https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

Multi-scale spatial validation and probability calibration of pixel-based landslide susceptibility modeling in the northern Peruvian Andes

Wendy Quiroz[1]

[1] Instituto Geofísico del Perú (IGP), Dirección de Ciencias de la Tierra Sólida, Lima, Peru
Correspondence to: Wendy Quiroz (wquiroz@igp.gob.pe)

Abstract

Landslides are recurrent geohazards in Andean regions, causing significant impacts on infrastructure and local communities. In spatially structured terrains, model reliability hinges on the definition of pseudo-absence samples and the treatment of spatial dependence during validation. This study evaluates pixel-based rotational landslide susceptibility in the province of Huancabamba (Piura, northern Peru) using a Random Forest classifier and seven conditioning factors derived from a photogrammetric digital elevation model and lithological data at 10 m resolution.

The landslide inventory consists of 25 field-mapped rotational landslides compiled from geomorphological surveys and high-resolution photogrammetric products. Pseudo-absence samples were selected outside mapped polygons using a buffered exclusion zone to reduce label uncertainty, and a balanced sampling scheme (1:1) was adopted. To obtain spatially realistic performance estimates, model evaluation was conducted using spatial block cross-validation with block sizes ranging from 600 to 1500 m. This provides a clear view of how spatial partitioning affects discrimination and calibration, alongside the model's stability throughout the validation folds.

Results show that discrimination performance decreases systematically as spatial block size increases, indicating that conventional random validation may overestimate predictive capacity due to spatial autocorrelation. A block size of 900 m provided a compromise between spatial independence and fold stability. Permutation importance computed under spatially independent folds identified lithology and elevation as the dominant predictors of rotational landslide occurrence, followed by aspect and topographic wetness index. Calibration metrics (Brier score and Expected Calibration Error) indicated moderate but stable reliability of susceptibility scores across spatial configurations.

The resulting susceptibility map shows spatial patterns consistent with the geomorphological setting and the mapped inventory, with high susceptibility concentrated in steep slopes developed over weak lithological units. These findings indicate that integrating spatial validation, calibration, and constrained sampling improves the reliability of pixel-based modelling in this Andean setting.

Keywords: landslide susceptibility; Random Forest; spatial cross-validation; pseudo-absence sampling; Andean

1. Introduction

Landslides represent one of the most recurrent and damaging geohazards in mountainous regions worldwide, producing significant impacts on infrastructure, ecosystems, and population safety. In

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

the Andes, slope instability is strongly controlled by the interaction of steep relief, heterogeneous lithological conditions, and marked climatic variability, which together favor the recurrence of mass-movement processes (Hermanns et al., 2012; Guzzetti et al., 2012). In northern Peru, particularly in the province of Huancabamba (Piura region), rotational landslides frequently affect transportation corridors, agricultural areas, and expanding peri-urban zones, increasing territorial vulnerability and socioeconomic losses.

Landslide susceptibility mapping (LSM) is widely used in risk reduction strategies and land-use planning, particularly in mountainous regions (Fell et al., 2008; Van Westen et al., 2008). Susceptibility assessment aims to estimate the spatial probability of landslide occurrence based on the relationship between past events and terrain conditioning factors (Corominas et al., 2014). The shift toward machine learning in landslide susceptibility modeling stems from the need to manage intricate, nonlinear interactions within high-dimensional multivariate datasets. Among them, Random Forest has demonstrated robust predictive performance across diverse geomorphological contexts (Breiman, 2001; Catani et al., 2013; Hong et al., 2019).

However, model performance does not depend only on the choice of algorithm. A critical issue in susceptibility modelling is the evaluation procedure itself. In spatially structured environments, the assumption of independence between training and validation samples is often violated due to spatial autocorrelation, which can lead to overly optimistic performance estimates and limited generalization capability (Chung and Fabbri, 2003; Roberts et al., 2016; Meyer et al., 2019; Valavi et al., 2019). This issue is particularly relevant for pixel-based susceptibility models, where neighboring pixels often share similar terrain attributes and may inadvertently introduce spatial leakage between training and validation subsets.

Spatially explicit cross-validation strategies, such as block cross-validation, have been proposed to mitigate spatial leakage by partitioning data into geographically independent subsets. Nevertheless, the influence of block size on predictive performance, stability across folds, and probabilistic reliability remains insufficiently explored in landslide susceptibility studies. Furthermore, while discrimination metrics such as the area under the ROC curve (AUC) are commonly reported, considerably less attention has been given to probabilistic calibration and predictive stability, which are essential for translating susceptibility maps into operational risk management tools. In susceptibility modelling, predicted values are often interpreted as continuous susceptibility scores rather than absolute probabilities, particularly when balanced sampling strategies are adopted. Therefore, in addition to discrimination metrics, calibration measures such as the Brier score (Brier, 1950; Alvioli et al., 2024) and the expected calibration error (ECE; Guo et al., 2017) were used to evaluate the reliability of the predicted susceptibility scores.

In this context, the study evaluates pixel-based rotational landslide susceptibility in the province of Huancabamba (northern Peru) under a spatially explicit validation framework designed to reduce spatial leakage and improve generalization assessment (Roberts et al., 2016; Meyer et al., 2019). Using Random Forest and terrain conditioning factors derived on a consistent 10 m grid, we compare spatial block cross-validation schemes to examine their influence on predictive discrimination and stability across folds. Although spatial cross-validation has been increasingly recommended in ecological and environmental modelling, systematic multi-scale evaluation of block size effects on both discrimination and probabilistic calibration remains rare in pixel-based landslide susceptibility studies.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

In addition to conventional discrimination metrics, probabilistic calibration metrics are incorporated to assess reliability, together with an analysis of predictor contributions using permutation importance computed within spatially independent folds. By focusing on spatial realism and score reliability, this study aims to improve the methodological consistency of landslide susceptibility modelling in complex mountainous environments.

2. Study area

The study area is located in the province of Huancabamba, Piura region, northern Peru, within the western flank of the northern Andes (Figure 1). The analyzed sector covers approximately 32 km² and comprises mountainous terrain characterized by steep slopes, deeply incised valleys, and pronounced relief energy. Elevations range from approximately 1,870 to 2,805 m a.s.l., reflecting strong altitudinal gradients and complex topographic conditions conducive to slope instability (Hermanns et al., 2012; Reichenbach et al., 2018).

Geologically, the area is composed predominantly of volcanic and sedimentary units of Cenozoic age, locally overlain by colluvial and alluvial deposits. These lithological formations are frequently affected by fracturing, weathering, and structural discontinuities, which reduce rock mass strength and increase susceptibility to gravitational processes. The regional geological framework is consistent with the national geological map of Peru compiled by INGEMMET (INGEMMET, 2021). Regional assessments report diverse mass movement mechanisms, including rotational landslides, rockfalls, soil creep, and ground settlements, following the classification of slope movements proposed by Gonzáles de Vallejo (2002). The predominance of rotational failures in the mapped inventory reflects the geomorphological and mechanical characteristics of the local lithological framework.

Climatically, Huancabamba is influenced by a seasonal precipitation regime associated with the austral summer, with rainfall concentrated between December and March. Intense and prolonged rainfall events act as major triggering factors for slope instability, particularly in steep terrains composed of low-cohesion materials (Guzzetti et al., 2008; Gariano and Guzzetti, 2016). In addition, the region is periodically affected by large-scale climatic anomalies linked to the El Niño–Southern Oscillation (ENSO), which may substantially increase rainfall intensity and duration and consequently enhance landslide occurrence (Takahashi et al., 2019; Lavado-Casimiro et al., 2013).

Land use within the study area is predominantly agricultural, interspersed with rural settlements and transportation infrastructure. Anthropogenic modifications, including road cuts, terracing, and vegetation removal, locally alter drainage patterns and slope geometry, further contributing to instability processes (Van Westen et al., 2008).

Geodynamic investigations conducted by the Instituto Geofísico del Perú (IGP, 2022) documented active and ancient mass movement processes affecting the urban sector, and GNSS monitoring between 2018 and 2022 confirmed measurable ground displacements associated with ongoing slope deformation. The mapped inventory comprises 25 predominantly rotational landslides with areas ranging from 0.84 to 82.62 ha (mean area ≈ 16.7 ha). The characteristic spatial footprint of the mapped landslides provides a geomorphological scale reference that supports pixel-based susceptibility modeling at 10 m resolution.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

In addition, the average event size informed the selection and sensitivity analysis of spatial validation units, ensuring that block dimensions exceeded the typical landslide footprint while minimizing spatial leakage.

The combination of pronounced topographic gradients, heterogeneous lithological conditions, seasonal climatic forcing, documented ground deformation, and anthropogenic disturbance makes the study area an appropriate natural laboratory for evaluating spatially realistic susceptibility modeling in Andean environments.

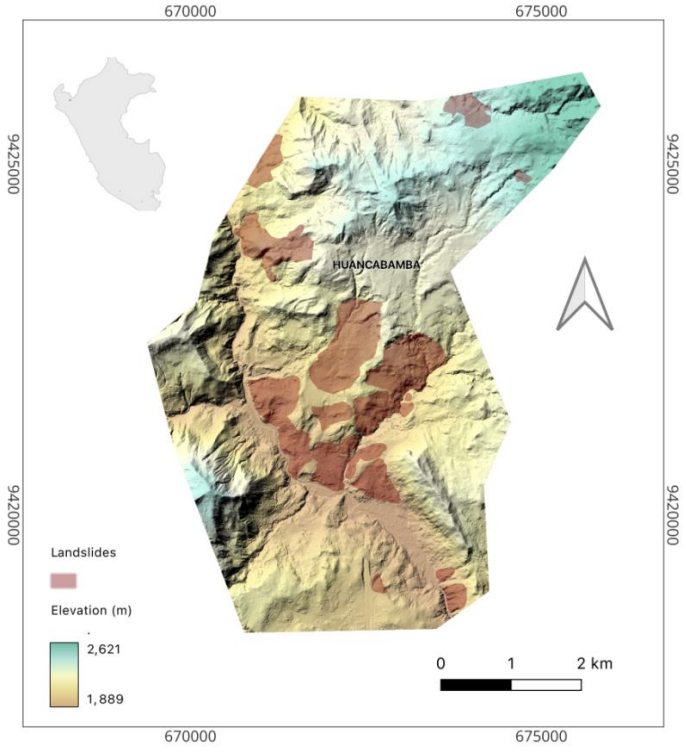

Figure 1. Study area and landslide inventory in the Huancabamba region, northern Peru. The background shows the digital elevation model (DEM) with hillshade to emphasize the regional topography. Red polygons represent the landslide inventory used in this study. The inset map indicates the location of the study area within Peru.

3. Materials and methods

3.1. Landslide inventory

The landslide inventory used in this study was compiled within the framework of the research project “Evaluación geofísica y geodinámica de los deslizamientos de tierra que afectan la seguridad física de la ciudad de Huancabamba, Piura”, conducted by the Instituto Geofísico del Perú (IGP) between 2018 and 2022. Detailed geomorphological mapping was carried out in 2018 through systematic field surveys supported by high-resolution photogrammetric products. A

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

consistent and process-based landslide inventory constitutes a fundamental prerequisite for susceptibility modeling and strongly conditions model validity (Reichenbach et al., 2018).

A total of 25 rotational landslide polygons were identified and delineated based on diagnostic geomorphic features, including main scarps, displaced material, tension cracks, hummocky topography, and disturbed drainage patterns. These polygons correspond to active instability sectors documented in regional geodynamic assessments (IGP, 2022). Polygon boundaries were manually digitized from field-validated photogrammetric products to ensure spatial consistency and geomorphological reliability, following established best practices in landslide inventory mapping (Guzzetti et al., 2012).

Although the study area exhibits multiple types of mass movement processes—including rotational landslides, rockfalls, soil creep, and ground settlements (Gonzáles de Vallejo, 2002)—only rotational landslides were considered in the present analysis. This decision was adopted to maintain process homogeneity and reduce class heterogeneity during model training, as susceptibility modelling is sensitive to the mechanical and geomorphological diversity of slope processes (Reichenbach et al., 2018). The modelling framework therefore follows a binary presence–absence scheme, where pixels located within mapped rotational landslide polygons represent the positive class.

Inventory polygons were rasterized on a consistent 10 m grid to match the spatial resolution of the conditioning factors derived from the resampled DEM and ancillary datasets. All pixels falling within the rasterized landslide polygons were considered potential presence samples. These were subsequently integrated into the spatial sampling and cross-validation framework described in Sect. 3.4.

The inventory represents the spatial distribution of mapped active instability at the time of the 2018 survey and provides a geomorphologically consistent reference dataset for pixel-based susceptibility modelling within the analyzed sector. The overall workflow of the susceptibility modelling procedure is shown in Figure 2.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

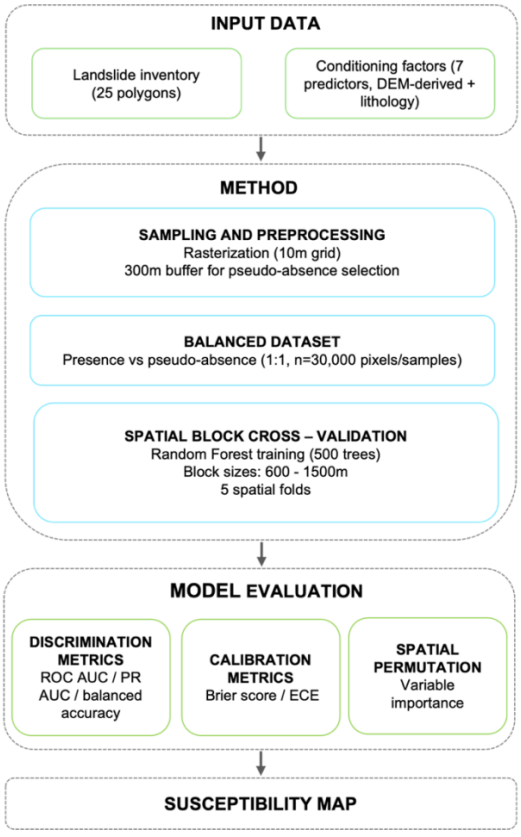

Figure 2. Workflow of the spatially explicit landslide susceptibility modelling framework. Landslide inventory polygons and terrain conditioning factors were rasterized on a consistent 10 m grid. Pseudo-absence samples were selected outside a 300 m buffer around mapped landslides to reduce label uncertainty, and a balanced dataset (1:1) was generated. Model evaluation was conducted using multi-scale spatial block cross-validation (600–1500 m) with five folds to ensure spatial independence between training and validation data. Discrimination metrics (ROC AUC, PR AUC, and balanced accuracy), calibration metrics (Brier score and expected calibration error), and spatial permutation importance were computed within each fold. The final Random Forest model was trained on the complete dataset to produce the landslide susceptibility map.

3.2. Conditioning factors

Based on geomorphological understanding of the study area and previous landslide susceptibility research, seven conditioning factors were selected to characterize terrain morphology and environmental variability: elevation, slope, plan curvature, slope aspect, topographic wetness index (TWI), surface roughness, and lithology (Van Westen et al., 2008; Reichenbach et al., 2018).

Topographic variables were derived from a UAV-based photogrammetric digital elevation model (DEM) with a native spatial resolution of 8 cm. Although this very high resolution provides detailed representation of micro-topographic features, such detail exceeds the characteristic scale

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

of mapped slope failures and may introduce noise unrelated to slope-scale instability processes. Therefore, the DEM was resampled to a spatial resolution of 10 m using bilinear interpolation prior to computing terrain derivatives. This resolution ensures compatibility with the geomorphological footprint of mapped landslides while reducing micro-scale variability.

The selected grid size was defined considering the size distribution of the mapped rotational landslides. The smallest mapped polygon exhibits an equivalent circular diameter of approximately 103 m, while the mean equivalent diameter is approximately 393 m. At 10 m resolution, even the smallest event is represented by multiple raster cells, preserving slope-scale representation while maintaining computational efficiency and methodological consistency with regional susceptibility studies (Reichenbach et al., 2018).

Elevation, slope, plan curvature, and aspect were directly derived from the resampled DEM using standard terrain analysis tools in QGIS. Slope was computed in degrees and subsequently converted to radians where required for hydrological indices. The topographic wetness index (TWI) was calculated using the SAGA GIS implementation based on flow accumulation and slope derived from the DEM. Surface roughness was computed using the roughness algorithm implemented in QGIS from the same resampled DEM, which measures local elevation variability within a moving window. Lithology was derived from the national geological map of Peru (INGEMMET, 2021), revised and interpreted within the framework of the IGP geodynamic investigation (IGP, 2022), and rasterized to the 10 m grid using nearest-neighbour resampling in order to preserve categorical information. The lithological variable captures contrasts in material composition and mechanical behavior across the study area.

All conditioning factors were aligned to a common spatial resolution (10 m), spatial extent, and coordinate reference system (WGS 84 / UTM zone 17S) prior to sampling and model training, ensuring consistent pixel-wise predictor extraction and compatibility with the spatial cross-validation framework.

3.3. Pseudo-absence selection

In landslide susceptibility modelling, the absence of mapped landslides does not necessarily imply geomorphological stability, particularly in mountainous environments where latent or unrecorded instabilities may be present (Reichenbach et al., 2018). Consequently, the definition of pseudo-absence samples represents a critical methodological step to reduce label variability and prevent contamination between presence and absence classes (Barbet-Massin et al., 2012; Meyer et al., 2019; Guo et al., 2024).

In this study, pseudo-absence samples were randomly selected from areas located outside mapped rotational landslide polygons. To reduce the likelihood of including pixels potentially affected by boundary variability or adjacent slope deformation, a 300 m buffer was applied around all mapped landslide polygons prior to pseudo-absence extraction. Candidate negative samples were drawn exclusively from areas beyond this exclusion zone.

The buffer distance (300 m) was defined considering the characteristic scale of mapped landslides. The smallest mapped polygon exhibits an equivalent circular diameter of approximately 103 m, while the mean equivalent diameter is approximately 393 m. A 300 m exclusion distance therefore reduces the probability of selecting ambiguous negative samples in the immediate vicinity of

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

mapped instability sectors, while preserving a sufficiently large domain for pseudo-absence sampling across the study area.

To avoid class imbalance effects in the Random Forest classifier and ensure stable discrimination performance, a balanced sampling scheme was adopted. An equal number of presence and pseudo-absence samples were used during model training (1:1 ratio), resulting in a total dataset of 30,000 samples. This balanced configuration facilitates interpretation of discrimination metrics and reduces bias associated with highly skewed class prevalence (Breiman, 2001; Japkowicz and Stephen, 2002). Because a balanced sampling strategy was adopted, the predicted model outputs were interpreted as relative susceptibility scores rather than absolute probabilities of landslide occurrence. This approach is commonly used in susceptibility modelling to improve classifier stability while maintaining reliable ranking of susceptible areas (Reichenbach et al., 2018).

Sampling was conducted prior to spatial block partitioning, and spatial separation between training and validation data was subsequently enforced through block cross-validation (Sect. 3.4). This strategy prioritizes spatial separation, label clarity, and methodological transparency, while allowing the influence of spatial structure on predictive performance to be explicitly evaluated.

This procedure ensures that spatial independence between training and validation subsets is enforced at the block level while preserving a consistent sampling strategy across block configurations.

3.4. Spatial block cross-validation

Conventional random cross-validation assumes independence among samples, an assumption frequently violated in spatially structured environmental datasets due to spatial autocorrelation (Roberts et al., 2016; Meyer et al., 2019). In pixel-based landslide susceptibility models, neighboring pixels often share similar geomorphological and environmental characteristics, which may lead to spatial leakage between training and validation subsets and consequently to overoptimistic performance estimates (Samodra et al., 2024).

To mitigate this issue, a spatial block cross-validation framework was implemented. The study area was partitioned into non-overlapping square blocks arranged on a regular grid. Entire blocks were assigned to either training or validation subsets, ensuring geographical separation between samples used for model calibration and evaluation. Five spatial folds were generated, and block assignment to folds was performed using a fixed random seed to ensure full reproducibility while maintaining spatial contiguity within each block. No clustering or aggregation of neighboring blocks into larger fold regions was applied; instead, fold membership was determined at the block level.

Block size selection was informed by the characteristic spatial scale of mapped landslides. The smallest mapped rotational landslide exhibits an equivalent circular diameter of approximately 103 m, while the mean equivalent diameter is approximately 393 m. Block dimensions substantially larger than these characteristic footprints reduce the probability that individual landslide systems are spatially split between training and validation subsets.

Rather than adopting a single block size a priori, a multi-scale sensitivity analysis was conducted. Several spatial block sizes between 600 and 1500 m were tested in order to evaluate the sensitivity of model performance to the spatial partitioning scheme.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

The tested range of block sizes was selected to exceed the mean mapped landslide footprint (~393 m) while allowing evaluation of model performance across increasing levels of spatial independence.

For each block configuration, model training was performed exclusively on spatially distinct training blocks, and performance metrics were computed on geographically independent validation blocks. In configurations where a validation fold contained only one class (presence or absence), discrimination metrics requiring two classes (ROC AUC and PR AUC) were not computed for that fold and were excluded from aggregated statistics. This conservative treatment prevents artificial inflation or distortion of discrimination performance under highly imbalanced spatial partitions.

This validation framework allows assessment of discrimination and calibration (Brier score and ECE) while maintaining spatial independence during generalization. It provides a spatially consistent basis for evaluating pixel-based susceptibility models in the Andes context.

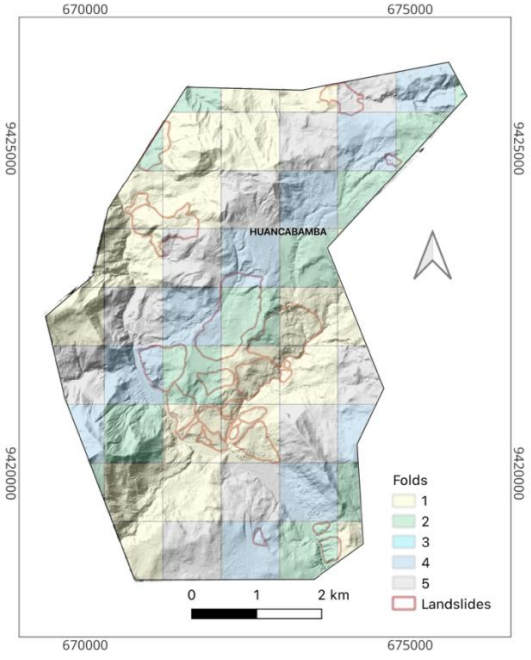

Figure 3. Spatial block cross-validation configuration using a 900 m block size. The study area was partitioned into non-overlapping square blocks assigned to five spatial folds, providing geographical separation between training and validation samples.

Table 1. Spatial cross-validation configuration.

Parameter	Primary configuration	Sensitivity analysis
Block size	900 m	600–1500 m
Number of folds	5	5
Pseudo-absence buffer	300 m	300 m
Class balance	1:1 (balanced)	1:1
Total sample size	30,000	30,000

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

3.5. Random Forest model specification

Landslide susceptibility was modeled using the Random Forest (RF) algorithm (Breiman, 2001), an ensemble-based classifier that constructs multiple decision trees through bootstrap aggregation and random feature selection at each split. RF has been widely adopted in landslide susceptibility studies due to its ability to capture nonlinear relationships, accommodate multicollinearity among predictors, and maintain robust predictive performance under complex geomorphological settings (Catani et al., 2013; Hong et al., 2019; Reichenbach et al., 2018).

Random Forest is well-suited for pixel-based susceptibility modelling in mountain terrain, where predictor variables often display complex spatial patterns and non-linear interactions.

The RF model was implemented using 500 trees (n_estimators = 500). This value was selected based on an explicit convergence analysis conducted under the spatial block cross-validation framework. Model performance metrics (ROC AUC and Brier score) exhibited stabilization beyond approximately 300–400 trees, while marginal improvements between 500 and larger ensembles were negligible (<0.002 in ROC AUC). Therefore, 500 trees were retained as a balance between predictive stability and computational efficiency.

Trees were grown without imposing a maximum depth (max_depth = None), allowing recursive partitioning until terminal nodes became pure or contained fewer than the minimum number of samples required for splitting. The minimum number of samples per leaf was set to one (min_samples_leaf = 1), following common practice in environmental modelling applications where flexible partitioning of complex predictor space is desirable (Breiman, 2001).

Because internal out-of-bag (OOB) error estimates may underestimate generalization error in spatially structured datasets, predictive performance was evaluated exclusively under spatial block cross-validation (Sect. 3.4). Block sizes ranging from 600 m to 1500 m were assessed to quantify the influence of spatial separation on discrimination performance, probabilistic reliability, and model stability.

All models were trained using a fixed random seed to ensure reproducibility of bootstrap sampling and fold allocation. Given the balanced sampling strategy (1:1 presence to pseudo-absence ratio), no additional class weighting was applied.

Predictor importance was evaluated using two complementary approaches. First, mean decrease in impurity (MDI) was computed for the full model. Second, and more conservatively, spatial permutation importance was calculated within each spatial validation fold. This fold-wise permutation procedure quantifies the reduction in predictive performance when the association between a predictor and the response is randomly disrupted under spatially independent validation conditions, thereby providing a more robust assessment of variable contribution in the presence of spatial autocorrelation.

In addition to conventional discrimination metrics (ROC AUC and PR AUC), probabilistic reliability was explicitly evaluated through the Brier score and the Expected Calibration Error (ECE). This evaluation allows assessment of both discrimination and calibration metrics, improving the interpretability of susceptibility scores.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

Finally, after model evaluation under spatial cross-validation, a final Random Forest model with the selected hyperparameters was trained using the complete balanced dataset (30,000 samples) to generate the continuous landslide susceptibility map for the study area.

3.6. Model evaluation and spatial validation

Model performance was evaluated exclusively under the spatial block cross-validation framework described in Sect. 3.4. For each block configuration, predictive performance metrics were computed independently for each spatial fold and subsequently summarized using mean and standard deviation values to quantify inter-fold variability and model stability.

Discrimination capacity was assessed using the area under the receiver operating characteristic curve (ROC AUC) and the area under the precision–recall curve (PR AUC). ROC AUC provides a threshold-independent measure of separability between presence and pseudo-absence classes, whereas PR AUC offers complementary information focused on the positive (landslide) class and is particularly informative under spatially variable class prevalence.

To evaluate classification balance at a fixed decision threshold of 0.5 applied to the susceptibility scores, balanced accuracy was computed. This metric accounts for both sensitivity and specificity and reduces bias associated with uneven class distribution across spatial folds.

Beyond discrimination metrics, calibration of the predicted susceptibility scores was explicitly evaluated. The Brier score (Brier, 1950) was used to quantify the mean squared difference between predicted susceptibility scores and observed outcomes, thereby measuring overall probabilistic accuracy. In addition, the Expected Calibration Error (ECE; Guo et al., 2017, Dormann et al., 2020) was calculated using 10 prediction bins to evaluate agreement between predicted probabilities and observed event frequencies across the probability spectrum.

Because the modelling dataset was constructed using balanced sampling (1:1 presence and pseudo-absence), the Random Forest outputs were interpreted as susceptibility scores rather than absolute event probabilities. Calibration metrics were therefore used to evaluate the reliability of these scores under spatially independent validation. The results should not be interpreted as unconditional landslide probabilities at the landscape scale.

In spatial folds containing only one class (presence or pseudo-absence), discrimination metrics requiring two classes (ROC AUC and PR AUC) were not computed and were excluded from aggregated statistics. This conservative approach prevents artificial inflation or distortion of performance estimates under highly imbalanced spatial partitions.

This evaluation approach assesses discrimination capacity, probabilistic calibration, and predictive stability under increasing spatial independence.

4. Results

4.1. Multi-scale spatial validation results

Model performance varied systematically as a function of spatial block size. Increasing block size imposed stricter spatial separation between training and validation samples, thereby reducing spatial information sharing and providing a more conservative estimate of generalization

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

performance (Figure 4). Block sizes between 600 and 1500 m were evaluated to quantify the effect of increasing spatial independence constraints on model performance.

Under the 600 m block configuration, the model achieved the highest discrimination performance (mean ROC AUC ≈ 0.91), accompanied by relatively low inter-fold variability. However, this configuration likely permits partial spatial dependence between neighboring blocks, potentially resulting in optimistic performance estimates.

At 750 m block size, discrimination metrics remained high (ROC AUC ≈ 0.90), while maintaining stable fold representation of both classes. Based on the sensitivity analysis, the 900 m block size was selected as the primary spatial validation configuration because it provided a balance between spatial independence and sufficient sample size within each fold. As block size increased to 900 m, discrimination performance decreased (ROC AUC ≈ 0.85), reflecting reduced spatial information leakage and more realistic generalization constraints (Figure 4).

Further increases in block size (1050–1500 m) led to additional variability across folds. In particular, for block sizes ≥1200 m, some validation folds contained only one class (either presence or pseudo-absence), preventing computation of ROC AUC and PR AUC in those folds. This behavior reflects the spatial clustering of mapped rotational landslides and highlights the trade-off between spatial independence and statistical representativeness.

Across block configurations, probabilistic calibration metrics exhibited moderate variation. Brier scores ranged approximately between 0.13 and 0.16, while Expected Calibration Error (ECE) values remained below 0.18 for all configurations, indicating acceptable reliability of susceptibility scores under spatially explicit validation.

The 900 m configuration represents a balanced compromise between spatial independence and fold stability. It avoids excessive class degeneration observed in larger block sizes while substantially reducing potential spatial leakage relative to smaller partitions. For this reason, the 900 m configuration was retained as the primary spatial validation reference in subsequent analyses (vertical dashed line in Figure 4).

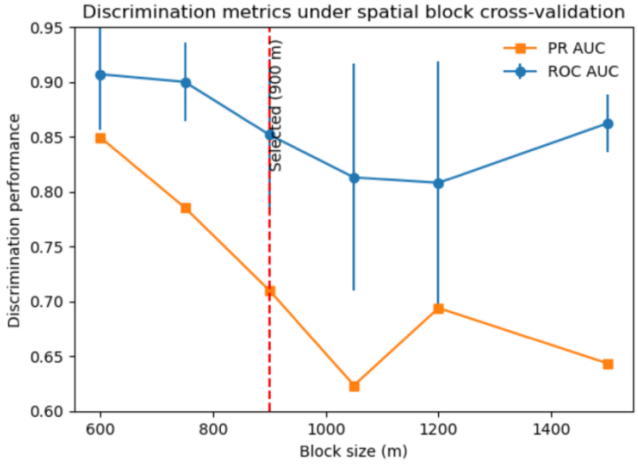

Figure 4. Discrimination performance under spatial block cross-validation. Mean ROC AUC and PR AUC values are shown across increasing spatial block sizes (600–1500 m). Error bars represent

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

inter-fold standard deviation. The dashed vertical line indicates the 900 m configuration selected as the primary spatial validation reference.

4.2. Variable importance under spatial permutation

Variable contribution was evaluated using spatial permutation importance computed within each spatial validation fold. Unlike mean decrease in impurity (MDI), which is derived from internal tree splits and may be biased in the presence of correlated predictors, permutation importance quantifies the reduction in predictive performance when the relationship between a given predictor and the response variable is randomly disrupted under spatially independent validation conditions.

Under the 900 m block configuration, lithology and elevation consistently exhibited the highest importance values across folds. Lithology showed the strongest average contribution (mean importance ≈ 0.23), followed by elevation (≈ 0.19), although elevation displayed greater inter-fold variability. These results indicate that geological contrasts and large-scale topographic gradients exert primary control on the spatial distribution of rotational landslides in the study area.

Slope aspect (direction) showed moderate importance (≈ 0.016), suggesting that orientation-dependent processes—potentially related to differential insolation, moisture retention, or structural control—play a secondary but non-negligible role. In contrast, curvature, slope gradient, and roughness exhibited low importance values, with roughness occasionally yielding near-zero or slightly negative permutation effects, indicating limited independent predictive contribution once lithology and elevation are accounted for.

Comparable patterns were observed under the 750 m configuration, although importance values exhibited greater variability for lithology and elevation. The relative ranking of predictors remained stable across block sizes, reinforcing the robustness of the dominant geomorphological controls identified.

The dominance of lithology highlights the role of material properties and structural discontinuities in governing rotational failure development, while elevation likely acts as a proxy for broader geomorphic setting and relief energy. The reduced contribution of fine-scale morphometric derivatives suggests that, at 10 m resolution and under spatially independent validation, large-scale geological and topographic gradients exert stronger predictive influence than local micro-topographic variability.

Spatial permutation analysis indicates consistent dominance of lithology and elevation across independent folds, supporting the geomorphological coherence of the susceptibility model.

Table 2. Permutation-based variable importance obtained from the Random Forest model. Importance values represent the mean decrease in model performance after random permutation of each predictor.

Variable	Mean Importance	Std
Lithology	0.234	0.030
Elevation	0.191	0.062
Aspect	0.016	0.005
Topographic wetness index (TWI)	0.004	0.005
Plan curvature	0.001	0.002
Slope	0.0005	0.003
Surface roughness	-0.0007	0.007

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

Importance was computed using permutation importance based on the decrease in model performance after random shuffling of each predictor.

4.3. Susceptibility map and spatial patterns

The final landslide susceptibility map was generated using the Random Forest model trained on the complete balanced dataset (30,000 samples) with the selected hyperparameters (500 trees). The resulting output represents a continuous susceptibility score (0-1), where higher values indicate higher relative susceptibility (Figure 5). In susceptibility modelling, predicted values are often interpreted as continuous susceptibility scores rather than absolute probabilities, particularly when balanced sampling strategies are adopted.

Spatially, high-susceptibility zones are predominantly concentrated in sectors characterized by steep slopes developed over lithological units previously identified as mechanically weaker or structurally fractured (Figure 5). These areas coincide with mapped rotational landslides and extend along geomorphological alignments controlled by geological contrasts and relief energy.

Intermediate susceptibility values are distributed along transitional terrain zones, often corresponding to slope breaks, valley flanks, and lithological boundaries. These sectors may represent areas with favorable conditioning factors but lacking direct evidence of current instability in the mapped inventory.

Low-susceptibility zones are primarily located in relatively gentle slopes, valley bottoms, and areas underlain by more competent lithological units (Figure 5). The spatial coherence of low-probability areas supports the internal consistency of the model and suggests that the dominant conditioning factors exert meaningful geomorphological control.

The susceptibility surface exhibits smooth spatial gradients rather than isolated pixel-level artifacts, reflecting the influence of large-scale predictors such as lithology and elevation. This pattern is consistent with the spatial permutation importance analysis, which identified geological and topographic gradients as primary controls of instability.

The spatial distribution of predicted susceptibility is consistent with observed geomorphological patterns and mapped instability sectors.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

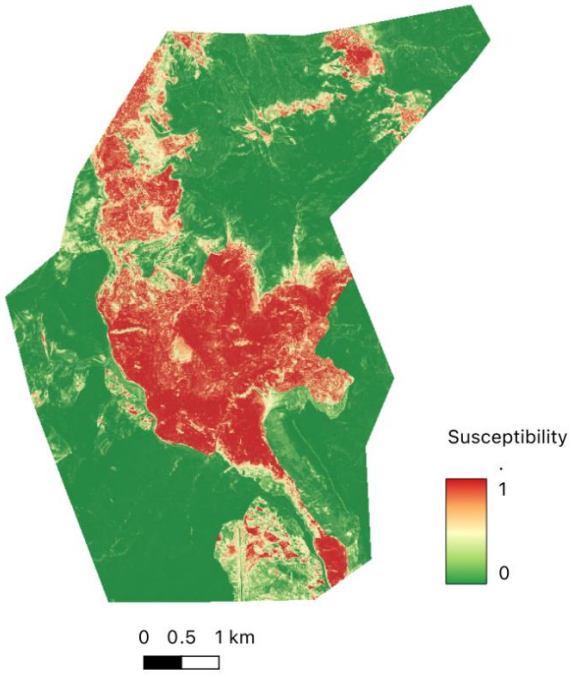

Figure 5. Landslide susceptibility map of the Huancabamba study area derived from the Random Forest model. High susceptibility values (red) coincide predominantly with steep slopes and weaker lithological units, whereas low susceptibility areas (green) occur in relatively stable terrain.

4.4. Calibration and performance variability

In addition to discrimination metrics, calibration measures were used to evaluate the reliability of the predicted susceptibility scores. Model stability was assessed through the variability of performance metrics across spatial cross-validation folds. The Brier score and the expected calibration error (ECE) were computed to quantify the agreement between predicted scores and observed landslide occurrence, while the standard deviation of performance metrics across folds was used as an indicator of model stability under different spatial partitions.

Across spatial block configurations, Brier scores ranged between approximately 0.13 and 0.16, indicating moderate calibration accuracy of the susceptibility scores under spatially independent validation (Figure 6). Expected Calibration Error (ECE) values remained below 0.18 for all evaluated block sizes, suggesting acceptable calibration reliability.

Calibration metrics exhibited moderate but non-monotonic variation with increasing spatial block size (Figure 6). While discrimination performance decreased as spatial independence increased (Sect. 4.1), calibration indicators remained comparatively stable, without systematic degradation.

For block sizes ≥1200 m, some validation folds contained only a single class, reflecting the spatial clustering of mapped rotational landslides. Although this structural constraint increased variability in discrimination metrics, calibration values remained within a relatively narrow range.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

The observed variability across folds reflects the sensitivity of the model to spatial partitioning and provides an estimate of the robustness of the susceptibility patterns under different spatial independence settings.

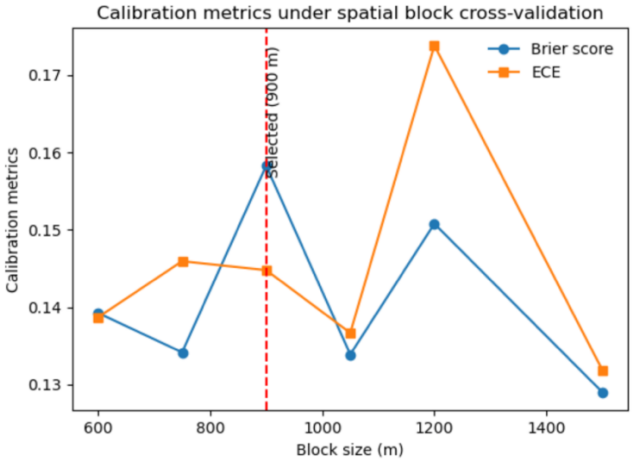

Figure 6. Calibration performance across spatial block sizes. Brier score and Expected Calibration Error (ECE) computed under spatial block cross-validation. The dashed line indicates the selected 900 m configuration.

5. Discussion

5.1 Influence of spatial scale on model generalization

Pixel-based susceptibility models are particularly prone to inflated validation scores when spatial autocorrelation is ignored, because neighbouring pixels share similar predictor values and can cause spatial leakage under random cross-validation (Roberts et al., 2016; Meyer et al., 2019; Valavi et al., 2019; Knevels et al., 2023).

The multi-scale spatial validation framework implemented in this study reveals a clear trade-off between discrimination performance and spatial independence. Smaller block sizes (600–750 m) yielded higher ROC AUC values, suggesting strong apparent predictive capacity. However, such configurations may allow residual spatial dependence between training and validation subsets, thereby inflating discrimination metrics through spatial leakage—a well-known issue when conventional cross-validation assumptions are violated in spatially structured environmental datasets (Roberts et al., 2016; Meyer et al., 2019; Valavi et al., 2019).

As block size increased, discrimination performance systematically decreased and inter-fold variability increased, consistent with progressively stricter spatial independence constraints and reduced spatial autocorrelation between training and validation samples (Roberts et al., 2016; Valavi et al., 2019). For block sizes ≥1200 m, the occurrence of validation folds containing only one class further indicates a clustered spatial distribution of rotational landslides within the study area. This outcome illustrates an important practical limitation of strict spatial partitioning: while larger blocks increase geographical separation, they may compromise statistical representativeness

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

when the number of independent spatial units becomes small (Valavi et al., 2019; Roberts et al., 2016).

The 900 m configuration represents a geomorphologically informed compromise between spatial independence and fold stability. This scale exceeds the characteristic footprint of mapped landslides (mean equivalent diameter ≈ 393 m), reducing the likelihood of splitting individual instability systems across folds while preserving a sufficient number of blocks to maintain class representation.

These results show that model performance depends strongly on the spatial scale used for validation. Reporting only a single validation configuration, without testing alternative block sizes, may therefore obscure the influence of spatial autocorrelation on apparent predictive performance (Meyer et al., 2019; Roberts et al., 2016; Reichenbach et al., 2018).

5.2 Geomorphological controls and stability of dominant predictors

Spatial permutation importance analysis consistently identified lithology and elevation as the dominant predictors of rotational landslide occurrence across spatial block configurations. This finding highlights the primary role of geological contrasts and large-scale topographic gradients in controlling slope instability within the study area.

Lithology exhibited the strongest and most stable contribution under spatially independent validation. Unlike internal tree-based importance metrics, spatial permutation importance evaluates predictor contribution under geographically separated validation subsets, thereby reducing bias introduced by spatial autocorrelation. The dominance of lithology suggests that material properties, structural discontinuities, and differential weathering are the most influential predictors of landslide occurrence in the study area. This result is consistent with previous landslide susceptibility studies emphasizing the fundamental role of geological framework in slope instability processes (Van Westen et al., 2008; Reichenbach et al., 2018).

Elevation also showed strong predictive contribution, although with greater inter-fold variability. Rather than acting as a purely geometric descriptor, elevation likely captures broader geomorphological setting, relief energy, and associated environmental gradients. In Andean terrain, elevation may integrate multiple process-related controls, including slope steepness patterns, hydrological gradients, and long-term landscape evolution (Guzzetti et al., 2008; Gariano and Guzzetti, 2016). Its importance therefore reflects both direct and indirect conditioning influences.

In contrast, local morphometric derivatives such as curvature, slope gradient, and roughness exhibited limited independent contribution under spatial permutation testing. While such variables are frequently reported as relevant predictors in conventional susceptibility studies, their reduced importance under spatially explicit validation suggests that part of their apparent predictive power may be associated with localized spatial clustering rather than robust generalizable structure. This observation aligns with recent critiques emphasizing the need to reassess variable importance under spatially independent frameworks (Meyer et al., 2019; Roberts et al., 2016).

The relative ranking of predictors remained stable across block sizes. Even under more rigorous spatial separation, lithology and elevation continued to dominate. This suggests that the identified susceptibility patterns reflect consistent geomorphological controls rather than being solely driven by spatial autocorrelation.

5.3 Probabilistic reliability and implications for hazard assessment

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

While discrimination metrics such as ROC AUC are widely used to evaluate landslide susceptibility models, they do not assess whether predicted probabilities correspond to observed event frequencies. In operational risk management contexts, however, the reliability of probability estimates is as critical as class separability. A model capable of distinguishing presence from absence may still provide poorly calibrated probability values, potentially misleading decision-making processes.

The evaluation framework adopted in this study incorporated probabilistic reliability metrics, including the Brier score and the Expected Calibration Error (ECE), under spatially explicit validation. Across block configurations, calibration metrics indicated moderate and relatively stable probabilistic agreement, even as discrimination performance decreased under stricter spatial independence constraints. This behavior suggests that although spatial separation reduces apparent separability, the model maintains consistent probability structure.

Calibration curves revealed minor deviations at higher susceptibility bins, where predicted probabilities slightly overestimated observed frequencies. Such behavior is not uncommon in ensemble classifiers trained under balanced sampling schemes and may reflect the interaction between class balance and spatial clustering. Importantly, the magnitude of calibration deviation remained limited across spatial scales, indicating that predicted susceptibility values retain meaningful probabilistic interpretation.

From a hazard management perspective, probabilistic reliability enhances the practical value of susceptibility maps. Well-calibrated susceptibility scores allow thresholds for monitoring or intervention to be defined based on meaningful risk levels rather than solely on the relative ranking of susceptibility values. In mountainous environments characterized by spatially clustered instabilities, ensuring calibration of the predicted scores under spatially independent validation strengthens confidence in the applicability of susceptibility outputs to territorial planning and mitigation strategies.

Evaluating discrimination, calibration, and spatial stability together provides a more complete assessment of model estimates.

5.4 Limitations and future research directions

Despite the methodological rigor adopted in this study, several limitations should be acknowledged.

First, the landslide inventory is based on field mapping conducted in 2018 and represents active instability sectors documented during that survey period. Although geomorphologically consistent, the inventory does not capture potential future failures or temporal variability in slope activity. Incorporating multi-temporal inventories could further refine susceptibility calibration and improve long-term generalization assessment.

Second, only rotational landslides were included in the modelling framework in order to preserve process homogeneity. While this approach reduces class heterogeneity and enhances interpretability, it limits the applicability of the resulting susceptibility map to other mass movement types such as rockfalls or shallow translational slides. Future research could explore multi-class or process-specific modelling frameworks under spatially explicit validation.

Third, the balanced sampling strategy (1:1 presence to pseudo-absence ratio) facilitates discrimination assessment but does not reflect true spatial prevalence of landslides within the study area. Although probabilistic calibration metrics were evaluated under this configuration, further

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

work could examine the effect of prevalence-aware sampling strategies on probability interpretation.

Finally, the spatial block cross-validation approach, while reducing spatial leakage, is inherently sensitive to the number and configuration of available spatial units. In relatively small study areas (~32 km²), strict spatial independence constraints may reduce fold representativeness, as observed in larger block configurations. Expanding the framework to broader regional scales would allow evaluation of model behavior under increased spatial heterogeneity.

Future research incorporating multi-temporal data broader spatial coverage, and additional physically based predictors could improve evaluation of susceptibility models under spatially independent validation.

6. Conclusions

This study evaluated pixel-based rotational landslide susceptibility in the Huancabamba sector of the northern Peruvian Andes under a spatially explicit validation framework designed to reduce spatial leakage and provide realistic estimates of model generalization.

Model performance was strongly dependent on spatial partition scale. Discrimination decreased systematically from mean ROC AUC ≈ 0.91 at 600 m block size to ≈ 0.85 at 900 m, highlighting the influence of spatial autocorrelation on apparent predictive capacity. Larger block sizes (≥1200 m) increased inter-fold variability and led to class imbalance in some folds, illustrating the trade-off between spatial independence and statistical representativeness in relatively small study areas (~32 km²). A 900 m block size was retained as a compromise, exceeding the mean landslide footprint (~393 m) while preserving fold stability.

Spatial permutation importance computed under geographically independent folds consistently identified lithology (mean importance ≈ 0.23) and elevation (≈ 0.19) as dominant predictors. Their stable ranking across block configurations indicates that large-scale geological contrasts and topographic gradients exert primary control on the spatial distribution of rotational failures in the study area. In contrast, fine-scale morphometric derivatives showed limited independent contribution once spatial dependence was constrained.

Beyond discrimination metrics, calibration assessment indicated moderate but stable reliability of susceptibility scores across spatial configurations (Brier score ≈ 0.13–0.16; ECE < 0.18). Although discrimination declined under stricter spatial separation, calibration remained comparatively consistent, suggesting that predicted scores preserve meaningful probabilistic structure under spatially independent validation.

These results demonstrate that validation scale substantially affects both apparent discrimination and stability of pixel-based susceptibility models. Reporting a single validation configuration may therefore obscure the influence of spatial autocorrelation on performance estimates. Multi-scale spatial cross-validation provides a transparent means of quantifying this effect and selecting a validation scale consistent with geomorphological process dimensions.

The modelling approach developed in this study is particularly suited to rotational landslides under similar geomorphological conditions. Extension to other mass movement types would require process-tailored sampling strategies and predictor selection.

Data availability.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

The datasets used in this study include a landslide inventory and several terrain conditioning factors derived from remote sensing products and thematic cartography. Because some of the raster layers are relatively large, the complete dataset is not hosted in the code repository. The data supporting the results presented in this study can be obtained from the author upon reasonable request.

Code availability.

The notebook and scripts used for data preprocessing, Random Forest modelling, spatial block cross-validation, and performance evaluation are maintained in a GitHub repository prepared for this study. The repository can be made available upon request during the review process and will be made publicly accessible after acceptance.

Author contributions.

Wendy Quiroz designed the study, prepared the datasets, developed the methodology, carried out the modelling, analysed the results, and wrote the manuscript.

Competing interests.

The author declares that there are no competing interests.

Acknowledgements.

The author gratefully acknowledges the support of the Instituto Geofísico del Perú (IGP) during the development of this research and thanks colleagues and collaborators for helpful discussions that contributed to the preparation of this work.

References

Alvioli, M., et al.: A benchmark dataset and workflow for landslide susceptibility zonation, Earth-Science Reviews, 258, 104927, https://doi.org/10.1016/j.earscirev.2024.104927, 2024.

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W.: Selecting pseudo-absences for species distribution models: how, where and how many?, Methods Ecol. Evol., 3, 327–338, https://doi.org/10.1111/j.2041-210X.2011.00172.x, 2012.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brier, G. W.: Verification of forecasts expressed in terms of probability, Mon. Weather Rev. 78, 1–3, 1950

Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V.: Landslide susceptibility estimation by Random Forests technique: sensitivity and scaling issues, Nat. Hazards Earth Syst. Sci., 13, 2815–2831, https://doi.org/10.5194/nhess-13-2815-2013, 2013.

Chung, C. J. F. and Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping, Nat. Hazards, 30, 451–472, https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b, 2003.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., and Pastor, M.: Recommendations for the quantitative analysis of landslide risk, Bull. Eng. Geol. Environ., 73, 209–263, https://doi.org/10.1007/s10064-013-0538-8, 2014.

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M., Ciuti, S., Elith, J., Gerstner, K., Guisan, A., Keil, P., Lahoz-Monfort, J. J., Pollock, L. J., Reineking, B., Roberts, D. R., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Wood, S.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

N., and Hartig, F.: Calibration of probability predictions from machine-learning and statistical models, Global Ecology and Biogeography, 29, 760–776, https://doi.org/10.1111/geb.13070, 2020.
Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning, Eng. Geol., 102, 85–98, https://doi.org/10.1016/j.enggeo.2008.03.014, 2008.
Gariano, S. L., and Guzzetti, F.: Landslides in a changing climate, Earth-Science Reviews, 162, 227–252, https://doi.org/10.1016/j.earscirev.2016.08.011, 2016.
Gonzáles de Vallejo, L., Ferrer, M., Ortuño, L., and Otero, C.: Ingeniería geológica, Pearson Educación, Madrid, 2002.
Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q.: On Calibration of Modern Neural Networks, in: Proceedings of the 34th International Conference on Machine Learning (ICML), 2017. Available from https://proceedings.mlr.press/v70/guo17a.html.
Guo, Z., Tian, B., Zhu, Y., et al.: How do the landslide and non-landslide sampling strategies impact landslide susceptibility assessment? — A catchment-scale case study from China, Journal of Rock Mechanics and Geotechnical Engineering, 16, 877–894, https://doi.org/10.1016/j.jrmge.2023.07.026, 2024.
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.
Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T.: Landslide inventory maps: new tools for an old problem, Earth-Sci. Rev., 112, 42–66, https://doi.org/10.1016/j.earscirev.2012.02.001, 2012.
Hermanns, R. L., Oppikofer, T., Anda, E., Blikra, L. H., Crosta, G. B., and Dahle, H.: Landslide dams in the Andes of Peru – hazard and risk assessment, Landslides, 9, 157–172, https://doi.org/10.1007/s10346-011-0300-1, 2012.
Hong, H., Pradhan, B., Xu, C., and Tien Bui, D.: Spatial prediction of landslide hazard using Random Forest, Gradient Boosting Machine, and Naïve Bayes Tree models, Geomorphology, 330, 15–29, https://doi.org/10.1016/j.geomorph.2018.12.021, 2019.
INGEMMET: Mapa geológico del Perú, escala 1:1 000 000, Instituto Geológico, Minero y Metalúrgico, Lima, 2021.
Instituto Geofísico del Perú (IGP): Huancabamba: Evaluación geofísica y geodinámica de los deslizamientos de tierra que afectan la seguridad física de la ciudad de Huancabamba, Institutional report, 2022.
Japkowicz, N., and Stephen, S.: The class imbalance problem: A systematic study, Intelligent Data Analysis, 6(5), 429–449, https://doi.org/10.3233/IDA-2002-6504, 2002.
Knevels, R., Petschko, H., Proske, H., Leopold, P., Mishra, A. N., Maraun, D., and Brenning, A.: Assessing uncertainties in landslide susceptibility predictions in a changing environment (Styrian Basin, Austria), Nat. Hazards Earth Syst. Sci., 23, 205–229, https://doi.org/10.5194/nhess-23-205-2023, 2023.
Lavado-Casimiro, W. S., Felipe, O., Silvestre, E., and Bourrel, L.: ENSO impact on hydrology in Peru, Advances in Geosciences, 33, 33–39, https://doi.org/10.5194/adgeo-33-33-2013, 2013.
Meyer, H., Reudenbach, C., Wöllauer, S., and Nauss, T.: Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction, Ecol. Modell., 411, 108815, https://doi.org/10.1016/j.ecolmodel.2019.108815, 2019.

https://doi.org/10.5194/egusphere-2026-1318
Preprint. Discussion started: 17 March 2026

735 Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-
736 based landslide susceptibility models, Earth-Sci. Rev., 180, 60–91,
737 https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.
738 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S.,
739 Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and
740 Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or
741 phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2016.
742 Samodra, G., Ngadisih, and Nugroho, F. S.: *Benchmarking data handling strategies for landslide*
743 *susceptibility modeling using random forest workflows*, Artificial Intelligence in Geosciences, 5,
744 100093, https://doi.org/10.1016/j.aiig.2024.100093, 2024.
745 Takahashi, K., and Martínez, A. G.: The very strong coastal El Niño in 1925 in the far-eastern
746 Pacific, Climate Dynamics, 52, 7389–7415, https://doi.org/10.1007/s00382-017-3702-1, 2019.
747 Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. BlockCV: An R package for
748 generating spatially or environmentally separated folds for k-fold cross-validation of species
749 distribution models, Methods Ecol. Evol., 10, 225–232, https://doi.org/10.1111/2041-210X.13107,
750 2019.
751 Van Westen, C. J., Castellanos, E., and Kuriakose, S. L.: Spatial data for landslide susceptibility,
752 hazard, and vulnerability assessment: An overview, Eng. Geol., 102, 112–131,
753 https://doi.org/10.1016/j.enggeo.2008.03.010, 2008.