Referee #2

RC2: 'Comment on egusphere-2026-1599', Anonymous Referee #2, 21 Jul 2026

We sincerely thank the reviewer for the thorough and constructive evaluation of our manuscript. In response to the comments, we expanded the literature review and more clearly defined the novelty of the CWI-CSM framework with respect to existing physically based and distributed slope-stability models. We analyzed the cohesionless case in the sensitivity analysis, and better justified the assumptions concerning homogeneous soil properties and failure depth. We also clarified the calibration procedure and its uncertainties, supported the assumed failure-depth distribution with field observations, and explicitly discussed the current limitations of model validation. Finally, we better defined the applicability of the framework to LEWSs and expanded the discussion of limitations and future developments.

Below, we detail a point-by-point response to the referee's comments and outline the corresponding revisions made to the manuscript.

Our reply is in italic. Indications of the changes proposed for the manuscript are highlighted in red.

Referee #2: The manuscript is based on a simplified model for the stability analysis of rainfall-induced shallow landslides, relying on the infinite slope assumption and the limit equilibrium method (Montgomery and Dietrich, 1994). The authors propose an extension of the classical approach by introducing a gradual transition from partially saturated to fully saturated soil conditions over an increasingly thick soil layer, using the Soil Water Retention Curve (SWRC) to represent the partially saturated state.

The study falls within the broad framework of physically based, spatially distributed slope stability models developed to reproduce the triggering conditions of rainfall-induced shallow landslides and to generate susceptibility maps at different spatial scales.

The topic is undoubtedly relevant and of considerable interest within this well-established field of research. However, several aspects of the study raise significant concerns.

General comments

1. The literature review overlooks several recently developed physically based models capable of producing time-varying susceptibility maps while explicitly accounting for soil mechanical properties, rainfall, partial/full saturation conditions, water infiltration processes, and root reinforcement (e.g., TRIGRS, SLIP, SUSHI, SOSlope, GEOTOP-Fs, CRITERIA3D). The authors are encouraged to consider the following recent review papers:
    a. Murgia, I.; Giadrossich, F.; Mao, Z.; Cohen, D.; Capra, G.F.; Schwarz, M. (2022) Modeling shallow landslides and root reinforcement: A review. Ecol. Eng. 2022, 181, 106671.
    b. Sannino, G.; Bordoni, M.; Bittelli, M.; Meisina, C.; Tomei, F.; Valentino, R. (2024). Deterministic Physically Based Distributed Models for Rainfall-Induced Shallow Landslides. Geosciences, 14(10), 255.

Reply: *This comment is similar to that of referee #1. We reply here in a similar way.*

We agree that the original literature review did not adequately represent the recent generation of physically based, spatially distributed models that account for partial saturation, infiltration processes and root reinforcement. The Introduction has been revised accordingly (please see changes in Introduction in red), and both suggested works (Murgia et al., 2022; Sannino et al., 2024), and additional references suggested by other reviewers (e.g. Simoni et al., 2008; Capparelli and Versace, 2011; Rossi et al., 2013; Lizárraga and Buscarnera, 2018; Medina et al., 2021; Masi et al., 2023; Sannino et al., 2025), have been added and cited.

Additionally, we would like, to make explicit a distinction that was insufficiently clear in the original submission. Models such as TRIGRS, SLIP, SUSHI, GEOtop-FS, HIRESSS and CRITERIA-3D are forward, time-marching simulators: given a rainfall or infiltration forcing and an initial condition, they propagate the hydrological state through the soil profile and return the temporal evolution of the factor of safety. Their output is a time series, or a time-varying map, of stability, and obtaining it requires a hydrological model, a rainfall input, and the parameters that both require.

The present work does not simulate. It inverts the limit-equilibrium condition FoS = 1 analytically, cell by cell, to obtain the hydrological state that must be reached before failure becomes possible. Its output is a static field of state-variable thresholds - the Critical Wetness Index where the failure plane lies within the saturated zone, and the Critical Soil Moisture where it lies within the unsaturated zone, which is independent of any particular rainfall scenario and of any particular hydrological model. No hydrological simulation, no rainfall time series, and no landslide-triggering database is required to derive it, which is precisely what makes the approach applicable in the data-scarce settings where empirical thresholds cannot be constructed.

The two families of approaches are therefore complementary rather than competing: the thresholds derived here can be evaluated against the hydrological state predicted by any of the models the Referee cites, or against monitored soil moisture and groundwater levels. To the best of our knowledge, an inversion of this kind spanning the full continuum from residual moisture to full saturation has not previously been proposed; the classical inversion of Montgomery and Dietrich (1994) is restricted to groundwater-controlled, steady-state conditions and therefore cannot represent the stabilising contribution of matric suction.

This is now stated explicitly in the revised Introduction and in the new Section 5.1 of the Discussion, as follows:

1 Introduction (L36-L47): "*From an operational standpoint, hazard assessment and early warning systems increasingly rely on threshold-based approaches, which integrate rainfall information with hydrological indices, soil moisture data and real-time monitoring to anticipate failure conditions and issue alerts (Distefano et al., 2023; Mirus et al., 2018; Mirus et al., 2025; Palazzolo et al., 2025; Piciullo et al., 2018; Segoni et al., 2018a, b; Zhang et al., 2026) However, most the operational rainfall and soil moisture thresholds are essentially derived through empirical and data driven methods, which require high quality and reliable databases consisting of landslide events with the corresponding triggering hydrometeorological variables (i.e. rainfall, soil moisture) (Bogaard and Greco, 2018; Guzzetti et al., 2020). Additionally, empirical and data-driven rainfall thresholds do not explicitly represent hydromechanical processes, limiting their ability to anticipate failures under evolving climatic and hydrological conditions (Guzzetti et al., 2020). Significant advances have been made in this direction, with methodologies that have improved, for example, the way how the empirical thresholds are derived (e.g. using Artificial Neural Network, ANN, tools), the inclusion of antecedent conditions, and the use of reanalysis data (e.g. Mirus et al., 2018; Mirus et al., 2025; Greco et al., 2023; Palazzolo et al., 2025, DiStefano et al., 2023, Zhang et al., 2026)".*

1 Introduction (L58-L63): *"No hydrological model is needed to solve the inversion, therefore this approach differs from the subsequent implementations that couple advanced spatially- distributed modeling of hydrological processes with the stability model (Arnone et al., 2016; Baum et al., 2008, 2010; Brilli et al., 2025; Capparelli and Versace, 2011; Di Biagio et al., 2024; Dietrich et al., 1995; Godt et al., 2008; Iverson, 2000; Lepore et al., 2013; Masi et al., 2023; Medina et al., 2021; Rossi et al., 2013; Sannino et al., 2024; Sannino et al., 2025; Simoni et al., 2008; Wu and Sidle, 1995).*

1 Introduction (L77-L87): *"... however, none of them provide an inversion approach which identifies critical thresholds because their primary objective is to predict the temporal evolution of slope stability under prescribed rainfall or infiltration forcing.*

5.1 [Discussion] Interpretation of the proposed framework (L518-524): "*Numerous physically based slope stability models already incorporate unsaturated soil mechanics to simulate rainfall infiltration and the transient evolution of slope stability (Baum et al., 2008; Baum et al., 2010; Brilli et al., 2025; Capparelli and Versace 2011; Di Biagio et al., 2024; Dietrich et al., 1995; Godt et al., 2008; Iverson 2000; Lepore et al., 2013; Masi et al., 2023; Medina et al., 2021; Rossi et al., 2013; Sannino et al., 2024; Sannino et al., 2025; Simoni et al., 2008; Wu and Sidle, 1995). The contribution of the present study is therefore not the inclusion of matric suction itself, but the analytical derivation of physically-based hydrological thresholds that explicitly accounts for the continuum of the hydrological states. The thresholds are expressed in terms of Critical Wetness Index (CWI) and Critical Soil Moisture (CSM), by providing a spatial characterization of the critical groundwater and soil-moisture conditions potentially associated with slope failure".*

5.1 [Discussion] Interpretation of the proposed framework (L529-531): "*Our approach directly targets the underlying failure mechanism, by defining analytically the critical groundwater and soil-moisture states required to satisfy the stability criterion across the full hydrological continuum, i.e. from unsaturated to saturated conditions.*

5.2 [Discussion] Implications for Landslide Early Warning Systems (L550-556): "*Operational LEWSs commonly rely on empirical or data-driven rainfall threshold approaches (Guzzetti et al., 2008; Bogaard and Greco, 2018; Segoni et al., 2018; Guzzetti et al., 2020). Although significant advances have been made in terms of methods and integration with hydrological variables and antecedent conditions (e.g. Mirus et al., 2018; Greco et al., 2023; Palazzolo et al., 2025, DiStefano et al., 2023, Zhang et al., 2026), the definition of such type of thresholds strictly depend on the quality of landslide archive and the corresponding hydro-meteorological dataset, not always available. The proposed CWI–CSM thresholds are completely physically-based, thus particularly useful when reliable data is limited or not available to carry out empirical studies".*

2. The combination of Equations (1) – (6) and the conceptual framework illustrated in Figure 2 leads to an overly simplified classification of shallow landslide triggering conditions, neglecting several complex processes related to rainfall infiltration and, more importantly, heterogeneous stratigraphic conditions. Authors should justify this choice explaining any advantage of this approach.

Reply: *We thank the referee for this observation. The simplifications reflect the primary objective of the study, which is not to reproduce the complete hydrological evolution leading to failure, but rather to analytically derive the critical hydrological states associated with slope instability in a simple manner, that can be easily applied even at regional scale. By adopting a simplified formulation, the proposed framework provides closed-form thresholds at any depth requiring substantially fewer input data than fully transient physically based models. The real complex conditions, for example in terms of heterogeneous stratigraphic soils, may easily be taken into account when data are available and/or at the further operation step, when doing the cross-matching of the thresholds with the real hydrological evolution. This separation between threshold definition and hydrological simulation allows the derived thresholds to be coupled with different hydrological models or observational datasets while preserving computational efficiency and physical interpretability.*

We have stated this simplification as follows:

3.2 [Case study] Topographic, land use and soil characterization (L254): *"As a further simplification, parameter values are kept constant with depth"*

5.2 [Discussion] Limitations and future perspectives (L585-589): *"Model performance and sensitivity analyses further highlight the influence of uncertainties in geotechnical parameters and failure depth. This reflects the choice of assuming homogeneous stratigraphic conditions and homogeneous mechanical parameters across the basin. However, the real complex conditions, for example of heterogeneous stratigraphic soils, may be easily taken into account when data are available".*

3. The sensitivity analysis does not consider the case of zero effective cohesion, although this condition is subsequently included (and then effectively overlooked again) during model calibration, where cohesion is assumed to vary between 0 and 35 kPa. Zero cohesion is a very common condition in surficial soils and should therefore receive much greater attention.

Reply: *Yes, we agree that the zero-cohesion case was not explicitly illustrated in the sensitivity analysis, Although cohesion values from 0 to 35 kPa were already included in the calibration procedure. Following the reviewer's suggestion, we derived the maps evaluated for c = 0 kPa across the considered failure-plane depths, z, reported in Figure 1 of this document. The resulting spatial stability patterns of conditionally stable areas are insensitive to z. This is because, for c=0, at the critical conditions (FoS = 1), the Eq. 1 of the paper loses the dependency on z either* $\frac{h_w}{z} = 0$ *and* $\frac{h_w}{z} = 1$*, which are the conditions used to assess ω* and ω**, respectively. Only the values of CSM slightly change across the depths (please see Eq. 4 of the paper)*

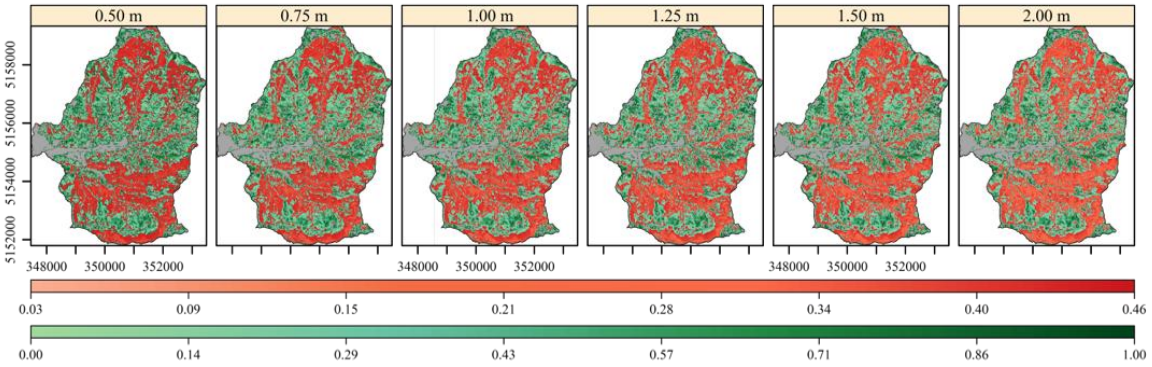


Figure 1 Sensitivity analysis of the spatial stability classification for zero effective cohesion (c = 0 Pa) across the considered depths to the potential failure plane. Shades of red indicate regions where slope instability is governed by unsaturated soil-moisture conditions (red graduate colorbar, CSM), while shades of green indicate regions where instability is controlled by the wetness index (graduate green colorbar, CWI). Gray areas denote unconditionally stable terrain at the respective failure-plane depth.

We added a sentence in the paper as follows:

3.4 [Case study] Sensitivity analysis (L282-284): "*The cohesionless case (c = 0 kPa) was also considered but not reported here, since it turns in a particular case in which the stability regions are not sensitive to z, because the equation that gives the ω* and ω** loses its dependency on z".*

4. In the sensitivity analysis, the ranges of shear strength angle values are reported with two decimal places. This level of precision merely reflects numerical computations and does not have any meaningful physical significance for this parameter.

Reply: *We thank the referee for this observation. Accordingly, we have revised the manuscript by rounding these values to one decimal place throughout the text and figures.*

5. The entire analysis assumes a single homogeneous soil type across the whole study area. Over an area of 28.62 km² (the Pontaiba catchment), the soil is assumed to exhibit identical shear strength properties and identical hydraulic parameters of the soil water retention curve (Table 1). This assumption appears overly restrictive and difficult to generalize.

Reply: *We thank the referee for this important observation. We agree that representing the basin using a single set of soil hydraulic and mechanical properties is a simplification, and we have clarified the basis and implications of this assumption in the revised manuscript.*

As described in Lines 217–226 of the original manuscript, the homogeneous parameterization was not assumed a priori. We first characterized the spatial soil information using two widely used global soil databases, SoilGrids (Hengl et al., 2017) and the Harmonized World Soil Database version 2.0 (HWSD v2; FAO, 2023). SoilGrids identified predominantly loam and clay-loam textures within the Pontaiba catchment, whereas HWSD classified the basin uniformly as loam. We extracted the spatially available sand, silt, and clay fractions from the two datasets and calculated their representative compositions, which were then plotted on the USDA soil-texture triangle (Fig. 1). The resulting compositions are consistent with the loam classification, while the mean composition of the SoilGrids-derived clay-loam pixels also plots close to the loam field. Together with the dominance of loam in the SoilGrids classification and the uniform loam classification in HWSD, these results support the adoption of a uniform loam texture for the basin. Thus, the use of a single soil texture was a data-supported simplification rather than an assumption made independently of the available soil information.

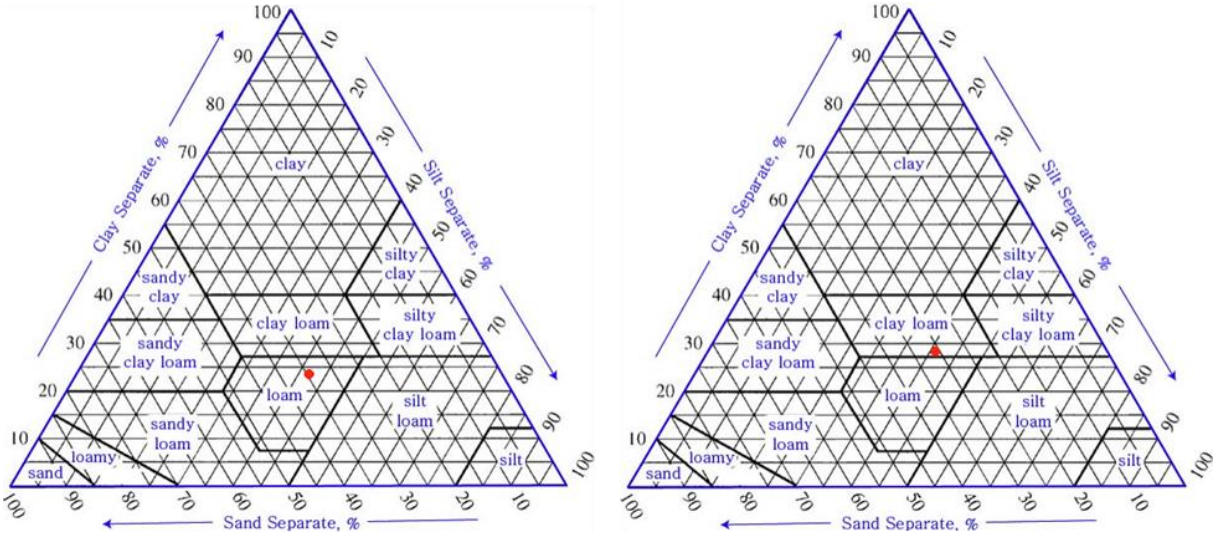


Figure 2 USDA soil-texture triangle showing the mean sand, silt, and clay compositions of the SoilGrids (left) and HWSD derived soil classes within the Pontaiba catchment. The mean composition of the clay-loam pixels (red point) according to HWSD plots close to the loam field, supporting the adoption of a representative loam texture for the subsequent hydro-mechanical characterization.

This was used to characterize the hydrological parameters. With regard to the shear strength properties and mechanical parameters, clearly this is a simplification due to the absence of data and the associated uncertainty that characterize these applications (Arnone et al., 2016; Burton et al., 1998). A further advance of the

methodology might be the inclusion of the level of uncertainty. However, note that the proposed framework can readily accommodate spatially distributed properties wherever such information is available.

We have clarified this point in the Discussion and explicitly acknowledge the homogeneous-soil assumption as a limitation of the present application. See reply of comment 2 and the following:

3.2 [Case study] Topographic, land use and soil characterization (L236-247): *"The soil is characterized using two global soil databases: SoilGrids (Hengl et al., 2017), developed by ISRIC – World Soil Information, and the Harmonized World Soil Database v2.0 (HWSD v2; FAO, 2023). SoilGrids provides gridded estimates (~250 m resolution) of USDA soil texture classes. Within the study basin, SoilGrids indicated two dominant soil classes, loam and clay loam, whereas HWSD classified the basin uniformly as loam. To evaluate the consistency of the available soil information, the sand, silt, and clay fractions associated with the spatial soil data were extracted across the basin and their representative compositions were calculated. The resulting compositions were evaluated using the USDA soil-texture classification. The mean composition associated with the SoilGrids-derived clay-loam pixels was close to the loam field, while the composition derived from the HWSD soil information was also consistent with a loam texture. Together with the dominance of loam in the SoilGrids classification and the uniform loam classification in HWSD, these results support the adoption of a uniform loam texture for the basin. Consequently, a single loam texture was adopted for the subsequent hydro-mechanical characterization. For the selected soil texture, the hydrological parameters were derived using the pedotransfer functions implemented in ROSETTA software (Schaap et al., 2001, 2004; Zhang and Schaap, 2017)."*

5.2 [Discussion] Limitations and future perspectives (L585-594): *Model performance and sensitivity analyses further highlight the influence of uncertainties in geotechnical parameters and failure depth. This reflects the choice of assuming homogeneous stratigraphic conditions and homogeneous mechanical parameters across the basin. However, the real complex conditions, for example of heterogeneous stratigraphic soils, may be easily taken into account when data are available. Morphometric approaches or remote sensing-based methodologies for estimating spatially distributed soil thickness (e.g., Zizioli et al., 2013, Yurdakul et al., 2025) provide a useful alternative, where direct observations of soil depth are unavailable, that might be considered in further development. Additionally, the application showed that the distributed character of the obtained maps derives largely from terrain morphology; this reflects the resolution of the available soil data rather than a structural limitation of the method. In catchments with strong lithological or pedological contrasts the framework would be expected to produce threshold patterns not aligned with topography. Its behavior in such settings remains to be tested.*

6. The most critical aspect of the manuscript concerns the calibration of the proposed model, which ultimately lacks a proper validation. The frequency distribution of the depths assigned to the 48 mapped landslides (Figure 8) appears rather arbitrary. The authors justify this choice by referring to field surveys carried out on landslides that occurred in a geomorphological setting similar to that of the study area (Lines 378–379). However, they also acknowledge that the actual or potential thickness of the soil involved in the landslides is unknown. This apparent inconsistency should be clarified.

Reply: *In this methodology we assumed that topography and all soil parameters are estimated from knowledge of the site except soil cohesion, which is important in the unsaturated zone. If data on landslide depths were available, together with landslide locations and soil moisture at failure, then cohesion could be obtained by simple inversion. In most places, though, we do not have depth of failure. The approach taken here is analogous to a constrained maximum likelihood exercise, numerically solved using a Monte Carlo method. We constrain the possible failure depths with a distribution from an analog site. The range of possible cohesion values is given by the nature of the soil. The goal is then to search the two-parameter space to maximize the probability of producing*

the observed landslide's locations in the data set. This is standard parameter estimation and it implies uncertainty, much like any regression of parameter estimate procedure (repeated from response to referee #1).

Like in most locations where data is available, the actual failure depths of the 48 inventoried landslides in the Pontaiba catchment are unknown. We believe that a reasonable approach is to assign a frequency distribution of landslides depths based on observations from a basin of similar characteristics but guided by local observations. The empirical distribution was constructed from reported failure depths of shallow rainfall-induced landslides in geomorphologically comparable terrain (Ramos-Scharrón et al., 2022). However, we refined and verified the validity of the distribution through field observations in our study area and comparison to other reported observations. (Del Fabbro et al., 2024; Bolla et al., 2026). Figure 2 reports the picture of two small landslides within the study area, which have been included in the supplementary material.



Figure 3 Field observations from the Pontaiba catchment showing small shallow failures and soil/regolith thicknesses of approximately 1.5 m (left), 1.0 m (centre), and 0.8 m (right) at selected locations. These observations provide field-based evidence of shallow soil mantles and support the plausibility of the failure-depth range adopted in the probabilistic calibration.

In this regard, see the following changes:

3.5 Model calibration procedure (L390-391): "*The empirical frequency distribution was then constructed from reported failure depths of shallow rainfall-induced landslides in the geomorphologically comparable terrain."*

Caption of Fig. 6 (Fig. 7 in the old version): *"Empirical frequency distribution of assigned depths to failure plane used in the probabilistic depth assignment procedure, based on reported failure depths from shallow rainfall-induced landslides in a geomorphologically comparable environment (Ramos-Scharrón et al., 2022), highlighting the emphasis on mid-range depths (with a peak at 1.25 m)".*

5.2 [Discussion] Implications for Landslide Early Warning Systems (L565-573): "*Specifically, in the procedure, we assumed that topography and all soil parameters are estimated from knowledge of the site except soil cohesion. If data on landslide depths were available, together with landslide locations and soil moisture at failure, then cohesion could be obtained by simple inversion. In most places, though, we do not have depth of failure. Then approach taken here is analogous to a constrained maximum likelihood exercise, numerically solved using a Monte Carlo method. We constrain the possible failure depths with a distribution from an analog site. The range of possible cohesion values is given by the nature of the soil. The goal is then to search the two-parameter space to maximize the probability of producing the observed landslide's locations in the data set. This is standard parameter estimation which implies uncertainty (Beven, 2006), much like any regression of parameter estimate procedure".*

Regarding the validation, we agree that this is the most critical aspect of our work. To investigate it, we carried out a spatially blocked k-fold cross-validation, (Roberts et al., 2017; Valavi et al., 2019), which reduces spatial dependence between calibration and validation data and thereby limits overly optimistic estimates of predictive performance. The basin was divided into 1-km spatial blocks, which were randomly assigned to 5 folds. Thus, all cells within a given spatial block were assigned to the same fold, while the landslide inventory was not used as a grouping constraint; individual landslide polygons could therefore extend across multiple blocks and folds, shown in Figure 4(a). For each of 10 independent block-to-fold assignments, each fold was held out in turn while the CWI/CSM threshold was calibrated for a fixed depth (in this example equal to 1.5 m), exclusively using the remaining four folds. The calibrated cohesion was converted to its corresponding critical slope angle, ω, and this threshold was then applied unchanged to the held-out fold. Youden's J was calculated separately for the training and held-out data, shown in Figure 4(b). The procedure should evaluate both how well the threshold fits the observations used for calibration, and how well the same threshold performs when transferred to spatially distinct portions of the basin. The obtained values (Figure 4 and Table 1) show that: (i) the calibrated cohesion is very stable across the repetitions (c = 6.66±0.09 kPa) and so also the corresponding critical slopes, equal to about 35.94° ± 0.48° (Table 1), confirming the robustness of the physically based procedure and that the calibrated values fit the observations; (ii) the mean Youden's J decreased only modestly from the training, Jtrain= 0.268, to the test, Jtest = 0.241, (iii) however, the Youden's J is very unstable during the test phase (held-out, Jt), going from Jtrain = 0.268 ± 0.048 to Jtest=0.241±0.170, meaning that, on average the fitting skill is stable but may vary substantially across space.*

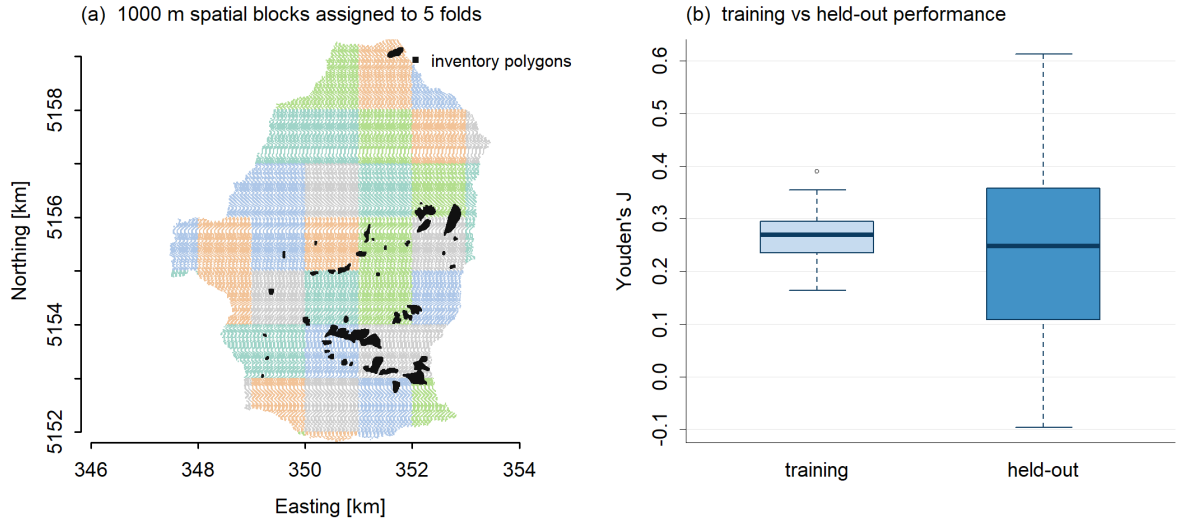


Figure 4 Spatially blocked cross-validation of the CWI/CSM threshold for depth equal to 1.5 m. (a) Spatial partitioning of the study area into 1-km blocks, randomly assigned to five folds. The landslide inventory is shown in black; landslide polygons were not constrained to individual folds and may therefore intersect multiple spatial blocks. The displayed partition represents one of the 10 independent block-to-fold assignments. (b) Distribution of Youden's J for model calibration on the training data and subsequent evaluation on spatially held-out folds across 10 repeated five-fold cross-validations (50-fold evaluations in total). For each fold, cohesion was calibrated using the training cells only, the corresponding critical slope angle (ω) was derived from the calibrated cohesion, and this fixed threshold was then applied without recalibration to the held-out cells. The comparison illustrates the change and increased variability in predictive performance when the calibrated threshold is transferred to spatially held-out data.*

Table 1 Summary of repeated spatially blocked five-fold cross-validation results for the CWI/CSM threshold framework. Values are reported as mean ± standard deviation across 50 fold evaluations from 10 independent repetitions.

Metric	Mean ± SD
Calibrated cohesion, *c*	6.66 ± 0.09 kPa
Critical slope, *ω**	35.94 ± 0.48°
Training Youden's *J*	0.268 ± 0.048
Held-out Youden's *J*	0.241 ± 0.170

However, the applied procedure has important criticisms that could affect the interpretation:

- *The calibration for each repetition is carried out for a fixed value of potential failure depth (e.g. h = 1.5m), without replicating rigorously the calibration procedure shown in the paper, which is more robust. This is the reason why the calibrated cohesion value is different (9.4 kPa vs 6.66 kPa).*
- *It is highly affected by the choice of the shape and number of the folds and repetitions, since the landslide polygons are not homogeneously distributed in space.*
- *The landslide polygons are not constrained to the random selection of the blocks, i.e. the same polygon can belong to both the calibration and the validation blocks.*

Therefore, we performed the analysis as an additional robustness check but did not include it as formal validation in the manuscript, because we do not consider this analysis to represent a fully independent validation of the proposed framework but rather may increase the level of uncertainties.

7. The soil thickness could be assessed by using morphological quantities. An example of these methods to assess the shallow soil thickness is reported in the following paper:
    a. Zizioli D., Meisina C., Valentino R., Montrasio L. (2013), Comparison between different approaches to modeling shallow landslide susceptibility: a case history in Oltrepo Pavese, Northern Italy, Natural Hazards and Earth System Sciences, 13, 559-573.

Reply: *We thank the referee for bringing this reference to our attention. We agree that morphometric approaches for estimating spatially distributed soil thickness, such as that proposed by Zizioli et al. (2013), provide a useful alternative where direct observations of soil depth are unavailable. We also have worked in this direction (see Yurdakul et al., 2025). With regard to the methodology development, the proposed analytical framework explicitly treats failure depth as a variable. Incorporating a spatially distributed soil-thickness model would require additional datasets, assumptions, and validation that would have increased the complexity and the uncertainty. The use of estimated soil thickness could be incorporated at an operational step.*

We have cited the suggested study in the Discussion and now acknowledge morphometric soil-thickness estimation as a promising direction for future developments of the proposed framework. See the following:

5.3 [Discussion] Limitations and future perspectives (L588-590): *"Morphometric approaches or remote sensing-based methodologies for estimating spatially distributed soil thickness (e.g., Zizioli et al., 2013, Yurdakul et al., 2025) provide a useful alternative, where direct observations of soil depth are unavailable, that might be considered in further development".*

8. In the ROC curve analysis (Figure 8), the Area Under the Curve (AUC) should also be reported, as it provides a standard quantitative measure of model performance.

Reply: *We agree that the Area Under the Receiver Operating Characteristic Curve (AUROC) is a standard measure of overall discriminatory performance. We used the Youden's J statistic because it provides a more appropriate information in our case: Youden's J is associated to the ROC and identifies the optimal operating point on the ROC curve by maximizing the combined sensitivity and specificity, whereas AUROC summarizes discrimination across all classification thresholds. In the present study, the ROC analysis is used specifically to identify the cohesion value associated with the optimal balance between sensitivity and specificity, for which the maximum Youden's J statistic provides the criterion used in our calibration procedure. We therefore retain Youden's J as the primary performance measure for parameter selection.*

9. Although the authors generate statistically based artificial scenarios, the uncertainty associated with two key parameters—soil cohesion and surficial soil thickness—introduces a substantial degree of randomness into the analysis, which remains largely unresolved.

Reply: *See point 6 above.*

10. The authors should more clearly emphasize the novelty and added value of the proposed model, particularly in comparison with other physically based distributed models available in the scientific literature, many of which already address several of the limitations identified by the authors themselves.
    a. Godt, J.W.; Baum, R.L.; Savage, W.Z.; Salciarini, D.; Schulz, W.H.; Harp, E.L. Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. Eng. Geol. 2008, 102, 214–226.
    b. Masi, E.B.; Tofani, V.; Rossi, G.; Cuomo, S.; Wu, W.; Salciarini, D.; Caporali, E.; Catani, F. Effects of roots cohesion on regional distributed slope stability modelling. Catena 2023, 222, 106853.
    c. Capparelli, G.; Versace, P. FLaIR and SUSHI: Two mathematical models for early warning of landslides induced by rainfall. Landslides 2011, 8, 67–79.
    d. Lizárraga, J.J.; Buscarnera, G. Safety factors to detect flowslides and slips in unsaturated shallow slopes. Géotechnique 2018, 68, 442–450.
    e. Simoni, S.; Zanotti, F.; Bertoldi, G.; Rigon, R. Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS. Hydrol. Process. 2008, 22, 532–545.
    f. Rossi, G.; Catani, F.; Leoni, L.; Segoni, S.; Tofani, V. HIRESSS: A physically based slope stability simulator for HPC applications. Nat. Hazards Earth Syst. Sci. 2013, 13, 151–166.
    g. Medina, V.; Hürlimann, M.; Guo, Z.; Lloret, A.; Vaunat, J. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. Catena 2021, 201, 105213.
    h. Sannino, G.; Tomei, F.; Bittelli, M.; Meisina, C.; Bordoni, M.; Valentino, R. (2025). A three-dimensional agro-hydrological model for predictive analysis of shallow landslides: CRITERIA-3D. Engineering Geology, 352, 108073.

Reply: *We thank the referee for the suggested references. Please, see reply 1 to referee #1 who raised the same observation. Additionally, we added the suggested references.*

Specific comments

11. **Line 57:** The expression “this additional cohesive strength” is more appropriately referred to as “apparent cohesion”.

Reply: *We have replaced the expression "additional cohesive strength" with "apparent cohesion" in the revised manuscript.*

12. **Line 90:** It should be clarified that the infinite slope assumption is applicable only to shallow landslides, where the thickness of the unstable soil layer is much smaller than the characteristic length of the landslide.

Reply: *We have clarified in the Introduction that the infinite-slope assumption is appropriate for shallow translational landslides, where the thickness of the unstable soil layer is much smaller than the characteristic length of the landslide. Please, see the following.*

2 Methodology (L97-98): *“When coupled with hydrological models, the most widely used FoS formulation is the one derived for the infinite slope model, valid for shallow translational landslides, where the thickness of the unstable soil layer is much smaller than the characteristic length of the landslide and grounded in the Mohr–Coulomb failure criterion (Montgomery and Dietrich, 1994; Terzaghi et al., 1996).*

13. **Line 104:** “This parameter represents the soil unit weight”. It should be clearly specified whether it refers to the saturated unit weight, the partially saturated unit weight, or the buoyant unit weight.

Reply: *In the present formulation, γ_s represents the saturated unit weight of the soil, assumed uniform from the potential failure plane to the ground surface. We revised the line. The saturated value was adopted in all the cases as a conservative assumption in the stability calculations.*

14. **Line 221:** How were the average percentages of clay, silt, and sand calculated?

Reply: *We thank the referee for requesting clarification. The average sand, silt, and clay contents were calculated separately for each SoilGrids-derived texture class. Specifically, all SoilGrids pixels within the basin were first grouped according to their assigned texture class (loam or clay loam), and the arithmetic mean of the sand, silt, and clay percentages was then calculated across the pixels belonging to each class. These mean compositions were subsequently plotted on the USDA soil-texture triangle to compare the representative textural composition of the two classes.*

We first characterized the spatial distribution of soil using two widely used global soil databases, SoilGrids (Hengl et al., 2017) and the Harmonized World Soil Database version 2.0 (HWSD v2; FAO, 2023). SoilGrids identified predominantly loam and clay-loam textures within the Pontaiba catchment. We extracted the sand, silt, and clay fractions from the available spatial soil information across the basin and calculated their representative compositions. These compositions were then plotted on the USDA soil-texture triangle (Figure 2 of this document). The resulting points fall within or close to the loam field, while loam also represents the dominant texture identified by SoilGrids. On this basis, a representative loam texture was adopted for the subsequent hydro-mechanical characterization and derivation of the hydraulic parameters. Thus, the use of a single representative soil texture was a data-supported simplification rather than an assumption made independently of the available soil information.

This was used to characterize the hydrological parameters. With regard to the shear strength properties and mechanical parameters, clearly this is a simplification due to the absence of data and the associated uncertainty that characterize these applications (Arnone et al., 2016; Burton et al., 1998). A further advance of the methodology might be the inclusion of the level of uncertainty. However, note that the proposed framework can readily accommodate spatially distributed properties wherever such information is available.

We have clarified this point in the Discussion and explicitly acknowledge the homogeneous-soil assumption as a limitation of the present application. See proposed changes in the reply to comment 5.

15. **Line 248:** The authors state that "there is no information on the time the landslides occurred." If the timing of the failures is unknown, how can landslide triggering be associated with a specific soil water content, given that neither the actual degree of saturation nor the rainfall conditions at failure are available?

The purpose of this effort is to develop a methodology that would indicate when failure is probable because certain critical parameters are reached. The data of when the observed landslides occurred is not known. All we can state is that the locations of failures are consistent with the methodology, given the maximum-likelihood-like parameter estimation. There is no attempt to use this methodology in a predictive mode with accounting of the evolution of the soil moisture. Nevertheless, the methodology could be driven by any hydrologic model and tested in the mode in the future.

We have further highlighted that the ISPRA-IFFI records are not time-stamped, that neither the soil moisture profile nor the groundwater configuration at the moment of failure are known, whereas validating CSM and CWI as operational triggering thresholds would require observations of the hydrological state concurrent with dated failures - a test we explicitly leave to future work.

3.3 [Case study] Landslide inventory (L264-270): *"…. the records are not time-stamped. …. As previously indicated, there is no information on the time the landslides occurred. Additionally, the ISPRA-IFFI does not include standardized measurements on depth to failure plane or slip surface or of the soil moisture profile and groundwater at the time of failure.*

16. **Line 293:** The authors state that "hydro-mechanical properties in Table 1 (including cohesion) are held constant." Does this mean that cohesion is fixed at 9 kPa?

Reply: *The cohesion value of 8 kPa was used only as a fixed reference value in the initial sensitivity analysis, where the objective was to isolate the influence of the assumed depth to the potential failure plane. In this analysis, cohesion was held constant at 8 k Pa while the failure-plane depth was varied from 0.50 to 2.00 m. Subsequently, cohesion was varied independently from 5 to 15 kPa in increments of 2.5 kPa to examine the sensitivity of the model response to this parameter. These analyses were performed prior to the calibration procedure and were intended to illustrate model sensitivity rather than to determine the final calibrated cohesion value.*

In the subsequent probabilistic calibration, cohesion was treated as an uncertain parameter and systematically varied over a broader range of 0 to 35 kPa together with the probabilistic realizations of failure depth. The Receiver Operating Characteristic (ROC)-based calibration yielded an optimum cohesion of 9,400 Pa (9.4 kPa), which was subsequently adopted for the calibrated stability analysis.

17. **Lines 352-361:** Figure 6 does not appear to provide any additional information beyond what is shown in the maps in the previous figures, other than a quantitative assessment that, however, does not seem to correspond to the physical reality of the study area or to the landslides observed.

Reply: *We thank the Referee for this observation. We have therefore removed the figure from the manuscript.*

18. **Lines 528–529:** The authors state that "The CSM and CWI metrics offer operational advantages for Early Warning Systems (EWS) because they are directly comparable with measurable or model-forecasted hydrological variables, such as soil moisture and groundwater levels." However, how do the authors envisage implementing an operational Early Warning System based on these forecasted variables when substantial uncertainties already affect the observed landslide dataset used for model calibration and validation?

Reply: *We thank the referee for this observation. As replied to referee #1 and #3, the aim of this step of the work was to develop a methodological framework to derive water content thresholds on physical bases that represent the minimum conditions to be met for potential triggering. Here we did not do forecasts. In an operational LEWS forecasts of hydrological variables could be compared with the identified thresholds to highlight potential landslide risk. Clearly, the identified thresholds are affected by the uncertainty of parameter estimates, as is the case with any model. See the following proposed changes:*

5.2 [Discussion] Implications for Landslide Early Warning Systems (L550-556): "*Operational LEWSs commonly rely on empirical or data-driven rainfall threshold approaches (Guzzetti et al., 2008; Bogaard and Greco, 2018; Segoni et al., 2018; Guzzetti et al., 2020). Although significant advances have been made in terms of methods and integration with hydrological variables and antecedent conditions (e.g. Mirus et al., 2018; Greco et al., 2023; Palazzolo et al., 2025, DiStefano et al., 2023, Zhang et al., 2026), the definition of such type of thresholds strictly depend on the quality of landslide archive and the corresponding hydro-meteorological dataset, not always available. The proposed CWI–CSM thresholds are completely physically-based, thus particularly useful when reliable data is limited or not available to carry out empirical studies".*

19. **Lines 553–557:** The manuscript does not adequately explain how the CWI and CSM indices can be effectively incorporated into a predictive modeling framework. The authors state that "In practice, these thresholds can be evaluated when soil moisture or groundwater conditions are available from field monitoring or hydrological model simulations driven by rainfall inputs, providing a physically interpretable basis for assessing slope stability under varying hydrological conditions." However, no explanation is provided regarding how rainfall observations or forecasts and groundwater table dynamics should be integrated into the proposed predictive framework.

Reply: *The proposed framework, which is not predictive, is limited to threshold values. The use in an operational manner can be implemented by simply comparing, grid by grid, the thresholds with forecasts of groundwater and soil moisture dynamics provided by hydrological models (e.g. tRIBS, SHE, and other distributed hydrological models) or observations. Please, see proposed changes:*

Abstract (L15-L17): "*The aim is deriving spatially consistent maps of thresholds that identify areas susceptible to rainfall-triggering landslides; the thresholds are therefore 'static' and not predictive"*

5.2 [Discussion] Implications for Landslide Early Warning Systems (L545-550): "*The CSM and CWI metrics have the potential to provide operational benefits for Landslide Early Warning Systems (LEWS) and may represent one component of a physically-based early warning system. Real time evolution of soil moisture and/or groundwater levels can be obtained through hydrological monitoring networks, remote sensing products, or spatially-distributed hydrological models capable of estimating the evolving hydrological conditions; measurable or model-forecasted hydrological variables could be then directly comparable with the critical thresholds derived through our methodology to ultimately derive maps of landslide hazards*"

5.2 [Discussion] Limitations and future perspectives (L595-599): "*Finally, future research should focus on integrating the proposed threshold framework with existing distributed, process-based hydrological models to evaluate transient hydrological conditions against the derived critical thresholds, validating the resulting threshold exceedances against instrumented slopes, and exploring how additional processes such as vegetation effects and dynamic root cohesion (Murgia et al., 2022) play a role and could be incorporated into the threshold framework.*"

20. **Lines 564–566:** The limitation identified by the authors and proposed as future work has already been addressed in other physically based models. See general comments No. 1 and 10.

Reply: *The referee refers to the sentence "Future research should therefore prioritize coupling the stability framework with distributed, process-based hydrological models to simulate transient responses, validating predictions against instrumented slopes, and incorporating additional processes such as vegetation effects and dynamic root cohesion".*

We agree that distributed process-based hydrological modelling, vegetation effects, and root reinforcement have already been incorporated into several physically based landslide models, as discussed in response to General Comments 1 and 10, including our past contributions. Rather, the proposed future work concerns their integration with the CWI–CSM threshold framework developed in this study. We have revised the manuscript to make it clearer.

21. It would be better to express soil effective cohesion in kPa instead of Pa.

Reply: *We thank the referee for this suggestion. We agree that expressing soil effective cohesion in kilopascals (kPa) can be convenient for geotechnical applications. We maintained kPa throughout the manuscript including on the figures.*

Main references

Arnone, E., Caracciolo, D., Noto, L. V., Preti, F., and Bras, R. L.: Modeling the hydrological and mechanical effect of roots on shallow landslides, Water Resour. Res., 52, 8590–8612, https://doi.org/10.1002/2015WR018227, 2016.

Baum, R. L., Savage, W. Z., and Godt, J. W.: TRIGRS - A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, Version 2.0, https://doi.org/10.3133/ofr20081159, 2008.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol. (Amst)., 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

Bolla, A., Del Fabbro, M., Peloso, C., Camerotto, I., Peranić, J., Zorčić, J., Vivoda Prodan, M., and Dugonjić Jovančević, S.: Hydro-mechanical characterisation of flysch-derived colluvial slopes affected by rainfall-induced landslide, in: Proceedings of the 21st International Conference on Soil Mechanics and Geotechnical Engineering (ICSMGE), Vienna, Austria, 14–19 June 2026, edited by: Pistrol, Adam, and Schweiger, ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, https://doi.org/10.53243/ICSMGE2026-1420, 2026.

Burton, A., Arkell, T. J., and Bathurst, J. C.: Field variability of landslide model parameters, Environmental Geology, 35, 100–114, https://doi.org/10.1007/s002540050297, 1998.

Capparelli, G. and Versace, P.: FLaIR and SUSHI: two mathematical models for early warning of landslides induced by rainfall, Landslides, 8, 67–79, https://doi.org/10.1007/s10346-010-0228-6, 2011.

Distefano, P., Peres, D. J., Piciullo, L., Palazzolo, N., Scandura, P., and Cancelliere, A.: Hydro-meteorological landslide triggering thresholds based on artificial neural networks using observed precipitation and ERA5-Land soil moisture, Landslides, 20, 2725–2739, https://doi.org/10.1007/s10346-023-02132-5, 2023.

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M.: Geographical landslide early warning systems, Earth. Sci. Rev., 200, 102973, https://doi.org/10.1016/j.earscirev.2019.102973, 2020.

Iverson, R. M.: Landslide triggering by rain infiltration, Water Resour. Res., 36, 1897–1910, https://doi.org/10.1029/2000WR900090, 2000.

Lepore, C., Arnone, E., Noto, L. V., Sivandran, G., and Bras, R. L.: Physically based modeling of rainfall-triggered landslides: a case study in the Luquillo forest, Puerto Rico, Hydrol. Earth Syst. Sci., 17, 3371–3387, https://doi.org/10.5194/hess-17-3371-2013, 2013.

Masi, E. B., Tofani, V., Rossi, G., Cuomo, S., Wu, W., Salciarini, D., Caporali, E., and Catani, F.: Effects of roots cohesion on regional distributed slope stability modelling, Catena (Amst)., 222, 106853, https://doi.org/10.1016/j.catena.2022.106853, 2023.

Medina, V., Hürlimann, M., Guo, Z., Lloret, A., and Vaunat, J.: Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale, Catena (Amst)., 201, 105213, https://doi.org/10.1016/j.catena.2021.105213, 2021.

Mirus, B. B., Morphew, M. D., and Smith, J. B.: Developing Hydro-Meteorological Thresholds for Shallow Landslide Initiation and Early Warning, Water (Basel)., 10, 1274, https://doi.org/10.3390/w10091274, 2018.

Mirus, B. B., T. Bogaard, R. Greco, and M. Stähli: Invited perspectives: Integrating hydrologic information into the next generation of landslide early warning systems. NHESS, 25, 169–182, doi.org/10.5194/nhess-25-169-2025, 2025

Montgomery, D. R. and Dietrich, W. E.: A physically based model for the topographic control on shallow landsliding, Water Resour. Res., 30, 1153–1171, https://doi.org/https://doi.org/10.1029/93WR02979, 1994.

Palazzolo, N., Cancelliere, A., Zofei, R. D., and Peres, D. J.: Use of delayed ERA5-Land soil moisture products for improving landslide early warning, Natural Hazards and Earth System Sciences, 25, 4907–4919, https://doi.org/10.5194/nhess-25-4907-2025, 2025.

Roberts, D. R., Bahn, V., Ciuti, S., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929. https://doi.org/10.1111/ecog.02881, 2017

Rossi, G., Catani, F., Leoni, L., Segoni, S., and Tofani, V.: HIRESSS: a physically based slope stability simulator for HPC applications, Natural Hazards and Earth System Sciences, 13, 151–166, https://doi.org/10.5194/nhess-13-151-2013, 2013.

Sannino, G., Tomei, F., Bittelli, M., Meisina, C., Bordoni, M., and Valentino, R.: A three-dimensional agro-hydrological model for predictive analysis of shallow landslides: CRITERIA-3D, Eng. Geol., 352, 108073, https://doi.org/10.1016/j.enggeo.2025.108073, 2025.

Simoni, S., Zanotti, F., Bertoldi, G., and Rigon, R.: Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS, Hydrol. Process., 22, 532–545, https://doi.org/10.1002/hyp.6886, 2008.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G.: blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225–232. https://doi.org/10.1111/2041-210X.13107, 2019.

Zhang, S., Pecoraro, G., Huang, D., Peng, J., Zhang, B., and Calvello, M.: From hydro-meteorological thresholds towards an operational warning model for landslides at regional scale: A real-case application, Eng. Geol., 362, 108542, https://doi.org/10.1016/j.enggeo.2026.108542, 2026.

Zizioli, D., Meisina, C., Valentino, R., and Montrasio, L.: Comparison between different approaches to modeling shallow landslide susceptibility: a case history in Oltrepo Pavese, Northern Italy, Natural Hazards and Earth System Sciences, 13, 559–573, https://doi.org/10.5194/nhess-13-559-2013, 2013.