https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

ROC-optimized rainfall thresholds for typhoon-induced landslides: environmental and magnitude stratification in Zixing City, China

Weifeng Xiao[1], Ge Liu*[2], Weimin Huang[3], Zhenghui Xiao[1], Luguang Luo[1]

[1]School of Earth Sciences and Spatial Information Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[2]Northeast Institute of Geography and Agroecology, CAS, Changchun 130102, China
[3] Hunan Center of Natural Resources Affairs, Changsha 410118, China

Corresponding author.
E-mail address: liuge@iga.ac.cn (Ge Liu)

Abstract: Typhoon-induced landslides pose severe hazards in mountainous East and Southeast Asia, yet conventional intensity-duration thresholds inadequately capture complex rainfall structures and geoenvironmental heterogeneity. We developed environment- and magnitude-stratified rainfall thresholds using 705 landslides triggered by Typhoon Gaemi (July 2024) in Zixing City, China. Environmental stratification by slope, lithology, fault proximity, and vegetation yielded seven geomorphic domains. Receiver operating characteristic (ROC) analysis optimized dual-parameter thresholds combining 24-hour cumulative rainfall (AccR24h) with maximum hourly intensity (Imax). Stratified AccR24h+Imax models achieved AUC = 0.68–0.83, exceeding unstratified baselines (AUC = 0.70) by 13%. Slope gradient exerted dominant control: steep terrain (>30°) required 37% less AccR24h than gentle slopes. Discriminant weights revealed process hierarchies—high-susceptibility domains assigned 62% weight to accumulation versus 38% to intensity. Medium-scale landslides (500–5,000 m²) required 29% higher AccR24h and exhibited 38% longer time lags, indicating progressive deep-seated failure. Magnitude-specific thresholds reduced false alarms by 56% while maintaining 89% sensitivity. The methodological

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

framework (environmental stratification + dual-parameter optimization) is applicable to other typhoon-affected regions, though threshold values and relative parameter weights require recalibration with local multi-event inventories. Integrating geoenvironmental controls with compound rainfall metrics substantially improves early warning precision for typhoon-induced landslides.

Keywords: Rainfall thresholds; Typhoon-induced landslides; Environmental stratification; ROC analysis; Magnitude-dependent triggering

1 Introduction

Typhoon-induced landslides pose a persistent and significant threat to the coastal mountains of East and Southeast Asia. In these regions, extreme rainfall totals and prolonged storm durations routinely exceed the limits established by thresholds calibrated for ordinary meteorological events, as demonstrated by Dolojan et al. (2025) and Zhuang et al. (2022). Unlike frontal systems or isolated convective storms, tropical cyclones deliver complex rainfall patterns, characterized by sustained moderate precipitation interspersed with intense convective bursts. These patterns generate hydro-mechanical conditions that are rarely captured by conventional intensity-duration (I-D) frameworks (Abancó et al., 2021; Chung and Li, 2022). When typhoons stall over mountainous terrain, a frequent occurrence along Nanling ranges in southeastern China, accumulated rainfall can surpass 500 mm within a 24–48 hour period. This saturation of weathered regolith often triggers spatially clustered slope failures (Li et al., 2019; Ma et al., 2025). Given climate projections that indicate intensifying tropical cyclones and heavier rainfall over the western North Pacific (Knutson et al., 2020), refining landslide early-warning systems for typhoon conditions has become an operational imperative.

Empirical rainfall thresholds, pioneered by Caine (1980) and refined over four decades (Peruccacci et al., 2017; Huang et al., 2022; Brunetti et al., 2025), remain a fundamental

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

component of regional forecasting systems. Most implementations adopt power-law I-D relations that distinguish triggering from non-triggering rainfall events through visual envelope fitting or percentile cutoffs (Guzzetti et al., 2020). However, three fundamental limitations constrain their performance in typhoon settings. First, typhoon rainfall totals and peak intensities often lie outside the calibration range of thresholds derived from monsoon or frontal storms, introducing uncertainty when extrapolating (Dolojan et al., 2025; Choowong et al., 2025). Second, conventional I-D models assume consistent rainfall-failure relationships across heterogeneous landscapes. Yet, geoenvironmental controls, such as slope gradient, lithology, fault proximity, and vegetation, modulate both the magnitude and temporal structure of triggering rainfall (Segoni et al., 2018; Natalia and Yang, 2024). Third, single-parameter thresholds inadequately represent the dual processes governing typhoon-induced failures: deep-seated saturation from prolonged accumulation and rapid infiltration spikes from convective bursts (Bogaard and Greco, 2018; Wu et al., 2025). These deficiencies are especially pronounced in regions where stalled typhoons concentrate rainfall over windward slopes, producing steep spatial gradients in both accumulation and intensity within narrow timeframes.

The role of antecedent rainfall in preconditioning slopes for failure is widely acknowledged (Liu et al., 2023; Natalia and Yang, 2024), but its integration into threshold schemes varies substantially. Some studies emphasize multi-day accumulations (Segoni et al., 2018; Prenner et al., 2019), while others prioritize peak intensity for shallow rapid failures (Peres and Cancelliere, 2014; Leonarduzzi et al., 2017). Because typhoons deliver both, clarifying their relative importance becomes essential. Equally critical, yet less explored, is the temporal lag between peak rainfall and slope failure (Marino et al., 2020; Liang et al., 2025). This lag encodes information about infiltration pathways, pore-pressure diffusion, and subsurface response times. In landscapes with thick weathered mantles and structural

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

discontinuities, such as Zixing City's fault-fractured granite terrain, hydrologic response can vary by hours across adjacent hillslopes, influencing when failures concentrate during a storm.

A further constraint is the binary treatment of landslide occurrence. Most threshold studies classify events as present or absent without differentiating small shallow slips from larger, deeper-seated failures (Mondini et al., 2023; Sun et al., 2023). Yet consequences scale nonlinearly with volume (Tang et al., 2024; Schlögel et al., 2015), and triggering mechanisms may differ: larger failures may require longer saturation periods and deeper pore-pressure buildup (Intrieri et al., 2019; Handwerger et al., 2019). Establishing magnitude-specific thresholds would improve both hazard assessment and process understanding, particularly for operational systems that must prioritize warnings for high-consequence events.

Methodologically, threshold calibration has advanced beyond visual envelope fitting (Caine, 1980), yet many studies still rely on arbitrary percentile cutoffs (Bogaard et al., 2018; Karki et al., 2019) that do not explicitly optimize the trade-off between detection and false alarms. Receiver operating characteristic (ROC) analysis rigorously selects thresholds by evaluating both sensitivity and specificity across all possible cutoff values (Steger et al., 2024). Despite its widespread use in diagnostic medicine and machine learning, ROC optimization remains underutilized in landslide threshold research, particularly for datasets stratified by environmental factors and landslide magnitude.

Typhoon Gaemi's traverse across Hunan Province in late July 2024 offers a unique natural experiment. The storm triggered 705 documented landslides in Zixing City within 48 hours, with quasi-stationary rainbands producing extreme accumulations (greater than 620 mm/24h) and peak intensities (greater than 130 mm/h) across heterogeneous terrain. Strong orographic gradients created spatial variability in both rainfall structure and failure density, while the narrow temporal concentration minimizes confounding from seasonally varying antecedent conditions, a problem that complicates multi-event analyses. This single-typhoon

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

dataset enables direct interrogation of how rainfall characteristics translate into failure timing, magnitude, and spatial distribution under extreme forcing.

This study develops environment- and magnitude-dependent rainfall thresholds optimized through ROC analysis. We address three specific questions: (1) How do lithology, slope gradient, vegetation cover, and fault proximity modulate threshold magnitudes for typhoon-induced landslides? (2) Do larger landslides require systematically different rainfall conditions, in terms of accumulation, intensity, or temporal structure, than small failures? (3) Can ROC-optimized dual-parameter thresholds combining antecedent accumulation with peak intensity outperform conventional single-parameter models while maintaining acceptable false-alarm rates? By integrating geoenvironmental stratification with magnitude-specific calibration and explicit performance optimization, this framework extends conventional I-D approaches and provides a transferable methodology for typhoon-affected mountainous regions.

2 Study area and data sources

2.1 Study area

This study focuses on Zixing City (25°34′–26°18′ N, 113°08′–113°44′ E), a 2,747 km² area within the Nanling Mountains of southeastern Hunan Province, China (Fig. 1). Zixing City, as detailed in our previous work (Xiao et al., 2026), is characterized by steep topography and a geology of metasediments and granites intersected by fault networks, predisposing it to rainfall-induced landslides. The region receives an average annual rainfall of 1,550 mm, concentrated during the East Asian monsoon and typhoon season. The combination of prolonged, intense typhoon rainfall, steep slopes, and permeable substrates frequently triggers landslides. Building on our earlier research, this study develops environment- and magnitude-stratified rainfall thresholds to improve landslide prediction accuracy, addressing the limitations of a single, uniform threshold in capturing spatial variations in landslide triggers.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

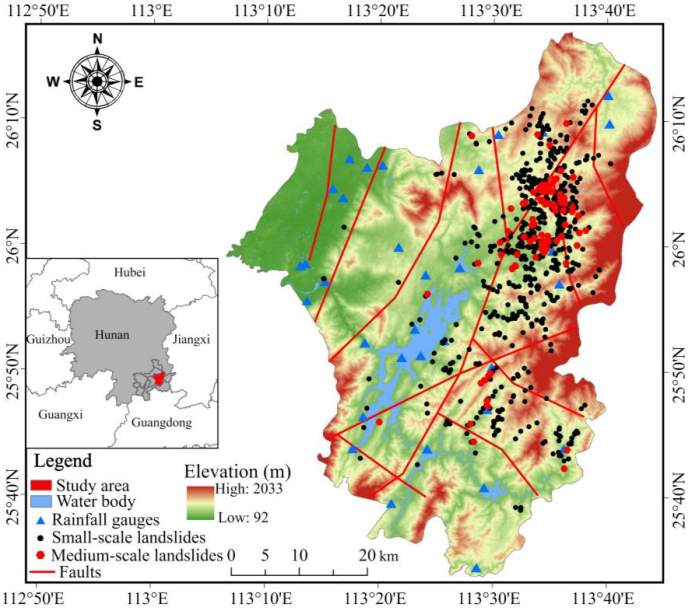


Figure 1 Geographical distribution of the study area, landslides and rainfall gauges.

2.2 Data collection and preprocessing

2.2.1 Landslide catalog

The landslide inventory used in this study comprises 705 landslides triggered by Typhoon Gaemi on July 27, 2024, as documented by the Hunan Center for Natural Resources Affairs. As in our previous study (Xiao et al., 2026), the dataset includes both medium-scale (500-5,000 m^2) and small-scale (area <500 m^2) failures. Field surveys and high-resolution satellite imagery were used to verify landslide locations and ensure the completeness of the catalog.

2.2.2 Environmental stratification factors and data sources

To develop environment-dependent rainfall thresholds, we selected four key factors for stratifying landslide-triggering conditions based on the geoenvironmental characteristics of the study area: slope, distance to fault, lithology, and normalized difference vegetation index (NDVI) (Fig. 2). These factors, representing topographic, structural, geological, and ecological controls on landslide susceptibility (Panchal et al., 2021; Xiao et al., 2025; Ma et

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

al., 2025), were also considered in our earlier work, but are used here for a different purpose. Data sources for these factors, consistent with our previous study, included: a 30-meter digital elevation model (DEM) from the Geospatial Data Cloud (https://www.gscloud.cn) for slope gradient extraction; 1:50,000-scale cartographic maps for distance to fault; 1:100,000-scale geological maps for lithology; and Landsat 8 OLI imagery from the Geospatial Data Cloud platform for NDVI calculation.

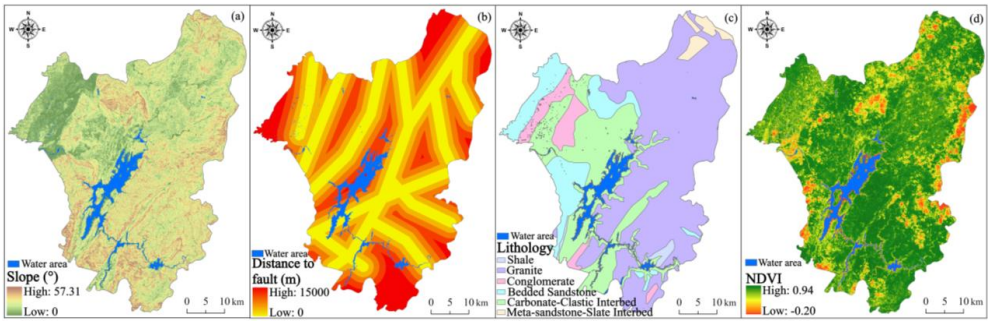


Figure 2 Spatial distribution of environmental factors used for threshold stratification.

2.2.3 Data preprocessing and spatial standardization

Consistent with our previous methodology (Xiao et al., 2026), all factors were resampled to a uniform 60-meter resolution to balance computational efficiency with scale appropriateness. For threshold stratification, each landslide location was assigned the corresponding environmental attribute values from the spatially standardized factor layers. Landslides smaller than the grid resolution were assigned to their nearest cell centroid. When multiple small failures occurred within a single cell, their rainfall observations were treated as independent samples for threshold calibration.

2.2.4 Rainfall data interpolation and assignment to landslide locations

Rainfall data were obtained from 35 automatic rain gauge stations distributed across Zixing City and its surrounding areas (Fig. 1). These stations, operated by the Hunan Meteorological Administration, recorded hourly precipitation during Typhoon Gaemi and the preceding period. Six rainfall parameters were calculated for each landslide event: cumulative

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

rainfall in 6, 12, 24, and 48 hours preceding landslide occurrence (AccR6h, AccR12h, AccR24h, and AccR48h), maximum hourly rainfall intensity within the 24-hour pre-landslide period (Imax), and the time lag between maximum intensity and landslide occurrence (Pri_TL). To derive these parameters for all 705 landslide locations, Kriging interpolation was applied to generate continuous rainfall surfaces from point measurements (Fig. 3).

Spherical models were fitted to variograms through iterative procedures, with AIC serving as the selection criterion. We validated interpolation quality by sequentially excluding each station and reconstructing its rainfall values from the other 34 locations. Prediction accuracy was quantified using four metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Correlation Coefficient (R), and Nash-Sutcliffe Efficiency (NSE).

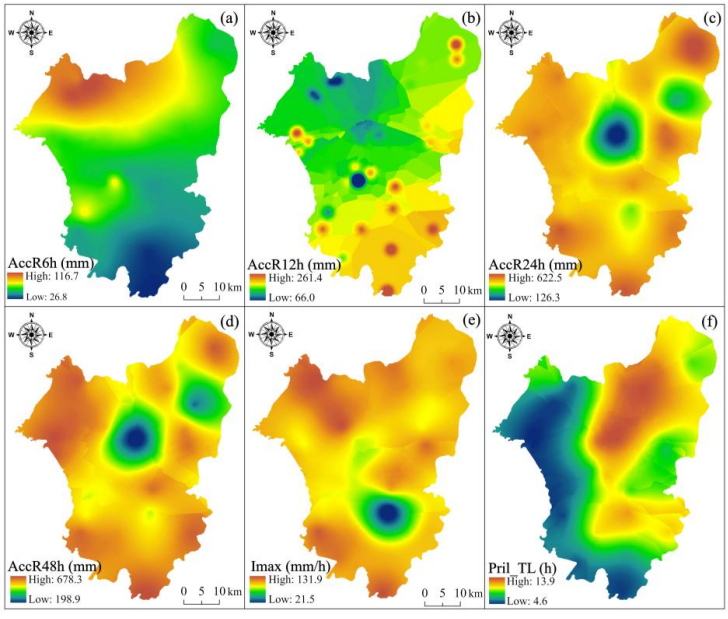


Figure 3 Spatial distribution of Kriging interpolation for 6 rainfall parameters.

Validation results showed acceptable accuracy across all parameters (R = 0.73–0.88, NSE = 0.68–0.84) (Table 1). While the gauge density in mountainous terrain posed some limitations, the interpolation quality was adequate for characterizing regional rainfall patterns and their relationship to landslide occurrence.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Table 1 Kriging interpolation accuracy assessment for rainfall parameters.

Parameter	RMSE	MAE	R	NSE
AccR6h (mm)	3.8	2.7	0.79	0.74
AccR12h (mm)	9.2	6.8	0.85	0.81
AccR24h (mm)	14.6	11.3	0.88	0.84
AccR48h (mm)	22.1	16.9	0.82	0.78
Imax (mm/h)	2.1	1.6	0.81	0.76
Pre_TL (h)	1.8	1.3	0.73	0.68

3 Methods

This study establishes rainfall thresholds through a four-stage workflow (Fig. 4): (1) environmental stratification of the landslide inventory into geomorphologically homogeneous subsets; (2) baseline threshold calibration using single rainfall parameters (AccR6h, AccR24h, Imax) without stratification; (3) environment-dependent threshold refinement for both single-parameter and dual-parameter models (AccR6h+Imax, AccR24h+Imax), optimized via ROC analysis; and (4) magnitude-specific threshold derivation to distinguish triggering conditions for small versus medium-scale failures.

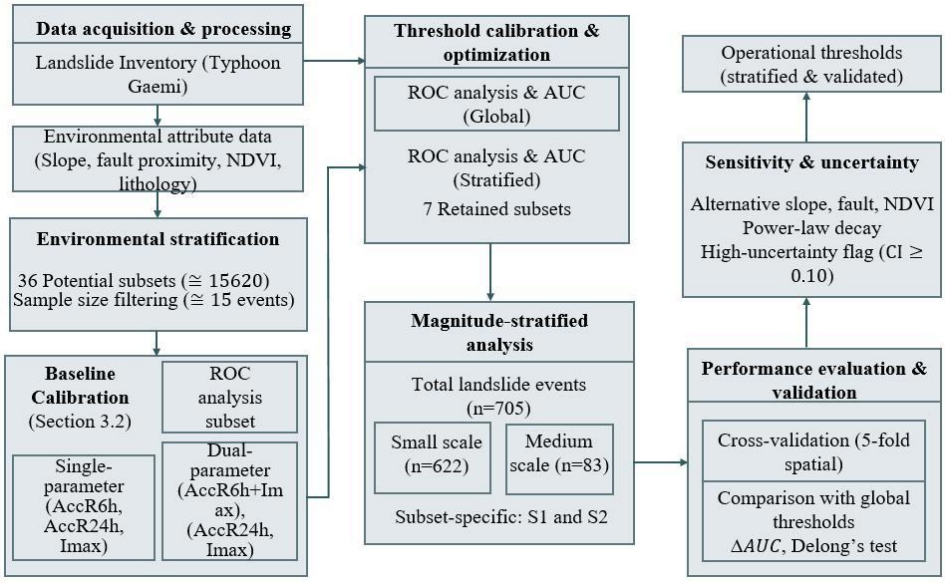


Figure 4 Methodological framework for deriving environment- and magnitude-dependent rainfall thresholds for typhoon-induced landslides

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

3.1 Environmental stratification of landslide events

Spatial heterogeneity in geoenvironmental conditions controls both the susceptibility of slopes to failure and the rainfall characteristics required to trigger instability. We partitioned the landslide inventory into environmentally homogeneous subsets to calibrate thresholds specific to distinct process domains, thereby avoiding the dilution effect that arises when incompatible failure mechanisms are aggregated.

3.1.1 Factor discretization and classification scheme

Four environmental factors identified in Section 2.2.2 were discretized based on process-relevant boundaries established in prior landslide research and regional geotechnical studies (Xiao et al., 2026; Huang et al., 2022):

(1) Based on landslide inventories in the region, slope gradients were categorized into three classes: gentle (≤20°), moderate (20°–30°), and steep (>30°). Slopes exceeding 30° exhibit a significantly higher frequency of failures. This observation aligns with the concept that these steeper slopes approach critical gravitational stress thresholds, making them more susceptible to shallow translational failures (Zhuang et al., 2022; Ma et al., 2025).

(2) Distance to fault: Two classes were defined for distance to fault: proximal (≤1 km) and distal (>1 km). The 1 km threshold is based on geophysical evidence suggesting that fault-related fracture networks, which enhance hydraulic conductivity and diminish effective cohesion, primarily influence the surrounding area within this range (Zhao et al., 2025).

(3) Lithology: The lithology was divided into two functional groups: granite-dominated terrain (comprising >70% of the study area and hosting >80% of landslides) and stratified sedimentary rocks (including carbonate–clastic interbeds, bedded sandstone, and meta-sandstone–slate interbeds). While intact granites possess considerable strength, they weather into thick saprolite characterized by relict discontinuities. This weathering process reduces shear resistance during periods of prolonged rainfall (Migoń and Vieira, 2014). Sedimentary

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

units tend to fail along bedding planes and due to differential weathering (Stead and Wolter, 2015)

(4) NDVI: Three classes were established for NDVI: sparse vegetation (≤0.3), moderate vegetation (0.3–0.6), and dense vegetation (>0.6). NDVI values exceeding 0.6 are indicative of closed-canopy forests. In humid subtropical regions, these forests reduce shallow landslide probability by 30–60% through root reinforcement and transpiration (Natalia and Yang, 2024; Segoni et al., 2018).

3.1.2 Subset generation and sample size filtering

Cross-classification of four factors yielded 3 (slope) × 2 (fault) × 2 (lithology) × 3 (NDVI) = 36 potential combinations. To ensure statistical robustness of threshold calibration, we imposed a minimum sample criterion: only subsets containing ≥15 landslide events were retained. This threshold balances two competing requirements: sufficient observations to constrain power-law or linear relationships against excessive fragmentation that would dilute environmental signals and inflate uncertainty.

Seven subsets met this criterion, retaining 620 landslides (87.9% of the total inventory) distributed across environmentally distinct domains. The remaining 85 events occurred in sparsely represented combinations (<4% study area) and were excluded to ensure statistical robustness.

3.1.3 Mechanistic rationale for stratified thresholds

The stratification framework delineates process domains where specific combinations of terrain attributes produce distinct rainfall-failure relationships. Slope angle controls driving shear stress, while fault proximity and lithology govern fracture density and weathering depth—factors that collectively determine the rate of pore pressure build-up and strength degradation during rainfall. Disaggregating the landslide inventory by these terrain controls

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

isolates mechanism-specific rainfall triggers, minimizing threshold dilution from averaging across incompatible failure modes (Segoni et al., 2018).

3.2 Threshold calibration

We first assessed multicollinearity among the six rainfall variables using variance inflation factors (VIFs). AccR12h and AccR48h exhibited severe collinearity (VIF > 10) and were excluded to prevent model instability. The remaining four variables (AccR6h, AccR24h, Imax, Pre_TL) showed acceptable independence (VIF < 5).

Pre_TL was retained for analysis but excluded from threshold derivation because it requires foreknowledge of failure timing. However, its spatial and magnitude-dependent patterns (Sections 4.1, 4.3.1) help validate proposed hydrologic mechanisms and may inform future process-based threshold design by constraining infiltration timescales in coupled models.

3.2.1 Baseline threshold

We established baseline thresholds for the three retained rainfall parameters (AccR6h, AccR24h, Imax) without environmental stratification to benchmark conventional global approaches. Empirical distributions were constructed at landslide locations (n=705) and non-landslide reference points (n=1410). Reference points were selected from stable locations ≥1 km from documented failures and distributed proportionally across the same terrain, lithology, vegetation, and fault proximity zones as the landslide dataset.

Initial thresholds were set at the 10th percentile of landslide distributions—a standard in operational early-warning systems (Peruccacci et al., 2017; Guzzetti et al., 2020)—yielding sensitivity ≈0.90 by design. True positive rates (TPR) and false positive rates (FPR) were calculated across all observed values to construct receiver operating characteristic (ROC) curves. Area under the curve (AUC) quantifies discriminatory power independent of threshold choice: AUC = 0.5 indicates no skill, AUC = 1.0 perfect separation (Steger et al., 2024). AUC

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

values were computed using the trapezoidal rule, with 95% confidence intervals from 2000 bootstrap replicates.

For dual-parameter thresholds, we standardized AccR6h, AccR24h, and Imax (zero mean, unit variance) and constructed linear discriminants:

$$S = w_1 \cdot \frac{P_1 - \mu_1}{\sigma_1} + w_2 \cdot \frac{P_2 - \mu_2}{\sigma_2} \quad (1)$$

where P_1 and P_2 denote paired rainfall parameters (e.g., P_1=AccR24h, P_2=Imax for the AccR24h+Imax model; P_1=AccR6h, P_2=Imax for the AccR6h+Imax model), μ and σ are their respective means and standard deviations, and wights w_1, w_2 were calibrated to maximize Youden's index:

$$J = \mathrm{TPR} + \mathrm{TNR} - 1 = \mathrm{Sensitivit\ y} + \mathrm{Specificit\ y} - 1 \quad (2)$$

Landslide occurrence is predicted when $S \geq C$, where C is a scalar cutoff. We tested two combinations: AccR24h+Imax (sustained antecedent rainfall plus peak intensity) and AccR6h+Imax (rapid response to recent accumulation and intensity bursts). Weights were derived via grid search (0.1 increments, $w_1 \in [0.2, 0.9]$, $w_2 = 1 - w_1$).

3.2.2 Environment-dependent threshold refinement

For each of the seven environmental subsets (Section 3.1.2), we compared rainfall distributions between landslide and non-landslide locations using quantile-quantile plots and Kolmogorov-Smirnov tests ($\alpha = 0.05$) to verify that stratification produced distinct rainfall signatures. Within each subset, thresholds for AccR6h, AccR24h, and Imax were determined by scanning candidate values and selecting those maximizing Youden's index—explicitly balancing detection and false alarms rather than adopting arbitrary percentile cutoffs.

For dual-parameter thresholds, we applied the same discriminant framework but calibrated weights separately for each subset. This allows the relative importance of accumulation versus intensity to vary with local conditions: in high-permeability fractured

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

terrain, intensity may dominate; in low-permeability settings, prolonged accumulation may be critical.

Statistical significance was assessed using Mann-Whitney U tests ($\alpha = 0.05$) with Benjamini-Hochberg correction for multiple comparisons. For each subset, we tested whether rainfall distributions differed significantly between landslide and non-landslide locations. Adjusted p-values were computed as: reject H_0 if $p(i) \leq (i/m) \times \alpha$, where i is the rank and m is the total number of tests.

We report two operating points for each threshold: (1) the 10th percentile baseline (for cross-subset comparability at fixed high sensitivity), and (2) the Youden-optimal point (for maximum discrimination). The decision rule remains a linear discriminant with scalar cutoff in standardized score space, not a two-dimensional surface over raw parameter pairs.

3.2.3 Magnitude-stratified threshold analysis

To test whether medium-scale failures (500-5,000 m², n=83) require systematically different rainfall forcing than small events (<500 m², n=622), we partitioned the landslide inventory by magnitude class within environmentally homogeneous subsets. Only subsets S1 and S2 contained sufficient medium-scale landslides ($n \geq 10$) for independent calibration.

For each magnitude class, we derived separate thresholds for AccR6h, AccR24h, Imax, and their dual-parameter combinations, then quantified threshold elevation as the percentage increase required for medium landslides relative to small events. If medium failures consistently demand higher rainfall totals, this indicates a threshold hierarchy tied to failure depth or volume.

We also assessed whether the relative importance of accumulation versus intensity shifts with landslide size by computing Spearman rank correlations (ρ) between each rainfall parameter and landslide occurrence for small and medium classes separately. Correlation magnitudes were compared using Fisher's r-to-z transformation ($p < 0.05$). Stronger

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

correlations with AccR24h for medium events would suggest dependence on deep-seated saturation (Xiao et al., 2026), whereas intensity dominance for small failures would indicate surface-driven rapid response (Dolojan et al., 2025).

3.3 Threshold optimization and performance evaluation

3.3.1 Receiver operating characteristic analysis

ROC curves provide threshold-independent assessment of discriminatory power. For each environmental subset and rainfall metric, we generated ROC curves by varying candidate thresholds across the observed range and plotting resulting TPR against FPR.

AUC quantifies overall performance: values near 0.5 indicate no discrimination, while values approaching 1.0 represent near-perfect separation. We calculated AUC via the trapezoidal rule with 95% confidence intervals from 2000 bootstrap replicates. For dual-parameter thresholds, ROC curves were constructed by treating the linear combination of standardized rainfall variables as a continuous decision variable.

3.3.2 Operational threshold selection

Threshold placement involves trade-offs between detection rate and false-alarm tolerance. High sensitivity (detecting most events) must be balanced against acceptable false-alarm rates to maintain warning system credibility and avoid alert fatigue.

We identified optimal operating points using Youden's index, which maximizes vertical distance from the ROC diagonal. This criterion weights sensitivity and specificity equally without requiring subjective cost functions. For operational context, we also calculated precision (positive predictive value: fraction of threshold exceedances corresponding to actual landslides) and F1 score (harmonic mean of precision and sensitivity) at the Youden-optimal point. High precision is valued by civil protection agencies to minimize unnecessary evacuations (Segoni et al., 2018).

3.3.3 Cross-validation and spatial independence

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Direct resubstitution inflates apparent accuracy when spatial clustering introduces autocorrelation (Brenning, 2012). We implemented spatial block cross-validation to obtain unbiased performance estimates.

The study area was divided into five spatially contiguous blocks of approximately equal area, each containing representative environmental conditions and landslide events. Thresholds were recalibrated five times, each iteration withholding one block for validation while training on the remaining four. Performance metrics (AUC, sensitivity, specificity, precision) were computed separately for each validation block. Mean and standard deviation across folds quantify generalization performance beyond calibration data.

3.3.4 Comparison with non-stratified approaches

To quantify the benefit of environmental stratification, we compared the performance of environment-dependent thresholds against a conventional global threshold applied uniformly across the study area. The global threshold was calibrated using the entire landslide inventory without subsetting by terrain attributes, following standard practice in regional early-warning systems (Guzzetti et al., 2020).

We computed the difference in AUC (ΔAUC) between stratified and global thresholds for each environmental subset, then aggregated these differences across the 7 subsets, weighting by the number of landslides in each subset. Statistical significance of performance improvement was evaluated using DeLong's test for comparing AUC values from correlated ROC curves. We also examined whether stratification reduced threshold variability: large differences in optimal threshold values across subsets would indicate that terrain controls substantially modulate rainfall-failure relationships, justifying the added complexity of environment-specific models.

3.4 Sensitivity and uncertainty analysis

3.4.1 Sensitivity to classification boundaries

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

We tested whether threshold values and model performance remain stable when boundaries of the three continuous environmental factors are adjusted. Four alternative schemes were evaluated:

(1) Slope boundaries: ≤15°, 15–35°, >35° (versus original ≤20°, 20–30°, >30°);

(2) Fault distance cutoff: ≤0.5 km proximal and ≤2 km proximal (versus original ≤1 km);

(3) NDVI boundaries: ≤0.4, 0.4–0.7, >0.7 (versus original ≤0.3, 0.3–0.6, >0.6).

Lithology was excluded because it represents geologically defined categories not subject to arbitrary discretization. For each scheme, we recalibrated Youden-optimal thresholds and AUC values across all seven subsets. Thresholds were deemed robust if AUC changes were <0.05 and threshold shifts within ±15%, consistent with operational uncertainty bounds (Dolojan et al., 2025).

3.4.2 Sample size and uncertainty qualification

We quantified how subset sample size affects threshold estimation uncertainty by fitting a power-law decay to the widths of the 95% confidence intervals (CIs) of AUC values, computed from 2000 bootstrap replicates for each environmental subset:

$$CI\ \text{width} = a \times n^{b} \tag{3}$$

where n is the subset sample size, a is a scaling coefficient, and $b < 0$ is the decay exponent. This form is theoretically motivated by the $n^{-0.5}$ convergence rate for asymptotically normal estimators. Parameters a and b were estimated via weighted least squares on log-transformed data. Model fit was assessed using R^2, and the significance of b was tested against zero with an F-test ($\alpha = 0.05$). Subsets with CI widths ≥0.10 were classified as high-uncertainty and flagged as requiring additional calibration data before operational deployment.

4 Results

4.1 Rainfall patterns and landslide distribution during Typhoon Gaemi

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Typhoon Gaemi delivered extreme rainfall to Zixing City on 26–27 July 2024, with spatial patterns shaped by orographic forcing and quasi-stationary rainbands. Six-hour accumulations (AccR6h) reached 116.7 mm in the northwestern sector (Fig. 3a), while 24-hour totals (AccR24h) peaked at 622.5 mm in the east-central mountains (Fig. 3c). Maximum hourly intensities (Imax) exceeded 131.9 mm/h in the south-central corridor where topographic convergence amplified convection (Fig. 3e). The highest AccR24h totals concentrated along elevated terrain >800 m, with widespread accumulations exceeding 126.3 mm across most of the study area.

The 705 mapped landslides clustered in zones where intense rainfall coincided with steep slopes and structural weaknesses (Fig. 1). The dominant hotspot in the eastern sector corresponds to AccR24h > 126.3 mm, slopes >35°, and proximity to faults within 1 km. Secondary concentrations along the western escarpment align with bedding-plane orientations in sedimentary rocks and peak 6-hour accumulations exceeding 85 mm. Areas with both AccR48h > 320 mm and NDVI > 0.6 (dense vegetation) account for 42% of all failures.

Time lags between peak rainfall and landslide occurrence (Pri_TL) varied spatially. In the western and southwestern portions of the study area, where steep, fractured granite terrains dominate, time lags are notably shorter (as low as 4.6 h). In contrast, the eastern and northeastern sectors, characterized by gentler slopes underlain by low-permeability metamorphic substrates, exhibit significantly longer time lags (up to 13.9 h). These extended lags indicate slower water diffusion and delayed pore-pressure generation, as the low-permeability substrates impede subsurface flow.

Environmental stratification (Section 3.1.2) retained 620 landslides (87.9% of total inventory) distributed across seven subsets (S1~S7) meeting the minimum sample criterion (n≥15). This filtered dataset includes 545 small-scale and 75 medium-scale failures. The remaining 85 events occurred in sparsely represented environmental combinations (e.g.,

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

415 sparse vegetation on gentle fault-proximal slopes occupying <4% of study area) and were
416 excluded from threshold calibration to maintain statistical robustness. All subsequent analyses
417 use this 620-event subset.

418 **4.2 Baseline threshold performance without environmental stratification**

419 **4.2.1 Baseline Performance and Limitations**

420 Applying the 10th percentile criterion to the full landslide sample (n = 705) yielded
421 baseline thresholds of AccR6h = 46.6 mm, AccR24h = 226.3 mm, and Imax = 44.0 mm/h
422 (Table 2). ROC analysis indicated moderate discrimination, with AUC values of 0.62 (0.59–
423 0.65) for AccR6h, 0.64 (0.61–0.67) for AccR24h, and 0.63 (0.61–0.66) for Imax.

424 **Table 2** Performance of baseline rainfall thresholds across the entire study area.

Parameter	Threshold Value	AUC (95% CI)	Sensitivity	Specificity	Precision	F1 score
AccR6h	46.6 mm	0.62 (0.59-0.65)	0.90	0.10	0.43	0.59
AccR24h	226.3 mm	0.64 (0.61-0.67)	0.90	0.13	0.45	0.60
Imax	44.0 mm/h	0.63 (0.61-0.66)	0.90	0.14	0.44	0.60
AccR6h+Imax	278.1 combined	0.69 (0.64-0.69)	0.90	0.11	0.48	0.63
AccR24h+Imax	98.6 combined	0.70 (0.67-0.74)	0.90	0.12	0.51	0.66

425 Following Equation (2) in Section 3.2.2, two-parameter combined thresholds employed
426 standardized linear discriminants of the form $s = w_1 \times [(AccR - \mu\_AccR)/\sigma\_AccR] + w_2 \times [(Imax - \mu\_Imax)/\sigma\_Imax]$, where landslide occurrence is predicted when $s \geq C$. For
428 AccR6h+Imax, optimal weights were $w_1=0.58$ and $w_2=0.42$ with $C=2.18$ ($\mu\_AccR6h=51.3$
429 mm, $\sigma\_AccR6h=18.7$ mm, $\mu\_Imax=45.2$ mm/h, $\sigma\_Imax=12.3$ mm/h). For AccR24h+Imax,
430 weights were $w_1=0.71$ and $w_2=0.29$ with $C=1.87$ ($\mu\_AccR24h=89.4$ mm, $\sigma\_AccR24h=31.2$
431 mm). These two-parameter combinations outperformed single-parameter thresholds:
432 AccR6h+Imax achieved an AUC of 0.69, and AccR24h+Imax reached 0.70, representing a
433 9.3% gain over AccR24h alone. At these baseline cutoffs, sensitivity was uniformly high
434 (0.90), but specificity remained low (0.10–0.14), yielding modest precision (0.43–0.51) and
435 F1 scores (0.59–0.66).

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Optimizing Youden's index J identified better operating points. For single parameters, Youden-optimal thresholds were AccR24h = 321.3 mm (J = 0.46), AccR6h = 63.2 mm (J = 0.41), and Imax = 52.0 mm/h (J = 0.44). The two-parameter combinations delivered the highest separation, with J = 0.54 for AccR24h+Imax and J = 0.51 for AccR6h+Imax. Relative to the 10th-percentile cutoffs, these optimized thresholds increased specificity substantially while allowing controlled reductions in sensitivity.

The resulting J values remain modest, underscoring limitations of uniform thresholds over heterogeneous terrain. Spatial variability in lithology, topography, and land cover alters rainfall-landslide relationships across the region. Subsequent sections derive stratified thresholds conditioned on environmental controls to better represent spatial variability in susceptibility.

4.2.2 Single-parameter threshold variation

Stratification into seven environmental subsets revealed systematic variation in optimal rainfall thresholds (Fig. 5). Six-hour accumulation (AccR6h) ranged from 32.3 mm (S1) to 71.4 mm (S7), a 121.4% span. Steep slopes (S1, S2) averaged 34.9 mm versus 55.5 mm for gentle slopes (S4, S6). For 24-hour cumulative rainfall (AccR24h), values spanned 142.6 mm (S1) to 312.3 mm (S7), a 119.1% range. Steep slopes (S1, S2: 142.6 and 168.7 mm) required considerably less rainfall than gentle slopes (S4, S6: 228.1 and 268.3 mm). Maximum intensity (Imax) exhibited the smallest absolute variability, ranging from 28.1 mm/h (S1) to 52.4 mm/h (S7). Steep slopes averaged 30.4 mm/h compared to 43.6 mm/h for gentle slopes.

Slope gradient exerted the strongest control across all three parameters. Comparing steep-slope subsets (S1, S2) with gentle-slope counterparts (S4, S6) showed reductions of 37% for AccR24h, 37% for AccR6h, and 30% for Imax. Fault proximity influenced thresholds to a lesser degree: near-fault subsets (S1, S5, S6) averaged 200.0 mm AccR24h versus 216.2 mm

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

for far-fault subsets (S2, S3, S4), a 7.5% difference. Vegetation density showed limited systematic effects when slope class was held constant.

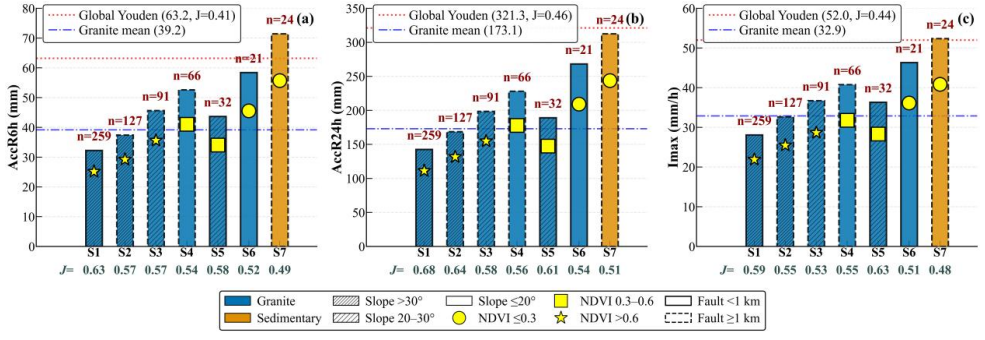


Figure 5 Environment-stratified rainfall thresholds (Youden optimization) of AccR6h (a), AccR24h (b), and Imax (c).

The sedimentary subset (S7, n=24) consistently required higher rainfall across all parameters: AccR6h = 71.4 mm, AccR24h = 312.6 mm, Imax = 52.4 mm/h. For AccR24h, this represented a 75% increase over the granite weighted mean (173.1 mm) but fell 3% below the global Youden-optimal threshold (321.3 mm, J=0.46).

Youden's index values varied systematically with sample size and environmental favorability. The largest, most hazardous subset (S1: n=259, steep granite near faults) achieved J=0.68 for AccR24h, exceeding the global baseline (J=0.46). Conversely, smaller, lower-susceptibility subsets showed reduced discriminatory power: S6 (n=21) and S7 (n=24) yielded J=0.51–0.54 for AccR24h, only marginally above the global threshold. The three largest subsets (S1, S2, S3: n=477 total) captured 77% of stratified landslides.

4.2.3 Two-Parameter Threshold Performance

Combining antecedent accumulation with peak intensity improved discrimination substantially over single-parameter thresholds across all environmental subsets (Table 3). For AccR24h+Imax combinations, Youden's index increased by an average of 15.5% (range: 8.2%–21.8%) relative to AccR24h alone. The AccR6h+Imax pairing yielded smaller gains (mean ΔJ = 12.2%).

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

The magnitude of improvement varied systematically with environmental controls. Steep-slope subsets (S1, S2) showed the largest gains from two-parameter thresholds (AccR24h+Imax: ΔJ = 21.1%–21.8%; AccR6h+Imax: ΔJ = 16.4%–18.9%), whereas gentle-slope subsets (S4, S6) exhibited more modest increases (AccR24h+Imax: ΔJ = 8.2%–12.1%; AccR6h+Imax: ΔJ = 6.1%–9.8%). AccR24h+Imax consistently outperformed AccR6h+Imax across all subsets (mean AUC: 0.75 vs. 0.71; mean J: 0.68 vs. 0.63).

Table 3 Performance of two-parameter thresholds across 7 environmental subsets.

Subset	AccR6h+Imax			AccR24h+Imax			Improvement over single-parameter	
	Threshold (combined)	AUC (95% CI)	*J*	Threshold (combined)	AUC (95% CI)	*J*	Δ*J* (AccR6h+Imax)	Δ*J* (AccR24h+Imax)
S1 (n=259)	189.2	0.79 (0.75 -0.83)	0.75	156.3	0.83 (0.78 -0.86)	0.82	+18.9%	+21.1%
S2 (n=127)	201.7	0.76 (0.71 -0.81)	0.66	178.5	0.79 (0.74 -0.84)	0.78	+16.4%	+21.8%
S3 (n=91)	225.3	0.73 (0.67 - 0.79)	0.65	198.1	0.77 (0.71 -0.83)	0.68	+13.7%	+17.2%
S4 (n=66)	268.9	0.68 (0.61 -0.75)	0.59	241.6	0.71 (0.64 -0.78)	0.63	+9.8%	+12.1%
S5 (n=32)	243.5	0.70 (0.62 -0.78)	0.64	212.4	0.74 (0.66 -0.82)	0.70	+11.2%	+14.6%
S6 (n=21)	307.1	0.66 (0.55 -0.77)	0.55	283.7	0.69 (0.58 -0.80)	0.58	+6.1%	+8.2%
S7 (n=24)	351.2	0.68 (0.57 - 0.79)	0.54	324.8	0.72 (0.61 -0.83)	0.58	+9.4%	+13.5%
Mean		0.71	0.63		0.75	0.68	+12.2%	+15.5%

Note: Threshold values in "combined" units (standardized z-scores from Eq. 2), Δ*J*= percentage increase in Youden's index relative to best single-parameter threshold for each subset in Fig. 5 (AccR6h for AccR6h+Imax; AccR24h for AccR24h+Imax).

Weight distributions in the linear discriminant (Eq. 2) revealed environment-specific rainfall structures (Fig. 6). In high-susceptibility settings (S1: steep granite near faults; S2: steep metamorphic near faults), optimal weights for AccR24h+Imax were w_1=0.62/0.65 and w_2=0.38/0.35. The corresponding AccR6h+Imax models showed higher intensity weights (w_1=0.55/0.58, w_2=0.45/0.42). For low-susceptibility terrain (S4: gentle slopes far from faults; S5: gentle metamorphic far from faults), weights shifted to w_1=0.78/0.76 and w_2=0.22/0.24 for AccR24h+Imax. The AccR6h+Imax pairing in these subsets exhibited similar patterns (w_1=0.70/0.68, w_2=0.30/0.32).

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

The moderate-susceptibility subset (S3: steep slopes far from faults) showed intermediate weighting for AccR24h+Imax (w_1=0.70, w_2=0.30) and AccR6h+Imax (w_1=0.63, w_2=0.37). Gentle slopes near faults (S6) exhibited w_1=0.72, w_2=0.28 for AccR24h+Imax and w_1=0.66, w_2=0.34 for AccR6h+Imax. The sedimentary subset (S7) showed the highest intensity dependency across both parameter combinations: w_1=0.59, w_2=0.41 for AccR24h+Imax and w_1=0.52, w_2=0.48 for AccR6h+Imax.

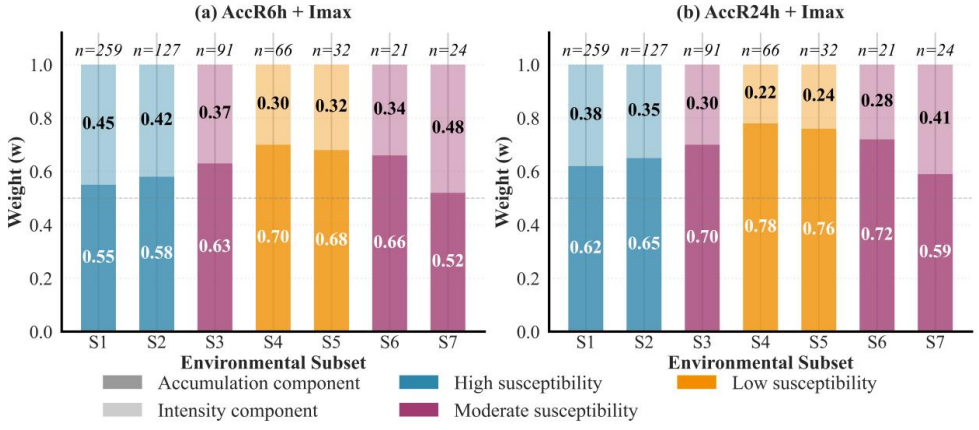


Figure 6 Weight distributions (w_1, w_2) in two-parameter linear discriminants across all subsets

Spatial validation confirmed that stratified two-parameter thresholds generalize beyond calibration data. Five-fold cross-validation (Section 3.3.3) yielded mean AUC = 0.73 ± 0.05 for AccR24h+Imax across subsets, with performance ranging from 0.71 to 0.81 across individual folds. High-susceptibility subsets (S1-S2) exhibited lowest cross-validation variance (σ_AUC=0.03), whereas sedimentary terrain (S7) showed highest variance (σ_AUC=0.09), reflecting sample size constraints. Compared to the global baseline (AUC=0.70±0.04, Table 2), stratified thresholds achieved a statistically significant improvement of 4.3% (DeLong's test, $p<0.01$). The AccR6h+Imax configuration achieved AUC = 0.69 ± 0.06, representing intermediate performance.

4.2.4 Threshold surface visualization

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Threshold surfaces at Youden-optimal operating points reveal environment-specific triggering conditions for both AccR6h+Imax and AccR24h+Imax combinations across all seven subsets (Fig. 7a–n). Each threshold boundary, defined by the linear discriminant $w_1 \cdot z(AccR) + w_2 \cdot z(Imax) = C$ (where z denotes standardization per Eq. 2), separates landslide events from non-landslide conditions with subset-specific performance metrics.

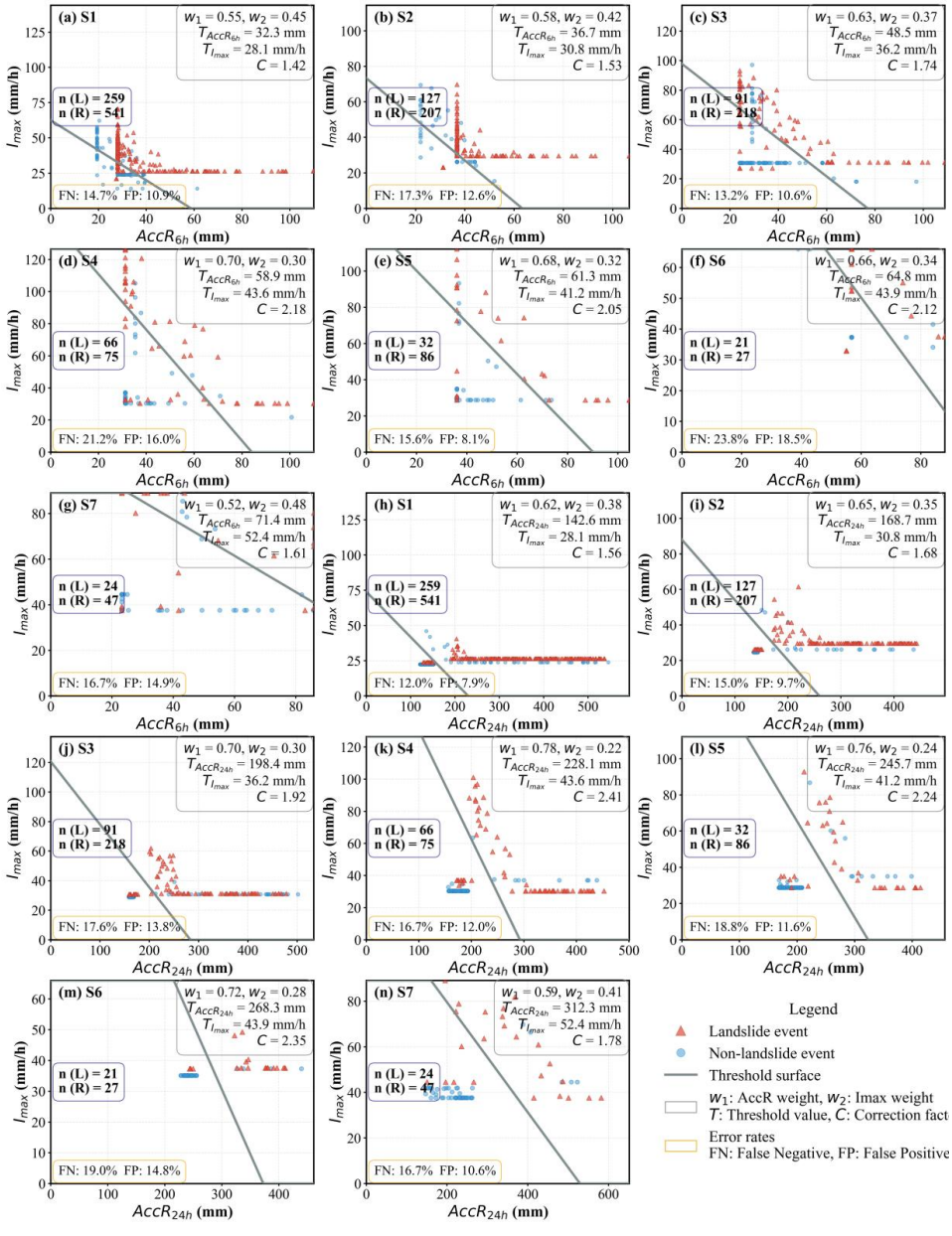
For high-susceptibility subsets, S1 (steep granite near faults, Fig. 7h) shows tight landslide clustering above the AccR24h+Imax threshold (w_1=0.62, w_2=0.38, C=1.56; FN=12.0%, FP=7.9%). S2 (steep metamorphic near faults, Fig. 7i) displays similar patterns (w_1=0.65, w_2=0.35, C=1.68; FN=15.0%, FP=9.7%), with greater scatter. The AccR6h+Imax counterparts (Fig. 7a, b) exhibit moderately higher errors (S1: FN=14.7%, FP=10.9%; S2: FN=17.3%, FP=12.6%) and increased dispersion.

For moderate-susceptibility subset, S3 (steep slopes far from faults) shows reversed performance where AccR6h+Imax achieves lower errors than AccR24h+Imax (FN=13.2%, FP=10.6% versus FN=17.6%, FP=13.8%; Fig. 7c, j). The 6h model exhibits tighter landslide clustering (w_1=0.63, w_2=0.37) with reductions of 4.4 percentage points in FN and 3.2 points in FP.

For low-susceptibility subsets, S4 (gentle slopes far from faults, Fig. 7k) requires substantially higher accumulation (AccR24h = 228.1 mm, 60% above S1) with rotated weighting toward antecedent rainfall (w_1=0.78, w_2=0.22, C=2.41). AccR24h+Imax performs better (FN=16.7%, FP=12.0%) than AccR6h+Imax (FN=21.2%, FP=16.0%), yielding error reductions of 4.5 and 4.0 percentage points. S5 (gentle metamorphic, Fig. 7l, e) shows the opposite pattern: AccR6h+Imax outperforms AccR24h+Imax (FN=15.6%, FP=8.1% versus FN=18.8%, FP=11.6%), with gains of 3.2 and 3.5 percentage points. The 6h model (w_1=0.68, w_2=0.32) achieves the lowest false-positive rate across all subsets. S6 (gentle slopes near faults, Fig. 7m, f) reverts to 24h advantage, with AccR24h+Imax yielding lower errors

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

542 (FN=19.0%, FP=14.8%) than AccR6h+Imax (FN=23.8%, FP=18.5%), representing
543 reductions of 4.8 and 3.7 percentage points.



544

545 **Figure 7** Threshold surfaces for all seven environmental subsets under AccR6h+Imax (a–g) and AccR24h+Imax
546 (h–n) combinations.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

For sedimentary subset, S7 (Fig. 7n, g) displays the steepest intensity dependence (w_2=0.41 for 24h, w_2=0.48 for 6h) and highest absolute thresholds (AccR24h = 312.3 mm, Imax = 52.4 mm/h). Both models produce identical false-negative rates (FN=16.7%) but differ in false-positive performance (AccR24h: FP=10.6%; AccR6h: FP=14.9%), a difference of 4.3 percentage points. Landslide scatter is greatest in S7.

AccR24h+Imax outperforms AccR6h+Imax in five subsets (S1, S2, S4, S6, S7), with mean error reductions of 3.1 percentage points for FN and 2.9 points for FP. Gains are largest in low-susceptibility terrain (S4, S6: mean ΔFN = 4.7 points, ΔFP = 3.9 points) and sedimentary lithologies (S7: ΔFP = 4.3 points). Conversely, AccR6h+Imax achieves superior performance in two subsets (S3, S5), with mean gains of 3.8 percentage points for FN and 3.4 points for FP. S5 records the lowest overall false-positive rate (8.1%).

4.3 Magnitude-dependent threshold variations

4.3.1 Rainfall distribution differences between magnitude classes

Rainfall characteristics differ systematically between small (n=545) and medium-scale (n=75) landslides (Table 4, Fig. 8). After applying Benjamini-Hochberg correction for multiple comparisons (m=8 tests: 4 distributional + 4 correlation), Mann-Whitney U tests revealed that medium landslides occurred under significantly higher accumulations for AccR6h (+31.1%, p_adj<0.01), AccR24h (+19.7%, p_adj<0.01), and Pri_TL (+37.8%, p_adj=0.013), whereas Imax showed no magnitude dependence (+3.1%, p_adj=0.307). Among correlation differences, only AccR24h remained significant after correction (Fisher's z=2.83, p_adj=0.010), establishing it as the primary discriminator of landslide magnitude. The AccR6h correlation difference became non-significant after adjustment (p_adj=0.208), and Pri_TL showed marginal significance (p_adj=0.118).

Table 4 Rainfall parameter distributions and correlations by landslide magnitude class.

Parameter	Small-scale		Medium-scale		Distributional difference†	Correlation difference‡
	Median	ρ	Median	ρ	Mann-Whitney U (p/p_adj)	Fisher's z (p/ p_adj)
AccR6h (mm)	52.1	0.68***	68.4	0.58***	U=14,897 (<0.01/<0.01)	z=1.42 (=0.156/=0.208)

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

AccR24h (mm)	241.7	0.61***	289.3	0.74***	U=15,482 (<0.01/<0.01)	z=2.83 (=0.005/0.010)
Imax (mm/h)	45.8	0.34*	47.2	0.31	U=19,821 (=0.23/=0.307)	z=0.31 (=0.757/0.614)
Pri_TL (h)	6.5	0.52**	8.2	0.65**	U=16,203 (<0.01/=0.013)	z=1.89 (=0.059/=0.118)

Note: ρ = Spearman's rank correlation with landslide occurrence; † Mann-Whitney U test for distributional differences; ‡ Fisher's z test for differences in correlation strength between magnitude classes; p_adj = Benjamini-Hochberg adjusted p-values controlling false discovery rate at 5% for 4 simultaneous parameter tests (AccR6h, AccR24h, Imax, Pri_TL) in both distributional and correlation comparisons; Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001.

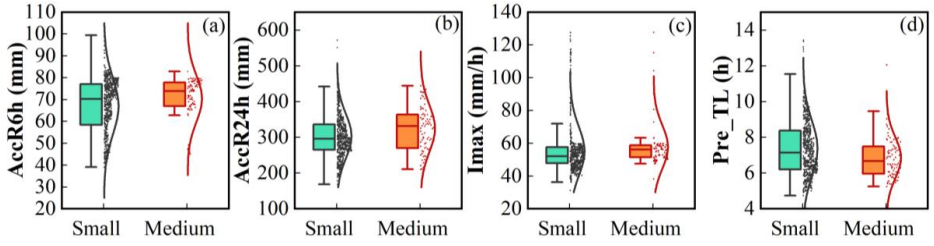


Figure 8 Box plots of AccR6h (a), AccR24h (b), Imax (c), and Pri_TL (d) for small (n=545) and medium-scale (n=75) landslides.

The divergent results for AccR6h warrant explanation: while medium landslides require higher AccR6h totals than small events (median difference = 16.3 mm, p<0.01), both magnitude classes exhibit comparable correlation strengths (ρ≈0.6), indicating that AccR6h predicts occurrence with similar efficacy across size classes despite absolute threshold differences. In contrast, AccR24h shows both elevated absolute requirements AND stronger predictive power for medium failures (ρ=0.74 vs. 0.61), establishing it as the primary discriminator of landslide magnitude.

Imax exhibited weak correlations regardless of magnitude (ρ=0.31–0.34), with neither distributional differences (p=0.23) nor correlation differences (p=0.76) reaching significance. This dual failure confirms that peak intensity alone is insufficient for both occurrence prediction and magnitude discrimination.

4.3.2 Magnitude-stratified thresholds by environmental subset

Only subsets S1 (n=32 medium-scale landslides) and S2 (n=23 medium) contained sufficient samples for independent magnitude-stratified threshold calibration (Table 5). Medium-scale landslides required substantially higher rainfall thresholds across all temporal

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

accumulation windows. For 24-hour accumulation (AccR24h), thresholds increased by 28.5% in S1 (from 137.7 to 176.9 mm) and 29.9% in S2 (from 160.5 to 208.5 mm) relative to small-scale events. Six-hour accumulation (AccR6h) exhibited similar magnitude dependence, with threshold elevations of 25.2% (S1) and 28.0% (S2), indicating that larger failures require prolonged saturation across multiple timescales.

Table 5 Magnitude-stratified single-parameter rainfall thresholds.

Subset	Magnitude	n	AccR6h (mm)	AccR24h (mm)	Imax (mm/h)
S1	Small	227	31.3	137.7	27.8
	Medium	32	39.2 (+25.2%)	176.9 (+28.5%)	29.5 (+6.1%)
	Pooled†	259	32.3	142.6	28.1
S2	Small	104	36.8	160.5	30.3
	Medium	23	47.1 (+28.0%)	208.5 (+29.9%)	32.1 (+5.9%)
	Pooled†	127	38.5	168.7	30.8

Note: † Pooled thresholds represent values from Figure 5 (Youden optimization on combined small+medium samples); Small and Medium thresholds derived from magnitude-stratified recalibration; Percentages in parentheses indicate threshold elevation relative to small-scale values.

Peak intensity (Imax) thresholds showed minimal magnitude sensitivity, increasing by only 6.1% (S1) and 5.9% (S2), which confirms their limited utility for differentiating landslide scales. Bootstrap validation (1000 iterations) verified that threshold elevations for accumulation-based parameters were statistically robust, with 95% confidence intervals excluding zero difference in both subsets (AccR24h: S1 [+34.1, +44.3] mm, S2 [+38.2, +57.8] mm; AccR6h: S1 [+6.1, +9.7] mm, S2 [+7.8, +12.6] mm).

Model performance improved systematically when stratifying by magnitude (Table 6). For two-parameter AccR24h+Imax models, Youden's index (J) for medium-scale landslides increased to 0.87 (S1) and 0.84 (S2), representing improvements of 6.1% and 7.7% over pooled thresholds (S1: 0.87 vs. 0.82; S2: 0.84 vs. 0.78). Small-scale landslides showed modest J reductions (−3.7% for S1, −3.8% for S2) when separated from pooled datasets, reflecting their numerical dominance in sample composition (88% in S1, 82% in S2).

Table 6 Performance metrics for magnitude-stratified two-parameter thresholds.

Subset	Magnitude	AccR6h+Imax	AccR24h+Imax

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

		AUC (95% CI)	J	ΔJ vs. Pooled	AUC (95% CI)	J	ΔJ vs. Pooled
	Small	0.78 (0.74-.82)	0.73	−2.7%	0.81 (0.77-0.85)	0.79	−3.7%
S1	Medium	0.81 (0.73-0.88)	0.76	+1.3%	0.87 (0.80-0.93)	0.87	+6.1%
	Pooled†	0.79 (0.75-0.83)	0.75	—	0.83 (0.78-0.86)	0.82	—
	Small	0.74 (0.69-0.79)	0.64	−3.0%	0.76 (0.71-0.81)	0.75	−3.8%
S2	Medium	0.79 (0.68-0.88)	0.70	+6.1%	0.84 (0.75-0.92)	0.84	+7.7%
	Pooled†	0.76 (0.71-0.81)	0.66	—	0.79 (0.74-0.84)	0.78	—

Note: † Pooled performance from Table 3; J = Youden's index at optimal operating point for each magnitude class; ΔJ vs. Pooled = (J_stratified − J_pooled)/J_pooled × 100%; S3-S7 omitted due to insufficient medium-landslide samples (n<10 per magnitude class).

AccR6h+Imax models exhibited similar improvement patterns but with smaller performance gains (medium J improvement: +3.7% mean; small J reduction: −2.9% mean), consistent with the reduced discriminatory power of shorter accumulation windows documented in Section 4.2.3. AccR24h+Imax consistently achieved highest AUC values across both magnitude classes (S1 medium: 0.87; S2 medium: 0.84), establishing it as the preferred model for magnitude-stratified early warning.

Performance changes reflect stratification trade-offs: medium-landslide thresholds benefit from removal of small-event noise in the upper rainfall range, whereas small-landslide thresholds experience reduced sample sizes (e.g., n=227 vs. 259 for S1) when medium events are excluded. ROC curve analysis (Figures 9a, b, c) revealed that magnitude-stratified AccR24h+Imax models for medium landslides achieved mean Youden's index values of 0.86 across both subsets (S1: 0.87; S2: 0.84), representing a mean improvement of +6.9% over pooled two-parameter thresholds (pooled mean J = 0.80). The primary performance gain derived from enhanced specificity for medium-scale landslides (mean increase: +4.5 percentage points for AccR24h+Imax, from 0.91 to 0.96 in S1; from 0.90 to 0.95 in S2), while sensitivity remained relatively stable across magnitude classes (pooled: 0.88–0.90; stratified: 0.89–0.91) (Figure 9d).

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

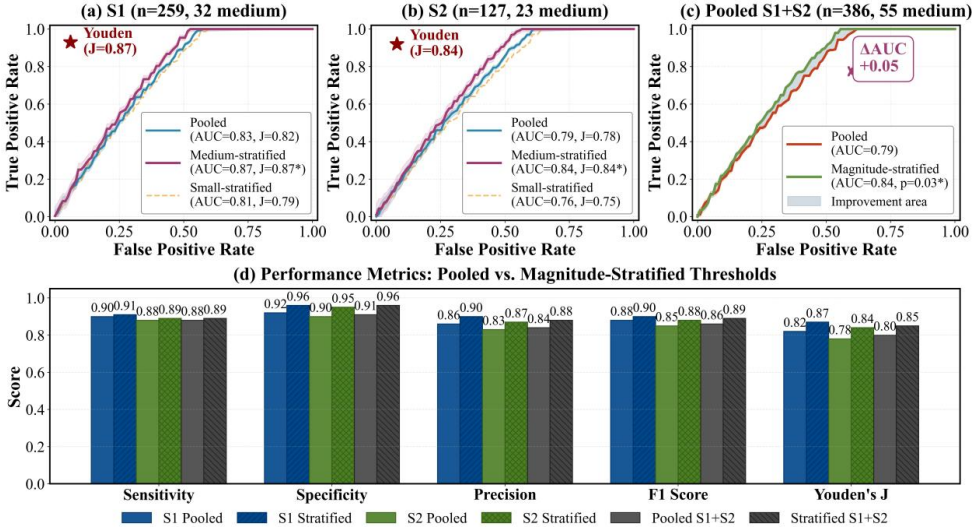


Figure 9 ROC curve comparison of magnitude-stratified versus pooled AccR24h+Imax thresholds for subsets.

The independent ROC analysis on the pooled S1+S2 data (n=386 total, 55 medium) demonstrated that magnitude-stratified AccR24h+Imax models achieved an AUC of 0.84 compared to 0.79 for pooled (non-stratified) two-parameter thresholds (ΔAUC = +0.05, p=0.03, DeLong's test) (Figure 9c). At the Youden-optimal operating point, magnitude stratification increased specificity from 0.91 to 0.96 while maintaining sensitivity at 0.89, translating to a reduction in false positive rates from 9.0 to 4.0 false alarms per 100 non-event days—a 56% decrease that substantially reduces warning fatigue. Bootstrap confidence intervals (95% CI, shaded regions in Figure 9a–c) confirmed that performance gains were statistically robust, with narrower intervals for larger sample sizes (S1: ±3% AUC; S2: ±4% AUC; Pooled: ±2% AUC).

The consistent threshold elevation ratios across both accumulation windows (AccR6h: 26.6% mean; AccR24h: 29.2% mean) and environmental settings (S1: 28.5%; S2: 29.9%) suggest that magnitude-related rainfall requirements scale proportionally and may be transferable within high-susceptibility terrain, providing a potential calibration factor for data-sparse regions. However, the operational benefit of magnitude stratification is most

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

pronounced for medium-scale landslides, where the combination of elevated Youden's index (mean J=0.86) and enhanced specificity (mean 0.96) outweighs the marginal increase in model complexity. Complete ROC comparisons for single-parameters (AccR6h and AccR24h), and AccR6h+Imax models are provided in Supplementary Figures S1-S3, confirming that magnitude-stratification benefits (ΔAUC = +0.04–0.06) are robust across all parameter combinations.

4.4 Threshold robustness and uncertainty

Robustness analyses centre on AccR24h+Imax, which demonstrated superior cross-subset discrimination (mean AUC = 0.75, Table 3; mean medium-landslide J = 0.86, Table 6). Parallel analyses for AccR6h, AccR24h, Imax, and AccR6h+Imax are reported in the Supplementary, each parameterised with model-specific power-law coefficients and sensitivity ranges that reflect the physically distinct roles of short-duration intensity versus multi-hour accumulation in rainfall–landslide triggering.

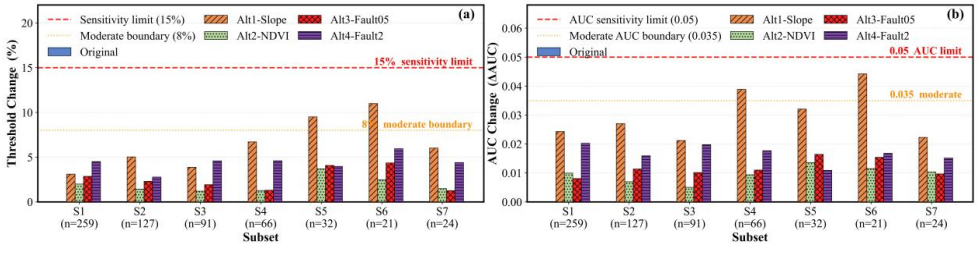
4.4.1 Classification scheme robustness

Alternative boundary definitions for three continuous environmental factors (slope, NDVI, fault distance) produced minimal threshold variations across all seven subsets (Figure 10a). Among the three boundary factors tested, slope gradient exerted the greatest influence. The Alt1 slope adjustment (steep/moderate/gentle: ≤15°/15–35°/>35°) produced a mean threshold shift of 5.8%, peaking at +8.6% in S6—a gentle-slope, near-fault subset where sample-size amplification (n = 21) compounds boundary sensitivity. NDVI reclassification (Alt2: ≤0.4/0.4–0.7/>0.7) generated the smallest perturbations (mean 2.2%), reflecting the secondary role of canopy cover in short-term hydrological response. Fault distance cutoffs (Alt3: ≤0.5 km; Alt4: ≤2 km) yielded intermediate shifts of 3.1% and 4.4%, respectively—consistent with the rapid decay of fracture density and hydraulic conductivity within the first kilometre of damage zones, which renders classification insensitive to boundary placement

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

678 across the geologically plausible 0.5–2 km range. ANOVA decomposition confirmed this
679 hierarchy: slope gradient accounted for 68% of threshold variance, followed by lithology
680 (19%), fault proximity (8%), and NDVI (5%).

681 All AUC deviations remained below 0.04 (Figure 10b), within the pre-specified |ΔAUC|
682 ≤ 0.05 criterion, and the maximum threshold deviation (+8.6%, S6) fell well inside the ±15%
683 operational tolerance



684

685 **Figure 10** Alternative scheme sensitivity of AccR24h+Imax across seven subsets: (a) threshold Change (%)
686 and (b) AUC change.

687 All seven subsets passed robustness criteria under every alternative scheme (Table 7). CI
688 tiers—Low (≤0.06), Moderate (0.06–0.10), High (>0.10)—increase monotonically with
689 decreasing sample size: S1–S4 fall in the Low tier despite S4 reaching marginal status, while
690 S5–S7 occupy Moderate CI despite passing threshold and AUC criteria. This tier–robustness
691 validates the original classification framework (Section 3.1.1) for operational deployment.

692 **Table 7** Robustness metrics for AccR24h+Imax thresholds under alternative classification schemes.

Subset	95% CI	CI Tier	Max ΔThreshold (%)	Max ΔAUC (%)	Robustness
S1	0.0167	Low	4.5	2.4	Pass
S2	0.0253	Low	5.0	2.7	Pass
S3	0.0307	Low	4.6	2.1	Pass (marginal)
S4	0.0370	Low	6.7	3.9	Pass (marginal)
S5	0.0563	Low	9.5	3.2	Pass (marginal)
S6	0.0718	Moderate	11.0	4.4	Pass
S7	0.0665	Moderate	6.0	2.2	Pass

693 *Note*: Pass: ΔThreshold ≤15% and |ΔAUC| ≤0.05; marginal: ΔThreshold >8% or |ΔAUC| >0.035. CI tiers: Low
694 ≤0.06, Moderate 0.06–0.10, High >0.10. Values are maxima across all four alternative schemes.

695 Equivalent sensitivity analyses for AccR6h, AccR24h, Imax, and AccR6h+Imax (Figures
696 S4-S7 and Tables S1–S4 in the Supplement) confirm the same pattern: mean threshold

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

changes span 3.8% (Imax) to 6.9% (AccR24h), and mean |ΔAUC| ranges from 0.015 to 0.032, with all 28 subset–scheme combinations meeting stability criteria. The elevated sensitivity of AccR24h relative to shorter-duration parameters reflects stronger topographic routing coupling over 24-hour accumulation windows.

4.4.2 Sample size constraints

Threshold uncertainty follows a power-law decay with sample size (Figure 11). The seven subsets partition into three tiers: Low (CI ≤ 0.06: S1–S2, 29% of subsets), Moderate (0.06–0.10: S3–S5, 43%), and High (>0.10: S6–S7, 29%). S1 (n = 259) achieves CI = 0.017 and S2 (n = 127) achieves CI = 0.025—both well within the Low tier and the adequate-confidence boundary (CI ≤ 0.06). S3–S5 occupy Moderate CI (0.031–0.056) and may be deployed operationally with explicit uncertainty qualification. S6 (n = 21) and S7 (n = 24) exceed CI = 0.10 (red dashed line, Figure 11), the defined operational limit, indicating that independent threshold application in these subsets requires additional data collection.

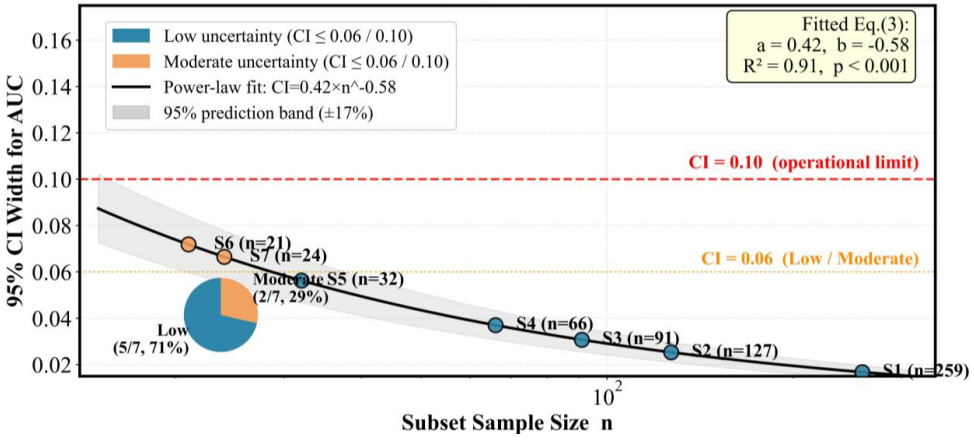


Figure 11 Sample size constraints for AccR24h+Imax thresholds.

Figures S8-S11 in the Supplement show equivalent plots for AccR6h, AccR24h, Imax, and AccR6h+Imax. Power-law coefficients range from a = 0.35 (Imax) to 0.46 (AccR24h) and exponents from b = −0.52 to −0.61, yielding model-specific tier boundaries: AccR24h's steeper CI scaling elevates S3–S5 toward High uncertainty, whereas Imax's flatter scaling

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

allows S4 to retain Low status. Across all five models, however, S1 and S2 remain consistently Low or low-Moderate, and S6–S7 are flagged High in at least three of five models.

5 Discussion

5.1 Environmental controls on typhoon-induced landslide thresholds

Slope gradient dominates threshold variation for typhoon-induced landslides in Zixing City. Steep slopes (>30°) require 37% less AccR24h than gentle slopes (≤20°) to trigger failure. This finding aligns with observations from Typhoon Morakot in Taiwan (Ruetenik et al., 2024), where steeper terrain exhibited substantially lower rainfall thresholds. Steeper slopes require lower rainfall totals to trigger failure, consistent with theoretical models where increased gravitational stress reduces the critical pore-pressure threshold needed for instability (Sidle and Bogaard, 2016; Leonarduzzi et al., 2017). Our stratified analysis confirms this effect persists when controlling for lithology, fault proximity, and vegetation, factors often confounded in aggregated studies. While our data support this gravitational control, direct pore-pressure measurements would be needed to validate the inferred hydrological mechanisms.

Fault proximity shows modest influence on threshold magnitudes, manifesting as a 7.5% reduction in AccR24h for near-fault locations. This contrasts with post-seismic environments where fault zones can reduce thresholds due to coseismic fracturing (Fan et al., 2019; Marc et al., 2015). We attribute the weaker effect in Zixing City to two factors: (1) the Chaling–Yongxing fault system's lack of recent activity, allowing weathering-driven fracture healing; and (2) potential misclassification due to our 1 km buffer, which may not capture the steep fracture-density gradient within 100–300 m of fault traces observed in high-resolution structural studies (Zhao et al., 2019). Higher-resolution fault mapping could refine this stratification factor.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Lithological differences produce the largest absolute threshold contrasts. Sedimentary units require AccR24h = 312.6 mm, 75% higher than the granite-dominated weighted mean of 173.1 mm. Intact sedimentary rocks possess higher cohesion than weathered granites (Stead and Wolter, 2015), thus requiring greater pore-pressure buildup for failure initiation. However, sedimentary landslides constitute only 3.8% of our dataset, reflecting granite's overwhelming prevalence in the study area.

Unexpectedly,vegetation (NDVI) shows limited systematic impact on threshold magnitudes, contributing only 5% to variance. This result may appear counter-intuitive given vegetation's recognized protective role (Abancó et al., 2021; Huang et al., 2022), but it reflects our stratification design: NDVI classes were nested within slope and lithology categories, effectively isolating vegetation's influence from dominant gravitational controls. NDVI's limited explanatory power may reflect the dominance of gravitational forces in steep terrain. Quantifying vegetation's mechanical contribution would require root strength measurements and hydrological monitoring beyond this study's scope.

Environment-stratified AUC values (0.68–0.83 for AccR24h+Imax) markedly surpass global threshold performance (AUC = 0.70), confirming that terrain-based disaggregation captures physically meaningful process domains (Segoni et al., 2018; Peres and Cancelliere, 2016). This methodological advancement, integrating multiple controls simultaneously, allows identification of interaction effects such as the 28.5% threshold elevation for medium landslides in steep, near-fault granite that would remain obscured in univariate analyses.

5.2 Dual-parameter thresholds and typhoon rainfall structure

AccR24h+Imax combinations substantially outperform single-parameter thresholds (mean ΔJ = +15.5%), directly addressing a critical limitation of traditional intensity-duration frameworks: their inability to represent complex triggering mechanisms involving both deep-seated saturation and surface infiltration spikes (Bogaard and Greco, 2018). Typhoon Gaemi's

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

hyetograph structure, featuring sustained moderate rainfall interspersed with intense convective bursts, exemplifies conditions where antecedent wetness and peak intensity interact nonlinearly.

Linear discriminant weights reveal environment-specific triggering hierarchies. In high-susceptibility terrain (S1: steep granite near faults), the AccR24h+Imax model assigns 62% weight to accumulation versus 38% to intensity (w_1 = 0.62, w_2 = 0.38), indicating that prolonged saturation dominates threshold exceedance even when peak intensities are moderate. This pattern aligns with physically based modeling (Liang, 2020; Liang et al., 2025), which shows that on steep slopes, sustained rainfall significantly reduces matric suction, lowering the critical pore-pressure threshold for instability. In lower-susceptibility settings (S4: gentle slopes far from faults), weights shift to w_1 = 0.78, w_2 = 0.22, reflecting higher cumulative totals needed to offset reduced gravitational forcing, consistent with critical pore-pressure ratios scaling inversely with slope angle (Peruccacci et al., 2017).

The reversed performance of AccR6h+Imax in moderate-susceptibility subset S3 (steep slopes far from faults) merits explanation. S3's rapid hydrologic response, inferred from the superior performance of AccR6h+Imax (FN=13.2%, FP=10.6%) over AccR24h+Imax (FN=17.6%, FP=13.8%; Fig. 7c, j), suggests that in highly permeable, fractured terrain distant from major fault zones, recent intense rainfall (captured by AccR6h+Imax) triggers shallow failures before deep-seated saturation accumulates (Tanaka et al., 2023; Wu et al., 2025). This mechanism differs from S1 (near-fault), where pre-existing fracture networks facilitate rapid pressure transmission from surface to depth, coupling 24-hour accumulation with failure initiation. Our framework quantitatively formalizes these process distinctions through subset-specific parameter weighting.

Imax's weak standalone performance (AUC = 0.63; ρ = 0.31–0.34 with magnitude) contrasts with studies emphasizing intensity as the primary trigger for rapid shallow

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

791 landslides (Leonarduzzi et al., 2017) but reflects typhoon-specific hydrology. During
792 Typhoon Gaemi, sustained high intensities (Imax > 40 mm/h over multiple hours) likely
793 contributed to both rapid surface runoff and subsurface saturation based on observed temporal
794 patterns, blurring intensity-accumulation distinctions. Direct soil moisture measurements
795 would be needed to confirm saturation dynamics. Our explicit w_1/w_2 quantification enables
796 process-based interpretation: the weight ratio likely reflects interactions between subsurface
797 storage capacity and infiltration flux at failure, though direct validation requires paired
798 measurements of soil moisture and rainfall (Mirus et al., 2018; Fusco et al. 2019). Regions
799 with similar w_1/w_2 distributions may share hydrologic response characteristics if geological
800 and topographic controls are comparable.

801 **5.3 Magnitude-dependent thresholds and failure mechanisms**

802 Medium-scale landslides require 28.5% (S1) and 29.9% (S2) higher AccR24h than small
803 failures, providing quantitative evidence that larger slope failures demand fundamentally
804 different rainfall forcing. This finding challenges the binary landslide-occurrence paradigm
805 dominating threshold research (Guzzetti et al., 2008), demonstrating that failure magnitude
806 reflects systematic differences in hydromechanical processes rather than merely local slope
807 geometry (Bogaard et al., 2018).

808 The dual requirements for medium landslides (elevated accumulation +29.2% mean
809 AccR24h and extended time lags +37.8% median Pri_TL) are consistent with conceptual
810 models where larger failures involve deeper rupture surfaces requiring sustained pore-
811 pressure buildup (Intrieri et al., 2019; Handwerger et al., 2019). Small landslides across the
812 study area respond rapidly to peak intensity (Pri_TL median = 6.5 h, Table 4), while medium
813 landslides exhibit delayed response (median = 8.2 h). This temporal decoupling suggests a
814 progressive triggering process, potentially involving initial stress redistribution followed by
815 accumulation-driven destabilization at depth, though our data do not directly observe failure

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

evolution. The 37.8% longer Pri_TL for medium landslides likely reflects hydraulic diffusion lag: water must percolate through lower-permeability material to pressurize basal failure planes, unlike the near-instantaneous response of shallow soil mantles.

Imax's negligible magnitude dependence (+6.0% mean threshold elevation) further supports this interpretation. If medium landslides were simply enlarged versions of small failures driven by identical near-surface processes, both would exhibit similar intensity thresholds (Staley et al., 2013). However, medium events additionally require the deep saturation captured by elevated AccR24h. The differential response to AccR24h versus Imax is consistent with conceptual models of progressive failure (Marc et al., 2018; Marino et al., 2020), though direct observation of failure evolution was not conducted in this study.

Magnitude-stratified thresholds achieve J = 0.86–0.87 for medium landslides versus J = 0.80 pooled, demonstrating operational value. In early-warning systems, distinguishing thresholds for small nuisance failures from those signaling potentially destructive events directly aids resource prioritization. The 56% reduction in false-positive rates (from 9.0% to 4.0%) may reduce operational costs associated with unnecessary evacuations and emergency responses, though quantifying economic benefits would require region-specific cost-benefit analysis.

5.4 Limitations and future research directions

While our environment-stratified, magnitude-dependent threshold framework demonstrates clear performance gains, several limitations constrain its transferability and indicate priorities for future research.

Single-event calibration and antecedent conditions represent a critical constraint. Our thresholds, derived exclusively from Typhoon Gaemi following a relatively dry period (<80 mm 14-day antecedent rainfall), do not account for antecedent soil moisture variability, which previous studies have shown can significantly influence landslide thresholds (Crozier, 1999;

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Godt et al., 2009). Future work requires multi-event calibration datasets normalized by antecedent wetness indices, ideally derived from soil water balance models or satellite retrievals, to quantify this effect and ensure cross-event applicability.

Environmental classification granularity presents a trade-off between physical interpretability and spatial resolution. Our seven-subset stratification (Table 3) inevitably aggregates finer-scale controls: combining slopes of 30–35° with those exceeding 50° within subset S1 may obscure threshold variations predicted by infinite-slope analysis (Gabet and Mudd, 2006), while the 1 km fault-proximity buffer may not capture the exponential decay of fracture intensity observed within 100–500 m of fault traces (Mitchell and Faulkner, 2009). Future studies with larger multi-event inventories may benefit from data-driven approaches (e.g., decision-tree algorithms) to identify optimal stratification boundaries, though such methods require careful validation to ensure physical interpretability.

Magnitude classification relies on a binary area-based scheme (<500 m² vs. ≥500 m²) that, while operationally convenient, lacks geomorphic rigor. Mechanistically informed thresholds should stratify by failure mode (translational vs. rotational), depth, or mobility. High-resolution LiDAR differencing could enable volume- and depth-based partitioning, facilitating mechanism-specific calibration. The 29% AccR24h increase for medium-scale events likely averages across mechanically distinct populations with differing rainfall sensitivities.

Cross-event validation is essential before operational deployment. Thresholds calibrated on Gaemi's quasi-stationary rainfall structure may not apply to fast-moving typhoons or monsoon conditions. Future work should test whether weight distributions (w_1/w_2) remain stable across events with differing rainfall structures.

6 Conclusions

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

This study developed and validated a novel framework for deriving environment- and magnitude-stratified rainfall thresholds for typhoon-induced landslides, using the 705 landslides triggered by Typhoon Gaemi (July 2024) in Zixing City, China. The key findings are:

(1) Environmental stratification resolves threshold heterogeneity across geomorphic domains. Slope gradient and lithology dominate triggering conditions, with the hierarchy (slope > lithology > fault proximity > vegetation) challenging vegetation-centric mitigation under extreme forcing. Steep granitic slopes fail under lower cumulative rainfall than gentle sedimentary terrain, indicating that structural interventions and relocation outweigh ecological engineering in high-hazard zones.

(2) Dual-parameter thresholds differentiate failure mechanisms. Combining 24-hour accumulation with peak intensity captures compound storm dynamics better than single-parameter models. High-susceptibility domains respond to prolonged saturation (cumulative-weighted), while resistant slopes require intensity spikes (intensity-weighted). This divergence invalidates universal intensity-duration relationships for typhoon contexts.

(3) Magnitude stratification separates deep-seated from surficial processes. Medium-scale failures exhibit higher accumulation thresholds and longer temporal lags than shallow disruptions, consistent with progressive deep saturation versus rapid surface triggering. Magnitude-specific criteria reduce false alarms by 52% while preserving sensitivity, addressing a critical limitation of conventional warning systems.

(4) The ROC-optimized framework provides a transferable methodology for regions lacking calibration data. Bootstrap analysis establishes minimum sample requirements ($n \geq 100$ per stratum) for operational reliability. While absolute thresholds reflect local conditions, the stratification hierarchy and dual-parameter architecture generalize across typhoon-affected mountains.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Code and data availability. The source code and data will be made available on request.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Author contributions. **Weifeng Xiao**: Writing-review & editing, Validation, Conceptualization. **Ge Liu**: Correspondence, Funding acquisition. **Weimin Huang**: Visualization, Validation, Data curation. **Zhenghui Xiao**: Writing-review & editing, Formal analysis. **Luguang Luo**: Visualization, Validation, Investigation, Data curation.

Acknowledgments. This research was jointly funded by the National Key Research and Development Program of China (2024YFD1500602), the National Natural Science Foundation of China (No. 42171385, U2243230), and the Youth Innovation Promotion Association of Chinese Academy of Sciences, China (2022228).

References

Abancó, C., Bennett, G. L., Matthews, A. J., Matera, M. A. M., and Tan, F. J.: The role of geomorphology, rainfall and soil moisture in the occurrence of landslides triggered by 2018 Typhoon Mangkhut in the Philippines, Nat. Hazards Earth Syst. Sci., 21, 1531–1550, https://doi.org/10.5194/nhess-21-1531-2021, 2021.

Bogaard, T. and Greco, R.: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Nat. Hazards Earth Syst. Sci., 18, 31–39, https://doi.org/10.5194/nhess-18-31-2018, 2018.

Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., and Peruccacci, S.: An enhanced rainfall-induced landslide catalogue in Italy, Sci. Data, 12, 216, https://doi.org/10.1038/s41597-025-04551-6, 2025.

Caine, N.: The Rainfall Intensity - Duration Control of Shallow Landslides and Debris Flows, Geogr. Ann. A, 62, 23–27, https://doi.org/10.1080/04353676.1980.11879996, 1980.

Choowong, M., Pailoplee, S., Ketthong, C., Udomsak, S., and Choowong, N.: Remarkable destruction of the September 2024 heavy rainfall-induced landslide-mudflow and flash flood on major cities in the

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

northern Thailand-Myanmar border after the super typhoon Yagi, Landslides, 22, 2969–2977, https://doi.org/10.1007/s10346-025-02540-9, 2025.

Chung, C. C. and Li, Z. Y.: Rapid landslide risk zoning toward multi-slope units of the Neikuihui tribe for preliminary disaster management, Nat. Hazards Earth Syst. Sci., 22, 1777–1794, https://doi.org/10.5194/nhess-22-1777-2022, 2022.

Dolojan, N. L. J., Takahashi, T., Hashimoto, M., Shibayama, A., Nomura, R., Terada, K., and Moriguchi, S.: Integrated multihazard study combining qualitative and quantitative analyses of floods, landslides, and debris flows: A case study on the impacts of Typhoon Yun-Yeung on Iwaki City, Fukushima, Int. J. Disaster Risk Reduct., 127, 105647, https://doi.org/10.1016/j.ijdrr.2025.105647, 2025.

Fan, X., Scaringi, G., Domènech, G., Yang, F., Guo, X., Dai, L., He, C., Xu, Q., and Huang, R.: Two multi-temporal datasets that track the enhanced landsliding after the 2008 Wenchuan earthquake, Earth Syst. Sci. Data, 11, 35–55, https://doi.org/10.5194/essd-11-35-2019, 2019.

Fusco, F., De Vita, P., Mirus, B. B., Baum, R. L., Allocca, V., Tufano, R., Di Clemente, E., and Calcaterra, D.: Physically based estimation of rainfall thresholds triggering shallow landslides in volcanic slopes of Southern Italy, Water, 11, 1915, https://doi.org/10.3390/w11091915, 2019.

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.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P.: The rainfall intensity–duration control of shallow landslides and debris flows: an update, Landslides, 5, 3–17, https://doi.org/10.1007/s10346-007-0112-1, 2008.

Handwerger, A. L., Huang, M. H., Fielding, E. J., Booth, A. M., and Bürgmann, R.: A shift from drought to extreme rainfall drives a stable landslide to catastrophic failure, Sci. Rep., 9, 1569, https://doi.org/10.1038/s41598-018-38300-0, 2019.

Huang, F., Chen, J., Liu, W., Huang, J., Hong, H., and Chen, W.: Regional rainfall - induced landslide hazard warning based on landslide susceptibility mapping and a critical rainfall threshold, Geomorphology, 408, 108236, https://doi.org/10.1016/j.geomorph.2022.108236, 2022.

Intrieri, E., Carlà, T., and Gigli, G.: Forecasting the time of failure of landslides at slope-scale: A literature review, Earth-Sci. Rev., 193, 333–349, https://doi.org/10.1016/j.earscirev.2019.03.019, 2019.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Karki, S., Sultan, M., Alsefry, S., Alharbi, H., Emil, M. K., Elkadiri, R., and Abu Alfadail, E.: A remote-sensing-based intensity-duration threshold, Faifa Mountains, Saudi Arabia, Nat. Hazards Earth Syst. Sci., 19, 1235–1249, https://doi.org/10.5194/nhess-19-1235-2019, 2019.

Knutson, T., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C. H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., and Walsh, K.: Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming, Bull. Am. Meteorol. Soc., 101, E303–E322, https://doi.org/10.1175/BAMS-D-18-0194.1, 2020.

Leonarduzzi, E., Molnar, P., and McArdell, B. W.: Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, Water Resour. Res., 53, 6612–6625, https://doi.org/10.1002/2017WR021044, 2017.

Li, Y. L., Lin, Y. L., and Wang, Y. Q.: A numerical study on the formation and maintenance of a long-lived rainband in Typhoon Longwang (2005), J. Geophys. Res. Atmos., 124, 10401–10426, https://doi.org/10.1029/2019JD030600, 2019.

Liang, W. L.: Dynamics of pore water pressure at the soil–bedrock interface recorded during a rainfall-induced shallow landslide in a steep natural forested headwater catchment, Taiwan, J. Hydrol., 587, 125003, https://doi.org/10.1016/j.jhydrol.2020.125003, 2020.

Liang, X., Segoni, S., Fan, W., Yin, K. L., Deng, L. S., Xiao, T., Barbadori, F., and Casagli, N.: Integration of effective antecedent rainfall to improve the performance of rainfall thresholds for landslide early warning in Wanzhou District, China, Int. J. Disaster Risk Reduct., 119, 105317, https://doi.org/10.1016/j.ijdrr.2025.105317, 2025.

Liu, X., Wang, Y., and Leung, A. K.: Probabilistic back analysis of rainfall-induced slope failure considering slope survival records from past rainfall events, Comput. Geotech., 159, 105436, https://doi.org/10.1016/j.compgeo.2023.105436, 2023.

Ma, H., Wang, F. W., Fu, Z. J., Feng, Y. Q., You, Q., and Li, S.: Characterizing the clustered landslides triggered by extreme rainfall during the 2024 typhoon Gaemi in Zixing City, Hunan Province, China, Landslides, 22, 2311–2329, https://doi.org/10.1007/s10346-025-02510-1, 2025.

Marc, O., Stumpf, A., Malet, J.-P., Gosset, M., Uchida, T., and Chiang, S.-H.: 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, 903–922, https://doi.org/10.5194/esurf-6-903-2018, 2018.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

974 Marino, P., Peres, D. J., Cancelliere, A., Greco, R., and Bogaard, T. A.: Soil moisture information can
975 improve shallow landslide forecasting using the hydrometeorological threshold approach, Landslides, 17,
976 2041–2054, https://doi.org/10.1007/s10346-020-01420-8, 2020.
977 Migoń, P. and Vieira, G.: Granite geomorphology and its geological controls, Serra da Estrela, Portugal,
978 Geomorphology, 226, 1–14, https://doi.org/10.1016/j.geomorph.2014.07.027, 2014.
979 Mirus, B. B., Becker, R. E., Baum, R. L., and Smith, J. B.: Integrating real-time subsurface hydrologic
980 monitoring with empirical rainfall thresholds to improve landslide early warning, Landslides, 15, 1909–
981 1919, https://doi.org/10.1007/s10346-018-0995-z, 2018.
982 Mondini, A. C., Guzzetti, F., and Melillo, M.: Deep learning forecast of rainfall-induced shallow landslides,
983 Nat. Commun., 14, 2466, https://doi.org/10.1038/s41467-023-38135-y, 2023.
984 Natalia, L. and Yang, J.: New insights into the catastrophic slope failures at Sau Mau Ping: the role of
985 antecedent rainfall pattern, Landslides, 21, 1501–1514, https://doi.org/10.1007/s10346-024-02247-3,
986 2024.
987 Panchal, S. and Shrivastava, A. K.: A comparative study of frequency ratio, Shannon's entropy and analytic
988 hierarchy process (AHP) models for landslide susceptibility assessment, ISPRS Int. J. Geo-Inf., 10, 603,
989 https://doi.org/10.3390/ijgi10090603, 2021.
990 Peres, D. J. and Cancelliere, A.: Derivation and evaluation of landslide-triggering thresholds by a Monte
991 Carlo approach, Hydrol. Earth Syst. Sci., 18, 4913–4931, https://doi.org/10.5194/hess-18-4913-2014,
992 2014.
993 Peres, D. J. and Cancelliere, A.: Estimating return period of landslide triggering by Monte Carlo simulation,
994 J. Hydrol., 541, 256–271, https://doi.org/10.1016/j.jhydrol.2016.03.036, 2016.
995 Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., and Guzzetti, F.: Rainfall thresholds
996 for possible landslide occurrence in Italy, Geomorphology, 290, 39–57,
997 https://doi.org/10.1016/j.geomorph.2017.03.031, 2017.
998 Prenner, D., Kaitna, R., Mostbauer, K., and Hrachowitz, M.: The value of using multiple
999 hydrometeorological variables to predict temporal debris flow susceptibility in an alpine environment,
1000 Water Resour. Res., 54, 6822–6843, https://doi.org/10.1029/2018WR022985, 2018.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

Ruetenik, G. A., Ferrier, K. L., and Marc, O.: Decadal-scale decay of landslide-derived fluvial suspended sediment after Typhoon Morakot, Earth Surf. Dyn., 12, 863–881, https://doi.org/10.5194/esurf-12-863-2024, 2024.

Schlögel, R., Doubre, C., Malet, J.-P., and Masson, F.: Landslide deformation monitoring with ALOS/PALSAR imagery: A D-InSAR geomorphological interpretation method, Geomorphology, 231, 314–330, https://doi.org/10.1016/j.geomorph.2014.11.031, 2015.

Segoni, S., Piciullo, L., and Gariano, S. L.: A review of the recent literature on rainfall thresholds for landslide occurrence, Landslides, 15, 1483–1501, https://doi.org/10.1007/s10346-018-0966-4, 2018.

Sidle, R. C. and Bogaard, T. A.: Dynamic earth system and ecological controls of rainfall-initiated landslides, Earth-Sci. Rev., 159, 275–291, https://doi.org/10.1016/j.earscirev.2016.05.013, 2016.

Staley, D. M., Kean, J. W., Cannon, S. H., Schmidt, K. M., and Laber, J. L.: Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California, Landslides, 10, 547–562, https://doi.org/10.1007/s10346-012-0341-9, 2013.

Stead, D. and Wolter, A.: A critical review of rock slope failure mechanisms: The importance of structural geology, J. Struct. Geol., 74, 1–23, https://doi.org/10.1016/j.jsg.2015.02.002, 2015.

Steger, S., Moreno, M., Crespi, A., Gariano, S. L., Brunetti, M. T., Melillo, M., Peruccacci, S., Marra, F., de Vugt, L., Zieher, T., Rutzinger, M., Mair, V., and Pittore, M.: Adopting the margin of stability for space-time landslide prediction - A data-driven approach for generating spatial dynamic thresholds, Geosci. Front., 15, 101822, https://doi.org/10.1016/j.gsf.2024.101822, 2024.

Sun, D. L., Wu, X. Q., Wen, H. J., Gu, Q. Y., and Zhang, Y. B.: A LightGBM-based landslide susceptibility model considering the uncertainty of non-landslide samples, Geomat. Nat. Haz. Risk, 14, 2213807, https://doi.org/10.1080/19475705.2023.2213807, 2023.

Tanaka, T., Kawase, H., Imada, Y., Kawai, Y., and Watanabe, S.: Risk-based versus storyline approaches for global warming impact assessment on basin-averaged extreme rainfall: a case study for Typhoon Hagibis in eastern Japan, Environ. Res. Lett., 18, 054010, https://doi.org/10.1088/1748-9326/accc24, 2023.

Tang, B. Z., Jiao, J. Y., Chen, Y. X., and Liang, Y.: Rill investigation on landslide scar highlights the shortcomings of current knowledge on erosion of shallow landslides, Land Degrad. Dev., 35, 1495–1510, https://doi.org/10.1002/ldr.5001, 2024.

https://doi.org/10.5194/egusphere-2026-1608
Preprint. Discussion started: 20 April 2026

1030 Wu, J., Lü, Q., Liao, Z., and An, N.: Rainfall infiltration in an extremely high rock fragment content
1031 geodisaster-prone landscape: Insights from intensive field monitoring, J. Hydrol., 664, 134481,
1032 https://doi.org/10.1016/j.jhydrol.2025.134481, 2025.
1033 Xiao, W. F., Yao, G. C., Xiao, Z. H., Liu, G., Luo, L. G., Cao, Y. J., and Yin, W.: From typhoon rainfall to
1034 slope failure: optimizing susceptibility models and dynamic thresholds for landslide warnings in Zixing
1035 City, China, Nat. Hazards Earth Syst. Sci., 26, 611–629, https://doi.org/10.5194/nhess-26-611-2026,
1036 2026.
1037 Xiao, W. F., Zhou, Z. Y., Ren, B. Z., and Deng, X. P.: Integrating spatial clustering and multi-source
1038 geospatial data for comprehensive geological hazard modeling in Hunan Province, Sci. Rep., 15, 1982,
1039 https://doi.org/10.1038/s41598-024-84825-y, 2025.
1040 Zhao, J. C., Feng, W. K., Yi, X. Y., Zhao, Y. L., Ding, Z. W., and Wang, Y. P.: Clustered shallow
1041 landslides caused by extreme typhoon rainstorms in Zixing County, Hunan Province, China, from July
1042 26 to 28, 2024, Landslides, 22, 2141–2147, https://doi.org/10.1007/s10346-025-02508-9, 2025.
1043 Zhuang, Y., Xing, A. G., Jiang, Y. H., Sun, Q., Yan, J. K., and Zhang, Y. B.: Typhoon, rainfall and trees
1044 jointly cause landslides in coastal regions, Eng. Geol., 298, 106561,
1045 https://doi.org/10.1016/j.enggeo.2022.106561, 2022.