Quaternary Research (2026), 1–19
doi:10.1017/qua.2026.10101

Research Article

Sensitivity analysis workflow to address landslide susceptibility model variability

Hudson J. Koch[a,b], Jason M. Dortch[a,b], Brent Harrison[c] and Matthew M. Crawford[a,b]

[a]Department of Earth & Environmental Sciences, University of Kentucky, Lexington, USA; [b]Kentucky Geological Survey, University of Kentucky, Lexington, USA and [c]Department of Computer Science, University of Kentucky, Lexington, USA

Abstract

Contemporary landslide susceptibility modeling commonly utilizes machine learning (ML) approaches and algorithms. These modeling approaches typically use landslide inventories as a primary component, along with a conditioning feature selection method. Although model results may be sensitive to variability in input data, resulting effects are seldom investigated in detail. We leveraged a susceptibility approach based on logistic regression (LR) to assess the sensitivity of model results to landslide inventory size. Using an inventory of 1054 landslides, we evaluated sample size at 50-landslide increments (to the maximum possible value of 1050), generating 1000 random inventory subsets for each sample size (i.e., Monte Carlo simulation) to quantify variation in model performance due to randomized train and holdout selection. Initial visual inspection of susceptibility map outputs produced using these model results revealed extreme overprediction. The spurious models indicate a false correlation between model accuracy and realism. To address this, our sensitivity analysis workflow utilizes a performance metric correction factor and stricter feature selection to promote consistent, realistic, and reproducible models. The workflow presented here quantifies model variability and assesses the sensitivity of susceptibility maps to changing model inputs.

Keywords: landslides; machine learning; data-driven modeling

Introduction

Landscapes evolve due to tectonic and climatic forcing, ideally moving towards dynamic equilibrium (Gilbert, 1909; Hack, 1960), specifically balancing uplift with denudation (Whipple, 2004), leaving behind a complex record (i.e., topography; Dortch et al., 2011; Dietsch et al., 2015). Landslides play a significant role in transient topographic cyclicity over Quaternary time scales where successive changes in base level can drive vertical incision, both glacial and fluvial. The effect of landslides on regional- and catchment-scale denudation and landscape evolution is significant (DiBiase et al., 2010; Campforts et al., 2022), although sometimes difficult to quantify (Dortch et al., 2009), as landslides communicate fluvial incision to ridges through hillslope adjustment (Willett, 2010). Landslide activity is variable in frequency and magnitude across time and space and dependent on several geomorphic factors, such as hypsometry (Aristizábal and Korup, 2025), incision (Campforts et al., 2020), and sediment accumulation (Parker et al., 2016). Erosion rates inferred from landslide frequency–size distributions have been recorded at the level of, or greater than, those of the tectonic uplift rate (Malamud et al., 2004; Larsen and Montgomery, 2012; Booth et al., 2013). These processes can be interlinked; rejuvenated landscapes are typically associated with a relative increase in landsliding and the presence of many landslides in the catchments of the southern Appalachians led to identifying regional uplift in the Miocene (Gallen et al., 2013).

Landslides are an integral part of landscape evolution modeling, however, they are also a threat to humans and their built environments (Broeckx et al., 2020). The threat of landsliding turns modeling foci towards applied modeling of landslide susceptibility. Susceptibility models have the goal of modeling probability of occurrence of landslides over a landscape (Reichenbach et al., 2018), often employing some statistical method and an input dataset for training (i.e., data-driven approaches). Data-driven machine learning (ML) modeling of landslide susceptibility has a wide variety of parameterization options. These include the selection of modeling algorithm, hyperparameters, features (e.g., choosing slope angle or curvature as conditioning factors), and specifics of the training data (Reichenbach et al., 2018). ML applications for landslide susceptibility models are commonplace, and research continues to grow (Lima et al., 2022; Liu et al., 2022). The large variety of ML algorithms, landslide types (e.g., translational or rotational), and inventory characteristics considered by researchers are typically specific to an area of interest.

Due to a lack of a standardized workflow or benchmark metrics for model performance (Merghadi et al., 2020), the workflows used to develop regional models have limited comparability (Carrara et al., 1992; Lombardo and Mai, 2018; Hearn and Hart, 2019; Alvioli et al., 2024; Crawford et al., 2025). National and continental landslide susceptibility models (Lima et al., 2017; Broeckx et al., 2018; Mirus et al., 2024) offer useful comparisons across regions.

Corresponding author: Hudson J. Koch; Email: hkoch@uky.edu

Cite this article: Koch, H.J., Dortch, J.M., Harrison, B., Crawford, M.M., 2026. Sensitivity analysis workflow to address landslide susceptibility model variability. *Quaternary Research*, 1–19. https://doi.org/10.1017/qua.2026.10101

© The Author(s), 2026. Published by Cambridge University Press on behalf of Quaternary Research Center. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

However, these models still rely on finer scale, region-specific studies for data acquisition and more detailed analysis. Further, the results of large-scale models can still vary due to local data and modeling approaches, resulting in discontinuities between scales and regions. Moreover, studies (at any scale) of methods designed to improve model results typically focus on accuracy metrics, such as receiver operating characteristic curve (ROC) and area under the curve (AUC) values (Fawcett, 2006), and seldom report the practical feasibility of the susceptibility model results.

Early landslide susceptibility workflows leveraged landslide inventories to inform future landslide potential (Neuland, 1976; Taylor and Brabb, 1986), and this trend of data usage has not waned (Reichenbach et al., 2018). As data-driven landslide susceptibility modeling has become more commonplace, ML algorithms have been applied to improve workflows (Aleotti and Chowdhury, 1999; Lima et al., 2022). Much of the research on landslide susceptibility modeling, ML or otherwise, focuses on feature selection (Guzzetti et al., 2006; Jebur et al., 2014; Lima et al., 2017; Lombardo and Mai, 2018; Crawford et al., 2025). As an example, Goetz et al. (2015) tested the effects of input features, such as depth to bedrock and topographic wetness index (TWI), and their influence on model accuracy and performance. Often, studies that iteratively test different features and their subsequent effects also compare ML algorithms (Brenning, 2005; Jebur et al., 2014; Goetz et al., 2015; Pham et al., 2016; Achour and Pourghasemi, 2020; Boussouf et al., 2023; Bostan, 2024; Chowdhury et al., 2024). For example, Knevels et al. (2021) used land cover changes over time (among other features) as the basis for different generalized additive models. Some work assesses feature resolutions, weighing generalizations such as slope units (Carrara et al., 1992; Lima et al., 2021) or cell size resolutions (Gaidzik and Ramírez-Herrera, 2021; Bounab et al., 2022). Other assessments have considered the effects of landslide inventory size and parameterization (e.g., use of point or polygon data, choice of buffer size, pre-failure topography using event-based inventories, and the accuracy of landslide/non-landslide locations) on model results (Brenning, 2005; Hussin et al., 2015; Palamakumbure et al., 2015; Steger et al., 2017; Gaidzik and Ramírez-Herrera, 2021; Lima et al., 2021; Sahrane et al., 2023; Woodard et al., 2023; Crawford et al., 2025; Swallom et al., 2025). Lima et al. (2021) considered the effect of imprecise positional accuracy in inventory data and noted incoherent (i.e., implausible based on slope angle) results even with good model performance. Similarly, Nefeslioglu et al. (2008) showed variation and inaccuracies in model results despite high performance metrics and determined that these variations were related to the selection sampling of the landslide inventory.

Investigations into the importance of training datasets have not been limited to landslide susceptibility approaches. For instance, Hjort and Marmion (2008) considered input data of varying sizes across multiple geomorphological modeling approaches and showed that a minimum size requirement for the input data could be determined and that larger datasets could lead to overfitting and diminished model transferability. Similar concerns have arisen in other areas of scientific inquiry. For example, species distribution modeling assessments have also grappled with comparisons of algorithms and model specifics (Elith et al., 2006; Valavi et al., 2022) including specific issues of input data such as selecting negative sampling (Renner et al., 2015) or better preparation of contextual data for use in models (Gomes et al., 2018). Valavi et al. (2022) emphasized the importance of algorithm and hyperparameter selection that is fitting for the data and goals of the modeling efforts, as well as the relationship between positive and negative sample balance and spatial dependence. Some landslide susceptibility studies have also noted that training and validation data variations within the same datasets have yielded a wide range of map results and accuracies (Wang et al., 2013; Heckmann et al., 2014; Steger et al., 2016; Tsangaratos and Ilia, 2016; Mirus et al., 2024). Bounab et al. (2022) emphasized the importance of input data, noting the significant influence of data variation on model results.

Guns and Vanacker (2012) noted through Monte Carlo simulation of landslide inventory data that important features varied model to model and that results could not be ruled out as atypical outliers without this thorough assessment. Heckmann et al. (2014) performed a similar Monte Carlo analysis of non-initiation locations in debris flow modeling and noted that 1000 iterations were needed to account for the variability of results. Further, Broeckx et al. (2018) used a Monte Carlo approach to ensure optimal inventory selection. This referenced work represents the paucity of studies investigating how model parameterizations affect susceptibility map realism and accuracy. While these researchers have demonstrated potential issues with landslide data in modeling, many susceptibility assessments have been carried out without a substantive analysis of input inventory data, and understanding the effects of input data is a recognized area for improvement (Lima et al., 2022; Woodard et al., 2023).

Landslide susceptibility modeling typically relies on a number of common ML algorithms, specifically random forest ensembles such as bagged trees (BT; Breiman, 2001; Altman and Kryzwinski, 2017), as well as logistic regression (LR; Dobson and Barnett, 2008), Naïve Bayes (NB; Hastie et al., 2009), and support vector machine (SVM; Christianini and Shawe-Taylor, 2000; Huang and Zhao, 2018). For our study, we used a Monte Carlo implementation in an iterative workflow to quantify the effect of input inventory data, and limit the tested algorithms to BT, LR, NB, and SVM because they are the most commonly used in landslide-based research, ultimately selecting one for use in the proposed workflow.

This study assessed the cause-and-effect relationship between a general landslide inventory data and the selection and size of these datasets for use in data-driven susceptibility models. We hypothesized that these input data considerations alone have a considerable effect on model performance, and the resulting map accuracy and appearance. If true, the ability of models to be applicable in other regions or even to perform consistently within a single area of interest (i.e., transferability) cannot be determined. As such, it cannot be determined if performance improvements from altering workflows (e.g., adding feature selection to logistic regression) are statistically significant. Considering these issues, we developed a sensitivity workflow, beginning with an evaluation of inventory data variation to establish a baseline for model variability (i.e., the uncertainty associated with a susceptibility model AUC result). With this baseline, we established a quantifiable metric to assess whether adding features makes a meaningful difference to model performance and the workflow broadly.

This study uses a number of terms all relating to modeling steps, for clarity we specifically define: 'algorithm' for any type of ML model such as a logistic regression or Naïve Bayes classifier; 'feature' for any variation of a map layer such as minimum slope angle or average elevation used in the modeling; 'workflow' for the complete Monte Carlo susceptibility modeling process we propose in this work; 'simulation' or iteration for a single step of our workflow; and 'map output' for the susceptibility model outputs from the iterations of the workflow.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

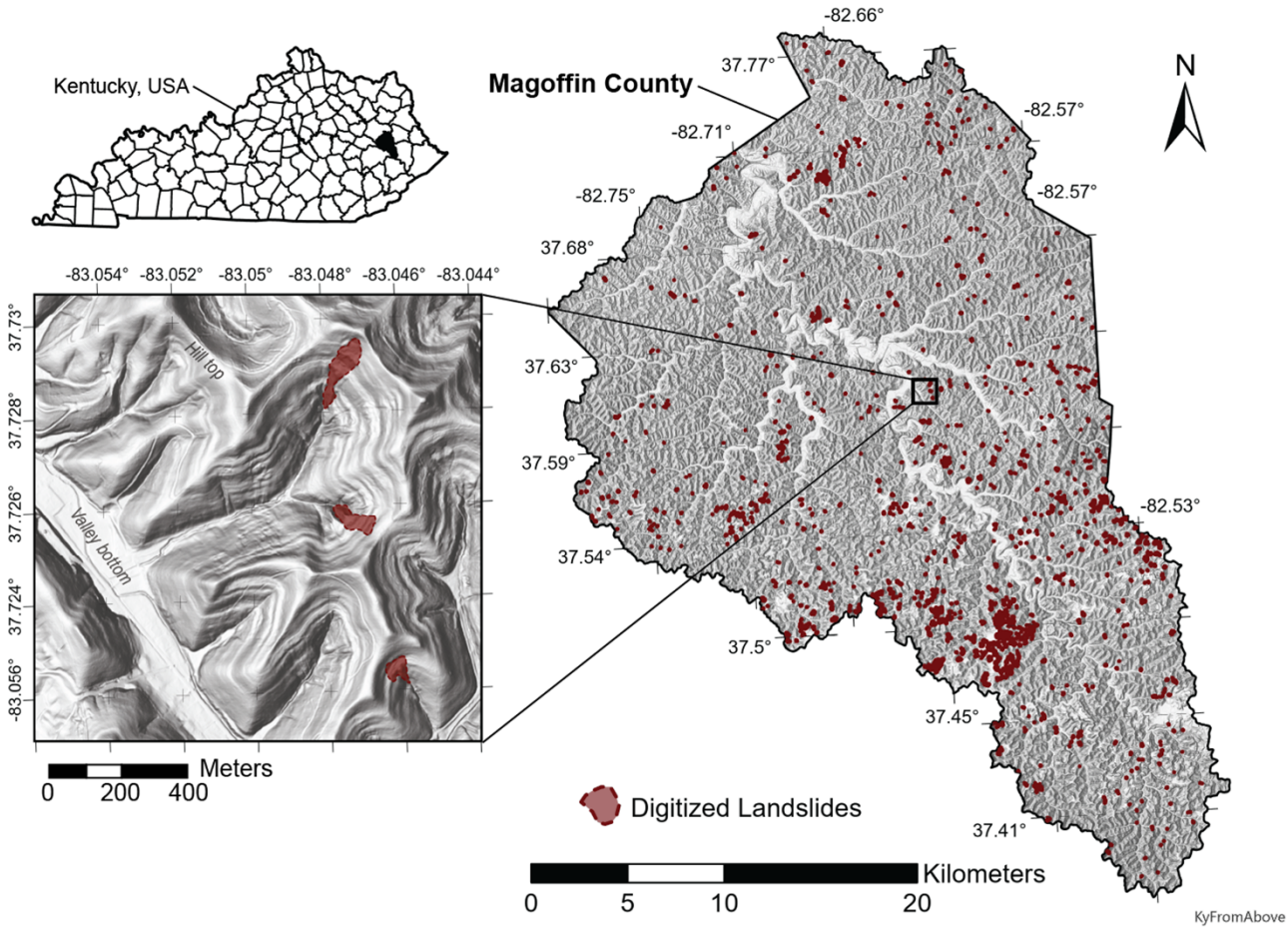


Figure 1. The study area of Magoffin County, Kentucky, USA, with corresponding landslide inventory polygons (n = 1054) overlying a 1.5-meter multidirectional hillshade. The left inset location is also used in Figures 2, 9, 10, 11, and 13. Dashed lines show latitude and longitude values in decimal degrees.

Study area and landslide inventory data

The landslide inventory from Magoffin County, Kentucky, USA (n = 1054; Crawford et al., 2021, 2024) was used to develop the methodology in this study (Fig. 1). Magoffin County lies near the edge of the Appalachian Plateau, a region characterized by topography varying from sharply dissected, sinuous valleys to rolling hills and terraced fluvial drainages. Lithology comprises nearly flat-lying, undifferentiated packages of sedimentary rocks (limestones, sandstones, shales, and underclays; Greb et al., 2009). Although low-magnitude earthquakes occur in the region, landslides in this region are predominantly triggered by rainfall (Crawford, 2014; Crawford et al., 2023) and occur in thin colluvial soils (Crawford et al., 2021). Age constraints for the landslide deposits collected in the inventory are lacking, and documentation for event-based inventory data connected to significant storms is minimal. However, there are many large landslide deposits that fill topographic hollows, which may indicate that those landslides correspond to older mass failure events (Parker et al., 2016; Medwedeff et al., 2020). In general, the landslides in the study area are planar and translational (Crawford et al., 2021).

The Magoffin County landslide inventory was mapped with a methodology modified from Burns and Madin (2009) and used 1.5-m Quality Level 3 LIDAR data (Heidemann, 2018) from 2011.

Methodology

Model contextual data and conditioning features

The LIDAR digital elevation model (DEM) was resampled from 1.5 m to 3 m, then smoothed using a 15-m window size. We used the DEM to develop five geomorphic layers: aspect, curvature, elevation, plan curvature, and slope (Fig. 2), with five statistics (minimum, maximum, mean, standard deviation, and sum) derived for each. With each statistic applied to each geomorphic layer, we obtained 25 total features (e.g., minimum slope, maximum slope, etc.). The data determination and subsequent modeling followed procedures similar to those used by Crawford et al. (2021). Values from the 25 features were extracted for the landslide polygons within the geographic information system (GIS) platform Esri ArcPro [10.3.3] and averaged. Additionally, feature data were extracted for an equal number of non-landslide polygons to create a balanced dataset (n = 2108).

Negative samples of non-landslide points are essential to algorithm training (Woodard et al., 2023), and although there is some debate on the best ratios of positive/negative samples (Oommen et al., 2011), it has been observed that a balanced number is good for model performance (Tanyu et al., 2021; Woodard et al., 2023); thus, we selected 1:1 as our ratio. Non-landslide polygons were created to be equal in area to the mean landslide polygon size, and non-landslide polygon locations were randomly selected (Create

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

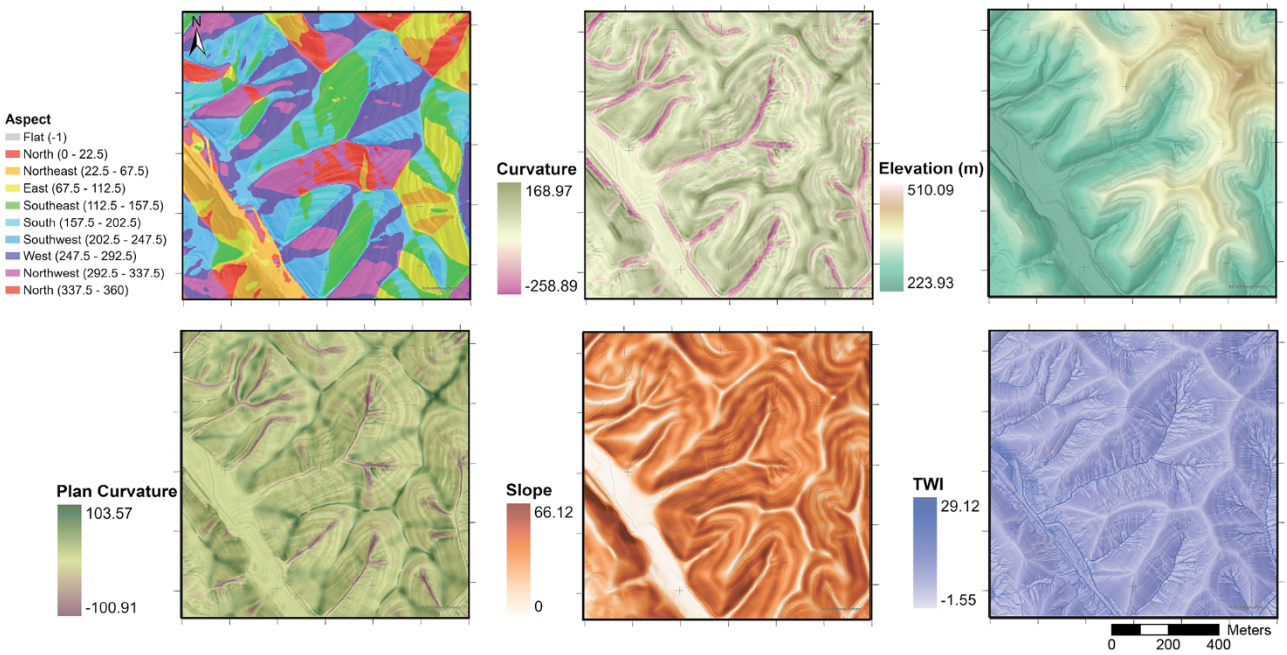


Figure 2. The five geomorphic layers (aspect, curvature, elevation, plan curvature, and slope) used as conditioning factors for the landslide susceptibility models. An additional geomorphic layer, topographic wetness index (TWI), was used in model performance comparisons, having been withheld from the workflow development and instead used to showcase workflow utility. All geomorphic layers were generated from the LIDAR-derived digital elevation model (DEM).

Random Points, Esri ArcPro); care was taken to avoid overlaps with identified landslides. Non-landslide points were reviewed to check for intersection with un-mapped landslides as well. Less than 2% (17 out of 1054) were found to be intersecting, or proximal to, a possible non-mapped landslide.

Motivation of study

To investigate the role that quantity of landslides as a data input has on model performance and map output, we identified inventory sizes (i.e., the number of landslides recorded) from the literature. These inventory sizes ranged from 14 to 600,000 but, for this analysis, were capped at 1054 based on the Magoffin County inventory size. A total of 64 studies had inventory sizes within range and were used to select subsets of the Magoffin County inventory to test the effect of inventory size on model results. By using the subsets of the Magoffin County inventory, we nullified many features that could cause results to differ between research groups and regions, as data were drawn from a single region with flat-lying stratigraphy, and the landslides were digitized using a specific protocol by a single mapper trained within one research group.

The subsets defined from the literature were used in a dual BT-LR ML workflow with the 25 aforementioned statistical geomorphic features, mirroring related work from the region (Crawford et al., 2021, 2025). Model performance was assessed by AUC values (Fawcett, 2006). Despite the recognized limitations of AUC values (Lobo et al., 2007; Goetz et al., 2015), AUC is a popular performance metric in landslide susceptibility assessment (Lima et al., 2022). As such, AUC values were used as a shorthand assessment for inventory size effects (Fig. 3).

Results from this preliminary assessment showed two visually distinct scatter groups within the landslide inventory sizes used in previously published studies, with a clear division seen at ~300 landslides. When ≤300 landslides are used for an inventory, there appears to be a positive trend weakly indicating an improvement to model AUC with more landslides. However, neither linear fit is statistically significant. Likewise, the combined array of landslide size as a single linear regression yielded a very poor fit ($R^2 = 0.001$, $p > 0.05$). Very weak correlations indicate that inventory size is not the sole driver of the variability in model performance. Moreover, the plateauing trend observed in AUC results from models using larger landslide inventories suggests that a greater proportion of variability in model results may be attributable to some characteristic other than inventory size. This initial foray into model sensitivity served as the motivation for a deeper analysis to elucidate how (and by how much) model accuracy varies within a single dataset.

Sensitivity analysis variations

Inventory sample variation

Expanding upon our preliminary investigation of sample-size effects, we sampled the Magoffin County landslide inventory at 50-landslide (100-data entries, including non-landslide polygons) increments ($n = 21$) for use in a Monte Carlo sensitivity assessment. This assessment intended to reproduce the variability of landslide inventory sizes seen in the preliminary assessment (Fig. 3) and used 1000 random iterations per sample size to generate a distribution of susceptibility maps. During workflow design, we experimented with different order of magnitude quantities of the Monte Carlo iterations, finding 100 achieved similar trends but stability was best seen at a minimum 1000 iterations (supporting Heckmann et al., 2014). Using 10,000 iterations did not change distributions significantly, thus was deemed computationally inefficient.

These distributions fully consider the variability possible, with holdout validation data (75/25% training/testing) of every increment being iteration-specific and randomly selected using the Mersenne Twister algorithm (Matsumoto and Nishimura, 1998).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

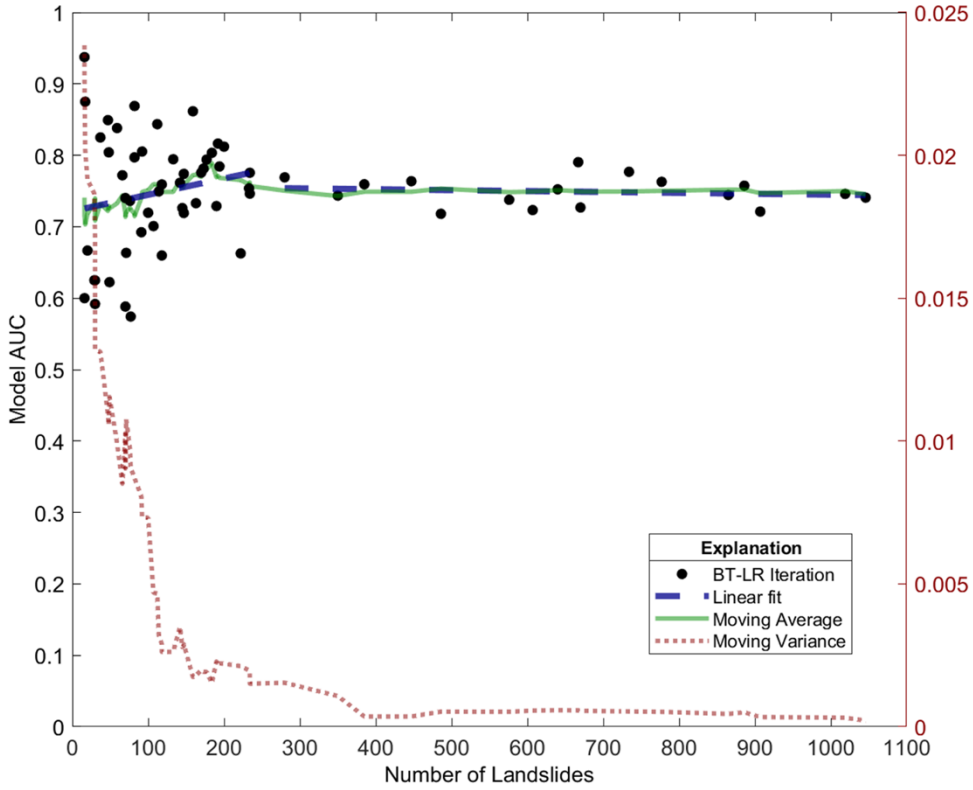


Figure 3. Dual bagged trees (BT) and logistic regression (LR) machine learning model AUC compared to landslide inventory size, with each point ($n = 64$) representing a single BT–LR model iteration using a subset inventory. The data are given linear fits, with division visually based on groups seen in scatter at approximately 300 landslides. Fits of the grouped data do not show statistically significant correlations between model AUC and landslide number. Moving average corroborates the poor linear fits, with the additional insight from the moving variance (right axis) being highest with the lower landslide counts and demonstrating convergence with higher landslide counts.

Due to uneven division of the total number of landslides in the Magoffin County inventory by the 50-landslide increment used to define the sample sizes, a few landslides were withheld from even the maximum sample size, ensuring unique sampling for each iteration. The distribution of model iteration performances was used to highlight the variability inherent in sample holdout for model training and testing.

Model approach variation

We performed Monte Carlo analysis with four common ML algorithms: BT, LR, NB, and SVM. Model performances for each sample size were assessed by their AUC values. Mean AUC values for all 1000 iterations at each increment were compared against the respective sample size. The point of this exercise was to assess variation of these common algorithms for variability to make an informed selection for further sensitivity analysis.

Sensitivity analysis workflow

Feature selection

LR is a commonly used learning technique (Heckmann et al., 2014; Lombardo and Mai, 2018; Reichenbach et al., 2018; Riegel et al., 2020; Lima et al., 2022) and uses a straightforward transfer function to calculate a probability of an event occurring, which is typically interpreted as a susceptibility map. Therefore, LR was selected to create a county-scale landslide susceptibility output map for each iteration.

Feature selection was required before model generation. Specific feature selection methods are common in landslide susceptibility modeling (Pradhan and Lee, 2010; Lombardo and Mai, 2018; Crawford et al., 2021; Chowdhury et al., 2024), where selection refers to a parsing technique on a starting pool of features to minimize negative effects on performance within the ML algorithm. We conducted feature selection by parsing the starting pool of 25 features with a collinearity matrix as a pre-modeling filter. Collinearity was assessed with a Pearson's rank correlation (Eq. 1; Kendall, 1970), and features were iteratively omitted until the group mean of means (μ_{group}) of feature absolute correlation values, ρ, was less than 0.2 and the standard deviation (σ_{group}) of the sum of absolute ρ values of all features was less than 0.2 (Eq. 2 and 3; Fig. 4).

$$\rho(a,b) = \frac{\sum_{i=1}^{n}\left(X_{a,i}-\bar{X}_a\right)\left(Y_{b,i}-\bar{Y}_b\right)}{\sqrt{\left\{\sum_{i=1}^{n}\left(X_{a,i}-\bar{X}_a\right)^2\sum_{j=1}^{n}\left(Y_{b,j}-\bar{Y}_b\right)^2\right\}}} \quad (1)$$

$$\mu_{group} = \frac{\sum(\sum_{i=1}^{n}|\rho_{X_i}|)}{n} \quad (2)$$

$$\sigma_{group} = \sqrt{\frac{\sum_{i=1}^{n}[(\sum_{i=1}^{n}|\rho_{X_i}|)-\mu_{group}]^2}{n}} \quad (3)$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

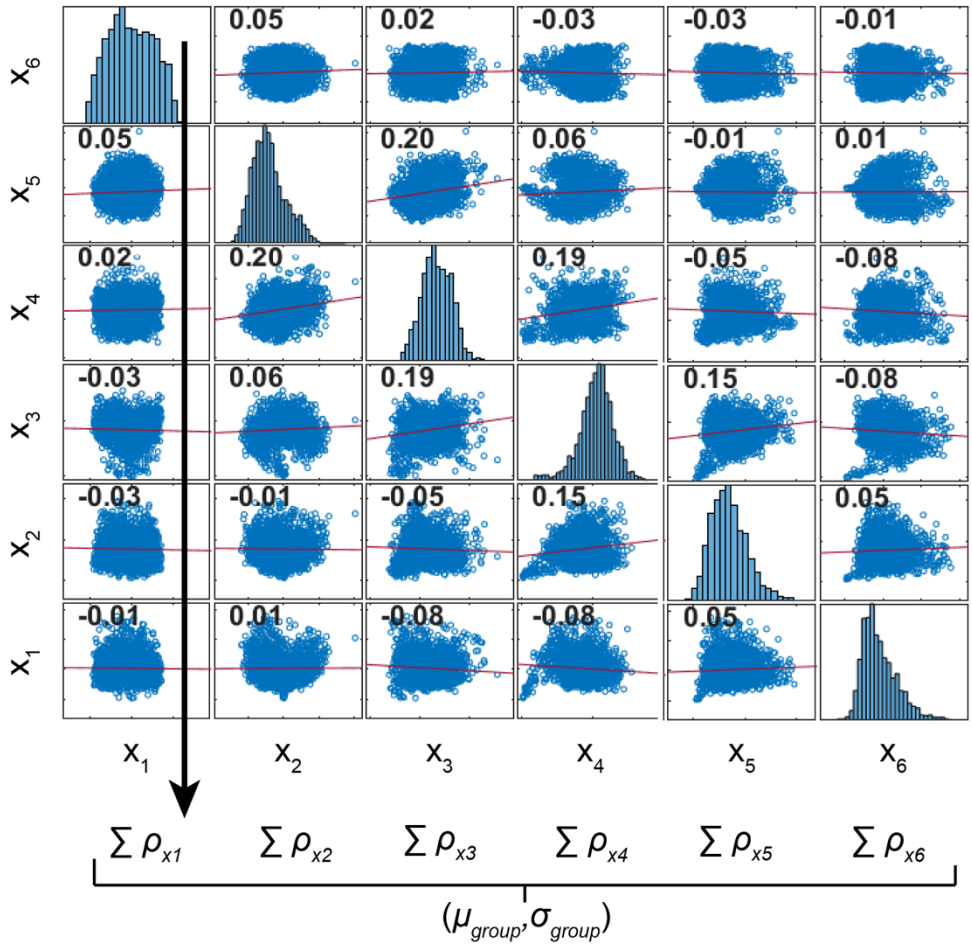


Figure 4. Visual example of the correlation-based feature selection methodology. Each column (feature) of the absolute values in the correlation plot was averaged and summed, with the mean of the full group (μ_{group}) of mean absolute ρ values and the standard deviation of the full group (σ_{group}) of summed absolute ρ values (unitless) compared against set thresholds. If the μ_{group} and σ_{group} were above the set threshold, the largest summed ρ column was omitted and the technique repeated.

The σ_{group} ρ values held more control over the feature selection than the μ_{group} ρ values, and the threshold from σ_{group} was determined for each inventory dataset by a stepwise sensitivity analysis. The σ_{group} ρ thresholds were iteratively assessed at increments of 0.1 from 0.1 to 1.0, and model results were compared to determine the best threshold.

Logistic regression

With features parsed, we used LR to model landslide susceptibility and create maps from a sigmoid transformation (Verhulst, 1838) of the modeled regression (Eq. 4; Crawford et al., 2021).

$$Probability = \frac{1}{1 + e^{[\beta_0+\beta_1 V_1+\beta_2 V_2+\ldots+\beta_x V_x]}} \quad (4)$$

Only features with significant p values (alpha of 0.05) were used in map creation.

Model AUC correction

Due to large variability in expected and observed model results and related AUC performances (Opitz, 2024), an interpretable control was applied to model results outside of the LR. Occasionally, models produce "rare logistic regression events", which are spurious model results (King and Zeng, 2001; Guns and Vanacker, 2012) in which unrealistic and uninterpretable landslide susceptibility maps are generated with high AUC values. Some correction to AUC to limit the effect of these rare logistic regression events is critical for a useful workflow.

Landslides are their own topographic negative feedback as they destroy topography. Topography and slope angle are positively skewed in a landscape (Romans et al., 2016) with more valley bottoms than steep slopes. Further, landslides act to reduce high topography and slope angles reducing the positive-skewed tail on the hypsometric curve (Strahler, 1952; Gallen et al., 2011), achieving normality (Montgomery, 2001; Fig. 5) chiefly through the shortening of the x-axis and secondarily though increased area of lower elevation topography. The normalization trend translates to landslide susceptibility models since topography and slope angle are typically the most significant components of many susceptibility workflows, and susceptibility maps generally are models of the entire landscape.

Although infrequently discussed, landslide susceptibility is inherently tied to timescale. The lack of landslide inventory chronology in most susceptibility studies precludes the identification of the timescale that resulting susceptibility maps are applicable to. Given the unknown landslide chronology (which

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

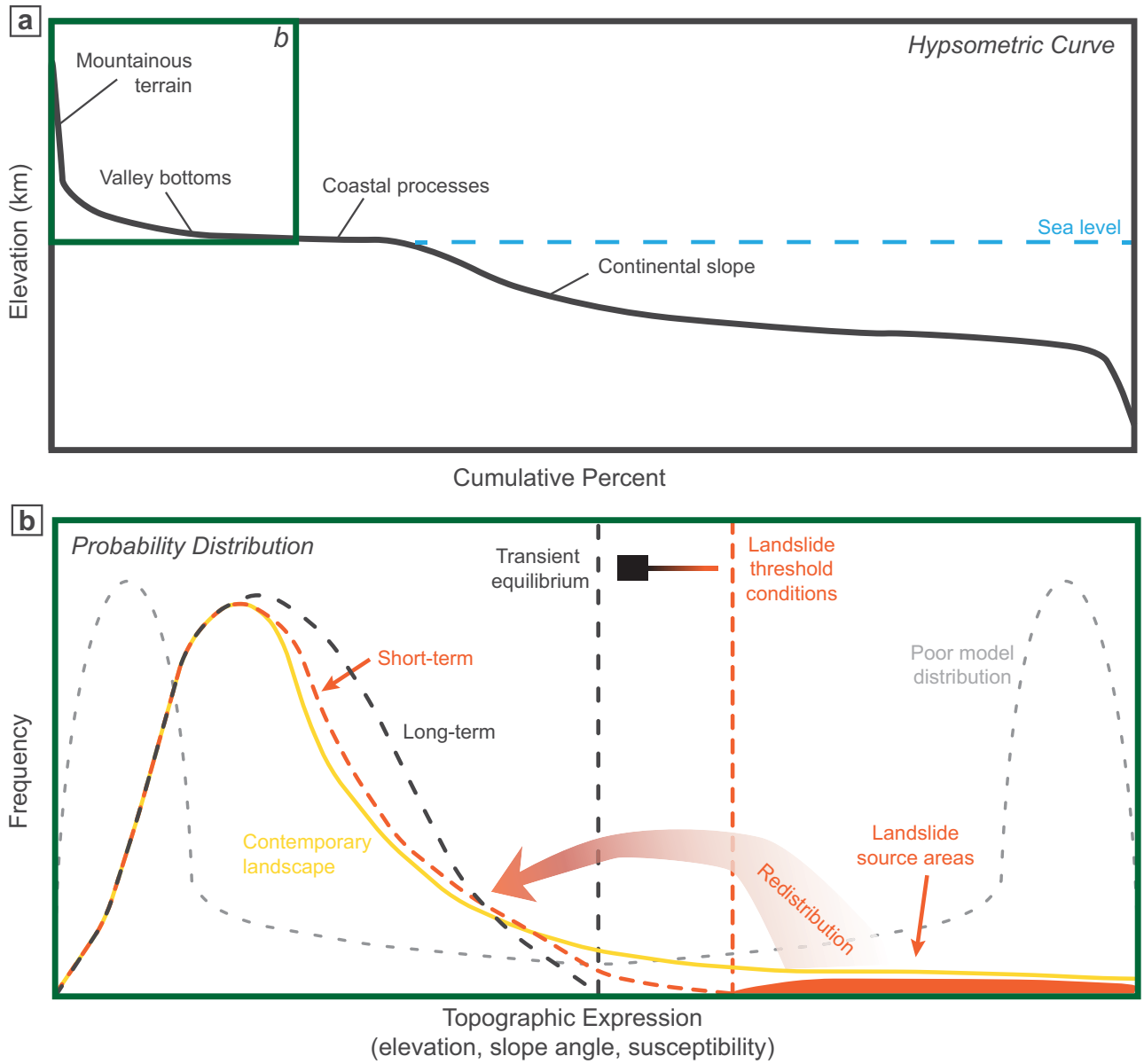


Figure 5. (**a**) Generalized hypsometric curve showing a cumulative distribution of topography. (**b**) Conceptual diagram depicting topographic expression (e.g., elevation, slope angle) for landscapes (positively skewed) as a probability distribution (effectively a rotation of the land region of the hypsometric curve). Depending on the time scale, the x-axes of the (topographic expression) are reduced as the landscape denudes moving skewed distributions to normalcy and landslide source areas into landslides as landslides destroy topography (peaks).

is likely a palimpsest of both long-term background localized events and short-term spatially distributed catastrophic events) with unknown deposit preservation potential, a normal distribution, that in theory is applicable over a medium-term timescale (e.g., Holocene to late Quaternary), would be the most robust assumption to make. As an applied example of this assumption, susceptibility maps with probability distributions that were closer to a normal distribution were considered more useful by Crawford et al. (2025) and, among other characteristics, helped in choosing those models over others that had higher accuracy metrics.

Regardless of the specific distribution chosen, early investigations showed that rare logistic events demonstrated extreme skewed or extreme bimodal distributions. Because it would be impractical to visually inspect tens of thousands of maps, we utilized a correction factor for AUC, where probability distributions (PD) of map outputs were compared to an idealized standard distribution curve $\varphi(z)$, which we characterized as having a mean of 0.5 and a standard deviation of 0.15 to represent the 0–1 probability ranges of the landslide susceptibility models. While output maps are not expected to exactly match a normal distribution, many natural processes follow the central limit theorem (e.g., Koch and Link, 1971; Dortch et al., 2022). We acknowledge that a positively skewed distribution would also effectively filter extreme map distribution events, however it would make little difference in model results and picking a non-arbitrary positively skewed distribution is difficult. The difference between the modeled landslide probabil-

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

Table 1. Thresholds of AUC and corrected AUC (cAUC) for model performance; AUC and correlation thresholds derived from de Hond et al. (2022) and Schober et al. (2018), respectively.

Performance Descriptor	AUC Threshold	Correlation Coefficient Threshold	cAUC Threshold (AUC × Correlation)
Moderate	0.7	0.5	0.35
Strong	0.8	0.7	0.56

ity distribution and idealized normal distribution determined the correction factor value (Eq. 5).

$$Correction\ factor = 1 - \left[\frac{|PD_{Normal} - PD_{Model}|}{PD_{Normal}}\right] \quad (5)$$

To categorize the AUC value distributions, thresholds for strong and moderate performance were assigned based on typical AUC thresholds (de Hond et al., 2022). For corrected AUC (cAUC), standard correlation thresholds (i.e., 0.7, 0.5; Schober et al., 2018) were multiplied against standard AUC thresholds (i.e., 0.8, 0.7) to propagate strong ($0.7 \times 0.8 = 0.56$) and moderate ($0.5 \times 0.7 = 0.35$) cutoffs for cAUC values (Table 1). Classification of the cAUC is one part of the performance assessment; however, we argue that it is also important to visually check some subset of susceptibility maps to guard against algorithm hallucination and ensure realistic results, especially during the threshold and parameter testing phases of workflow development.

Using a parallelization of MATLAB, we performed simulations with the described covariance feature parsing and Monte Carlo simulation, specifically with functions fitmnr, fitcnb, fitcsvm, TreeBagger, and corr. Geomorphic data preparation and resulting output map raster visualization were done in Esri ArcPro [3.4.3].

Comparing with additional features

To test the utility of the workflow for comparative analysis and assess the workflow in the intended use case, we performed the simulation over the full Magoffin inventory to evaluate TWI (Beven and Kirby, 1979; Grabs et al., 2009) and compare against previous full-inventory modeling results (Fig. 2). TWI has been found to be a significant feature for landslide susceptibility models by other feature-assessment studies (Jebur et al., 2014; Goetz et al., 2015; Crawford et al., 2025).

Results

Model approach variations

In the algorithm comparison test, ML algorithms performed similarly overall (Fig. 6). The LR and SVM had nearly identical mean AUC across increments. NB and BT had higher initial AUC but showed little improvement as the number of landslides increased.

Much like the preliminary assessment, the mean AUC values for each ML algorithm appeared to improve with more data, then effectively plateaued after approximately 300 landslides. Plateaus were seen across all approaches, although some models achieved higher AUC values overall. The lack of correlation between AUC and increased inventory size hints at deeper complexities in the modeling approach despite which ML algorithm is chosen.

Feature selection

A majority of the 25 features were preserved by starting with a less stringent group $\rho\ \sigma_{group}$ threshold (1.0). The large number of features led to model instability and rare logistic regression events. Thus, a stricter group $\rho\ \sigma_{group}$ threshold was determined. The effects of ρ were quantified using thresholds ranging from 0.1 to 1 ρ at

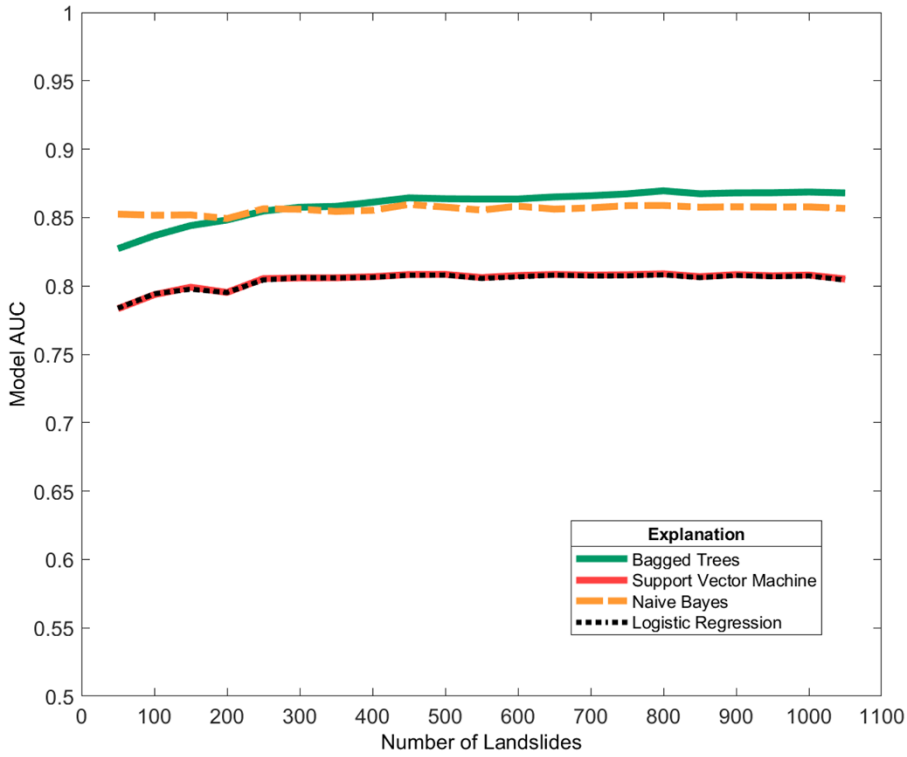


Figure 6. Mean model AUC for 1000 iterations with increasing landslide inventory size in four typical machine-learning approaches. The logistic regression (LR) and support vector machine (SVM) models had nearly overlapping mean AUC. While the mean AUC for the Naïve Bayes (NB) and bagged trees (BT) models were initially higher than those of the LR and SVM models, they likewise did not improve dramatically with the inclusion of more landslides.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

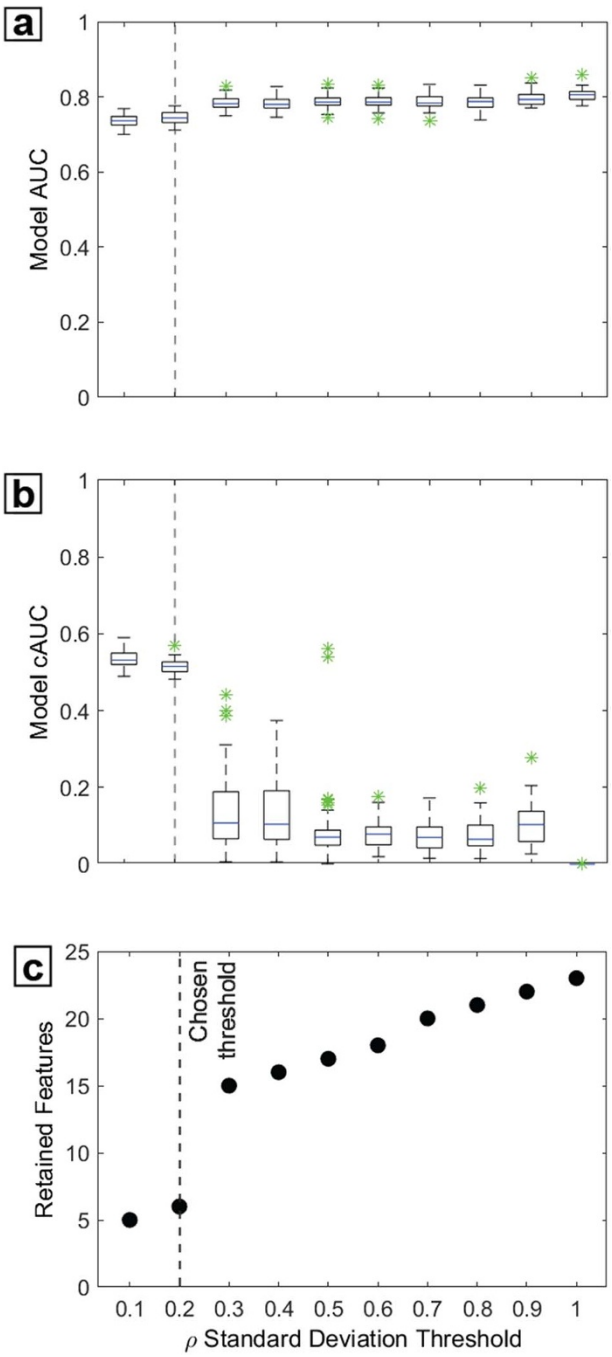


Figure 7. Results of threshold testing for the collinearity feature parsing. (**a**) Boxplots of 100 logistic regression (LR) iterations and their AUC distributions for each group ρ standard deviation threshold; (**b**) the same LR runs with distributions of the cAUC values from Eq. 5.; (**c**) the number of features included at each threshold. The dashed line represents the ρ standard deviation threshold chosen for the subsequent results. Green asterisks indicate outliers of the related boxplots.

0.1 increments on reduced Monte Carlo simulations (100 iterations) with AUC correction. The uncorrected AUC distributions for the model results indicated increased performance with higher (less stringent) thresholds and more features preserved; however, model results using cAUC (i.e., with correction factor) showed the highest and most stable performances at low ρ (more stringent) values (Fig. 7). We chose a threshold of 0.2 because it retained the most features before a precipitous drop in cAUC. Retaining as many layers as possible without significantly degrading performance is critical to capturing model result nuance (e.g., unimodal probability distribution instead of bimodal) and making more useful local susceptibility maps.

Simulation results

The 0.2 ρ σ_{group} threshold omitted 19 of the original 25 statistical geomorphic features but improved the cAUC distributions compared to higher ρ σ_{group} thresholds. The seven retained features were mean aspect, sum curvature, minimum elevation, standard deviation elevation, standard deviation plan curvature, and standard deviation slope. In comparison, Crawford et al. (2021) used the BT-based parsing methodology to yield 12 features for use in the LR, three of which were also retained in this study (sum curvature, standard deviation elevation, and standard deviation plan curvature).

Monte Carlo simulations with limited parsing and higher feature counts had less stable cAUC distributions compared to the reduced feature count simulations. That is to say, boxplots of the simulations with higher feature counts were characterized by more outliers, indicating that very few of the 1000 model iterations yielded useful maps (Fig. 8). Even after stability was reached through stricter feature parsing, at the maximum sample size, the cAUC values effectively ranged from 0.5 to 0.6 (~10%; Fig. 9).

Logistic regression map results

From the Monte Carlo simulations, the maximum and mean iterations for each sample size were calculated as county-wide susceptibility maps for visual inspection. Select sample sizes (50, 100, 500, and 1050 landslides) are shown in Figures 10 and 11. Figure 10 specifically highlights the map results (maximum AUC and maximum cAUC) from a minimally parsed feature pool, that of a ρ threshold of 1.0 (Fig. 8). While largely spurious and incomprehensible, output map results from the less parsed simulations are presented to emphasize the effectiveness of the correction and parsing methods. Figure 11 focuses on the parsed model (Fig. 9) and compares maximum AUC and maximum and mean cAUC values. With Figure 11 we show that with feature parsing and use of the correction, the mean cAUC and maximum cAUC were visually similar. While the distributions of the AUC values for the Monte Carlo iterations in the less-parsed models had mean values indicating strong model performances, Figure 10 details the lack of consistent or interpretable model results. Applying a correction factor for normality improved results but was overall ineffective with so many retained features, as the whole of the AUC/cAUC distribution produced incoherent model results despite good model performance. Figure 11 shows that with strict parsing, output maps were more visually realistic for both maximum uncorrected and corrected map results. Further, these maps were stable (largely unchanged and visually similar) for 500 or more landslides.

Figure 11 also denotes the stability of results across the distribution of cAUC values, indicating that with correction and stricter feature parsing using the lower ρ threshold of 0.2, map variability was minimized across iterations of the larger sample sizes in the Monte Carlo simulation. In sum, stricter parsing led to more con-

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

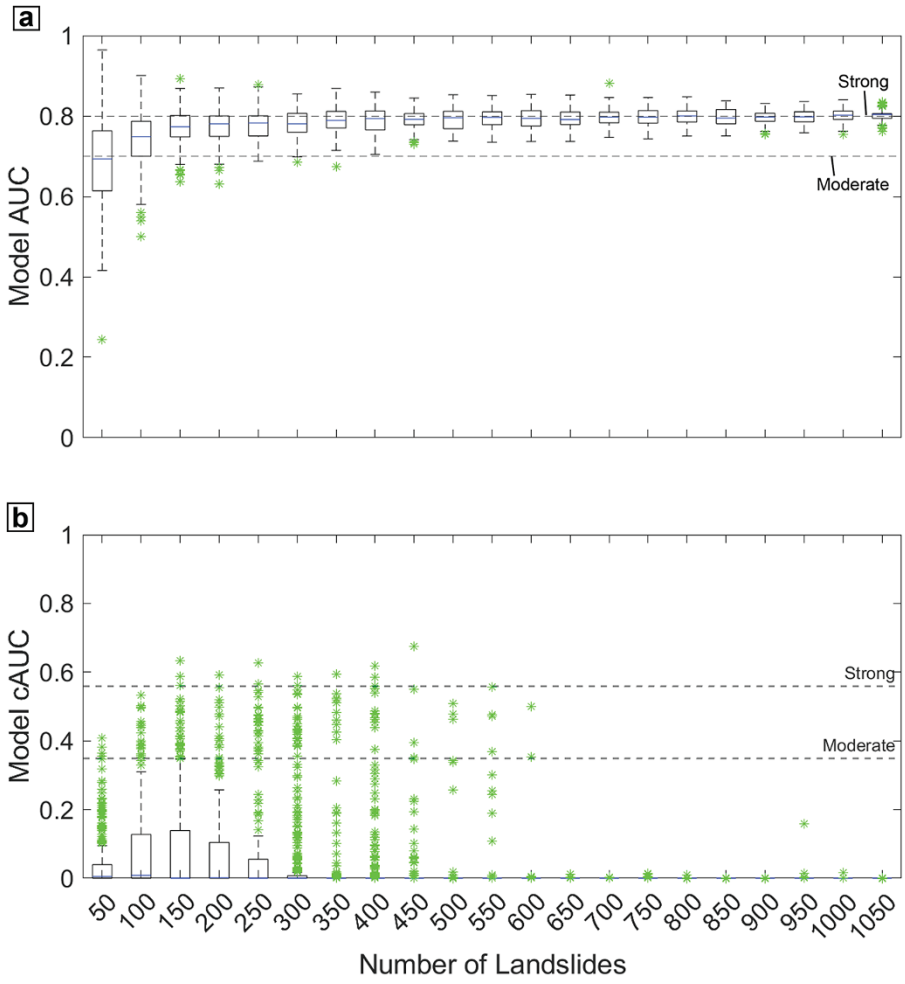


Figure 8. Boxplots for 1000-iteration Monte Carlo simulations of landslide inventory sizes tested using logistic regression and parsed with a ρ threshold of 1.0, which retained 23 features. (**a**) Uncorrected AUC distribution of the Monte Carlo simulation; (**b**) corrected AUC (cAUC) distribution. The thresholds for strong and moderate model performance are based on the AUC and cAUC thresholds shown in Table 1. Green asterisks indicate outliers of the related boxplots.

sistent map results and cAUC for landslide inventory sizes of 350 or greater, as shown in Figures 9b and 11.

Comparison with additional features

Using the full inventory (1054 landslides), TWI was added to the feature pool and the full workflow was run. The ρ threshold had to be manually increased to 0.3 because TWI features were otherwise omitted in the LR map making features. The results show that including TWI increased AUC but decreased cAUC compared to results without TWI (Fig. 12). This test validates both the choice of ρ and the necessity to apply corrections to AUC to determine the statistical significance and direction of change. The addition of TWI-based statistical geomorphic features did not yield a significant visual difference in the map results due to symbology. However, distinctions between the maps with and without TWI can be seen (Fig. 13), as many areas of lower probability without TWI increase to the highest probability category with TWI, indicating probability overprediction. The TWI-based workflow output maps show much sharper spatial contrast between areas of lower and higher probability due to the bimodal nature of the distribution. In general, more nuanced gradation between susceptibility classes lends towards better map utility (Crawford et al., 2025). This approximation of expert preference by utilizing more normally distributed map results is correlated with tighter cAUC distributions, which suggest stronger performance.

Discussion

Interpretation of results

Inventory size

The motivating investigation using subsets of the Magoffin inventory (Fig. 3) detailed that the inclusion of more landslides did not make a statistically significant improvement to AUC past 300 landslides. This plateauing effect was also noted to some extent in both the AUC and cAUC value distributions for the larger sample sizes. Plateauing has been observed previously in similar studies of data effects (Heckmann et al., 2014; Hussin et al., 2015) and is not specific to landslide susceptibility modeling approaches (Stockwell and Peterson, 2002; Valavi et al., 2022). We are unable to fully account for the accuracy plateau with an increase in landslide inventory size. Heckmann et al. (2014) showed in their own Monte Carlo simulation that additional landslide data yielded diminishing returns, attributing this to a breakdown in an autocorrelated range. Ultimately, Heckmann et al. (2014) recommended optimizing inventory size for usage in models.

The stabilization of model AUC and cAUC value distributions appears to match a rule-of-thumb estimate for minimum sample size based on the number of model features, as suggested by Green (1991). Green (1991) noted that 50 + 8x, where "x" represents the number of features, corresponds to the number of samples needed. Using this equation, an inventory size of 100 is needed for a model with six features. cAUC value distributions are less constrained at a sample size of 100 landslides (plus 100 non-landslides) than at

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

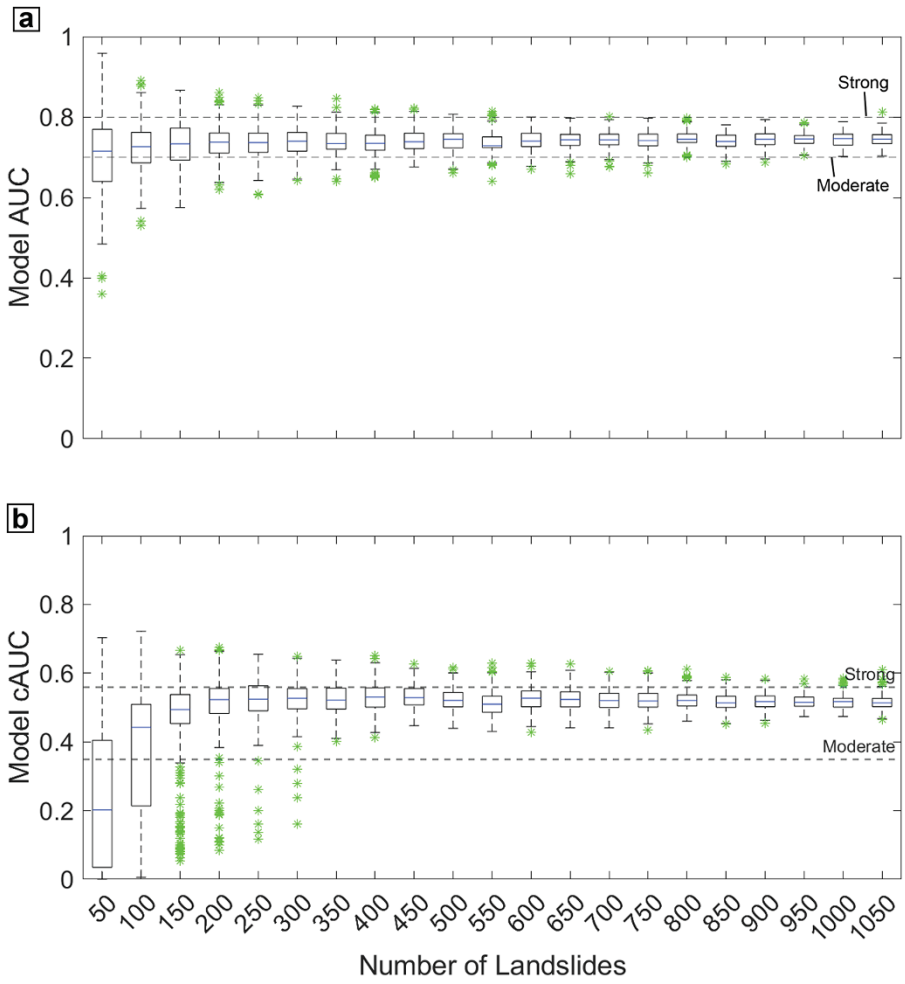


Figure 9. Boxplots for 1000-iteration Monte Carlo simulations of landslide inventory sizes tested using logistic regression and parsed with a ρ threshold of 0.2, which retained six features. (**a**) Standard distribution of the Monte Carlo simulation; (**b**) corrected AUC (cAUC) distribution. While the mean cAUC was initially very low for small landslide inventories, it stabilized rapidly at approximately 0.55 with an inventory size of only 200 landslides. Horizontal dashed lines represent thresholds for strong and moderate model performance based on typical AUC and cAUC divisions. By 350 landslides, model outliers were within the moderate performance zone. Green asterisks indicate outliers of the related boxplots.

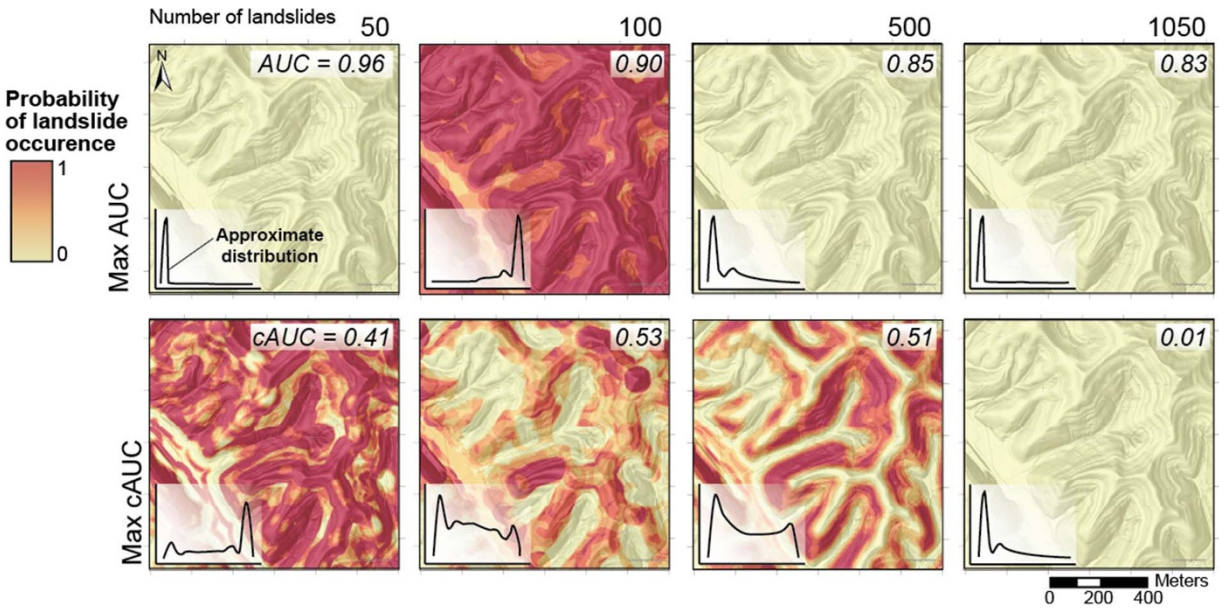


Figure 10. Landslide susceptibility maps for a small region in Magoffin County (location shown in Figure 1 inset). The maps display logistic regression probabilities from the unparsed Monte Carlo simulation iterations with the maximum AUC or maximum corrected AUC (cAUC) values for four select inventory sizes. AUC and cAUC values are shown in the upper right-hand corners of each map; probability distributions of county-level susceptibility are denoted by line graphs in the lower left-hand corners.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

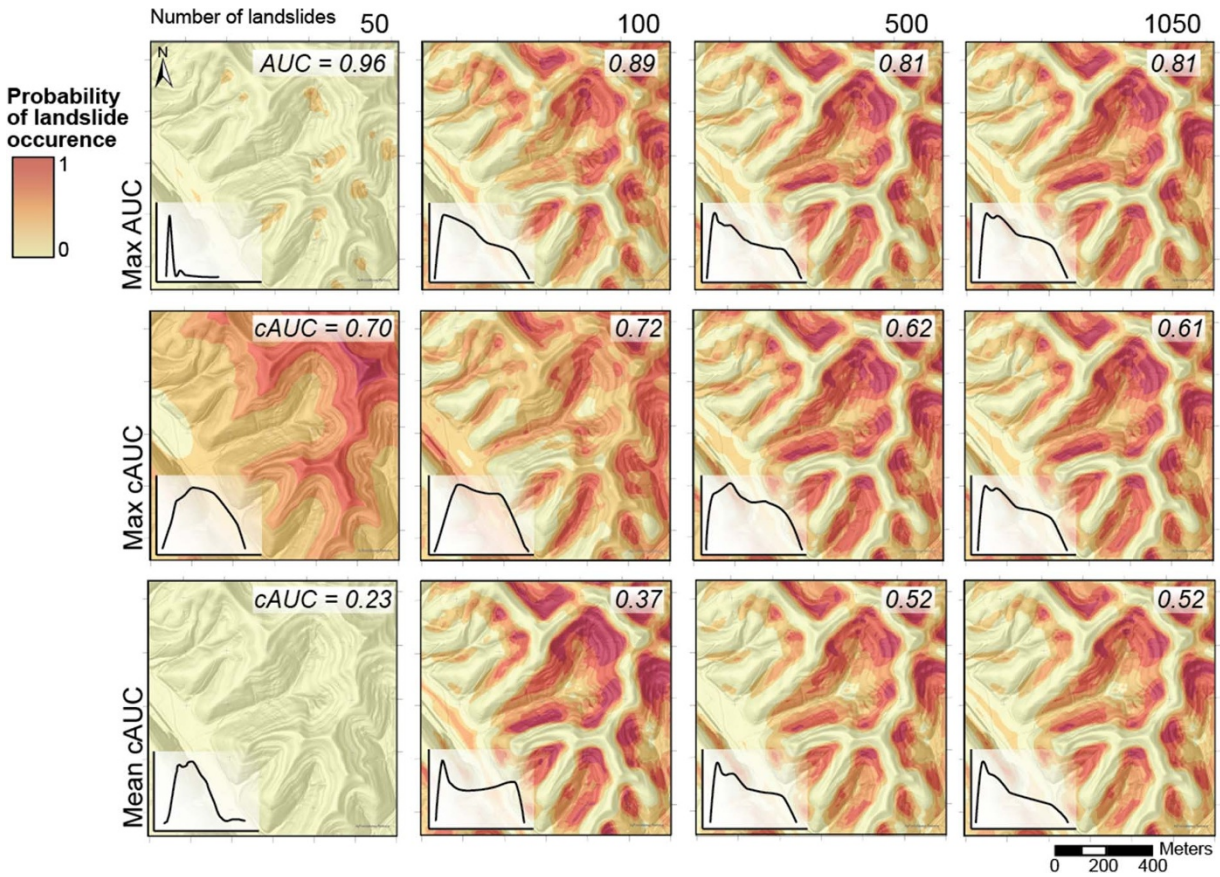


Figure 11. Landslide susceptibility maps for a small region in Magoffin County (location shown in Figure 1 inset). The maps display logistic regression probabilities from the strictly parsed (six retained features) Monte Carlo simulation iterations with the maximum AUC, maximum corrected AUC (cAUC), and mean cAUC values for four select inventory sizes. AUC and cAUC values are shown in the upper right-hand corners of each map; probability distributions of county-level susceptibility are denoted by line graphs in the lower left-hand corners. By 500 landslides, models are visually consistent with the full inventory for both AUC and cAUC.

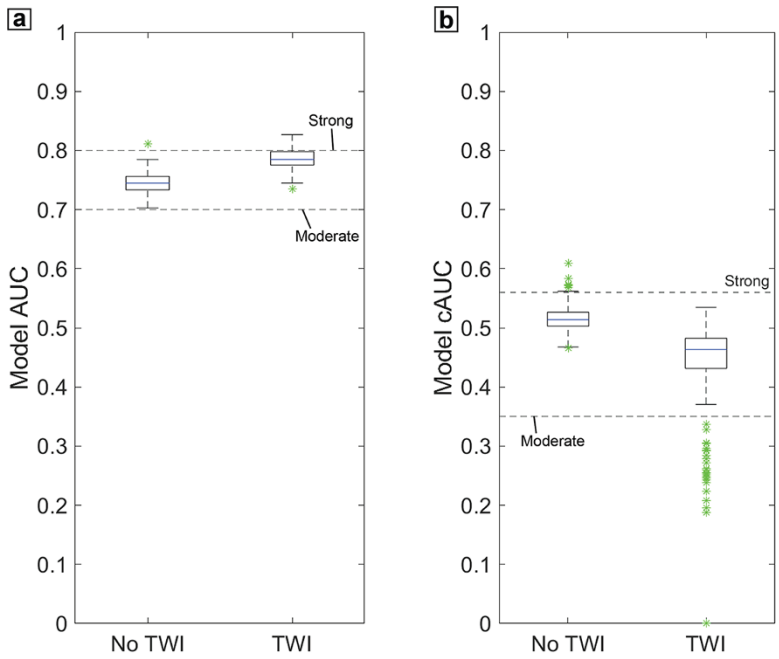


Figure 12. Boxplots for 1000-iteration Monte Carlo simulations for the full inventory (1054 landslides) in a logistic regression with (**a**) an uncorrected AUC value distribution comparing models with and without topographic wetness index (TWI), and (**b**) a corrected AUC (cAUC) value distribution of the same TWI comparison. The lower cAUC value outliers are likely a result of too many features present. Thresholds for strong and moderate model performance were based on the AUC and cAUC thresholds shown in Table 1. Green asterisks indicate outliers of the related boxplots.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

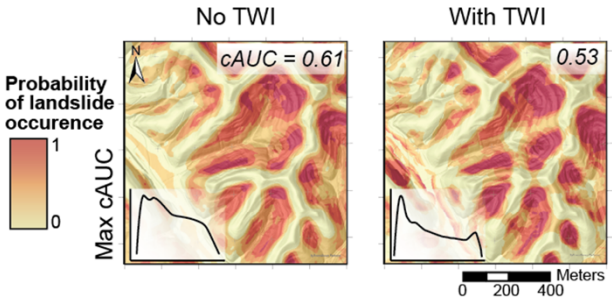


Figure 13. Landslide susceptibility maps for a small region in Magoffin County (location shown in Figure 1 inset). The maps display logistic regression probabilities from the best performing (maximum corrected AUC [cAUC]) Monte Carlo simulation iterations using feature pools with and without topographic wetness index (TWI). The cAUC values are shown in the upper right-hand corners of each map; probability distributions of county-level are denoted by line graphs in lower left-hand corners. When including TWI, there is in an increase in the total area of high probability; thus, the contrast between higher and lower landslide probabilities is much sharper.

larger sample sizes. Model performance began to fluctuate around a maximum cAUC with visually good (i.e., interpretable and close to normally distributed) maps starting around 500 landslides. With the correction and full feature parsing, a good map could be chosen with a minimum sample size as small as 100 landslides with our Monte Carlo approach; however, variance generally decreased as landslide inventory size increased, reaching a minimum variance of 10% with the full dataset. Mean AUC and cAUC both stabilized at approximately 350 landslides similar to the preliminary investigation (Fig. 3).

Box plot variability in Figures 8 and 9 clearly demonstrates that very small landslide inventory sizes do not produce stable map results, although the reduced variability of the AUC values in the increments $\geq$ 200 landslides does allude to some additional information provided by more landslide inventory data in our study area. As expected, 50 landslides are unable to produce a stable increment and subsequently useful output maps, even with correction. The minimum number of landslides needed to reach stability will likely vary between regions based on incision, steepness, sediment thickness, vegetation, climate, etc. Thus, we suggest that using an inventory that is as complete possible over varied terrain within the area of interest would be expected to yield more robust model results. Additionally, Figure 7 demonstrates that model AUC variability generally increased when more than six features were included. As models and features become more robust (i.e., using datasets not derived from a DEM that likely will pass the ρ threshold), the number of landslides required in an inventory may significantly increase before stabilizing. Ultimately, some variability in model results appears inherent in data sampling regardless of inventory size, thus there is a need to diversify the types of features used in the models or the quality of data (but not necessarily the number of examples) in the inventory.

Utility of feature parsing

Isolating the effects of ρ as a threshold for feature inclusion is an important step in LR analysis, demonstrated by Figure 6. It would be impractical to use too many features, the exact number seemingly variable, as highlighted by the probability distributions and maps of the unparsed Monte Carlo assessments. Our extreme case using 25 statistical geomorphic features showcases the utility of our workflow, as many researchers leverage some form of parsing technique. This level of feature parsing may be less important with other ML algorithms, such as various decision tree algorithms, but those approaches are less common in the landslide literature, and it is difficult to produce susceptibility maps from the results. While collinearity checking was effective when using DEM derivatives, we highlight the possibility that many more features may be retained if disparate features (i.e., not all DEM derivatives, such as soil thickness) are utilized. However, this may not be detrimental because disparate features may each be useful for fitting. If model results become overfit with the application of disparate features, other feature parsing techniques would need to be explored. DEM derivatives will likely remain important features due to the significance of high-quality DEMs and ease of generation in landslide susceptibility modeling (Brock et al., 2020). Despite the potential inflexibility of the collinearity parser, the presented feature parsing method still has utility as a parsing method for DEM derivatives.

Utility of correction factor

The boxplots of cAUC results demonstrate that the correction factor effectively isolated "rare logistic regression events," which are quantitively demonstrated to be relatively common when many input features are selected. These spurious map results (i.e. rare logistic regression events) (King and Zeng, 2001) are likely not a result of imbalanced datasets. The datasets used in this study were specifically balanced, but when using the full suite of 25 statistical geomorphic features, the ML approaches may have overpredicted susceptibility. This corroborates observations made by other landslide researchers (Steger et al., 2016; Lin et al., 2021). Rare logistic regression events largely disappeared when using models with reduced features. The corrected ρ threshold boxplots were decisive in determining where model stability was the highest. cAUC values did not follow typical benchmarks for performance—AUC values less than 0.7 indicate poor performance and less than 0.50 indicate random chance. We suggest cAUC value thresholds based on accepted AUC thresholds multiplied by a minimum correlation (0.8); thus, a cAUC of 0.56 (0.8×0.7) would provide a serviceable baseline for "strong" susceptibility results (Table 1). Both the mean and maximum cAUC stabilize above this threshold for quality map results.

Assessment of methodology

Feature selection

Using the same data as this study, Crawford et al. (2021) used a BT algorithm based parsing technique that reduced their LR features down to 12, with only eight used in map making. This produced similar map results, but the workflow requires use of an additional ML approach, in which the process of deciding which features are parsed can be uninterpretable and subject to data subsetting variability. In contrast, the iterative collinearity check and feature removal based on quantitative assessment of the behavior of ρ used here led to quality results while maintaining interpretability. While BT may be more robust and is preferred by some researchers (Merghadi et al., 2020), this approach is still vulnerable to lack of interpretability due to its internal "black box" nature. Relatedly, Tsangaratos and Ilia (2016) showed that NB achieved more stable results with less data, which is corroborated by this study (Fig. 6).

Despite these apparent improvements on LR and SVM, the BT and NB approaches experienced similar AUC plateaus as sample size increased. This workflow was calibrated using a limited pool of DEM-derived features. Susceptibility studies often focus

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

on DEMs and their derivatives for generating model input data (Reichenbach et al., 2018), and Maxwell et al. (2021) indicated that accurate probabilistic slope failure maps did not need non-terrain-based features. Inclusion of non-DEM specific datasets, such as root strength index, has been noted to improve results by as much as 8% (Swallom et al., 2025). However, these model improvements (like many others in the literature) may need to be reassessed using a workflow similar to that proposed here. With potential tweaking to the ρ threshold, features such as quantitative proxies of soil, lithology, and landcover can be included and tested for model improvement, in a similar fashion to TWI addition of this study (Figs. 12 and 13).

Correction factor

Our study demonstrates inherent data variability previously observed in the literature (Heckmann et al., 2014) and presents a consistent method to select the most reasonable map. AUC is not the sole indicator of a good model (Lobo et al., 2007; Goetz et al., 2015) and focusing on AUC can be detrimental to the production of realistic susceptibility maps. Expert opinion could be an alternative to AUC, but to identify the most realistic and accurate model among thousands, these subjective opinions would need to be objectively quantified. Crawford et al. (2025) showed that a more normally distributed map surpassed models with higher AUC values. Considering the timespan of these susceptibility maps based on the lack of age constraints, we found a normal distribution to be the best approximation for this inventory. Further, the input landslide inventories, even when derived from high-resolution LIDAR terrain interpretation, are likely not fully encompassing (Lima et al., 2021). As such, when incomplete datasets are used, it is advisable to select a more stable (i.e., normally distributed) susceptibility map because incomplete datasets can yield significant biases (Steger et al., 2016, 2017). Additionally, the normality-based correction factor cuts out the extreme probabilities in favor of non-extreme, more normal maps. Even if selected arbitrarily, the normal-distribution-based correction serves the purpose of a high- and low-probability filter, those end member probabilities that generally yield spurious maps despite high AUC values. AUC is widely recognized, generally interpretable, and calculatable; therefore, using an AUC-based correction factor increases the accessibility of the correction method.

We acknowledge that that a correction factor on model probabilities may cause model results to further deviate from true landslide probabilities as inferred from the inventory data. Because our workflow relies on comparison to an idealized probability distribution for the AUC correction, we started with a normal distribution as the benchmark. Note that despite choosing a normal distribution, the best performing cAUC maps have a minor to moderate positive skew, thus are not excluded (Fig. 11). Different comparative distributions could be explored further, but the literature provides little basis for distribution selection. Jai et al. (2021) found a positively skewed distribution inferred from global non-landslide areas; their observation could mean that a negatively skewed distribution would be a more suitable comparison probability distribution.

Observing that landslides are relatively sparse and unevenly distributed on the landscape (Huang et al., 2024), it may be that true static landslide susceptibility maps should highlight landslides on specific, relatively small portions of hillslopes. Therefore, the expected probability of a full landscape-scale landslide susceptibility distribution should be positively skewed toward the lower landslide probabilities that cover a majority of the landscape. Investigations into the optimal comparison probability distribution and its potential linkage to topography, lithology, average climate, and structure are left for future work.

Moreover, cAUC results may be low if the initial data/inventory are poorly behaved. The cAUC approach simply identifies the 'best' map of the iterations, which does not guarantee that it is good or realistic. Importantly, the baseline workflow presented here enables those future comparisons to be made by providing quantitative assessment of noise. By performing an AUC correction outside of the ML infrastructure, spurious and poor model results can be assessed and omitted in a systematic fashion. Further, application of a model correction factor can effectively identify features for omission (such as seen with our TWI test), leading to more realism and overall model stability. Detachment of the correction factor from the ML algorithm steps also avoids a workflow with a stronger stepwise approach. Stepwise LR, in which features are selected by additive t-testing within multiple LR, are sometimes used for data parsing (Van Den Eeckhaut et al., 2010; Heckmann et al., 2014). Repeat feature selection and further data sampling are not uncommon (Beguería, 2006; Guns and Vanacker, 2012). However, stepwise workflows are often considered by statisticians to be ineffective (Smith, 2018). Thus, we avoided stepwise workflows for feature and ML algorithm selection, instead comparing the full Monte Carlo simulation for each algorithm type.

Sensitivity approach

Many published landslide susceptibility studies utilizing ML are likely unreproducible given the 10% model variance on our full inventory simulation, which is attributable to the selection of train and test data. Reproducibility could be increased through standard reporting of the seed used by random number generators in various software packages. However, a variety of software packages, many without access to seed-setting functionality, are used by the community, and the same seeds and programs would need to be used to achieve the same results. Only robust stochastic or brute force simulations would be able to assess the number of landslides in an inventory to consistently obtain the best models.

Model results centered near the means of the AUC value distributions also produced spurious results when using the full suite of 25 features. Figure 11 shows that the mean susceptibility map with six features is similar to, but less normally distributed than, the maximum cAUC map. This suggests that susceptibility models were roughly stable across the distribution due to workflow filtering components (i.e., ρ thresholds and AUC correction). It is impossible to determine where a single model run, without context, would fit into the larger potential distribution or to claim that any single model run is the most realistic map result. Significant variability is seen in similar models (Maxwell et al., 2021; Crawford et al., 2025), demonstrating that variability in results is not only possible but has potentially been incorrectly attributed to the inclusion of new layers or methods.

The addition of TWI test demonstrates the importance and practical application of the approach outlined in this study. Adding TWI increased AUC by 2%, which is significant even after accounting for noise (variance) due to selection within the landslide inventory. This apparent improvement is confirmed by findings from several studies (Jebur et al., 2014; Goetz et al., 2015; Crawford et al., 2025). However, when comparing cAUC values, including TWI significantly decreased performance of the best model and reduced mean cAUC by 5%. The implications of this result are that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

many commonly tested features may not increase model performance outside of noise, and features that increase AUC may make susceptibility maps look more unrealistic (i.e., lower cAUC).

Future work

Now that a quantitative workflow to establish the 'best' map selection and uncertainty has been identified, many previously modeled ideas on landslide parameter suitability and modeling approaches need to be reassessed. This includes map averaging, hyperparameter specifics, algorithm selection, feature parsing, and negative sampling approaches. For example, it may be useful to explore aggregating a subset of map results into a single map, such as a top quartile range of cAUC for each sample size. Aggregating a subset of results around either the mean or median (Heckmann et al., 2014) would have been useful for AUC results but unnecessary for cAUC results because the correction factor removes extreme results.

Assessing the effects of landslide inventory size and feature selection in isolation potentially ignores a number of noteworthy model optimizations, such as hyperparameter specifications. The wide variety of hyperparameter specifications can increase with model complexity. For the LR model, hyperparameters are comparatively simple. However, hyperparameter variation was not done in this study because the number of models required for full assessment (21,000) would have increased exponentially (e.g., 21,000 × 30). However, it is possible to test models with complicated and significant hyperparameters using the workflow presented in this study to quantify the impact of hyperparameter settings. Algorithm selection considerations were also comparatively limited, we targeted specifically common and relatively simple algorithms to determine if there were differences to support our point that the issues presented are inherent in the data not in the algorithm. Specific modeling methods may be more robust against data variability (e.g., maximum entropy; Gomes et al., 2018; Pandey et al., 2020; Valvai et al., 2022) and targeted modeling frameworks that include hyperparameter considerations would also likely improve this workflow. An objective of this study was to quantify and account for variability from input inventory data. With that variability recognized, algorithm selection, hyperparameter testing, and input data differentiation can now be performed and accurately assessed for further model improvement. These tests do require a full simulation of the Monte Carlo workflow, to ensure that the changes in performance are not specific to a single train/test split.

Feature parsing was not performed on the individual subsamples of landslide data because this would have added complexity to the modeling workflow. Parsing techniques that depend on the data sampled are likely to be subject to the same variability observed in the model results, potentially at a larger scale. These assessments may continue progress towards representative landslide susceptibility models based on input data but may not lend towards modeling stability. Feature parsing based on random subsampling of the inventory will also likely be more sensitive to incomplete and inconsistent landslide inventories. Woodard and Mirus (2025) cited problems with landslide data limiting model performance. While this method attempts to control variability inherent in landslide datasets, further work on more robust characterization and parametrization of landslide inventories can only improve assessments like the one presented in this study.

In relation to landslide location inventory, the selection of non-landslide points can have a significant effect on model variability (Hu et al., 2020; Crawford et al., 2025). Random selection of non-landslide points was used in this work, despite the existence of more robust approaches to non-landslide point selection (e.g., Huang et al., 2020; Khabiri et al., 2023; Zhou et al., 2024; Koch et al., 2025). A weakness of non-landslide point selection is the potential of sampling of a landslide, especially if the landslide has not been mapped. In this study, all non-landslide points were visually accessed, which showed <2% of non-landslide locations were near or overlapped with an existing landslide. However, landslide inventories are often noted to be incomplete (Steger et al., 2016, 2017; Woodard and Mirus, 2025), and in some cases may contain false positives. Until mapping can be more certain, and robust review of landslide inventories are widespread, selection of positive and negative samples likely will contribute a small fraction of errors. Ultimately, non-landslide locations are complex considerations and likely need a full comparison of different selection approaches using the Monte Carlo workflow presented in this study.

Conclusions

Much of landslide susceptibility modeling leverages ML algorithms, often without knowing the extent to which the results are affected by the input data. These input data effects can significantly affect the output models, reducing reproducibility, comparability, transferability, and the efficacy of landslide susceptibility models used by researchers interested in landscape evolution, local emergency managers, and other end users.

Presented in this study is a workflow to quantify and correct output model variability from the effects of input data. We leveraged a Monte Carlo sensitivity analysis on a LR-based susceptibility approach. Outside of the model, a rank correlation-based feature selection method was employed, as well as an AUC correction on the output LR results. This workflow, including the application of a correction factor and feature selection method, promoted stability and helped select realistic map results that may have been overlooked due to lower performance metrics.

This study highlights the difficulties associated with pursuing performance metrics without considering map utility, bringing attention to the importance of producing usable results even if the chosen model does not have the highest AUC. Moreover, a quantitative metric is still necessary to compare results of thousands of models both within and between study areas—our presented cAUC has the potential to fill that role. Other than circumventing AUC skew and yielding practical map results, this sensitivity analysis delineates a benchmark of noise from inventory effects alone. Further alterations to the modeling workflow can be assessed for significance by comparison to this baseline. Models need to be supported to minimize deviation and fully leverage the available data—our approach is a step in that direction.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/qua.2026.10101.

Acknowledgments. Thank you to Bill Haneberg and Junfeng Zhu for critical questions and comments during the conceptualization of this workflow. Thank you to Meredith Swallom for the good suggestions and feedback on clarity. Thank you to Kentucky Geological Survey (KGS) editor Shelby Fulton for thorough and prompt review. Thank you to the anonymous reviewers and Quaternary Research editors for their suggestions and improvements. Thank you to National Science Foundation EPSCoR CLIMBS Project 2 (OIA 2344533) and KGS SP&M Lab for building capacity to run these memory-intensive simulations. This research was also supported by the Center for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

Land Surface Hazards (CLaSH), funded via National Science Foundation Cooperative Agreement EAR-2425607. We thank Stardust Dortch and Keiko Suzuki for feedback and support.

Author contributions. Conceptualization: H.J.K., J.M.D. Methodology: H.J.K., J.M.D., B.H. Software: H.J.K., J.M.D. Data Curation: H.J.K. Formal analysis: H.J.K Visualization: H.J.K Investigation: H.J.K., J.M.D. Validation: B.H. Resources: J.M.D. Supervision: J.M.D. Writing—Original draft: H.J.K., J.M.D. Writing—Review and editing: H.J.K., J.M.D., M.M.C., B.H.

Competing interests. The authors declare that they have no competing interests.

References

Achour, Y., Pourghasemi, H.R., 2020. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers* **11**, 871–883. https://doi.org/10.1016/j.gsf.2019.10.001.

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the Environment* **58**, 21–44. https://doi.org/10.1007/s100640050066.

Altman, N., Krzywinski, M., 2017. Ensemble methods: bagging and random forests. *Nature Methods* **14**, 933–934. https://doi.org/10.1038/nmeth.4438.

Alvioli, M., Loche, M., Jacobs, L., Grohmann, C.H., Abraham, M.T., Gupta, K., Satyam, N., et al., 2024. A benchmark dataset and workflow for landslide susceptibility zonation. *Earth-Science Reviews* **258**, 104927. https://doi.org/10.1016/j.earscirev.2024.104927.

Aristizábal, E., Korup, O., 2025. Linking landslide patterns to transient landscapes in the Northern Colombian Andes. *Journal of Geophysical Research: Earth Surface* **130**, e2024JF008027. https://doi.org/10.1029/2024JF008027.

Beguería, S., 2006. Changes in land cover and shallow landslide activity: a case study in the Spanish Pyrenees. *Geomorphology* **74**, 196–206. https://doi.org/10.1016/j.geomorph.2005.07.018.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* **24**, 43–69. https://doi.org/10.1080/02626667909491834.

Booth, A.M., Roering, J.J., Rempel, A.W., 2013. Topographic signatures and a general transport law for deep-seated landslides in a landscape evolution model. *Journal of Geophysical Research and Earth Surface* **118**, 603–624. https://doi.org/10.1002/jgrf.20051.

Bostan, T., 2024. Generating a landslide susceptibility map using integrated meta-heuristic optimization and machine learning models. *Sustainability* **16**, 9396. https://doi.org/10.3390/su16219396.

Bounab, A., Agharroud, K., El Kharim, Y., El Hamdouni, R., Faghloumi, L., 2022. The importance of investigating causative factors and training data selection for accurate landslide susceptibility assessment: the case of Ain Lahcen commune (Tetouan, Northern Morocco). *Geocarto International* **37**, 9967–9997. https://doi.org/10.1080/10106049.2022.2028905.

Boussouf, S., Fernández, T., Hart, A.B., 2023. Landslide susceptibility mapping using maximum entropy (MaxEnt) and geographically weighted logistic regression (GWLR) models in the Río Aguas catchment (Almería, SE Spain). *Natural Hazards* **117**, 207–235. https://doi.org/10.1007/s11069-023-05857-7.

Breiman, L., 2001. Random forests. *Machine Learning* **45**, 5–32. https://doi.org/10.1023/A:1010933404324.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences* **5**, 853–862. https://doi.org/10.5194/nhess-5-853-2005.

Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M., Brenning, A., 2020. The performance of landslide susceptibility models critically depends on the quality of digital elevation models. *Geomatics, Natural Hazards and Risk* **11**, 1075–1092. https://doi.org/10.1080/19475705.2020.1776403.

Broeckx, J., Vanmaercke, M., Duchateau, R., Poesen, J., 2018. A data-based landslide susceptibility map of Africa. *Earth-Science Reviews* **185**, 102–121. https://doi.org/10.1016/j.earscirev.2018.05.002.

Broeckx, J., Rossi, M., Lijnen, K., Campforts, B., Poesen, J., Vanmaercke, M., 2020. Landslide mobilization rates: a global analysis and model. *Earth-Science Reviews* **201**, 102972. https://doi.org/10.1016/j.earscirev.2019.102972.

Burns, W.J., Madin, I.P., 2009. Protocol for inventory mapping of landslide deposits from light detection and ranging (LIDAR) imagery. *Oregon Department of Geology and Mineral Industries Special Paper* **42**, 1–30. https://pubs.oregon.gov/dogami/dds/slido/sp-42_onscreen.pdf.

Campforts, B., Shobe, C.M., Steer, P., Vanmaercke, M., Lague, D., Braun, J., 2020. Hylands 1.0: A hybrid landscape evolution model to simulate the impact of landslides and landslide-derived sediment on landscape evolution. *Geoscientific Model Development* **13**, 3863–3886. https://doi.org/10.5194/gmd-13-3863-2020.

Campforts, B., Shobe, C.M., Overeem, I., Tucker, G.E., 2022. The art of landslides: how stochastic mass wasting shapes topography and influences landscape dynamics. *Journal of Geophysical Research: Earth Surface* **127**, e2022JF006745. https://doi.org/10.1029/2022JF006745.

Carrara, A., Cardinali, M., Guzzetti, F., 1992. Uncertainty in assessing landslide hazard and risk. *ITC Journal* **2**, 172–183.

Chowdhury, M.S., Rahman, M.N., Sheikh, M.S., Sayeid, M.A., Mahmud, K.H., Hafsa, B., 2024. GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. *Heliyon* e23424. https://doi.org/10.1016/j.heliyon.2023.e23424.

Christianini, N., Shawe-Taylor, J.C., 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.

Crawford, M.M., 2014. Kentucky Geological Survey landslide inventory: from design to application. *Kentucky Geological Survey Series 12 Information Circular* **31**, 1–18. https://doi.org/10.13023/kgs.ic31.12.

Crawford, M.M., Dortch, J.M., Koch, H.J., Killen, A.A., Zhu, J., Zhu, Y., Bryson, L.S., Haneberg, W.C., 2021. Using landslide-inventory mapping for a combined bagged-trees and logistic-regression approach to determining landslide susceptibility in eastern Kentucky, USA. *Quarterly Journal of Engineering Geology and Hydrogeology* **54**, qjegh2020-177. http://doi.org/10.1144/qjegh2020-177.

Crawford, M.M., Zhenming, W., Carpenter, N.S., Schmidt, J., Koch H., Dortch, J., 2023. Reconnaissance of landslides and debris flows associated with the July 2022 flooding in eastern Kentucky. *Kentucky Geological Survey Series 13 Report of Investigations* **13**, 1–14.

Crawford, M.M., Bibbins, E.M., Koch, H.J., 2024. Kentucky Geological Survey landslide inventory [2024-04]. *Kentucky Geological Survey Research Data*. https://doi.org/10.13023/kgs.data.2022.01.

Crawford, M.M., Koch, H.J., Dortch, J.M., 2025. Evaluating map quality and model performance through iterative statistics-based landslide susceptibility in eastern KY. *Natural Hazards* **121**, 11633–11661. https://doi.org/10.1007/s11069-025-07255-7.

de Hond, A.A.H., Steyerberg, E.W., van Calster, B., 2022. Interpreting area under the receiver operating characteristic curve. *The Lancet Digital Health* **4**, e853–e855.

Dobson, A.J., Barnett, A.G., 2008. *An Introduction to Generalized Linear Models*, 3rd ed. CRC Press, Boca Raton.

DiBiase, R.A., Whipple, K.X., Heimsath, A.M., Ouimet, W.B., 2010. Landscape form and millennial erosion rates in the San Gabriel Mountains, CA. *Earth and Planetary Science Letters* **289**, 134–144. https://doi.org/10.1016/j.epsl.2009.10.036.

Dietsch, C., Dortch, J.M., Reynhout, S.A., Owen, L.A., Caffee, M.W., 2015. Very slow erosion rates and landscape preservation across the southwestern slope of the Ladakh Range, India. *Earth Surface Processes and Landforms* **40**, 389–402. https://doi.org/10.1002/esp.3640.

Dortch, J.M., Owen, L.A., Haneberg, W.C., Caffee, M.W., Dietsch, C., Kamp, U., 2009. Nature and timing of large landslides in the Himalaya and Transhimalaya of northern India. *Quaternary Science Reviews* **28**, 1037–1054. https://doi.org/10.1016/j.quascirev.2008.05.002.

Dortch, J.M., Owen, L.A., Schoenbohm, L.M., Caffee, M.W., 2011. Asymmetrical erosion and morphological development of the central Ladakh Range, northern India. *Geomorphology* **135**, 167–180. https://doi.org/10.1016/j.geomorph.2011.08.014.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

Dortch, J.M., Tomkins, M.D., Saha, S., Murari, M.K., Schoenbohm, L.M., Curl, D., 2022. A tool for the ages: the Probabilistic Cosmogenic Age Analysis Tool (P-CAAT). *Quaternary Geochronology* **71**, 101323. https://doi.org/10.1016/j.quageo.2022.101323.

Elith, J., Graham, H.C., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A.J. Hijmans, R.J., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Gallen, S.F., Wegmann, K.W., Frankel, K.L., Hughes, S., Lewis, R.Q., Lyons, N., Paris, P., Ross, K., Bauer, J.B., Witt, A.C., 2011. Hillslope response to knickpoint migration in the Southern Appalachians: implications for the evolution of post-orogenic landscapes. *Earth Surface Processes and Landforms* **36**, 1254–1267. https://doi.org/10.1002/esp.2150.

Gallen, S.F., Wegmann, K.W., Bohnenstieh, D.R., 2013. Miocene rejuvenation of topographic relief in the southern Appalachians. *GSA Today* **23**, 4–10. https://doi.org/10.1130/GSATG163A.1.

Gaidzik, K., Ramírez-Herrera, M.T., 2021. The importance of input data on landslide susceptibility mapping. *Scientific Reports* **11**, 19334. https://doi.org/10.1038/s41598-021-98830-y.

Gilbert, G.K., 1909. The convexity of hilltops. *The Journal of Geology* **17**, 344–350.

Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences* **81**, 1–11. http://dx.doi.org/10.1016/j.cageo.2015.04.007.

Gomes, V.H.F., IJff, S.D., Raes, N., Amaral, I.L., Salamão, R.P., de Souza Coelho, L., de Almeida Matos, F.D., et al., 2018. Species distribution modelling: contrasting presence-only models with plot abundance data. *Scientific Reports* **8**, 1003. https://doi.org/10.1038/s41598-017-18927-1.

Grabs, T., Seibert, J., Bishop, K., Laudon, H., 2009. Modeling spatial patterns of saturated areas: a comparison of the topographic wetness index and a dynamic distributed model. *Journal of Hydrology* **373**, 15–23. https://doi.org/10.1016/j.jhydrol.2009.03.031.

Greb, S.F., Chesnut, D.R., Jr., Eble, C.F., Blake, B.M., 2009. The Pennsylvanian of the Appalachian Basin. In: Greb, S.F., Chesnut, D.R. Jr. (Eds.), Carboniferous Geology and Biostratigraphy of the Appalachian Basin and Black Warrior Basins. *Kentucky Geological Survey Series 12 Special Publication* **10**, 32–45.

Green, S.B., 1991. How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research* **26**, 499–510. https://doi.org/10.1207/s15327906mbr2603_7.

Guns, M., Vanacker, V., 2012. Logistic regression applied to natural hazards: rare event logistic regression with replications. *Natural Hazards and Earth System Sciences* **12**, 1937–1947. https://doi.org/10.5194/nhess-12-1937-2012.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. Estimating the quality of landslide susceptibility models. *Geomorphology* **81**, 166–184. http://dx.doi.org/10.1016/j.geomorph.2006.04.007.

Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York.

Hack, J.T., 1960. Interpretation of erosional topography in humid temperate regions. *American Journal of Science* **253-A**, 80–97.

Hearn, G.J., Hart, A.B., 2019. Landslide susceptibility mapping: a practitioner's view. *Bulletin of Engineering Geology and the Environment* **78**, 5811–5826. https://doi.org/10.1007/s10064-019-01506-1.

Heckmann, T., Gegg, K., Gegg, A., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. *Natural Hazards and Earth System Sciences* **14**, 259–278. https://doi.org/10.5194/nhess-14-259-2014.

Heidemann, H.K., 2018. Lidar base specification. *U.S. Geological Survey Techniques and Methods* **11-B4**, 1–101. https://doi.org/10.3133/tm11B4.

Hjort, J., Marmion, M., 2008. Effects of sample size on the accuracy of geomorphological models. *Geomorphology* **102**, 341–350. https://doi.org/10.1016/j.geomorph.2008.04.006.

Hu, Q., Zhou, Y., Wang, S., Wang, F., 2020. Machine learning and fractal theory models for landslide susceptibility mapping: case study from the Jinsha River Basin. *Geomorphology* **351**, 106975. https://doi.org/10.1016/j.geomorph.2019.106975.

Huang, F., Cao, Z., Jiang, S.H., Zhou, C., Huang, J., Guo, Z., 2020. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. *Landslides* **17**, 2919–2930. https://doi.org/10.1007/s10346-020-01473-9.

Huang, F., Xiong, H., Jiang, S.-H., Yao, C., Fan, X., Catani, F., Chang, Z., Zhou, X., Huang, J., Liu, K., 2024. Modelling landslide susceptibility prediction: a review and construction of semi-supervised imbalanced theory. *Earth-Science Reviews* **250**, 104700. https://doi.org/10.1016/j.earscirev.2024.104700.

Huang, Y., Zhao, L., 2018. Review on landslide susceptibility mapping using support vector machines. *Catena* **165**, 520–529. https://doi.org/10.1016/j.catena.2018.03.003.

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2015. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* **253**, 508–523. https://doi.org/10.1016/j.geomorph.2015.10.030.

Jebur, M.N., Pradhan, B., Tehrany, M.S., 2014. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment* **152**, 150–165. http://dx.doi.org/10.1016/j.rse.2014.05.013.

Jia, G., Alvioli, M., Gariano, S.L., Marchesini, I., Guzzetti, F., Tang, Q., 2021. A global landslide non-susceptibility map. *Geomorphology* **389**, 107804. https://doi.org/10.1016/j.geomorph.2021.107804.

Kendall, M.G., 1970. *Rank Correlation Methods*. Griffin, London.

Khabiri, S., Crawford, M.M., Koch, H.J., Haneberg, W.C., Zhu, Y., 2023. An assessment of negative samples and model structures in landslide susceptibility characterization based on Bayesian network models. *Remote Sensing* **15**, 3200. https://doi.org/10.3390/rs15123200.

King, G., Zeng, L., 2001. Logistic regression in rare events data. *Political Analysis* **9**, 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868.

Knevels, R., Brenning, A., Gingrich, S., Heiss, G., Lechner, T., Leopold, P., Plutzar, C., Proske, H., Petschko, H., 2021. Towards the use of land use legacies in landslide modeling: current challenges and future perspectives in an Austrian case study. *Land* **10**, 954. https://doi.org/10.3390/land10090954.

Koch, G.S., Link, R.F., 1971. *Statistical Analysis of Geological Data*, Vol. 1–2. Dover Publications, Mineola, NY.

Koch, H.J., Haneberg, W.C., Crawford, M.M., Dortch, J.M., Swallom, M.L., 2025. Evaluating the physics-based slope stability program PISA-m at a regional scale. *Kentucky Geological Survey Series 14 Report of Investigations* **1**, 1–22. https://doi.org/10.13023/kgs.14.ri.1.2025.

Larsen, I.J., Montgomery, D.R., 2012. Landslide erosion coupled to tectonics and river incision. *Nature Geoscience* **5**, 468–473. https://doi.org/10.1038/ngeo1479.

Lima, P., Steger, S., Glade, T., Tilch, N., Schwarz, L., Kociu, A., 2017. Landslide susceptibility mapping at national scale: a first attempt for Austria. In: Mikoš, M., Tiwari, B., Yin, Y., Sassa, K. (Eds.), *Advancing Culture of Living with Landslides*, Vol. 2. Springer, Cham, pp. 943–951. https://doi.org/10.1007/978-3-319-53498-5_107.

Lima, P., Steger, S., Glade, T., 2021. Counteracting flawed landslide data in statistically based landslide susceptibility modelling for very large areas: a national-scale assessment for Austria. *Landslides* **18**, 3531–3546. https://doi.org/10.1007/s10346-021-01693-7

Lima, P., Steger, S., Glade, T., Murillo-García, F.G., 2022. Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. *Journal of Mountain Science* **19**, 1670–1698. https://doi.org/10.1007/s11629-021-7254-9.

Lin, Q., Lima, P., Steger, S., Glade, T., Jiang, T., Zhang, J., Liu, T., Wang, Y., 2021. National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. *Geoscience Frontiers* **12**, 101248. https://doi.org/10.1016/j.gsf.2021.101248.

Liu, L.-L., Zhang, J., Li, J.-Z., Huang, F., Wang, L.-C., 2022. A bibliometric analysis of the landslide susceptibility research (1999–2021). *Geocarto International* **37**, 14309–14334. http://doi.org/10.1080/10106049.2022.2087753.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2007. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Lombardo, L., Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results. *Engineering Geology* **244**, 14–24. https://doi.org/10.1016/j.enggeo.2018.07.019.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslides, earthquakes, and erosion. *Earth and Planetary Science Letters* **229**, 45–59. https://doi.org/10.1016/j.epsl.2004.10.018.

Matsumoto, M., Nishimura, T., 1998. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation* **8**, 3–30. https://doi.org/10.1145/272991.272995.

Maxwell, A.E., Sharma, M., Kite, J.S., Donaldson, K.A., Maynard, S.M., Malay, C.M., 2021. Assessing the generalization of machine learning-based slope failure prediction to new geographic extents. *ISPRS International Journal of Geo-Information* **10**, 293. https://doi.org/10.3390/ijgi10050293.

Medwedeff, W.G., Clark, M.K., Zekkos, D., West, A.J., 2020. Characteristic landslide distributions: an investigation of landscape controls on landslide size. *Earth and Planetary Science Letters* **539**, 116203. https://doi.org/10.1016/j.epsl.2020.116203.

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. *Earth-Science Reviews* **207**, 103225. https://doi.org/10.1016/j.earscirev.2020.103225.

Mirus, B.B., Belair, G.M., Wood, N.J., Jones, J., Martinez, S.N., 2024. Parsimonious high-resolution landslide susceptibility modeling at continental scales. *AGU Advances* **5**, e2024AV001214. https://doi.org/10.1029/2024AV001214.

Montgomery, D.R., 2001. Slope distributions, threshold hillslopes, and steady-state topography. *American Journal of Science* **301**, 432–454. https://doi.org/10.2475/ajs.301.4-5.432.

Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology* **97**, 171–191. https://doi.org/10.1016/j.enggeo.2008.01.004.

Neuland, H., 1976. A prediction model of landslips. *Catena* **3**, 215–230. https://doi.org/10.1016/0341-8162(76)90011-4.

Oommen, T., Baise, L.G., Vogel, R.M., 2011. Sampling bias and class imbalance in maximum-likelihood logistic regression. *Mathematical Geosciences* **43**, 99–120. https://doi.org/10.1007/s11004-010-9311-8.

Opitz, J., 2024. Schroedinger's threshold: when the AUC doesn't predict accuracy. In: Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., Xue, N. (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. ELRA and ICCL, Torino, pp. 14400–14406.

Palamakumbure, D., Flentje, P., Stirling, D., 2015. Consideration of optimal pixel resolution in deriving landslide susceptibility zoning within the Sydney Basin, New South Wales, Australia. *Computers & Geosciences* **82**, 13–22. http://dx.doi.org/10.1016/j.cageo.2015.05.002.

Pandey, V.K., Pourghasemi, H.R., Sharma, M.C., 2020. Landslide susceptibility mapping using maximum entropy and support vector machine models along the Highway Corridor, Garhwal Himalaya. *Geocarto International* **35**, 168–187. https://doi.org/10.1080/10106049.2018.1510038

Parker, R.N., Hales, T.C., Mudd, S.M., Grieve, S.W.D., Constantine, J.A., 2016. Colluvium supply in humid regions limits the frequency of storm-triggered landslides. *Scientific Reports* **6**, 34438. https://doi.org/10.1038/srep34438.

Pham, B.T., Tien Bui, D., Dholakia, M.B., Prakash, I., Pham, H.V., 2016. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotechnical and Geological Engineering* **34**, 1807–1824. https://doi.org/10.1007/s10706-016-9990-0.

Pradhan, B., Lee, S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software* **25**, 747–759. https://doi.org/10.1016/j.envsoft.2009.10.016.

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

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution* **6**, 366–379. https://doi.org/10.1111/2041-210X.12352.

Riegel, R.P., Alves, D.D., Schmidt, B.C., de Oliveira, G.G., Haetinger, C., Osório, D.M.M., Rodrigues, M.A.S., de Quevedo, D.M., 2020. Assessment of susceptibility to landslides through geographic information systems and the logistic regression model. *Natural Hazards* **103**, 497–511. https://doi.org/10.1007/s11069-020-03997-8.

Romans, B.W., Castelltort, S., Covault, J.A., Fildani, A., Walsh, J.P., 2016. Environmental signal propagation in sedimentary systems across timescales. *Earth-Science Reviews* **153**, 7–29. https://doi.org/10.1016/j.earscirev.2015.07.012.

Sahrane, R., Bounab, A., El Kharim, Y., 2023. Investigating the effects of landslides inventory completeness on susceptibility mapping and frequency-area distributions: case of Taounate Province, Northern Morocco. *Catena* **220B**, 106737. https://doi.org/10.1016/j.catena.2022.106737.

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* **126**, 1763–1768. https://doi.org/10.1213/ANE.0000000000002864.

Smith, G., 2018. Step away from stepwise. *Journal of Big Data* **5**, 32. https://doi.org/10.1186/s40537-018-0143-6.

Steger, S., Brenning, A., Bell, R., Petschko, H., Glade, T., 2016. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* **262**, 8–23. http://dx.doi.org/10.1016/j.geomorph.2016.03.015.

Steger, S., Brenning, A., Bell, R., Glade, T., 2017. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* **14**, 1767–1781. https://doi.org/10.1007/s10346-017-0820-0.

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**, 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X.

Strahler, A.N., 1952. Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin* **63**, 1117–1142. https://doi.org/10.1130/0016-7606(1952)63[1117:HAAOET]2.0.CO;2.

Swallom, M.L., Koch, H.J., Dortch, J.M., Crawford, M.M., Thigpen, J.R., Andrews, W.M., 2025. Evaluating root strength index as an indicator of landslide-prone slopes in eastern Kentucky. *Landslides* **22**, 567–578. https://doi.org/10.1007/s10346-024-02384-9.

Tanyu, B.F., Abbaspour, A., Alimohammadlou, Y., Tecuci, G., 2021. Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets. *Catena* **203**, 105355. https://doi.org/10.1016/j.catena.2021.105355.

Taylor, F., Brabb, E.E., 1986. Map showing the status of landslide inventory and susceptibility mapping in California. *U.S. Geological Survey Open-File Report* **86-100**, 1–39. https://doi.org/10.3133/ofr86100.

Tsangaratos, P., Ilia, I., 2016. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *Catena* **145**, 164–179. https://doi.org/10.1016/j.catena.2016.06.004.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs* **92**, e01486. https://doi.org/10.1002/ecm.1486.

Van Den Eeckhaut, M., Marre, A., Poesen, J., 2010. Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101

Geomorphology **115**, 141–155. https://doi.org/10.1016/j.geomorph.2009.09.042.

Verhulst, P.F., 1838. Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique* **20**, 113–121.

Wang, L.-J., Sawada, K., Moriguchi, S., 2013. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Computers & Geosciences* **57**, 81–92. https://doi.org/10.1016/j.cageo.2013.04.006.

Whipple, K.X., 2004. Bedrock rivers and the geomorphology of active orogens. *Annual Review of Earth and Planetary Sciences* **32**, 151–185. https://doi.org/10.1146/annurev.earth.32.101802.120356.

Willett, S.D., 2010. Erosion on a line. *Tectonophysics* **484**, 168–180. https://doi.org/10.1016/j.tecto.2009.09.011.

Woodard, J.B., Mirus, B.B., 2025. Overcoming the data limitations in landslide susceptibility modeling. *Science Advances* **11**, eadt1541. https://doi.org/10.1126/sciadv.adt1541.

Woodard, J.B., Mirus, B.B., Crawford, M.M., Or, D., Leshchinsky, B.A., Allstadt, K.E., Wood, N.J., 2023. Mapping landslide susceptibility over large regions with limited data. *Journal of Geophysical Research: Earth Surface* **128**, e2022JF006810. https://doi.org/10.1029/2022JF006810.

Zhou, C., Wang, Y., Cao, Y., Singh, R.P., Ahmed, B., Motagh, M., Wang, Y., Chen, L., Tan, G., Li, S., 2024. Enhancing landslide susceptibility modelling through a novel non-landslide sampling method and ensemble learning technique: *Geocarto International* **39**, 2327463. https://doi.org/10.1080/10106049.2024.2327463.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/qua.2026.10101