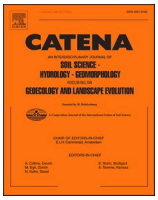
Catena 245 (2024) 108344

Contents lists available at ScienceDirect

Catena

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



Spatial patterns of landslides in a modest topography of the Ozark and Ouachita Mountains, USA

Netra R. Regmi [a,*], Jacob I. Walter [a], Junle Jiang [b], Annalise M. Orban [a], Nicholas W. Hayman [a]

[a] Oklahoma Geological Survey, The University of Oklahoma, Norman, OK 73019, USA
[b] School of Geosciences, The University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

Keywords:
Landslides
Ozark and Ouachita Mountains
Machine Learning
Random Forest Modeling
Landslide Susceptibility

ABSTRACT

Controls on landslides vary as a function of landscape and regional activity. For example, low-relief, woodland regions have slope gradients, soil types, and substrate lithologies that contrast with steeper mountainous regions prone to rock fall and debris flows. Similarly, regional variations in precipitation, earthquakes, and other impacts on landslide surfaces create regional variations in landslide properties. While the controls on landslide characteristics have been extensively studied for high-relief coastal and tectonically active regions, controls on low-relief landslides have received comparatively less attention. We focus here on a part of the Ozark and Ouachita Mountains in the US southern mid-continent to explore such characteristics of landslides and potential controls in low-relief regions. The area exhibits frequent landslides in soil-covered low-relief forested hillslopes. We evaluated the frequency-size scaling of landslides occurred during periods of different earthquake frequency and precipitation amount (pre- and post-2005). We also produced maps of landslide susceptibility based on random forest machine learning applied to remotely sensed data. We found that landslides are clustered mostly in upland hillslopes, and that small landslides dominate the area, quantified by a landslide frequency-size distribution fitting a double Pareto curve. Additionally, the overall landslide frequency, and potentially the portion of smaller landslides relative to the larger ones, significantly increased after 2005, the period during which the area also experienced increased induced seismicity and extreme storm events. Approximately 94 % of historical landslides were within random-forest-classified high-landslide probability (probability > 0.5) zones, coinciding with moderate to steep ($18° \pm 9°$) and convergent upland slopes underlain by shale and sandstone. Anomalously high frequency landslides appear to result from triggering by extreme weather, human-induced earthquake activity, and human-induced hillslope modification.

1. Introduction

Landslides are complex natural phenomena that pose a serious hazard to society and play a significant role in the evolution of landscapes. Landslides occur across a range of spatial and temporal scales. A wide range of triggering events such as earthquakes, large storms, rapid snowmelt, volcanic eruptions, and other forcings increase their frequency and magnitude (Dunham et al., 2022; Jibson and Tanyaş, 2020; Kargel et al., 2016; Tatard et al., 2010; Thomas et al., 2004; Yamagishi and Yamazaki, 2018). Yet, landslides also depend on regional tectonics, climate, geomorphology, as well as intrinsic hillslope characteristics such as hillslope geometry and substrate strength (Bellugi et al., 2021; Medwedeff et al., 2020; Samia et al., 2017).

The most studied landslide inventories worldwide come from high-relief, tectonically active mountainous and coastal regions (Kirschbaum et al., 2010; Medwedeff et al., 2020). High-relief mountains provide enough relief and slope gradient (Fig. 1) to generate large and long-runout landslides (Korup et al., 2007), which are typically the most societally-impactful geohazards (Emberson et al., 2020; Petley, 2012). Hillslopes in these regions tend to be near threshold conditions, a critical slope gradient or height at which hillslopes fail readily (Blöthe et al., 2015; Clarke and Burbank, 2010; Korup et al., 2007; Larsen et al., 2010). On the other hand, the sensitivity and associated hazards of landslides in tectonically stable, low-relief vegetated mountains in intraplate settings are less well-studied. In these environments, hillslopes tend to be relatively gentle (Fig. 1) where soil production rates can exceed erosion rates (Heimsath et al., 1997; Ouimet et al., 2009) which results in hillslopes generally being covered by thick soil. These

* Corresponding author.
E-mail address: netraregmi@ou.edu (N.R. Regmi).

https://doi.org/10.1016/j.catena.2024.108344
Received 23 February 2024; Received in revised form 27 June 2024; Accepted 23 August 2024
Available online 27 August 2024

hillslopes can also generate landslides when slight changes in the rate and/or the magnitude of the triggering mechanism influences the hillslope strength (Casadei et al., 2003; Lacroix et al., 2020; Pánek et al., 2019; Regmi and Walter, 2020a; Roering et al., 2001; Van Den Eeckhaut et al., 2007). The heterogeneity in soil-geomorphic and anthropogenic conditions in this environment cause challenges in understanding the slide characteristics and how they compare relative to high-relief, tectonically active regions.

Here we focus on the relatively modest topography of the Ozark and Ouachita Mountains of Oklahoma and Arkansas to explore the characteristics of landslides in low-relief regions. These mountains show relatively low relief and gentle hillslopes compared to the major high-relief mountain ranges across the world (Figs. 1 and 2, Fig. S1 in Supplementary Information S1). The area receives frequent storms (Oklahoma Climatological Survey, 2018) and has experienced increased low-magnitude induced seismicity over the last decade (Walter et al., 2020) (Fig. S2 in Supplementary Information S1). To date, however, there is a paucity of information on: the type and spatiotemporal pattern of landslide size, major hillslope characteristics that predispose landslides, and landslide susceptibility across the broader region encompassing the Ozark and Ouachita Mountains. Additionally, there remains no clear sense of potentially distinctive nucleation mechanisms for landslides in soil-covered regions, where the role of creep and shallow slides may be enhanced (Ferdowsi et al., 2018; Roering et al., 2001). We hypothesize that a disproportionately large number of landslides occur in this and similar regions given favorable intrinsic hillslope characteristics (bedrock lithology and soil) coupled with sufficient triggering mechanisms (precipitation, seismicity and human induced hillslope modification). To test this hypothesis we map landslides across a large part of the Ozark and Ouachita Mountains, determine their frequency-size (area) relationship, and compare results with published frequency-size datasets from different climatic and tectonic regions across the world. We determine the temporal frequency to assess if the recent change in the rate and magnitude of triggering forces has any influence on the landslide frequency. Additionally, we rank landslide predisposing factors to determine the major intrinsic controls on landslides and prepare a high-resolution map of landslide susceptibility to assess the likelihood of landslides in different soil-geomorphic environments. We then evaluate the association of these controls with landslide susceptibility in the study area and assess how these controls are different from the primary controls of landslides in different climatic and tectonic regions.

There are qualitative, deterministic, and statistical limitations in handling the large datasets needed to fully probe the spatiotemporal patterns of landslides (Booth et al., 2009; Carrara and Merenda, 1976; Dietrich et al., 2001; Nefeslioglu et al., 2008; Pack et al., 1998; Regmi et al., 2010b). Therefore, we employ a random forest machine learning algorithm to characterize landslide-topography associations. The algorithm has the potential to map landslide susceptibility across a large region with diverse geomorphology, soil and climate. Our study thus is a new example to complement previous work using machine learning methods such as regression tree, support vector machine, random forest, and neural networks in landslide susceptibility mapping (Huang et al., 2020; Liu et al., 2023; Liu et al., 2024; Merghadi et al., 2020; Taalab et al., 2018; Youssef et al., 2016). These computational methods: (i) handle large volumes of data, (ii) facilitate sampling design, factor selection, and model optimization, (iii) minimize model overfitting and (iv) solve for the nonlinear relationships between landslides and forcing factors (Kuhn et al., 2020; Wu et al., 2019). Using random forest method, we develop a high-resolution map of landslide susceptibility in the focus area in Oklahoma and Arkansas. Through the mapping we deduce several key properties of landslides in the study area, providing insight into controls on landslides in similar environments worldwide. Furthermore, our work demonstrates a workflow that could be expanded to include other regions so that spatiotemporal landslide relationships could be studied with the same lens.

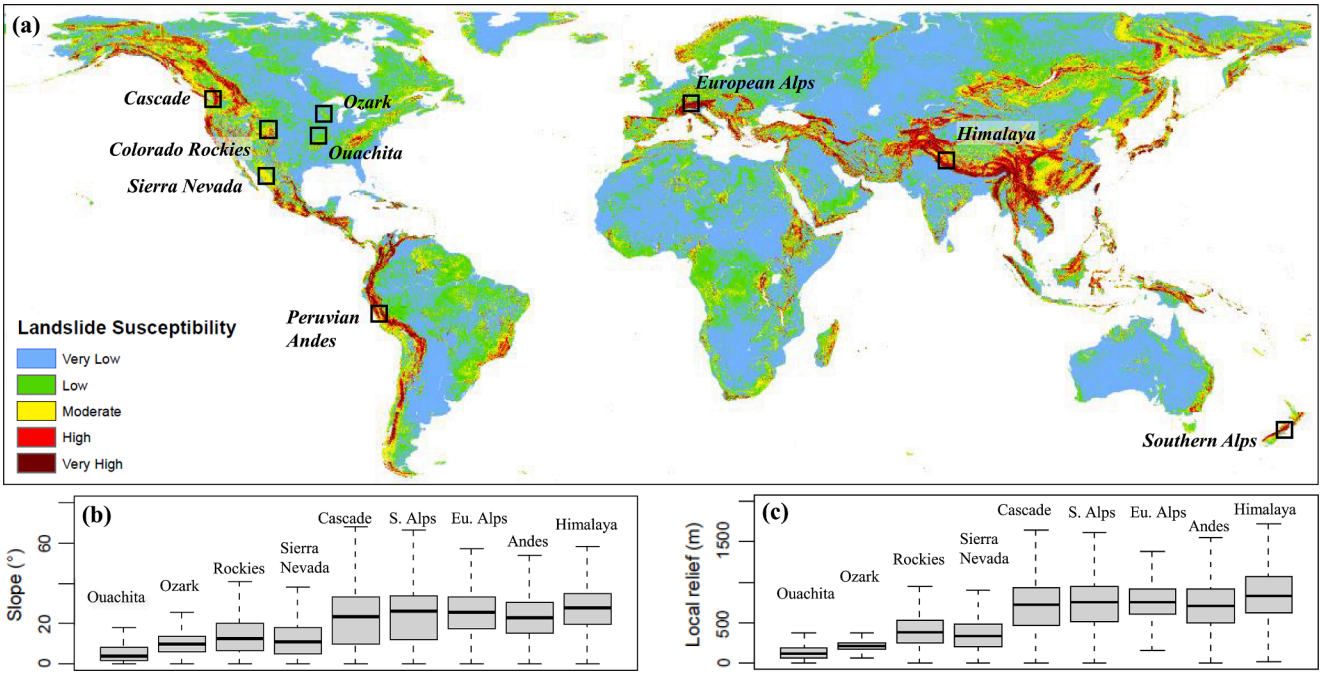


Fig. 1. (a) Landslide susceptibility map of the globe. (b, c) Distribution of slope gradient and local relief within selected areas across the globe (black rectangle in Fig. a). Note high-relief mountains along plate boundaries and coastal regions (Sierra Nevada, Cascade, Southern Alps, European Alps, Peruvian Andes, and Himalaya) fall within very high landslide susceptibility (red color). Large number of landslides including fatal ones have been reported from these regions. The susceptibility map was developed based on the approach presented in Stanley and Kirschbaum (2017) and Emberson et al. (2020) (map source: NASA). The study area (Ozark and Ouachita Mountains) shows significantly smaller distribution of hillslope gradient and local relief compared to other mountain ranges, and falls within moderate (yellow) to low susceptibility (green) region. Both slope gradient and local relief (range of elevation within 1 km radius circle) were mapped from SRTM 90 m DEM.

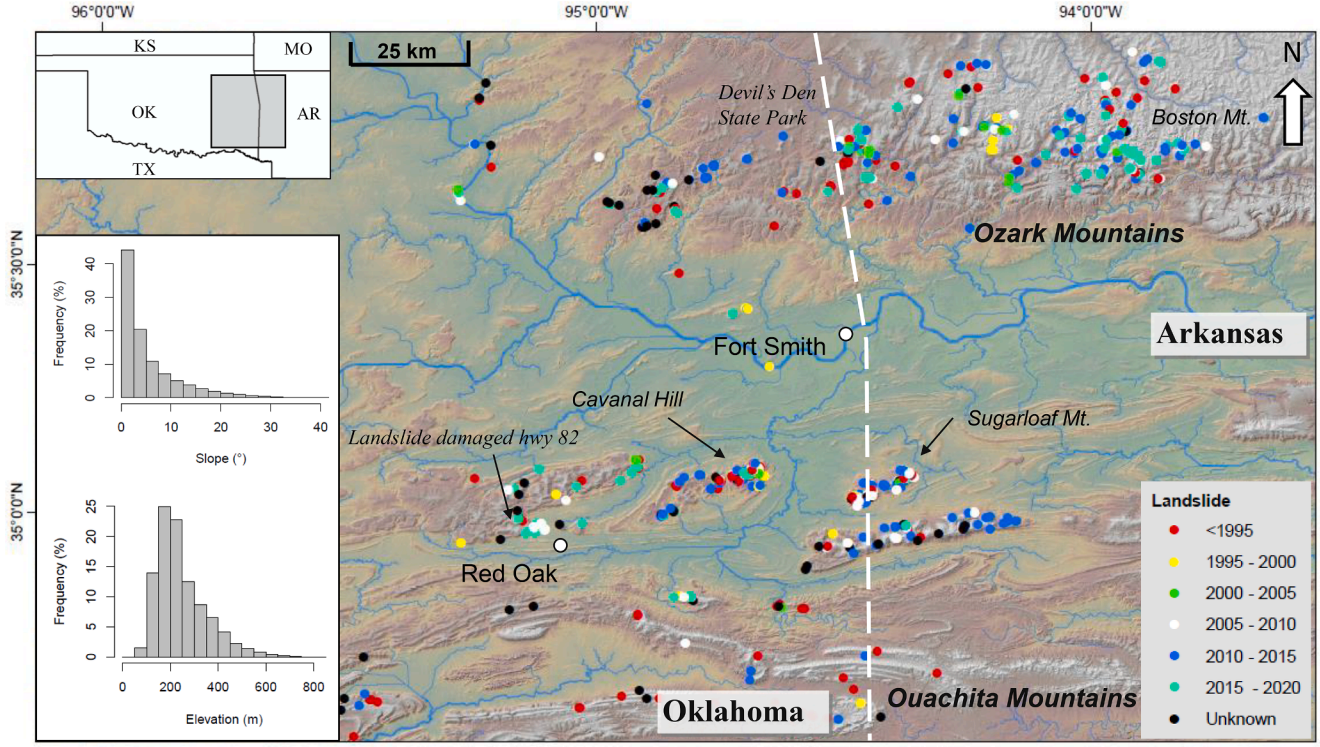


Fig. 2. Location of the study. The study area comprises parts of the Ozark and Ouachita Mountains in eastern Oklahoma and western Arkansas. Locations of landslides were mapped from satellite and aerial imagery acquired since 1995 and archived in Google Earth®. The frequency and size of these landslides are shown in Fig. 5. The inset maps show the distribution of hillslope gradient and elevation determined from 10 m USGS NED DEM. KS: Kansas; OK: Oklahoma; TX: Texas; AR: Arkansas; and MO: Missouri.

2. Study area

2.1. Location, climate, and geomorphology

The study area is located in the Ozark and Ouachita Mountains in eastern Oklahoma and western Arkansas (Fig. 2). Encompassing ~8,000 km^2, the area is characterized by a humid subtropical climate with the annual average precipitation ranging from ~1200 to ~1300 mm with maximum precipitation during summer (June, July and August) and minimum precipitation during winter (December, January and February) (Johnson and Luza, 2008; Oklahoma Climatological Survey, 2018). The vegetation is dominated by oak-pine forest in upland areas

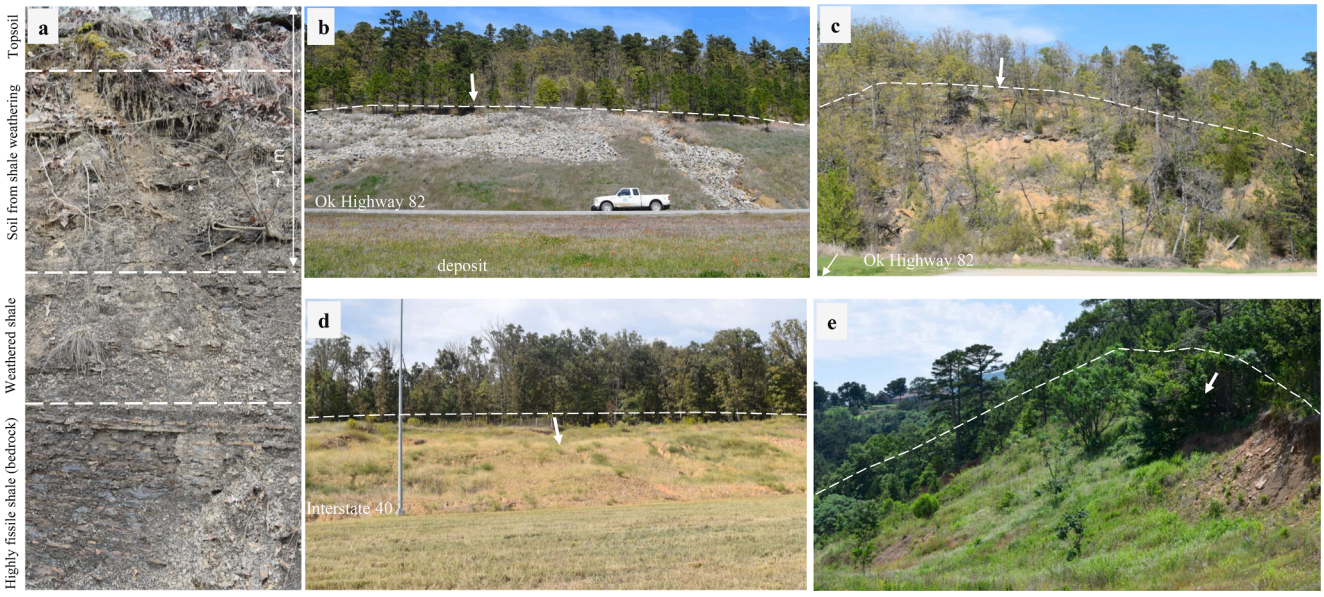


Fig. 3. (a) A soil profile observed along a road cut slope in Cavanal Hill. (b, c) Shallow landslides near Red Oak along the Oklahoma State Highway 82. (d) A shallow landslide along the Interstate 40 in the Ozark. (e) A shallow landslide located close to the Interstate 40 in Cavanal Hill. Top soil in the soil profile is mostly dominated by fine sandy loam (USDA Natural Resources Conservation Services, 2016). The first landslide (b), currently stabilized using riprap and other engineering techniques, completely damaged Oklahoma State Highway 82 and a forest road. All these landslides directly or indirectly affected the traffic.

and post-oak blackjack forest in lowland areas, along with trees and shrubs including flowering dogwood, and highbush and lowbush blueberries (Johnson and Luza, 2008).

The Ozark and Ouachita Mountains are deformed by many faults and the surface geology is dominated by moderately to highly weathered bedrock lithologies (Haley et al., 1993; Heran et al., 2003). The elevation ranges from 80 m to 840 m and the landscape consists of flat to gently sloping lowlands, and hilly and mountainous upland topography (Fig. 2). The lithology is primarily shale and sandstone with chert, limestone and colluvial deposits in lesser amounts (Haley et al., 1993; Heran et al., 2003). Upland slopes underlain by shale are mostly gentle and predominantly covered by regolith (hereafter called soil), whereas slopes underlain by sandstone exhibit relatively steep topography with talus deposits at the steep slope base. The topography causes several major drainage basins in the area to exhibit dendritic drainage patterns.

Soil-mantled slopes in the study area develop primarily by the weathering of shale, as well as colluvium deposited by erosion and landslides (Fig. 3). The texture of the topsoil is dominated by loam and silty loam in the upland slopes of the Ozark and Ouachita Mountains and clay and clay loams in the lowlands (Phillips et al., 2005; USDA Natural Resources Conservation Service, 2016). The average depth of the soil has been observed as ~1 m in upland landscapes of the Ouachita Mountains (Fig. 3a), particularly along the zones of topographic convergence (i.e., hillslope hollows or topographic depressions) (Phillips et al., 2005; Regmi and Walter, 2020a), and similar soil depths can be expected across the majority of the soil-mantled hillslopes in the Ozark Mountains.

2.2. Mass movement

The area exhibits clear evidence of various mass movement and erosional features including recent shallow landslides, rills, gullies, old deep-seated landslides, and incised channels (Fig. 3, Figs. S3 – S5 in Supplementary Information S1) (Regmi and Walter, 2020a). Shallow landslides, defined here as landslides with maximum depth of slip surface less than the tree root depth, are classified following Varnes (1978). They include soil creep, soil slides, debris flows, debris slides and rockslides. In general, soil creep and soil slides occur mostly on soil-mantled surfaces underlain by shale, debris flows occur along the zones of topographic convergence underlain mostly by shale, and debris slides occur mostly in colluvial deposits. Rockslides and rock falls also occur, particularly on steep slopes formed by sandstone exposure and contribute to the development of talus and colluvial deposits at the base of the slope.

A small number of deep-seated large landslides also occur across the study area. The deep-seated landslides occur mostly along the edges of the upland plateau underlain by resistant sandstone caps and appear to be related to the structural failure of the bedrock (Regmi and Walter, 2020a; Taalab et al., 2018). These landslides seem to be of ages ranging from hundreds- to thousands of years old as indicated by relatively dense vegetation and smooth surfaces with well-evolved hydrological network in comparison to recent and active slides (LaHusen et al., 2016). Recent shallow landslides are observed along the scarp and within the deposits of some of the deep-seated landslides (Fig. S3 in Supplementary Information S1).

Landslides in Arkansas and Oklahoma can be socially and economically disruptive (Fig. 3). For example, a landslide that occurred near Sallisaw in the southern Ozarks during a rainstorm of April 2020 damaged over 2,000 feet of water lines and left ~2,500 people without running water for a week (NEWS, 2020). A rockslide that occurred south of Devil's Den State Park (Fig. 2) in the Ozark Mountains of Arkansas during severe storms of June 2019 damaged Arkansas State Highway 220 and temporarily interrupted traffic (NEWS, 2019). Similarly, a large landslide that occurred north of Red Oak (Figs. 2 and 3b) in the Ouachita Mountains of Oklahoma prior to 2003, severely damaged Oklahoma State Highway 82 and a forest road (Cerato et al., 2014).

A few hillslope geomorphic studies have been conducted in these mountains (Hayes, 1971; He et al., 2014; Luza and Johnson, 2005; Oakes, 1952; Rahimi et al., 2021; Regmi and Walter, 2020a; Thaler and Covington, 2016). For example, Oakes (1952) notes eyewitness reports of landslides near Cavanal Hill shortly after the M 5.5 El Reno earthquake on April 9, 1952. The earthquake occurred ~300 km from Cavanal Hill. Regmi and Walter (2020a) determined the frequency of historical earthquakes that impart >2 kPa dynamic stresses for eastern Oklahoma since 1900 and found an increased potential for dynamic stress perturbation of landslide activity since ~2010. They also evaluated the effectiness of using surface roughness derived from high-resolution LiDAR topographic data in mapping landslides. Similarly, He et al. (2014) prepared a regional-scale 30 m cell-sized raster map of landslide susceptibility for the entire state of Oklahoma using an expert-based weighted linear combination of a number of factors that can predispose landslides.

3. Materials and methods

3.1. Overview

The methodology of this study consists of two parts. First, we prepare an inventory of landslides and associated characteristics including age, area, and runout distance from high-resolution time series satellite and aerial imagery acquired since 1995 and archived in Google Earth®, and 1 m bare-earth USGS 3DEP LiDAR topographic data. Field reconnaissance was carried out to verify the map of landslides and underlying soil and lithology. Second, we identify factors characterizing hillslope environments that can predispose landslides (Table 1) and can be used as independent variables of landslides (hereafter are called environmental covariates) in the random forest landslide susceptibility modeling effort from 10 m USGS NED elevation data, statewide integrated 1:500,000 scale USGS geological map, USDA soil map, USGS landcover map, and road and highway map (Fig. 4).

3.2. Landslide mapping and frequency-size characterization

Landslides were mapped over Google Earth imagery and a shaded relief image derived from 1 m LiDAR digital elevation model (DEM) in a GIS platform. The imagery-based mapping was carried out by qualitatively identifying the distinguishing tone, shape, size, and texture of landslide scarps and displaced materials. Challenges to this mapping include: (1) small-sized landslides are undetectable in low-resolution older images, (2) indistinct boundaries between highly vegetated landslide and non-landslide areas, (3) difficulties in identifying type (i.e., rock or soil dominated, and recent or old) and size of the material displaced by a landslide because of the vegetation cover and resolution of the images, and (4) similar appearance of rockslides and bare rock cliff. These issues in most of the cases were resolved by employing three-dimensional visualization of LiDAR DEM and shaded relief image in ArcGIS 10.7.1® (ESRI, Redland, CA). Ages and temporal frequencies of landslides in 5-year interval were determined from the time series imagery. The location, type, and activity of some of the landslides were verified by field mapping. All these attributes were linked with spatial information of the landslides and stored in ArcGIS® for calculation of landslide areas and runouts.

Runout length is computed as the distance from the head to the toe of a landslide along the path of the movement. The measurement is carried out manually by tracing the path of the landslide movement on georeferenced satellite images. Landslide frequency-size distribution is derived from Gaussian kernel-estimated probability density functions of landslide area (Larsen et al., 2010; Regmi et al., 2014b) using the bandwidth determined based on Silverman Rule of Thumb (Silverman, 1984). The probability density estimates of landslide areas are then fit by double Pareto distributions (Fig. S6 in Supplementary Information S1) following Stark & Hovius (2001) using a non-linear least-squares

Table 1
Types, source, and significance of landslide predisposing factors (covariates) used in landslide susceptibility mapping. Covariate acronyms are listed within brackets.

Data type	Covariates	Significance	Data source
Landslide inventory		Location of historical landslides	Field surveys, Google Earth ® images, and USGS LiDAR DEM (https://nationalmap.gov/elevation.html)
Geologic	Geological map (geo)	Characteristics of the slope material	USGS (https://pubs.usgs.gov/of/2005/1351/)
	Proximity to fault (df)	Co-seismic landslide triggering	
Landcover	Landcover (lc)	Root reinforcement of soil, surface runoff regulation	USGS (https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap)
Soil	Soil texture (st)	Shear strength of soil	USDA (https://websoilsurvey.sc.egov.usda.gov/)
Topographic	Elevation (elev)	Climate, vegetation, and potential energy	USGS NED DEM (https://nationalmap.gov/elevation.html)
	Slope (slp)	Overland and sub-surface flow velocity	
	Aspect (asp)	Solar insolation, evapo-transpiration, flora and fauna distribution and abundance	
	Plan curvature (plc)	Converging, diverging flow, soil water content, and soil characteristics	
	Profile Curvature (prc)	Flow acceleration, erosion/deposition potential	
	Tangent Curvature (tc)	Erosion/deposition	
Water related	Flow Accumulation (fac)	Runoff velocity, runoff volume, and potential energy	
	Surface roughness (sr)	Water flow, water pounding potential	
	SAGA Wetness index (swi)	Soil water content	
	Stream power index (spi)	Erosive power of water flow	
	Proximity to streams (ds)	Susceptible to hillslope undercutting	
Anthropogenic	Highway and roads (dh)	Landslide triggering by the road cutting and vibration generated by the vehicles	OKMaps and Arkansas Dept. of Transportation (https://okmaps.org/ogi/; https://gis.arkansas.gov/)

optimization method in LMFIT package (Newville et al., 2023). The goodness-of-fit statistics include the coefficient of determination (R^2), chi-square, Akaike information criterion, and Bayesian information criterion. Many studies show landslide frequency-size relation follows a double Pareto distribution, defined by two power-law relationships, first with positive power-exponent (β) representing small-sized landslides, and the second with negative power-exponent (α) representing large landslides, separated by a rollover in relationship curve (t) (Frattini and Crosta, 2013; Guthrie and Evans, 2004; Hurst et al., 2013; Medwedeff et al., 2020; Regmi et al., 2014b; Stark and Hovius, 2001; Tanyaş et al., 2019; Van Den Eeckhaut et al., 2007). These attributes together can characterize the overall distribution of landslide size and can be used to compare the sensitivity of different-sized landslides in different regions. We compared our power-law exponents with similar datasets from different climatic and tectonic regions to characterize how the size distribution of our landslides compares with global landslides.

3.3. Landslide susceptibility mapping

3.3.1. Overview

We prepare a landslide susceptibility map by assessing relationships between landslides and landslide predisposing factors (covariates) (Table 1), and implementing the random forest algorithm (Fig. 4). In this process, we: (i) prepare maps of covariates and convert them into raster GIS formats (10 m × 10 m cell size), (ii) develop a sampling design to prepare the dataset needed for training and testing landslide susceptibility models, (iii) preprocess the dataset, determine optimum samples, model parameters and input covariates, (iv) test the validity of training models, (v) and implement the best model in mapping landslide susceptibility (Fig. 4).

3.3.2. Environmental covariates

Based on the observation of landslides in the field and previous work conducted in eastern Oklahoma and western Arkansas (He et al., 2014; Regmi and Walter, 2020a) as well as worldwide (Ayalew and Yamagishi, 2005; Lee et al., 2004; Pourghasemi et al., 2012; Regmi et al., 2010a; Regmi et al., 2010b; Yalcin, 2008), we identified 16 factors including geologic, eco-hydrogeomorphic, and anthropogenic environments that can predispose landslides, and used them as covariates of landslides in landslide susceptibility modeling effort (Table 1, Figs. S7 and S8 in Supplementary Information S1).

We used USGS integrated geological maps of Oklahoma and Arkansas (Stoeser et al., 2005) to develop two covariates: a map of lithology and a map showing distance from faults (Table 1). The geological map of Oklahoma is a 1:500,000 scale map prepared by Heran et al., (2003), and the map of Arkansas is a 1:500,000 scale map prepared by Haley et al. (1993). Both maps were merged based on primary rock types and then converted the merged map into a 10 m cell-sized raster to match the resolution of other covariate rasters. These coarse-scale maps were adopted because of the lack of higher resolution maps across the entire study area.

Soil properties were extracted from the United States Department of Agriculture (USDA) State Soil Geographic (STATSGO) dataset (USDA Natural Resources Conservation Service, 2016). The data consist of generalized map unit polygons characterizing homogenous soil characteristics, and tabular data consisting of various properties of soil in each map unit. Using these data, we prepared a map showing classes of soil texture determined based on the percent distribution of sand, silt and clay sized particles in the soil and converted the map into a 10 m cell-sized raster; the approach follows several previous examples in the landslide literature (He et al., 2014; Hong et al., 2007; Lee and Min, 2001).

The land cover data used in this study is USGS GAP/LANDFIRE National Terrestrial Ecosystems 2011 raster data of 30 m resolution (USGS, 2016). The dataset was created using Landsat satellite imagery, and includes a detailed vegetation and land use classification system. We resampled these data into a 10 m cell-sized raster and used as a landslide covariate.

Many landslides were observed near roads and highways. The potential triggers of these landslides could be traffic-induced slope vibration as well as road-cutting activities (He et al., 2014; Maguigan et al., 2015; Pachauri and Pant, 1992). We assumed that landslides tend to occur closer to the highways and roads, so a map of distance from roads can be used as a landslide predictor. We merged roads and highways data of Oklahoma and Arkansas (from OKMaps: https://okmaps.org/ogi/search.aspx and the Arkansas Department of Transportation: https://gis.arkansas.gov/product-category/data/transportation/) and prepared a raster map of 10 m cell size showing distance from roads.

Geomorphic and hydrologic data layers (Table 1) were developed from 10 m NED elevation data in ArcGIS® and SAGA® (http://www.saga-gis.org/en/index.html). These datasets included slope, curvature

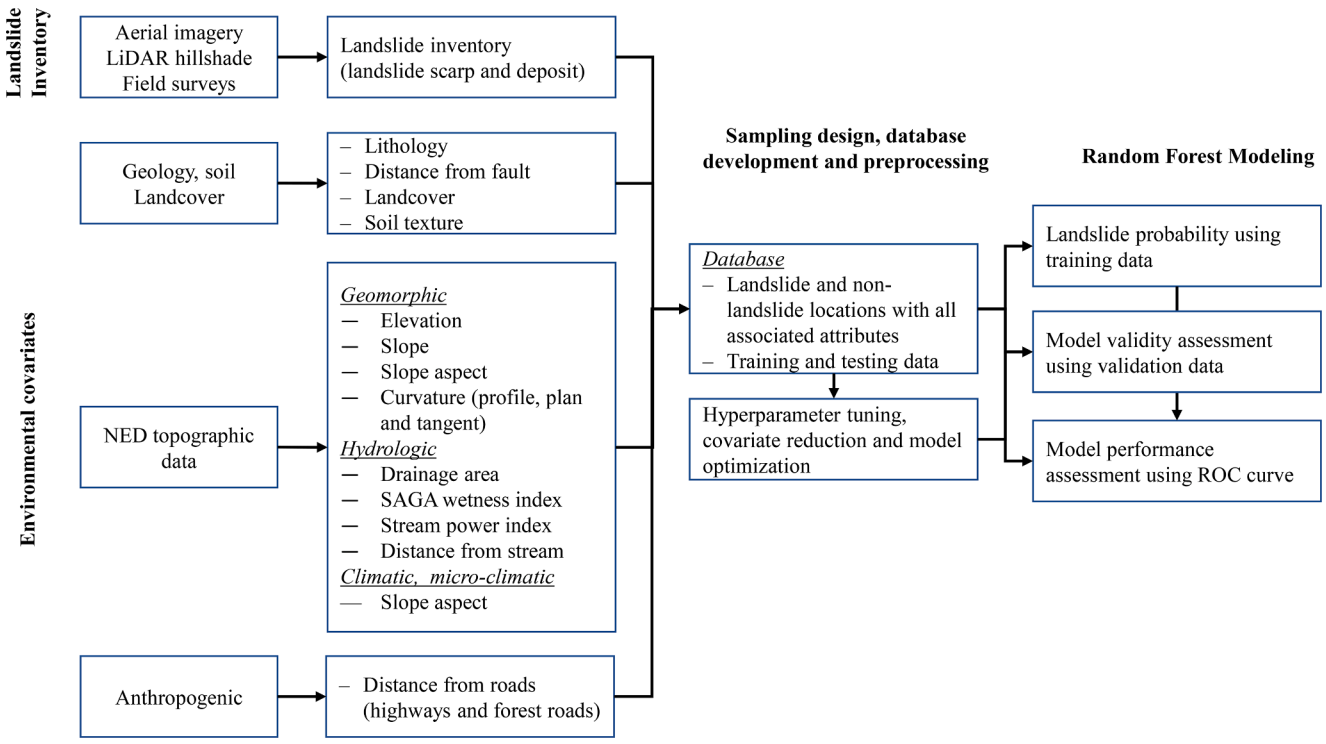


Fig. 4. Flowchart of methodology for mapping landslide susceptibility based on random forest classification.

and surface roughness as geomorphic covariates, aspect as a climatic covariate, SAGA wetness index, stream power index, flow accumulation and stream network as hydrologic covariates. Slope, curvature and aspect were calculated from 10 m NED elevation data. Surface roughness was quantified as a standard deviation of slope within a 30 m × 30 m moving window following Frankel and Dolan (2007). The variable is considered a predictor of landslides because studies show standard deviation of slope can effectively characterize rough surface topography of soil creep and landslides (LaHusen et al., 2016; Regmi and Walter, 2020a). Hydrological factors including flow accumulation (upstream contributing area) and stream network maps were developed from NED data. Similarly, SAGA wetness index, a natural logarithmic ratio of upstream contributing area and tangent of local slope, and stream power index, a product of contributing area and tangent of local slope (Böhner et al., 2006) were computed in SAGA GIS®.

3.3.3. Sampling design and data preprocessing

The primary focus of the sampling design is to create a landslide binary (presence and absence) database along with associated covariate attributes that can be used to train and validate models. We sampled landslide data from entire cells of landslide scarps and an equal number of non-landslide data from the areas devoid of landslides excluding broad lowland riverine plains and valley bottoms. The sampling methodology follows previous studies (Ayalew and Yamagishi, 2005; Dou et al., 2020; Nefeslioglu et al., 2008; Regmi et al., 2014a; Yilmaz, 2010). All covariate values were then extracted to these samples and the entire population was revised to eliminate covariate values associated with the DEM errors, such as striping artifacts.

Next, the entire population was divided into training (70 % of the total) and validation datasets (30 % of the total) where each of the datasets consists of an equal number of landslide and non-landslide data. The reason for using a 70/30 split between training and validation data is that 70 % of the data is anticipated to be sufficient for generating stable training models that uncover the underlying patterns and relationships within the data. The remaining 30 % is reserved specifically for independently validating these models. Both datasets were further pre-processed using the Caret package in R® (Kuhn et al., 2020). For example, the missing values were handled using a bootstrap aggregation imputation technique via the bagImpute function, and covariates that are irrelevant or redundant from the analysis were removed using recursive feature elimination (RFE) technique (Kuhn et al., 2020). RFE performs covariate selection by iteratively training a model, ranking covariates, and then removing the lowest-ranking covariate in a recursive fashion. The accuracy obtained after each RFE was then compared to select optimal input covariates. The selected datasets were then implemented for random forest hyperparameter tuning and landslide susceptibility model development, described in sections 3.3.4 and 3.3.5.

3.3.4. Random forest modeling

The random forest algorithm generates several decision trees by building a bootstrapped sample from observations, and aggregates the decision made by each of the trees to provide one single prediction. The prediction output is the majority of decisions from these trees in case of classification, and the average of the outputs from an individual tree in case of regression (Breiman et al., 1984). The algorithm consists of three user defined parameters, also termed “hyperparameters,” including the number of trees, the minimum number of nodes in a tree, and the number of the covariates to split at each node. The split at a node is based on a random subsample of covariates in such a way that the process minimizes the regression or classification error. Nodes continue to be split until no further improvement in error is achieved. Omitted observations, termed the “out-of-bag” sample, are used to estimate the prediction error. The performance of the models depends on a combination of these hyperparameters (Cutler et al., 2012). The number of covariates at each node affects both the correlation and strength of each individual tree, and thus influences the model performance. Similarly, a smaller number of trees can result in relatively unstable and less accurate models compared to those developed with larger number of trees. However, the computation time increases with the number of trees.

The random forest algorithm can handle both continuous (i.e. slope) and categorical covariates (i.e. geologic unit names or rock types). The algorithm can quantify the contribution of each of the covariates

towards the model prediction by accounting for interactions and non-linearity between covariates (Díaz-Uriarte and De Andres, 2006; Veronesi and Hurni, 2014). Because the algorithm obtains results from a large number of trees where each tree is trained on a unique bootstrap subsample of the data, the approach reduces the effect of correlation amongst covariates and minimizes the noise in the data (Strobl et al., 2007; Svetnik et al., 2003).

The approach has some limitations too. For example, the predictive capability of the model is solely dependent on the quality, quantity, and the spatial distribution of the observations used to train the model. Thus, stable models depend on the optimal samples as well as sufficient input covariates. The model results need to be verified by enough independent observations (validation data) as well as expert-based evaluation of the model results (i.e. landslide susceptibility map). In addition, the algorithm provides little understanding of decision process of each individual tree and therefore the relationship between landslide and covariates cannot be examined individually for every tree in the forest.

3.3.5. Hyperparameter tuning and model development

We implemented a grid search-based algorithm available in the Caret package (Kuhn et al., 2020) to determine the optimum number of trees and the maximum number of covariates to split at nodes. Models were developed using different combinations of numbers of trees (50, 100, 200, 300, 400 and 500) and covariates to split, ranging from 1 to the total number of covariates identified by the RFE. The Gini index, which calculates the purity of the classification of a dependent variable (i.e., landslide presence / absence) for a given covariate (Cutler et al., 2012), was used as a measure of the best covariate split selection, and trees were grown until no further splitting improves the model.

The hyperparameter combination and RFE-selected covariates that resulted in the best prediction accuracy were used for final landslide susceptibility model development. In doing so, training models were developed implementing five-fold cross validation technique. The approach divides the dataset into five equal subsets, and during each run it uses four sets (80 %) to develop a training model and tests the accuracy of the model using the fifth set (20 %). In such a way, the algorithm develops five models and independently tests the accuracy of each model. The final model is the average of all models. The validity of the final model was then assessed by the validation dataset.

The performance of all models were determined based on the receiver operating characteristic (ROC) curve analysis. ROC curve analysis is a statistical method that can be used to distinguish between two classes of response to an event, such as presence or absence of landslides (Egan and Egan, 1975; Gorsevski et al., 2006; Søreide, 2009). The curve is a plot of the probability of a correctly predicted response (true positive rate or sensitivity) versus the probability of a falsely predicted response (false positive rate or 1− specificity) as the cutoff probability varies. The area under the ROC curve (AUC) ranges from 0.5 to 1. If all the responses are correctly predicted, AUC would equal 1. Similarly, the performance of the prediction can be evaluated by computing "overall accuracy," which is a ratio of the frequency of correctly predicted assessments determined based on a probability threshold (i.e., 0.5 in this study) to the total number of assessments.

4. Results

4.1. Landslide characteristics

Altogether, 717 landslides were mapped. The landslides are primarily shallow and clustered in many upland regions (Fig. 2), particularly in Boston Mountain in the Ozark and Cavanal Hill and Sugarloaf Mountain in the Ouachita Mountains. The landslide area ranges from 30 m^2 to 340,000 m^2 with an arithmetic average of ~7,300 m^2, median of 2,200 m^2 and standard deviation of 20,700 m^2. In logarithmic distribution, the average and standard deviation of landslide area are computed as 3.32 m^2 and 0.7 m^2, respectively (Fig. 5a). The landslide runout ranges from 20 m to 500 m with an arithmetic average of ~90 m and standard deviation of 85 m (Fig. 5b). Additionally, the temporal frequency of landslides computed in the five-year interval shows the number of landslides increased significantly since 2005 (Fig. 5c).

The frequency-size scaling based on the double Pareto fitting shows the power-law exponents for large landslides (α) and small-sized landslides (β) as 2.24 and 1.07 for the entire landslide inventory, 2.59 and 1.04 for pre-2005 landslides, and 2.34 and 1.08 for post-2005 landslides (Fig. 6a and Table S1 in Supplementary Information S2). These relationships are separated by a landslide area of 7,301 m^2, 10,808 m^2 and 7,003 m^2, respectively (Fig. 6a). The probabilities of largest landslides exhibit notable deviations from the best-fitting double Pareto distributions. The frequency of smaller landslides relative to larger landslides increases after 2005. In each case, the exponent β was significantly smaller than that the average of β values ($\beta = 2.46$) reported for many regions across the world, whereas the exponent α was found slightly larger than the average of α values ($\alpha = 1.80$) reported worldwide (Fig. 6, Table 2, Table S2 in Supplementary Information S2).

4.2. Landslide-covariate relation, susceptibility model development and validation

Comparing the distribution of lithology, soil texture, landcover and various hillslope geometries across the entire landscape and within

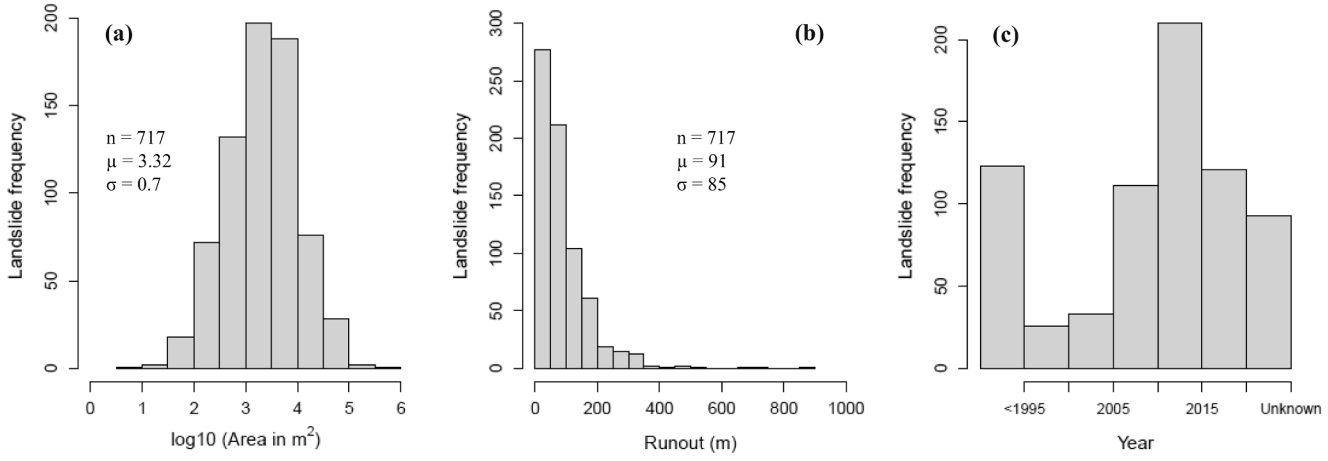


Fig. 5. Histograms showing frequency distribution of landslide properties: (a) areas, (b) runout lengths, and (c) temporal frequency. Areas are in logarithmic scale. Note the frequency of landslides increased from 2005 to 2015, and slightly decreased during 2015–2020. n = number of observation, μ = mean, and σ = standard deviation.

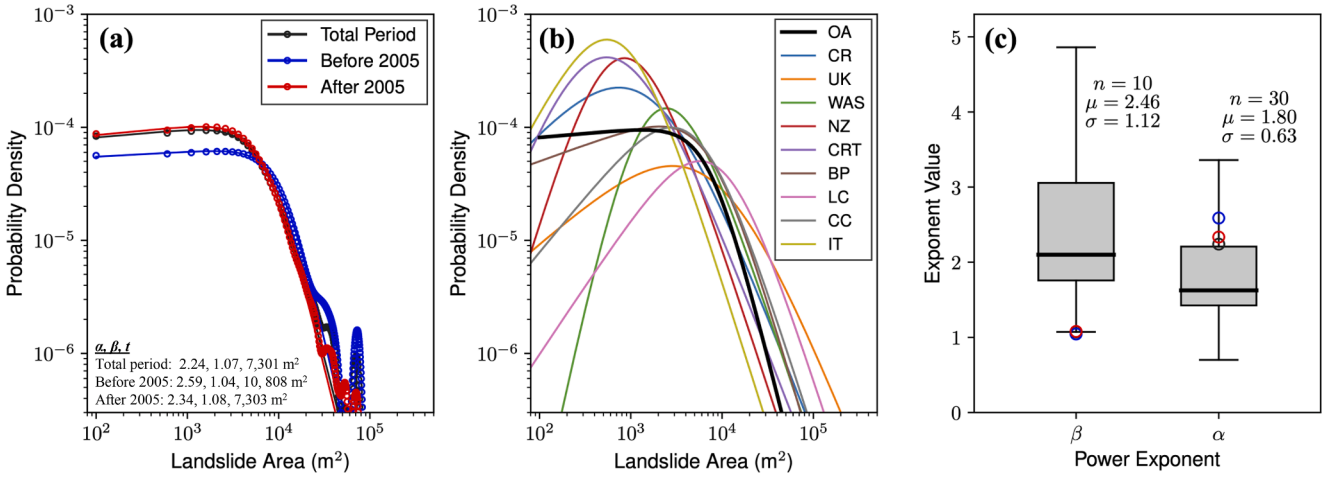


Fig. 6. Landslide frequency-size scaling based on double Pareto distributions. (a) Probability density estimates (open circles) and best-fitting double Pareto curves (solid lines) are shown for pre-2005 (blue), post-2005 (red) and total landslides in the Ozark and Ouachita Mountains. Details of curve fitting statistics and double Pareto parameters are provided in Table S1 in Supplementary Information S2. (b) The best-fitting double Pareto distributions for total landslide inventories in this study (OA) and other selected regions around the world. The legend acronyms and published β, α, and t parameter values for other regions are listed in Table 2. (c) Boxplots showing power-law exponents suggested for small-sized landslides (β) and large-sized landslides (α) across the world (see details in Table S2 in Supplementary Information S2). The best-fitting parameters from (a) are shown as open circles. The number of observations, n, the mean, μ, and standard deviation, σ, are marked for each exponent.

Table 2
Parameters of double Pareto distribution of landslide area in highly landslided regions across the globe. NA: Not available.

Location	α	β	Rollover (t)	Number	Reference
Ozark – Ouachita (OA)	1.70	1.15	4000	717	This Study
Colorado Rockies, USA (CR)	1.10	1.90	1600	735	Regmi et al., 2014b
United Kingdom (UK)	1.01	1.71	8009	8452	Hurst et al., 2013
Western Alps, New Zealand (WAS)	1.44	4.86	1781	5086	Stark and Hovius, 2001
Whataroa, New Zealand (NZ)	1.48	3.74	797	3986	Stark and Hovius, 2001
Central Range, Taiwan (CRT)	1.11	3.22	519	1086	Stark and Hovius, 2001
Brooks Peninsula, Canada (BP)	1.77	1.29	7380	201	Guthrie and Evans, 2004
Loughborough, Canada (LC)	1.51	2.21	9301	92	Guthrie and Evans, 2004
Clayoquot, Canada (CC)	1.54	1.99	5152	1109	Guthrie and Evans, 2004
Val di Fassa, Eastern Italian Alps (IT)	1.57	2.56	750	NA	Frattini and Crosta, 2013

landslides (Figs. S7 and S8 in Supplementary Information S1) shows that landslides are over-represented in sandstone and shale, fine sandy loam soil, forest and woodland, moderate elevation (300–550 m), moderate to steep slopes (10°–30°), N and NE facing slopes, concave and convex slopes, moderate wetness index, high stream power index, moderate to high surface roughness (2.5°–10°), and in close proximity to faults, streams and roads. The mean decrease in Gini (MDG) values of all the covariates included in the model suggest that the major covariates of landslides ranked in order of importance from highest to lowest are elevation, slope, distance from faults and roads, stream power index, geology, surface roughness, soil texture, and distance from streams (Fig. S9 in Supplementary Information S1). The significance of other covariates, such as aspect, wetness index, curvature, flow accumulation and land cover, appears to be relatively minimal. This is well-supported by our field observations of a large number of landslides in moderate to steep upland slopes underlain by shale and sandstone (also see Figs. S7 and S8 in Supplementary Information S1).

Overall accuracies and areas under the ROC curve (AUC) of five-fold cross-validated models developed using RFE selected covariates (9 out of 16) and the optimum hyperparameters (Figs. S9 and S10 in Supplementary Information S1, Table S3 in Supplementary Information S2) are not significantly different (Table 3). Similarly, validation of the final model, which is an average of the cross-validated models by completely independent dataset (validation data), also resulted in similar prediction performance (Table 4). The overall accuracy represents the ratio of the number of landslide and non-landslide observations correctly predicted based on a probability threshold of 0.5 (i.e., probability >0.5 in Fig. 8 are landslides) to the total number of assessments. The model truly predicted ~94 % of the total landslide and non-landslide cells and falsely predicted the remaining 6 % cells (Table 4). The median prediction probability values of truly and falsely predicted cells are computed as ~0.94 and ~0.65, respectively (Fig. S11 in Supplementary Information S1). All these outputs indicate that the models are valid and reasonably accurate for mapping landslide susceptibility.

4.3. *Landslide susceptibility map*

The overall performance of the model, as well as the occurrence of historic landslides in high probability zones of the landslide probability map (Figs. 8 and 9, Tables 3 and 4), imply that the model predicted most of the observed landslides. The distribution of modeled landslide probability across the entire area shows a left-skewed distribution (Fig. 10a) with most of the cells having probability <0.2. Similarly, the distribution of probability values within landslides shows a right-skewed

Table 3
Training and testing accuracies and areas under the ROC curves of five-fold classification of the final model.

	Training			Testing		
Fold	accuracy	ROC	freq	accuracy	ROC	freq
1	0.934	0.984	11,770	0.935	0.987	2942
2	0.936	0.984	11,770	0.928	0.985	2942
3	0.934	0.985	11,769	0.934	0.983	2943
4	0.934	0.985	11,769	0.935	0.982	2943
5	0.933	0.984	11,770	0.938	0.985	2942
Mean	0.934	0.984	11,770	0.934	0.985	2942

Table 4
Confusion matrix, overall accuracy and area under the curve (AUC) of ROC analysis of final model of landslide susceptibility developed using the training dataset (70% of the total dataset) and validated by the validation dataset (30% of the total population). Predicted classes are landslide absence (0) and presence (1).

Model		Confusion matrix and errors				
		0	1	class error	accuracy	AUC
Training (70 % data)	0	6688	668	0.09	0.94	0.98
	1	211	7145	0.03		
Validation (30 % data)	0	2860	96	0.03	0.94	0.98
	1	294	3058	0.09		

distribution where most of the landslide cells have probability >0.6. A significant number of landslide cells, particularly in lowland gentle hillslopes, resulted probability 0.2–0.6. Overall, the landslide probability is low (<0.2) in lowland near-planar landscape, moderate (0.2–0.6) in lowland gentle slopes and high (>0.6) in upland slopes (Fig. 10a, b and c). The high landslide probability areas in upland slopes mostly coincide with concave and convex slopes (Fig. 10d) and hummocky or rough surface topography (Fig. 10e). The landslide susceptibility map was therefore classified into high, medium and low susceptibility (Fig. 11) based on cut off probability values of <0.2, 0.2–0.6 and >0.6, respectively. The resulting map included most of the steep upland slopes underlain by shale and sandstone, zones of topographic convergence (i.e., bedrock hollows and topographic depressions), and hillslopes near faults, roads, and streams as highly landslide susceptible slopes (Figs. 8, 9, and 11).

5. Discussion

5.1. Landslide frequency and size

We consider that the observed number of landslides is reasonably high for the modest topography making up low to moderate relief hills separated by wide intervening stream valleys (Figs. 1 and 2, Fig. S1 in Supplementary Information S1). The small value of β (power-law exponent for small landslides) relative to α (power-law exponent for large landslides) and the rollover in the distribution at 7,300 m^2 from the double Pareto frequency-size plot of total landslides (Fig. 6a) suggest that landslides in Oklahoma and Arkansas are dominated by small landslides (<7,300 m^2 area). In addition, the smaller β exponent compared to that of nine landslide inventories from different parts of the world (Table 2, Fig. 6b) suggests that the occurrence probability of small-sized landslides in our study is relatively high. Our α value is slightly larger than that the average of 30 landslide inventories around the world (Fig. 6c), however the value is qualitatively similar to that of many high-relief and or tectonically active areas (Fig. 6, Table 2, Table S2 in Supplementary Information S2), implying that the focus area in Oklahoma and Arkansas can host relatively large and damaging landslides. For example, the value is larger than that for Colorado Rockies, Southern and Western Alps, and Central Range in Taiwan (Table 2), and smaller than reported for Northridge, California (Sierra Nevada in Fig. 1), Challana Valley in Bolivia (Andes in Fig. 1), Arno River Basin and Umbria Marche in Italy, and Nepal Himalayas (Table S2 in Supplementary Information S2).

Undersampling of small-sized landslides is one possible reason for the rollover in the power-law relationship. Landslides in this study were mapped from time series of high-resolution satellite imagery (5-yr interval) and 1 m LiDAR topographic data. Such detailed mapping efforts should minimize the effect of undersampling in the landslide probability distribution. This can be one of the reasons that our β value is significantly smaller than that of many of the global landslide inventories. Nevertheless, the value of our α is similar to the α values reported for highly landslided regions around the world suggesting that the inventory exhibits a wide range of landslide sizes. The reason for such a wide range of landslide size in our study could be intrinsic to the fractured and weathered lithology and long-term weakening of the soil-mantled hillslopes by frequent extreme weather, induced seismicity and anthropogenic hillslope modification.

5.2. Physical controls and potential mechanism

Landslide-covariate relationships and random forest covariate rankings imply that major controls of landslides in this study are: moderate to steep upland slopes, fractured and weathered shale and sandstone, and presence of soil cover dominated by sandy and silty loam in many hillslopes (Figs. 7 and 10, Figs. S7 and S8 in Supplementary Information S1). A noteworthy observation is that the average gradient of the landslide scarp in our study was ~18° with a standard deviation of 9° (Fig. 7a). In high-relief regions, landslides typically have steeper

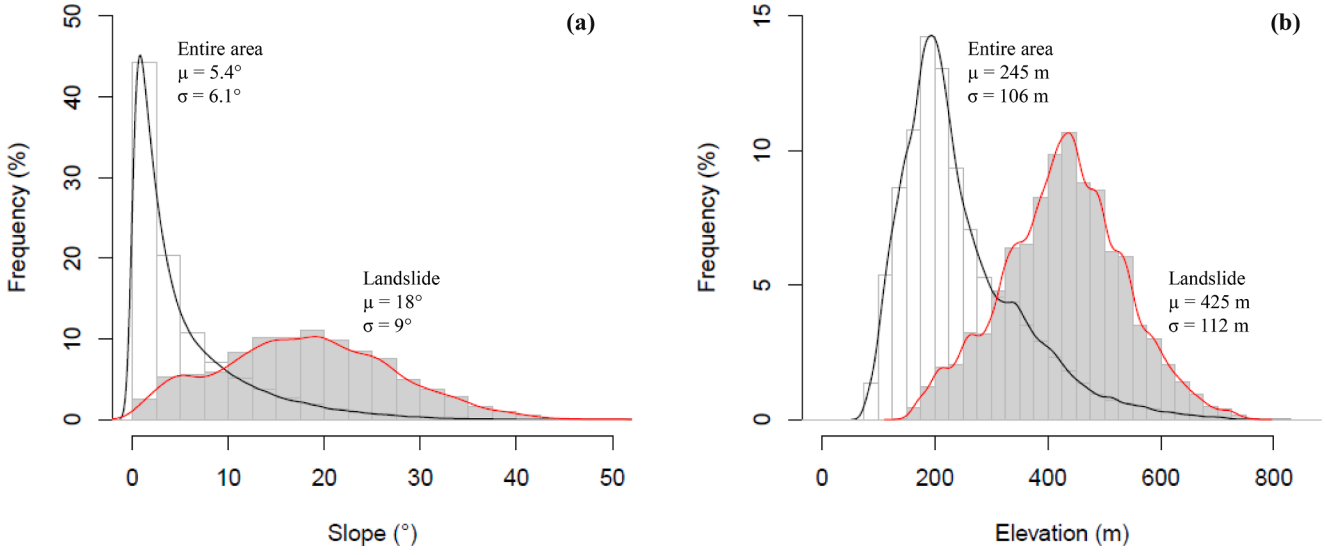


Fig. 7. Distribution of elevation and slope across the landscape and within landslide scarp. The mean, μ, and standard deviation, σ, values are listed. Note the landslides are dominated in moderate to steep slopes ($\mu = 18°$, $\sigma = 9°$) in upland areas ($\mu = 425$ m, $\sigma = 112$ m). The density curve envelopes are developed based on kernel density estimation (also shown in Fig. S8 in Supplementary Information S1).

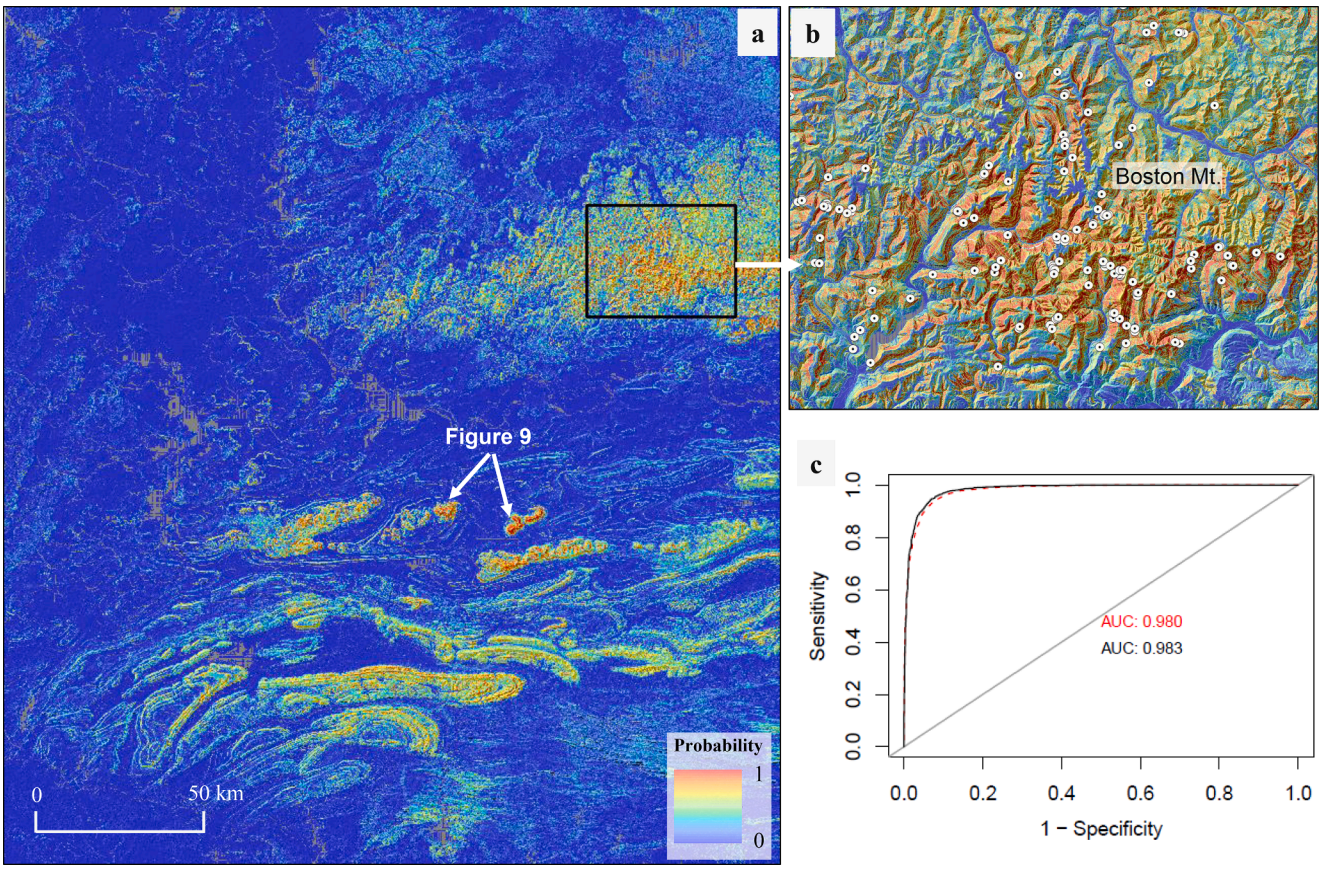


Fig. 8. (a, b) Landslide probability maps developed using training datasets (70% of total). (c) ROC curves developed by assessing the performance of the training model by training data (solid line) and validation data (dashed line). The landslide probability was computed by taking the ratios of decision tree votes that suggest a cell is a landslide to the sum of all votes.

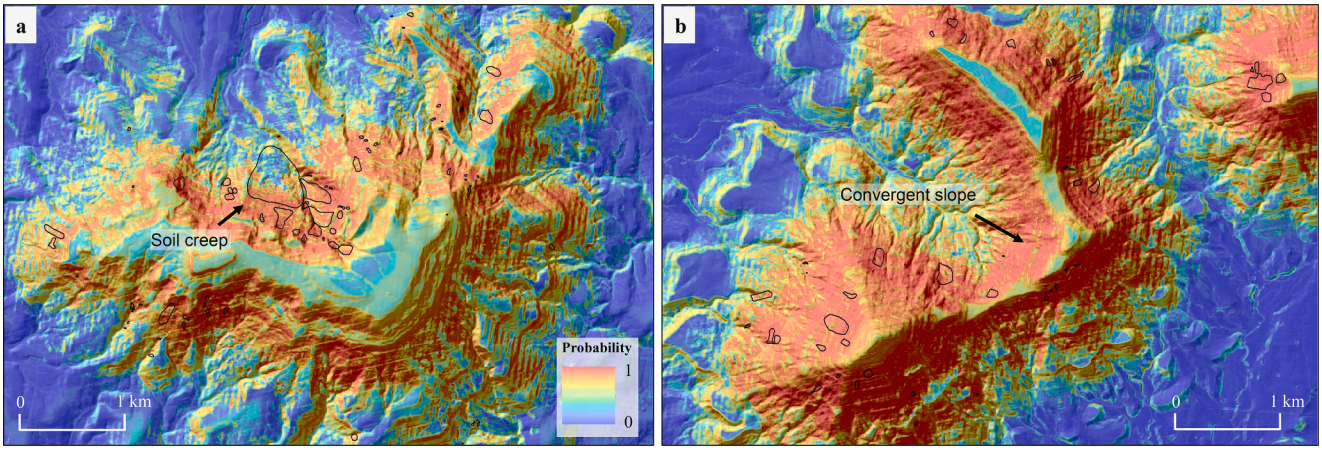


Fig. 9. Closer look of the landslide probability in: (a) Cavanal Hill, and (b) Sugarloaf Mountain. Polygons are the historic rapid landslides and arrows show soil creep areas shown in Fig. S3 in Supplementary Information S1. Note most of these hillslopes are dominated by soil creep.

average gradient (Larsen and Montgomery, 2012; Larsen et al., 2010; Regmi et al., 2014b). For example, a study conducted by Regmi et al. (2014b) shows that the average gradient of landslides in Colorado Rockies is ~26° with a standard deviation of 10°. Similarly, a study of more than 15,000 landslides across different regions in eastern Himalaya by Larsen and Montgomery (2012) shows that the mode of the landslide gradient in this region ranges from 34°–39°. The reason for landslides on shallower slopes in our study can be attributed to the characteristics of the soil cover (i.e. granular soil) and fractured and weathered bedrock lithology, which tends to have a smaller angle of repose compared to that of resistant bedrocks in high-relief regions. Additionally, the landslide susceptibility map shows high susceptibility along zones of topographic convergence (Figs. 9 and 10d) where soil thickness tends to be relatively high compared to the surroundings (Pelletier and Rasmussen, 2009). These areas can be sensitive to slight changes in soil–water flux as a function of precipitation as well as ground shaking (Lee and Kim, 2020). Furthermore, dissolution of limestone, a secondary rock type across the study area (Regmi and Walter, 2020b), and tree-root associated subsurface flow paths are some of the hillslope environments that can facilitate high infiltration,

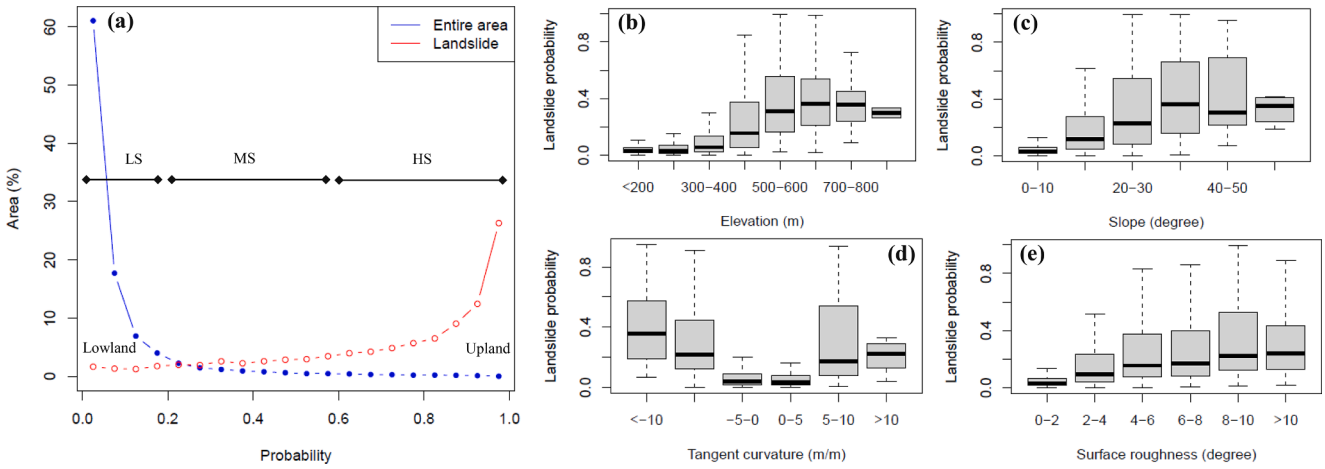


Fig. 10. Distribution of landslide probability values: (a) across the study area and within the landslide, and (b, c, d, e) within the classes of four hillslope geometries including elevation, slope, curvature and surface roughness. LS: Low Susceptibility, MS: Medium Susceptibility, HS: High Susceptibility.

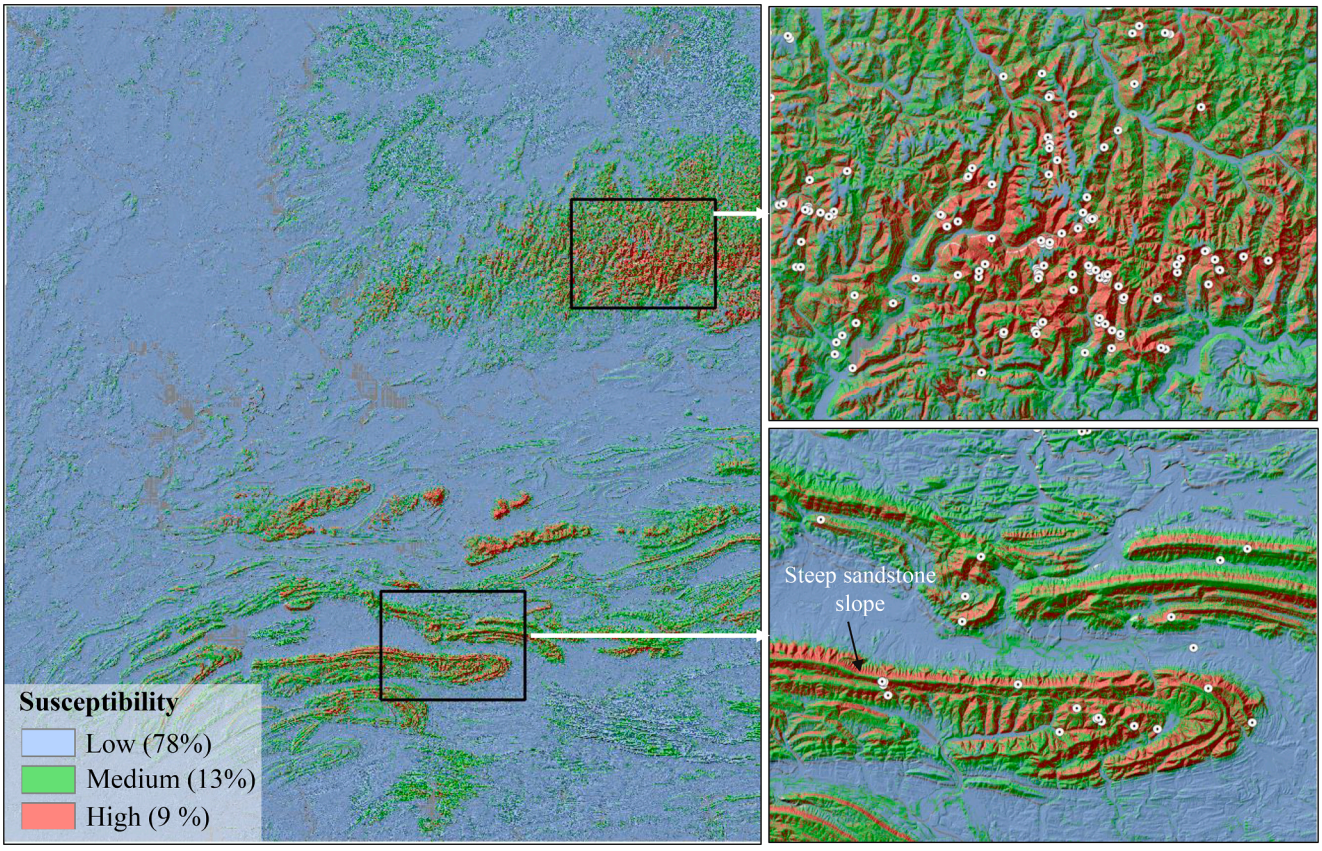


Fig. 11. A landslide susceptibility map developed by classifying landslide probability map (Fig. 8a). High, medium, and low susceptibility zones were classified as landslide probability > 0.6, 0.2–0.6 and < 0.2, respectively. The percentage within the bracket indicates the area coverage.

conductivity, and building up of pore-water pressure in soil, and, thereby, landslides during major storm and seismic events (Easter et al., 1991; Kay, 1998; Regmi and Walter, 2020b). We thus conclude that landsliding in the study area is controlled primarily by the combined effect of hillslope geometry, soil cover and underlying bedrock lithology.

Another notable observation is that the frequency of landslides (Fig. 5c and 6a) has increased significantly since 2005 (Fig. 5c). We believe that the increased induced seismicity in nearby north-central Oklahoma (Regmi and Walter, 2020a; Walter et al., 2020) and the frequent intense precipitation, such as the record precipitation of 2015 (Oklahoma Climatological Survey, 2018), are the major reasons for such a change in the rate of landsliding. The increased seismic shaking can contribute inelastic strain or granular energy that can cause the long-term weakening and slow deformation (i.e., creeping soil) of the soil-mantled hillslopes (Delchiaro et al., 2023). Eventually, in concert with precipitation or other triggering forces, this weakening can trigger landslides. Increased anthropogenic activities, such as urbanization, infrastructure development and land use/landcover change (Lopez et al., 2008; Phillips et al., 2020; Regmi and Walter, 2020a; Tappe et al., 2004) are additional possible factors responsible for triggering recent landslides in the study area. These changes in triggering forces can significantly influence the stability thresholds of previously stable

hillslopes (i.e. gentle slopes) or slowly moving slopes (soil creep) (Ferdowsi et al., 2018; Roering et al., 2001). These characteristics imply that the area would likely be more susceptible to landslides in the future and highlight the requirement of detailed landslide susceptibility mapping, continuous monitoring of highly unstable hillslopes, and implementation of landslide mitigation efforts. These types of low-relief landscapes have often been overlooked and less well-studied. As a result, potential changes in landslide susceptibility have gone unstudied. In other regions these types of shifts may be driven by interannual to decadal changes in urbanization and land use change, regional climate change, or shifts between seismic quiescence to high activity.

5.3. Landslide susceptibility map

The landslide susceptibility map developed here is based on the static covariates only, yet it successfully shows the locations of observed landslides within the high susceptibility zones. If the characteristics of input covariates and the landslide triggering forces do not change significantly in the future, it is very likely that future landslides would occur in areas of high susceptibility, as mapped by this model. The mapping performance (~94 %) of our model (Fig. 8, Tables 3 and 4) is satisfactory and as good as well-performing models implemented worldwide (Nefeslioglu et al., 2008; Neuhäuser and Terhorst, 2007; Regmi et al., 2014a). The high performance of the model could simply reflect the large number of landslide observations (717 landslides) and sampling landslide data from entire landslide scarps. Indeed, sampling from entire scarps was found better for mapping landslide susceptibility (Dou et al., 2020; Regmi et al., 2014a), potentially because scarp sampling can significantly reduce the variability in input covariate values, generate enough data required for the modeling and also introduce the role of landslide size in landslide susceptibility mapping.

The random forest workflow demonstrates a robust first-order identification of hazardous areas. Thus, the algorithm can be used as a tool to identify landslide sensitive areas for further detailed investigation, such as for continuous monitoring of high susceptibility regions implementing time-series remote sensing and geophysical techniques. Such an application would help us improve the model and further explore the feedback-response and potentially identify nucleation mechanisms of landslides in these hillslopes during extreme triggering events. The approach is data-driven supervised learning that can incorporate a large number of spatially continuous and categorical covariates in predicting landslides. In this view, the performance of the model can further be improved by including additional covariates, such as hillslope length and relief (Grieve et al., 2016), soil depth (Ochsner et al., 2019; Pelletier and Rasmussen, 2009), and landforms (Evans, 2012) as well as dynamic variables including soil moisture and precipitation (NASA, 2021). In addition, the model can be updated and improved with more landslides and higher resolution datasets, such as geology and soil physiochemical properties that may be available in the future.

6. Conclusions

Our work suggests that soil-mantled modest topography in an intraplate setting has the potential to occur landslides at rates and sizes that are similar to that of many tectonically active and high-relief regions. The increase in frequency of small-sized landslides after 2005, coinciding with Oklahoma's increased occurrence of low-magnitude induced sciesmity and extereme storm events, indicates that the changes in the rates and/or magnitude of triggering mechanisms influence the frequency and size distribution of landslides. Spatial and temporal frequency computed for this study imply landslides will occur frequently in the study area. The random forest approach and covariates employed here provide reasonable estimates of landslide susceptibility with performance as good as most of the qualitative and quantitative or statistical well-performed models implemented worldwide and highlights the effectiveness of the approach in mapping landslides. The model suggests elevation or relief, slope steepness, stream power, lithology, and soil characteristics are some of the major factors that play a significant role in the occurrence of landslides in the Ozark and Ouachita Mountains. Both observed and predicted landslides are found both in upland and lowland environments, particularly in steep slopes, zones of topographic convergence, shale and sandstone, and near roads and streams. The results obtained here, including the susceptibility map, quantification of each factor's contribution towards causing landslides, and landslide frequency-size characteristics are a vital dataset for hillslope monitoring, landslide mitigation, and slope management. These observations can also provide first-order insights on hillslope processes in low-relief intraplate regions.

CRediT authorship contribution statement

Netra R. Regmi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jacob I. Walter:** Writing – review & editing, Funding acquisition. **Junle Jiang:** Writing – review & editing, Funding acquisition. **Annalise M. Orban:** Writing – review & editing, Data curation. **Nicholas W. Hayman:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The research was funded by the FEMA Hazard Mitigation Grant Program under DR-4315-0030-OK via the Oklahoma Department of Emergency Management and partly contributed by NASA Earth Surface and Interior grant 80NSSC22K1723. We are grateful to David Brown, Russell Standridge and Francis Oyebanji of the Oklahoma Geological Survey for their valuable suggestions and assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.catena.2024.108344.

References

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65, 15–31.
Bellugi, D.G., Milledge, D.G., Cuffey, K.M., Dietrich, W.E., Larsen, L.G., 2021. Controls on the size distributions of shallow landslides. Proceedings of the National Academy of Sciences 118 (9), e2021855118.
Blöthe, J.H., Korup, O., Schwanghart, W., 2015. Large landslides lie low: Excess topography in the Himalaya-Karakoram ranges. Geology 43, 523–526.
Böhner, J., McCloy, K.R., Strobl, J., 2006. SAGA-analysis and modelling applications. Collect. Göttinger Geograp. Abhandlungen 115.
Booth, A.M., Roering, J.J., Perron, J.T., 2009. Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon. Geomorphology 109, 132–147.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. Taylor & Francis.
Carrara, A., Merenda, L., 1976. Landslide inventory in northern Calabria, southern Italy. Geol. Soc. Am. Bull. 87, 1153–1162.

Casadei, M., Dietrich, W., Miller, N., 2003. Testing a model for predicting the timing and location of shallow landslide initiation in soil-mantled landscapes. Earth Surf. Proc. Landforms 28, 925–950.
Cerato, A.B., Hong, Y., Yu, X., He, X., Tabet, W., 2014. Real time monitoring of slope stability in eastern Oklahoma. Oklahoma. Dept. of Transportation. Planning and Research Division, pp. 91.
Clarke, B.A., Burbank, D.W., 2010. Bedrock fracturing, threshold hillslopes, and limits to the magnitude of bedrock landslides. Earth Planet. Sci. Lett. 297, 577–586.
Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In: Zhang, C., Ma, Y. (Eds.), Ensemble Machine Learning. Springer, pp. 157–175.
Delchiaro, M., Della Seta, M., Martino, S., Nozaem, R., Moumeni, M., 2023. Tectonic deformation and landscape evolution inducing mass rock creep driven landslides: The Loumar case-study (Zagros Fold and Thrust Belt, Iran). Tectonophysics 846, 229655.
Díaz-Uriarte, R., De Andres, S.A., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinform. 7, 3.
Dietrich, W.E., Bellugi, D., De Asua, R.R., 2001. Validation of the shallow landslide model, SHALSTAB, for forest management. Water Sci. Appl. 2, 195–227.
Dou, J., et al., 2020. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Sci. Total Environ. 720, 137320.
Dunham, A.M., et al., 2022. Topographic control on ground motions and landslides from the 2015 Gorkha earthquake. Geophys. Res. Lett., e2022GL098582
Easter, K.W., Dixon, J.A., Hufschmidt, M.M. (Eds.), 1991. Watershed resources management: studies from Asia and the Pacific. Institute of Southeast Asian Studies.
Egan, J.P., Egan, J.P., 1975. Signal detection theory and ROC-analysis. Academic Press, New York.
Emberson, R., Kirschbaum, D., Stanley, T., 2020. New global characterisation of landslide exposure. Nat. Hazard. Earth Syst. Sci. 20, 3413–3424.
Evans, I.S., 2012. Geomorphometry and landform mapping: What is a landform? Geomorphology 137, 94–106.
Ferdowsi, B., Ortiz, C.P., Jerolmack, D.J., 2018. Glassy dynamics of landscape evolution. Proc. Natl. Acad. Sci. 115, 4827–4832.
Frankel, K.L., Dolan, J.F., 2007. Characterizing arid region alluvial fan surface roughness with airborne laser swath mapping digital topographic data. J. Geophys. Res.: Earth Surf. 112, F02025.
Frattini, P., Crosta, G.B., 2013. The role of material properties and landscape morphology on landslide size distributions. Earth Planet. Sci. Lett. 361, 310–319.
Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006. Spatial prediction of landslide hazard using logistic regression and ROC analysis. Transactions in GIS 10, 395–415.
Grieve, S.W., Mudd, S.M., Hurst, M.D., 2016. How long is a hillslope? Earth Surf. Proc. Landforms 41, 1039–1054.
Guthrie, R., Evans, S., 2004. Magnitude and frequency of landslides triggered by a storm event, Loughborough Inlet, British Columbia. Nat. Hazard. Earth Syst. Sci. 4, 475–483.
Haley, B. et al., 1993. Geologic map of Arkansas. 1: 500000 scale. Arkansas Geologic Commission, 1.
Hayes, C.J., 1971. Landslides and related phenomena pertaining to highway construction in Oklahoma. In: Environmental Aspects of Geology and Engineering in Oklahoma.
He, X., et al., 2014. Landslides susceptibility mapping in Oklahoma state using GIS-based weighted linear combination method. In: Sassa, K., Canuti, P., Yin, Y. (Eds.), Landslide Science for a Safer Geoenvironment. Springer, Cham, pp. 371–377.
Heimsath, A.M., Dietrich, W.E., Nishiizumi, K., Finkel, R.C., 1997. The soil production function and landscape equilibrium. Nature 388, 358–361.
Heran, W.D., Green, G.N., Stoeser, D.B., 2003. A digital geologic map database for the state of Oklahoma, USGS, Open-File Report 2003-247.
Hong, Y., Adler, R., Huffman, G., 2007. Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat. Hazard. 43, 245–256.
Huang, F., et al., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena 191, 104580.
Hurst, M., Ellis, M., Royse, K., Lee, K., Freeborough, K., 2013. Controls on the magnitude-frequency scaling of an inventory of secular landslides. Earth Surf. Dyn. 1, 67–78.
Jibson, R.W., Tanyaş, H., 2020. The influence of frequency and duration of seismic ground motion on the size of triggered landslides—a regional view. Eng. Geol. 273, 105671.
Johnson, K.S., Luza, K.V., 2008. Earth sciences and mineral resources of Oklahoma. Oklahoma Geological Survey, Norman, Oklahoma.
Kargel, J.S., et al., 2016. Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake. Science 351, aac8353.
Kay, B., 1998. Soil structure and organic carbon: a review. Soil Proc. Carbon Cycle 198, 169–197.
Kirschbaum, D.B., Adler, R., Hong, Y., Hill, S., Lerner-Lam, A., 2010. A global landslide catalog for hazard applications: method, results, and limitations. Nat. Hazard. 52, 561–575.
Korup, O., et al., 2007. Giant landslides, topography, and erosion. Earth Planet. Sci. Lett. 261, 578–589.
Kuhn, M. et al., 2020. caret: Classification and Regression Training. R package version 6.0-86. https://cran.r-project.org/web/packages/caret/caret.pdf.
Lacroix, P., Handwerger, A.L., Bièvre, G., 2020. Life and death of slow-moving landslides. Nat. Rev. Earth Environ. 1, 404–419.
LaHusen, S.R., Duvall, A.R., Booth, A.M., Montgomery, D.R., 2016. Surface roughness dating of long-runout landslides near Oso, Washington (USA), reveals persistent postglacial hillslope instability. Geology 44, 111–114.
Larsen, I.J., Montgomery, D.R., 2012. Landslide erosion coupled to tectonics and river incision. Nat. Geosci. 5, 468–473.
Larsen, I.J., Montgomery, D.R., Korup, O., 2010. Landslide erosion controlled by hillslope material. Nat. Geosci. 3, 247–251.
Lee, S.-R., Kim, Y.-T., 2020. Spatial probability assessment of landslide considering increases in pore-water pressure during rainfall and earthquakes: case studies at Atsuma and Mt. Umyeon. Catena 187, 104317.
Lee, S., Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environ. Geol. 40, 1095–1113.
Lee, S., Ryu, J.-H., Won, J.-S., Park, H.-J., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng. Geol. 71, 289–302.
Liu, S., Wang, L., Zhang, W., He, Y., Pijush, S., 2023. A comprehensive review of machine learning-based methods in landslide susceptibility mapping. Geol. J. 58, 2283–2301.
Liu, S. et al., 2024. Physics-informed optimization for a data-driven approach in landslide susceptibility evaluation. Journal of Rock Mechanics and Geotechnical Engineering, In Press.
Lopez, R.D., Nash, M.S., Heggem, D.T., Ebert, D.W., 2008. Watershed vulnerability predictions for the Ozarks using landscape models. J. Environ. Qual. 37, 1769–1780.
Luza, K.V., Johnson, K.S., 2005. Geologic Hazards in Oklahoma. Oklahoma Geology Notes 63, 52–70.
Maguigan, M., Dixon, J., Brye, K., 2015. Factors influencing the formation of shallow landslides in the Boston Mountains of northwest Arkansas, USA. Phys. Geography 36, 426–447.
Medwedeff, W.G., Clark, M.K., Zekkos, D., West, A.J., 2020. Characteristic landslide distributions: An investigation of landscape controls on landslide size. Earth Planet. Sci. Lett. 539, 116203.
Merghadi, A., et al., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews 207, 103225.
NASA, 2021. Predicting Landslide Hazards in Near Real-Time. https://earthobservatory.nasa.gov/images/92018/predicting-landslide-hazards-in-near-real-time.
Nefsliogluo, 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. Eng. Geol. 97, 171–191.
Neuhäuser, B., Terhorst, B., 2007. Landslide susceptibility assessment using "weights-of-evidence" applied to a study area at the Jurassic escarpment (SW-Germany). Geomorphology 86, 12–24.
NEWS, June 6, 2019. State Hwy. 220 South of Devil's Den closed due to rock slide. https://www.newsonline.com/article/news/local/outreach/back-to-school/state-hwy-220-south-of-devils-den-closed-due-to-rock-slide/27-968d6ae-9a0d-4e67-bed4-bc6443338974.
NEWS, April 13, 2020. Landslide damages water lines in Sequoyah County. https://www.newsonline.com/article/news/landslide-damages-water-lines-in-sequoyah-county/27-a1608768-c4f-4937-90c6-4b0e43687fc.
Newville, M. et al., 2023. lmfit/lmfit-py: 1.2.2. Zenodo.
Oakes, M.C., 1952. Landslides in Le Flore County. Oklahoma Geological Survey Publication, 12.
Ochsner, T.E., Linde, E., Haffner, M., Dong, J., 2019. Mesoscale soil moisture patterns revealed using a sparse in situ network and regression kriging. Water Res. Rese. 55, 4785–4800.
Oklahoma Climatological Survey, 2018. Climate of Oklahoma. https://climate.ok.gov/index.php/site/page/climate_of_oklahoma.
Ouimet, W.B., Whipple, K.X., Granger, D.E., 2009. Beyond threshold hillslopes: Channel adjustment to base-level fall in tectonically active mountain ranges. Geology 37, 579–582.
Pachauri, A., Pant, M., 1992. Landslide hazard mapping based on geological attributes. Eng. Geol. 32, 81–100.
Pack, R.T., Tarboton, D.G., Goodwin, C.N., 1998. The SINMAP approach to terrain stability mapping. in: Moore, D., Hungr, O. (Eds.), 8th Congress of the International Association of Engineering Geology, Vancouver, Canada, 21–25 Sept 1998.
Pánek, T., Březný, M., Kapustová, V., Lenart, J., Chalupa, V., 2019. Large landslides and deep-seated gravitational slope deformations in the Czech Flysch Carpathians: New LiDAR-based inventory. Geomorphology 346, 106852.
Pelletier, J.D., Rasmussen, C., 2009. Geomorphically based predictive mapping of soil thickness in upland watersheds. Water Resources Research 45, W09417.
Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40, 927–930.
Phillips, J.D., Marion, D.A., Luckow, K., Adams, K.R., 2005. Nonequilibrium regolith thickness in the Ouachita Mountains. J. Geol. 113, 325–340.
Phillips, J.D., Marion, D.A., Kilcoyne, K.G., 2020. Concentration and divergence of sediment in an erosional landscape. Geomorphology 367, 107281.
Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., 2012. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat. Hazard. 63, 965–996.
Rahimi, S., Wood, C.M., Bernhardt-Barry, M., 2021. The MHVSR technique as a rapid, cost-effective, and noninvasive method for landslide investigation: case studies of Sand Gap and Ozark, AR, USA. Landslides 18, 2705–2720.
Regmi, N.R., Walter, J.I., 2020a. Detailed mapping of shallow landslides in eastern Oklahoma and western Arkansas and potential triggering by Oklahoma earthquakes. Geomorphology 366, 106806.
Regmi, N.R., Giardino, J.R., Vitek, J.D., 2014b. Characteristics of landslides in western Colorado, USA. Landslides 11, 589–603.
Regmi, N.R., Walter, J.I., 2020b. Landslide Hazards in Eastern Oklahoma Mountains. Oklahoma Geology Notes 79, 5–16.
Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010a. Assessing susceptibility to landslides: using models to understand observed changes in slopes. Geomorphology 122, 25–38.

Regmi, N.R., Giardino, J.R., Vitek, J.D., Dangol, V., 2010b. Mapping landslide hazards in western Nepal: comparing qualitative and quantitative approaches. Environ. Eng. Geosci. 16, 127–142.
Regmi, N.R., Giardino, J.R., McDonald, E.V., Vitek, J.D., 2014a. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. Landslides 11, 247–262.
Roering, J.J., Kirchner, J.W., Sklar, L.S., Dietrich, W.E., 2001. Hillslope evolution by nonlinear creep and landsliding: An experimental study. Geology 29, 143–146.
Samia, J., et al., 2017. Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. Landslides 14, 547–558.
Silverman, B.W., 1984. Spline smoothing: the equivalent variable kernel method. Annals of Statistics 898–916.
Søreide, K., 2009. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. J. Clin. Pathol. 62, 1–5.
Stanley, T., Kirschbaum, D.B., 2017. A heuristic approach to global landslide susceptibility mapping. Nat. Hazard. 87, 145–164.
Stark, C.P., Hovius, N., 2001. The characterization of landslide size distributions. Geophys. Res. Lett. 28, 1091–1094.
Stoeser, D.B. et al., 2005. Preliminary integrated geologic map databases for the United States. US Geological Survey, Open-File Report, 1351.
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinform. 8, 25.
Svetnik, V., et al., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inform. Comp. Sci. 43, 1947–1958.
Taalab, K., Cheng, T., Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data 2, 159–178.
Tanyaş, H., van Westen, C.J., Allstadt, K.E., Jibson, R.W., 2019. Factors controlling landslide frequency–area distributions. Earth Surf. Proc. Landforms 44, 900–917.
Tappe, P.A., Weih Jr, R.C., Thill, R.E., Melchiors, M.A., Wigley, T.B., 2004. Lansdcape characterization of four watersheds under different forest management scenarios in the Ouachita Mountains of Arkansas, Ouachita and Ozark Mountains symposium: ecosystem management research. United States Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, USA, pp. 146-153.
Tatard, L., Grasso, J., Helmstetter, A., Garambois, S., 2010. Characterization and comparison of landslide triggering in different tectonic and climatic settings. J. Geophys. Res.: Earth Surf. 115, F04040.
Thaler, E., Covington, M.D., 2016. The influence of sandstone caprock material on bedrock channel steepness within a tectonically passive setting: Buffalo National River Basin, Arkansas, USA. J. Geophys. Res.: Earth Surf. 121, 1635–1650.
Thomas, G., Michael, J., 2004. A review of scale dependency in landslide hazard and risk analysis. in: Thomas, G., Michael, J. (Eds.), Landslide Hazard and Risk. John Wiley & Sons, Ltd, pp. 102-114.
USDA Natural Resources Conservation Service, 2016. U.S. General Soil Map (STATSGO2), https://data.nal.usda.gov/dataset/united-states-general-soil-map-statsgo2.
USGS, 2016. GAP/LANDFIRE National Terrestrial Ecosystems 2011. U.S. Geological Survey data release, https://doi.org/10.5066/F7ZS2TM0.
Van Den Eeckhaut, M., Poesen, J., Govers, G., Verstraeten, G., Demoulin, A., 2007. Characteristics of the size distribution of recent and historical landslides in a populated hilly region. Earth Planet. Sci. Lett. 256, 588–603.
Varnes, D.J., 1978. Slope movement types and processes. Special Report 176, 11–33.
Veronesi, F., Hurni, L., 2014. Random Forest with semantic tie points for classifying landforms and creating rigorous shaded relief representations. Geomorphology 224, 152–160.
Walter, J.I., et al., 2020. The Oklahoma geological survey statewide seismic network. Seismol. Res. Lett. 91, 611–621.
Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019. Machine learning-based predictive control of nonlinear processes. Part I: theory. AIChE J. 65, e16729.
Yalcin, A., 2008. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. Catena 72, 1–12.
Yamagishi, H., Yamazaki, F., 2018. Landslides by the 2018 Hokkaido Iburi-Tobu Earthquake on September 6. Landslides 15, 2521–2524.
Yilmaz, I., 2010. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. Environ. Earth Sci. 60, 505–519.
Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856.