2nd EGS Plinius Conference on Mediterranean Storms: Publication CNR GNDCI. vol. 2547, pp. 463–487.

Dahal, A., Huser, R., Lombardo, L., 2024a. At the junction between deep learning and statistics of extremes: formalizing the landslide hazard definition. J. Geophys. Res.: Mach. Learn. Comput. 1 (3), e2024JH000164.

Dahal, A., Tanyas, H., van Westen, C., van der Meijde, M., Mai, P.M., Huser, R., Lombardo, L., 2024b. Space–time landslide hazard modeling via ensemble neural networks. Nat. Hazards Earth Syst. Sci. 24 (3), 823–845.

Deijns, A.A., Dewitte, O., Thiery, W., d'Oreye, N., Malet, J.-P., Kervyn, F., 2022. Timing landslide and flash flood events from SAR satellite: a regionally applicable methodology illustrated in african cloud-covered tropical environments. Nat. Hazards Earth Syst. Sci. 22 (11), 3679–3700.

Dewitte, O., Dille, A., Depicker, A., Kubwimana, D., Maki Mateso, J.-C., Mugaruka Bibentyo, T., Uwihirwe, J., Monsieurs, E., 2021. Constraining landslide timing in a data-scarce context: from recent to very old processes in the tropical environment of the North Tanganyika-Kivu Rift region. Landslides 18 (1), 161–177.

Erener, A., Düzgün, H., 2012. Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? Environ. Earth Sci. 66 (3), 859–877.

Fang, Z., Tanyas, H., Gorum, T., Dahal, A., Wang, Y., Lombardo, L., 2023. Speech-recognition in landslide predictive modelling: A case for a next generation early warning system. Environ. Model. Softw. 170, 105833.

Fang, Z., Wang, Y., van Westen, C., Lombardo, L., 2024. Space–time landslide susceptibility modeling based on data-driven methods. Math. Geosci. 56 (6), 1335–1354.

Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and its associated cutoff point. Biom. J.: J. Math. Methods Biosci. 47 (4), 458–472.

Frattini, P., Crosta, G., Sosio, R., 2009. Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. Hydrol. Process.: Int. J. 23 (10), 1444–1460.

Fusco, F., Bordoni, M., Tufano, R., Vivaldi, V., Meisina, C., Valentino, R., Bittelli, M., De Vita, P., 2022. Hydrological regimes in different slope environments and implications on rainfall thresholds triggering shallow landslides. Nat. Hazards 114 (1), 907–939.

Glade, T., Crozier, M., Smith, P., 2000. Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "antecedent daily rainfall model". Pure Appl. Geophys. 157 (6), 1059–1079.

Goldsmith, J., Bobb, J., Crainiceanu, C., Caffo, B., Reich, D., 2011. Penalized functional regression. J. Comput. Graph. Statist. 20 (4), 830–851.

Greven, S., Scheipl, F., 2017. A general framework for functional regression modelling. Stat. Model. 17 (1–2), 1–35.

Guenni, L., Hutchinson, M.F., 1998. Spatial interpolation of the parameters of a rainfall model from ground-based data. J. Hydrol. 212, 335–347.

Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M., Melillo, M., 2020. Geographical landslide early warning systems. Earth-Sci. Rev. 200, 102973.

Harrison, L., Landsfeld, M., Husak, G., Davenport, F., Shukla, S., Turner, W., Peterson, P., Funk, C., 2022. Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. Sci. Data 9 (1), 375.

Hastie, T.J., 2017. Generalized additive models. Stat. Model. S 249–307.

Honda, R., Hiramatsu, Y., Kono, Y., Katagawa, H., 2008. Gravity anomalies and the geologic block structures in and around the aftershock area of the 2007 Noto Hanto Earthquake. Earth Planets Space 60 (2), 111–115.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression. vol. 398, John Wiley & Sons.

Huang, F., Chen, J., Liu, W., Huang, J., Hong, H., Chen, W., 2022. Regional rainfall-induced landslide hazard warning based on landslide susceptibility mapping and a critical rainfall threshold. Geomorphology 408, 108236.

Intrieri, E., Gigli, G., Casagli, N., Nadim, F., 2013. Brief communication "landslide early warning system: toolbox and general concepts". Nat. Hazards Earth Syst. Sci. 13 (1), 85–90.

Irasawa, M., Higaki, D., Koiwa, N., Takahashi, M., Okamoto, T., Yasuno, M., Tada, N., Nakajima, T., Arai, M., Ochiai, T., et al., 2017. Sediment disasters caused by Typhoon No. 10 on August 30, 2016 in Iwaizumi town and Miyako city, Iwate. J. Jpn. Soc. Eros. Control. Eng. 69 (6).

Irasawa, M., Koi, T., Tsou, C.-Y., Kato, N., Matsou, S., Arai, M., Kaibori, M., Yamada, T., Kasai, M., Wakahara, T., et al., 2020. October 2019 sediment disaster in the Tohoku region owing to typhoon no. 19 (tyhpoon hagibis). Int. J. Eros. Control. Eng. 13 (2), 48–55.

Kirschbaum, D., Stanley, T., Zhou, Y., 2015. Spatial and temporal analysis of a global landslide catalog. Geomorphology 249, 4–15.

Kosiorowski, D., 2014. Functional regression in short-term prediction of economic time series. Stat. Transit. New Ser. 15 (4), 611–626.

Lim, J., Santinelli, G., Dahal, A., Vrieling, A., Lombardo, L., 2024. An ensemble neural network approach for space–time landslide predictive modelling. Int. J. Appl. Earth Obs. Geoinf. 132, 104037.

Loche, M., Scaringi, G., Yunus, A.P., Catani, F., Tanyaş, H., Frodella, W., Fan, X., Lombardo, L., 2022. Surface temperature controls the pattern of post-earthquake landslide activity. Sci. Rep. 12 (1), 988.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., Huser, R., 2020. Space-time landslide predictive modelling. Earth-Sci. Rev. 209, 103318.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NeurIPS 2017, Curran Associates, Inc., pp. 4765–4774.

Marc, O., Oliveira, R.A.J., Gosset, M., Emberson, R., Malet, J.-P., 2022. Global assessment of the capability of satellite precipitation products to retrieve landslide-triggering extreme rainfall events. Earth Interact. 26 (1), 122–138.

Marc, O., Stumpf, A., Malet, J.-P., Gosset, M., Uchida, T., Chiang, S.-H., 2018. Initial insights from a global database of rainfall-induced landslide inventories: The weak influence of slope and strong influence of total storm rainfall. Earth Surf. Dyn. 6 (4), 903–922.

Martel, J.-L., Brissette, F.P., Lucas-Picher, P., Troin, M., Arsenault, R., 2021. Climate change and rainfall intensity–duration–frequency curves: Overview of science and guidelines for adaptation. J. Hydrol. Eng. 26 (10), 03121001.

McLean, M.W., Hooker, G., Staicu, A.-M., Scheipl, F., Ruppert, D., 2014. Functional generalized additive models. J. Comput. Graph. Statist. 23 (1), 249–269.

Mclean, M., Hooker, G., Staicu, A.-M., Scheipl, F., Ruppert, D., 2014. Functional generalized additive models. J. Comput. Graph. Stat.: Joint Publ. Am. Stat. Assoc. Inst. Math. Stat. Interface Found. North Am. 23, 249–269.

Mirus, B.B., Bogaard, T., Greco, R., Stähli, M., 2025. Invited perspectives: Integrating hydrologic information into the next generation of landslide early warning systems. Nat. Hazards Earth Syst. Sci. 25 (1), 169–182.

Mohammadreza, S., 2025. Scenario-driven GeoFutures: Exploring Japan's polycrisis through modeling urban dynamics and climatic risks, enhanced by geospatial analytics and machine intelligence.

Moreno, M., Lombardo, L., Crespi, A., Zellner, P.J., Mair, V., Pittore, M., van Westen, C., Steger, S., 2024. Space-time data-driven modeling of precipitation-induced shallow landslides in South Tyrol, Italy. Sci. Total Environ. 912, 169166.

Moreno, M., Lombardo, L., Steger, S., de Vugt, L., Zieher, T., Crespi, A., Marra, F., van Westen, C., Opitz, T., 2025. Functional regression for space-time prediction of precipitation-induced shallow landslides in South Tyrol, Italy. J. Geophys. Res.: Earth Surf. 130 (4), e2024JF008219.

Morss, R.E., Demuth, J.L., Lazo, J.K., 2008. Communicating uncertainty in weather forecasts: A survey of the US public. Weather. Forecast. 23 (5), 974–991.

Nocentini, N., Rosi, A., Segoni, S., Fanti, R., 2023. Towards landslide space-time forecasting through machine learning: the influence of rainfall parameters and model setting. Front. Earth Sci. 11, 1152130.

Onodera, T., Yoshinaka, R., Kazama, H., 1974. Slope failures caused by heavy rainfall in Japan. J. Jpn. Soc. Eng. Geol. 15 (4), 191–200.

Page, T., Beven, K.J., Hankin, B., Chappell, N.A., 2022. Interpolation of rainfall observations during extreme rainfall events in complex mountainous terrain. Hydrol. Process. 36 (11), e14758.

Palau, R.M., Berenguer, M., Hürlimann, M., Sempere-Torres, D., 2023. Implementation of hydrometeorological thresholds for regional landslide warning in Catalonia (NE Spain). Landslides 20 (10), 2039–2054.

Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40 (10), 927–930.

Piciullo, L., Tiranti, D., Pecoraro, G., Cepeda, J.M., Calvello, M., 2020. Standards for the performance assessment of territorial landslide early warning systems. Landslides 17 (11), 2533–2546.

Qiu, H., Li, Y., Zhu, Y., Ye, B., Yang, D., Liu, Y., Wei, Y., 2025. Do post-failure landslides become stable? Catena 249, 108699.

Qiu, H., Su, L., Tang, B., Yang, D., Ullah, M., Zhu, Y., Kamp, U., 2024. The effect of location and geometric properties of landslides caused by rainstorms and earthquakes. Earth Surf. Process. Landf. 49 (7), 2067–2079.

Rani, R., Dahal, A., Berti, M., Lombardo, L., 2026. Transformer, more than meets the eye: A deep learning approach to integrate rainfall time-series in multi-type landslide probability modelling. Geosci. Front. 102295.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Sci. Rev. 180, 60–91.

Reiss, P.T., Ogden, R.T., 2007. Functional principal component regression and functional partial least squares. J. Amer. Statist. Assoc. 102 (479), 984–996.

Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M.T., Peruccacci, S., Guzzetti, F., 2017. Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data. Geomorphology 285, 16–27.

Saito, H., Matsuyama, H., 2012. Catastrophic landslide disasters triggered by record-breaking rainfall in Japan: Their accurate detection with normalized soil water index in the Kii Peninsula for the year 2011. Sola 8, 81–84.

Saleem, N., Huq, M.E., Twumasi, N.Y.D., Javed, A., Sajjad, A., 2019. Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: a review. ISPRS Int. J. Geo-Information 8 (12), 545.

Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., 2019. Dynamic path dependent landslide susceptibility modelling. Nat. Hazards Earth Syst. Sci. Discuss. 2019, 1–20.

Satyaningsih, R., Jetten, V., Ettema, J., Sopaheluwakan, A., Lombardo, L., Nuryanto, D.E., 2023. Dynamic rainfall thresholds for landslide early warning in Progo Catchment, Java, Indonesia. Nat. Hazards 119 (3), 2133–2158.
Sedhain, S., van den Homberg, M., Teklesadik, A., van Aalst, M., Kerle, N., 2025. Evaluating impact-based forecasting models for tropical cyclone anticipatory action. Int. J. Disaster Risk Reduct. 105782.
Segoni, S., Piciullo, L., Gariano, S.L., 2018a. A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides 15 (8), 1483–1501.
Segoni, S., Tofani, V., Rosi, A., Catani, F., Casagli, N., 2018b. Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. Front. Earth Sci. 6, 85.
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., 2024. Adopting the margin of stability for space–time landslide prediction–A data-driven approach for generating spatial dynamic thresholds. Geosci. Front. 15 (5), 101822.
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., 2022. Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models. Nat. Hazards Earth Syst. Sci. Discuss. 2022, 1–38.
Titti, G., Hu, L., Festi, P., Elia, L., Borgatti, L., Lombardo, L., 2025. An updated version of the SZ-plugin: From space to space–time data-driven modeling in QGIS. Int. J. Appl. Earth Obs. Geoinf. 142, 104679.
Titti, G., Sarretta, A., Lombardo, L., Crema, S., Pasuto, A., Borgatti, L., 2022. Mapping susceptibility with open-source tools: a new plugin for QGIS. Front. Earth Sci. 10, 842425.
Uchida, T., Sato, T., Mizuno, M., Okamoto, A., 2012. The role of rainfall magnitude on landslide characteristics triggered by Typhoon Tales, 2011. Civ. Eng. J. 54, 10–13.
Uwihirwe, J., Hrachowitz, M., Bogaard, T.A., 2020. Landslide precipitation thresholds in Rwanda. Landslides 17 (10), 2469–2481.
Vollmer, M., Berchtold, C., Anniés, J., 2025. Co-creating solutions for disaster risk reduction in multi-country research projects–opportunities and challenges. Int. J. Disaster Risk Reduct. 105187.
Wang, N., Cheng, W., Marconcini, M., Bachofer, F., Liu, C., Xiong, J., Lombardo, L., 2022. Space-time susceptibility modeling of hydro-morphological processes at the Chinese national scale. Eng. Geol. 301, 106586.
Wang, N., Li, M., Zhang, H., Cheng, W., Du, C., Lombardo, L., 2024. Susceptibility modeling of hydro-morphological processes considered river topology. Geo-Spatial Inf. Sci. 1–20.
Wang, T., Yin, K., Wang, Z., Fang, Z., Dahal, A., Lombardo, L., 2025. Long and short-term perspectives on space–time landslide modelling. Int. J. Appl. Earth Obs. Geoinf. 142, 104694.
Wood, S.N., 2006. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics 62 (4), 1025–1036.
Wood, S.N., 2017. Generalized Additive Models: an Introduction with R. chapman and Hall/CRC.
Zafar, M.R., Khan, N., 2021. Deterministic local interpretable model-agnostic explanations for stable explainability. Mach. Learn. Knowl. Extr. 3 (3), 525–541.
Zhao, K., Qiu, H., Liu, Y., Liu, Z., Huangfu, W., Tang, B., Yang, D., Yang, G., 2025. Probability of rainfall-induced landslides coupled with effective-duration threshold and soil moisture. J. Hydrol.: Reg. Stud. 57, 102112.