Original Paper

Landslides
DOI 10.1007/s10346-023-02050-6
Received: 23 February 2022
Accepted: 23 February 2023

Tyler Rohan · Eitan Shelef · Ben Mirus · Tim Coleman

Prolonged influence of urbanization on landslide susceptibility

Abstract Landslides pose a threat to life and infrastructure and are influenced by anthropogenic modifications associated with land development. These modifications can affect susceptibility to landslides, and thus quantifying their influence on landslide occurrence can help design sustainable development efforts. Although landslide susceptibility has been shown to increase following urban expansion, the long-lasting effect of urbanization on landslide susceptibility remains largely unquantified. Hence, susceptibility maps developed based on inventories from non-urbanized areas may incorrectly evaluate the hazard in urbanized areas. To quantify this effect, we analyzed a landslide inventory from southwestern Pennsylvania, where the pulse of urbanization occurred more than a decade before the inventory was created. Using road density as a proxy for urbanization, the study area was divided into urbanized and non-urbanized areas. Susceptibility patterns were computed using statistical analyses of a post-urbanization landslide inventory together with maps of topographic, land cover, and geologic factors. A pre-urbanization landslide inventory was used as a control. Our findings indicate that urbanization has a decades-long effect on landslide susceptibility, where urbanized areas are generally more susceptible to landslides. In urbanized areas landslides are strongly associated with distance from roads and topographic curvature, whereas in non-urbanized landslides are strongly associated with stratigraphic formation and distance from streams. The consistent differences in susceptibility patterns between urbanized and non-urbanized areas indicate that urbanization has a long-lasting effect on landslide susceptibility and that susceptibility estimates should be made separately for these different environments to account for the persistent influence of urbanization.

Keywords Landslides · Landslide susceptibility · Urbanization · Hazard mapping

Introduction

The risk of landslide damage to life and infrastructure makes landslide susceptibility and hazard zoning crucial for proper land-use planning in urbanized areas. Urbanization, defined as increasing the density of population through urban settlement, is associated with deforestation and construction that typically reduce vegetation cover and modify hydrology and topography (Tubbs 1974; Zêzere et al. 1999; Glade 2003; Cascini et al. 2005; Van Den Eeckhaut et al. 2007; Papathoma-Koehle and Glade 2013). These modifications can alter topography, as well as surface and subsurface flow paths, which influence landslide susceptibility (e.g., Mirus et al. 2007; BeVille et al. 2010). Thus, quantifying the influence of urbanization on landslide occurrence can help guide the design of more sustainable infrastructure (Tarolli and Sofia 2016).

Landslide occurrence is governed by gravitational forces that transport earth materials downslope (Cruden and Varnes 1996). The location of a landslide depends on a multitude of landslide-related factors such as the magnitude of topographic slope, soil/rock properties and thicknesses, the inclination of layered rock units, reinforcement due to vegetation, and hydrologic factors that may reduce the frictional-resistance of soil and rock by increasing pore-pressure (Iverson 2000; Wang et al. 2013; Pfeil-McCullough et al. 2015; Bogaard and Greco 2016). These landslide-related factors may covary in linear and non-linear ways that influence the magnitude and likelihood of landslide occurrence. Because urbanization modifies some or all these factors, it also affects landslide susceptibility (Johnston et al. 2021).

Landslide susceptibility estimates can be quantitative or qualitative. Qualitative estimates use descriptive terms to categorize landslide susceptibility levels, whereas quantitative methods rely on statistical analyses to estimate probabilities of occurrence for landslides (Reichenbach et al. 2018). Statistical methods, such as conditional probability and machine learning approaches, can account for such non-linearities and produce accurate landslide susceptibility estimates over large areas at a high spatial resolution (Pourghasemi et al. 2012; Merghadi et al. 2020; Wang et al. 2021). These methods typically rely on a large dataset of landslide locations and utilize the covariance between landslide-related factors and the occurrence of landslides to weigh these factors and calculate the relative likelihood of landslide occurrence. The estimated likelihoods can be used to spatially map landslide susceptibility and guide future development and mitigation efforts (Zhang et al. 2017; Kim et al. 2018).

Studies that focus on the influence of different factors on landslide susceptibility (e.g., Dai and Lee 2002; Rohan et al. 2021; Zhou et al. 2020) often explore what geologic, topographic, and land-use factors affect landslide susceptibility (Glade 2003; Van Beek and Van Asch 2004; Wasowski et al. 2010; Chen and Huang 2013; Reichenbach et al. 2014; Pisano et al. 2017; Soma et al. 2017; Chen et al. 2019; Senanayake et al. 2020; Bernardie et al. 2021). Although land-use changes do influence landslide susceptibility, this influence depends on the local topographic, lithologic, and hydrologic conditions as well as on the rate of land-use changes (Soma and Kubota 2017; Chen et al. 2019). Previous studies (Braun et al. 2019; Simon et al. 2015) indicated that changes in landslide susceptibility may increase with the rate of urban expansion and depend on population and road density. Although landslide susceptibility has been shown to generally increase with the degree and rate of urbanization, studies have previously used only landslide inventories that were mapped in areas and times of urban expansion and thus primarily explored the immediate effect of urbanization on

landslide susceptibility (Alexander 1986; Smyth and Royle 2000; Sassa et al. 2004; Cascini et al. 2005; Dragićević et al. 2015; Frodella et al. 2018; Lee et al. 2018). Hence, the legacy influence of urbanization that occurred decades ago on landslide susceptibility remains largely unquantified.

A recent review of various national-scale landslide susceptibility maps reveals that southwestern Pennsylvania is consistently among the most susceptible areas in the conterminous United States (Mirus et al. 2020). The city of Pittsburgh, located at the heart of this region, is particularly susceptible to landslides (Pomeroy 1982; Ashland 2021). The history of southwestern Pennsylvania contains both periods of extreme population rise and fall. The study area encompasses four counties (Allegheny, Beaver, Washington, and Westmoreland), which all saw substantial population growth during 1890–1950 due to the flourishing local steel and coal-based economy. The population stabilized or dropped in each of the counties with the decline of the steel industry after 1950 (Giarratani and Houston 1989). Although the peak population generally occurred in the late 1950s, ~ 80% of the population growth occurred before 1940. Pittsburgh, the largest urban area in southwestern Pennsylvania, currently has a population that is 55.6% smaller than its size in 1950 (Winant 2021). The pulse of urban expansion in southwestern Pennsylvania between the 1890s and 1950s, and the lack of meaningful urban expansion since, presents a unique opportunity to explore the decades-long effect of urbanization on landslide susceptibility.

This research examines the prolonged influence of urbanization on landslide susceptibility by using pre- and post-urbanization landslide inventories from southwestern Pennsylvania, with road density as a proxy for the spatial and temporal pattern of urbanization. The pre-urbanization inventory is used as a control for potential biases in landslide mapping and analyses because it is not expected to differ between urbanized and non-urbanized areas. The post-urbanization inventory is used to quantify the difference in both landslide susceptibility estimates and in landslide-related factors between urbanized and non-urbanized areas (classified via road density).

Methods and study area

Study area

This study is focused on a large portion of southwestern Pennsylvania that includes Allegheny, Beaver, Washington, and Westmoreland Counties, with a combined area of 7993 km² (Fig. 1A). The area is primarily drained by two major rivers, the Allegheny and Monongahela, that merge to form the Ohio River at the center of the study area in the city of Pittsburgh. Much of the study area is located within the Allegheny Plateau section of the Appalachian Plateaus province. Bedrock units exposed in the study area are composed of horizontally bedded or slightly dipping sedimentary rocks. The surficial deposits of the area are primarily composed of sand, shale, alluvium, and gravel. Western Pennsylvanian strata are composed of cyclic sequences of sandstone, shale, claystone, coal, and limestone. Clay- and silt-rich soils overlie the bedrock and can be up to 30 m thick at the base of slopes. Stream and river valleys are often steep, with a local relief of tens to few hundred meters including steep soil-mantled slopes, rock/cliff faces, and steep riverbanks (a table of the range of local topographic factors is provided in supplementary information: Table S2). A large number of prehistoric slope failures formed landslide deposits that are common throughout the study area (Pomeroy 1982).Southwestern Pennsylvania's high susceptibility to landslides is due to a combination of topographic, climatic, lithologic, and anthropogenic factors. Precipitation (both rain and snow) is distributed throughout the year with a mean annual precipitation of 1006 mm (from 1900 to 2020; National Oceanic and Atmospheric Administration and National Centers for Environmental Information 2021). The majority of high intensity rainstorms associated with the triggering landslides occur between the months of June to October (Ashland 2021). Lithologic units referred to locally as the Pittsburgh "red beds" consist of shales and clay that are particularly susceptible for landslides. The urbanization and industrialization of the study area were associated with construction of roads, pipelines, railroad, extensive coal mining, and commercial and residential properties. This modification of the landscape results in decreased vegetation and increased impervious cover.

Road density

Road density, a widely used proxy for urbanization (Zope et al. 2016; Theodorou et al. 2021), was used to estimate the extent of urbanization and compare landslide susceptibility between urbanized and non-urbanized areas. The road density map calculated for this study area was generated by computing road density (total road length divided by area) over a circular kernel of 300-m diameter (~ 3 city blocks). The road density map was divided into urbanized and non-urbanized areas by using Gini impurity, a statistical metric commonly used in data classification (Archer 2010), to find a road density threshold that best classifies the values in the map into two groups that we associate with urbanized and non-urbanized areas (Fig. 1B). The road-density threshold produced through this method (3.3 km/km²) is within the typical range (2.5–6.1 km/km²) used in the literature to distinguish between urbanized and non-urbanized areas (McAdoo et al. 2018). These urbanized and non-urbanized areas can then be analyzed individually to quantify and compare landslide susceptibility as well as rank the importance of landslide-related factors between them.

The road density map is based on a digital road map produced by the Pennsylvania Division of Computer Services Geographic Information Systems Group (Pennsylvania Spatial Data Access (PASDA) 2007). Given that urbanization generally halted since the era of its peak development (~ late 1950s; Rappaport 2003), we assume that the current road map is generally similar to the road map at the time of this peak development. To evaluate this assumption, historical road maps ranging from 1947 to 1977 compiled by the Pennsylvania Department of Transportation (PennDOT) were visually compared with the digital road map from 2015. This inspection revealed that although some road construction occurred in southwestern Pennsylvania since the time of peak development, the majority of the construction has been associated with densification of previously urbanized areas or conversion of commercial/industrial zones into residential areas.

Digital elevation data

The Digital Elevation Model (DEM) used in this study was clipped 100 m away from the extent of the outermost landslide locations in a landslide inventory produced by the US Geologic Survey (USGS) and used in this study. The resolution of the DEM used for this

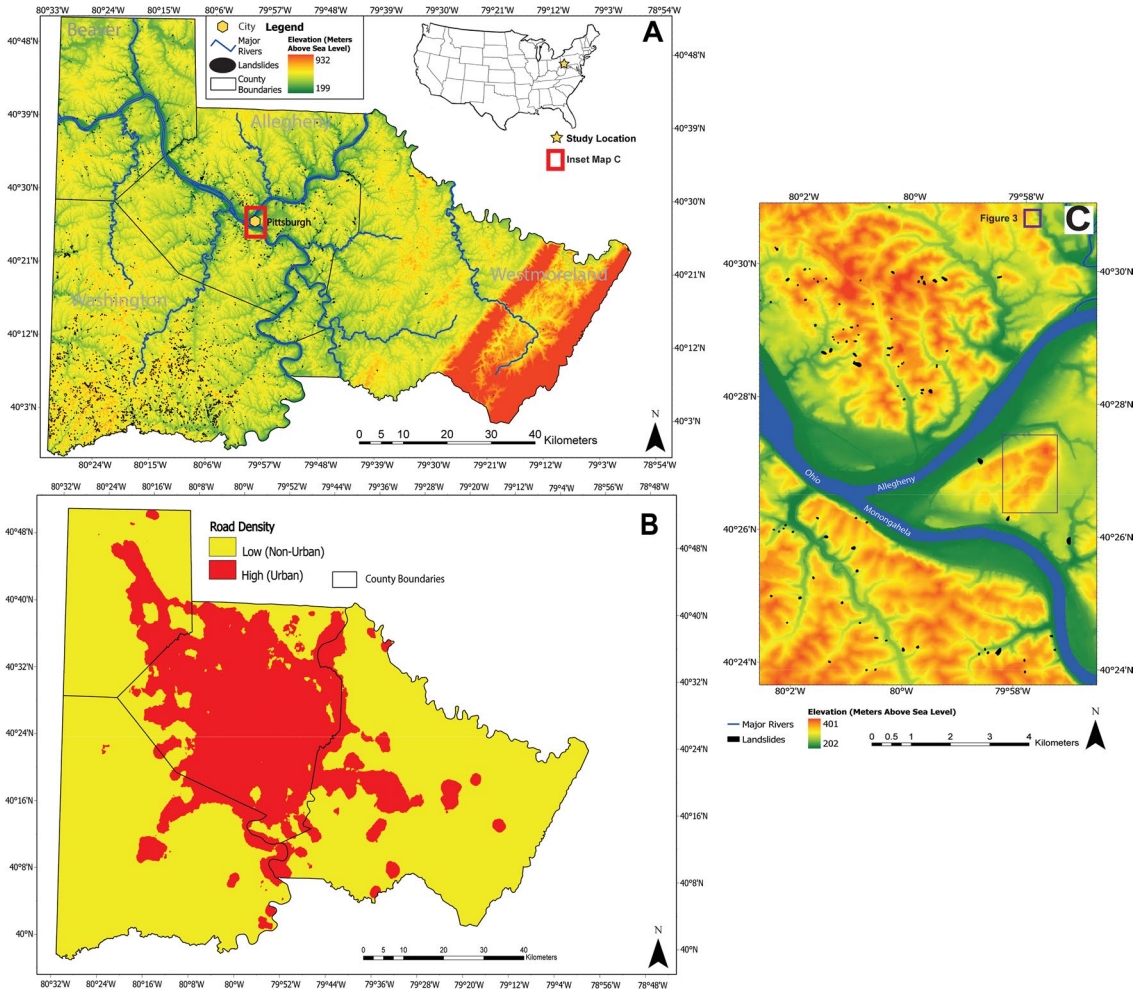

Fig. 1 **A** Elevation map of the study area overlain by county lines, major rivers, cities, and active landslides. **B** Map of urban (red) and non-urban (yellow) areas as separated by road density (threshold 3.3 km/km2). **C** An example zoomed in map of a subset of the study area showing major rivers and mapped landslides

study (1/3 arc-second, approximately 10 m) captures the scale of the mapped landslides (typically tens to hundreds of meters) and landslide-related factors at a resolution that can be efficiently analyzed. The DEM, obtained from the National Elevation Dataset (U. S. Geological Survey EROS Data Center 1999), is a seamless mosaic of best-available bare earth elevation data for the conterminous United States.

Landslide inventory

The landslide inventory used in this study is based on landslide maps produced by the USGS between the 1970s and 1980s (Briggs et al. 1975; Pomeroy 1982). These landslides were mapped as polygons via field reconnaissance combined with interpretation of aerial photographs from 1973 to 1975 (scale 1:24,000) (Pomeroy 1977; Fig. 1C). The landslides are defined as either active or old landslides. Active landslides are characterized by recent evidence of a landslide motion at the time of mapping and thus likely postdate the urbanization of the study area (Pomeroy 1977, 1982; Briggs et al. 1975). The active landslides are classified as shallow translational landslides that are typically less than 3 m thick as well as slumps of fill material generally associated with construction along hillslopes (Pomeroy 1982). The average area of a mapped active landslide in the inventory is 0.004 km^2 with maximum and minimum areas of 0.178 km^2 and 0.0001 km^2, respectively. Given that peak urbanization in the study area occurred during the late 1950s and that 80% of this urbanization had occurred by the 1940s, these active landslides formed or sustained more than a decade since urbanization reached its peak and approximately 3 decades since 80% of this peak urbanization had already taken place. Old landslides are inactive and defined based on hummocky landscape and deposits characteristics. They have formed since the Wisconsin Glaciation (Pomeroy 1982, and citation therein refer to landslide ages of 8.5–10 ka BP in the upper Ohio Valley based on ^{14}C ages) and likely predate the urbanization of the study area. The area of old landslides is larger than that of active ones (average, maximum, and minimum of 0.192, 1.56, and 0.094 km^2, respectively) and may represent the cumulative area of multiple small landslides that occurred over time in proximity to each other. A digitized dataset

of the active landslides is available through the Pennsylvania Spatial Data Access (Pennsylvania Spatial Data Access (PASDA) 2017). Old landslides were digitized as a part of this study from the aforementioned USGS landslide maps. To explore the association between urbanization and landslide occurrence, the inventory was divided into four classes: (1) Active landslides in urbanized areas (i.e., areas of high road density, hereafter Active-Urbanized) ($N = 1762$), (2) Active landslides in non-urbanized areas (i.e., areas of low road density, hereafter Active-Non-Urbanized) ($N = 1581$), (3) Old landslides in urbanized areas (hereafter Old-Urbanized) ($N = 1319$), and (4) Old landslides in non-urbanized areas (hereafter Old-Non-Urbanized) ($N = 1348$).

Landslide-related factors

To produce and compare landslide susceptibility estimates based on the four different landslide inventory classes, the landslides in each inventory were analyzed in the context of 11 topographic and environmental factors that may be associated with landslide occurrence (i.e., landslide-related factors). Seven topographic factors (slope, profile curvature, drainage area, relative location on hillslope, distance from nearest channel, elevation, and aspect) were computed from the aforementioned DEM. Slope was calculated from the DEM using an 8-connected neighborhood as the magnitude of the gradient vector and expressed in degrees. Profile curvature (1/meter) was calculated as the along-profile divergence of topographic slope using TopoToolBox (Schwanghart and Scherler 2014). Relative location on hillslope (between 0 and 1) was computed as the fraction of elevation relative to the hillslope relief, as estimated from the local relief value over a circular disk with a 200-m radius (a typical hillslope length in the study area). Aspect is calculated using the surface normal to the bicubic interpolation of the digital elevation data to identify the downslope direction of a pixel and is measured clockwise in azimuth degrees, where 0 and 360 are due north. All the topographic factors mentioned above are computed from the aforementioned National Elevation Dataset DEM at a 10-m resolution.

Non-topographic factors (distance to nearest road, stratigraphic formation, vegetation cover, land use) relied on various data sources and were resampled to the resolution of the DEM. Distance to roads was obtained from the same street centerlines used in the road density calculation that were mapped by Pennsylvania Division of Computer Services Geographic Information Systems Group. Distance to roads was computed as the Cartesian distance to the nearest road section. Distance to nearest stream was similarly computed using the Cartesian distance to the nearest stream, where streams are defined as DEM pixels having cumulative upstream drainage area larger than 10 km^2. Ten different lithologic units were derived from digital geologic maps (Miles et al. 2001). Vegetation cover, the percent of vegetation covering the area of a raster cell, was extracted from PASDA based on information collected in 2000. Land-use categories rely on photogrammetrically compiled information collected in 2015 (Yang et al. 2018) and are divided to the following nine classes: open water, forest, developed low intensity, developed medium intensity, developed high intensity, developed open space, shrub, barren, and pasture. The cumulative area of impermeable cover that drains to each DEM pixel was calculated using the land-use data in conjunction with flow accumulation, and was computed using TopoToolBox (Schwanghart and Scherler 2014). See supplementary information (Table S2) showing the range and mean of all landslide-related factors. Given that urbanization generally halted since the era of its peak development (~ 1950), we assume that current vegetation and landcover attributes are generally similar to those at the time of landslide mapping (1970s). To facilitate the susceptibility analysis, the maps of factors whose resolution differs from the 10-m resolution of the DEM (stratigraphic formation 125 m, vegetation cover 30 m, and land-use 50 m) were resampled to a 10-m resolution.

The association between landslides and the topographic and non-topographic factors is based on the areal extent of mapped landslide polygons. To do so, a binary map of landslide positions was produced by assigning a value of 1 for pixels within landslide polygons or that overlap with polygon boundaries and a value of 0 for pixels outside of such polygons. The values of topographic and non-topographic factors from all pixels contained within landslides are then used in our analysis.

Random forest and ROC validation

We used random forest and conditional probability to map landslide susceptibility and to explore the association of landslides with the aforementioned 11 factors. Random forest analysis (Hastie et al. 2009) is based on an assembly of decision trees and can be used to make probabilistic predictions and to rank the importance of factors that are associated with landslide occurrence (Maxwell et al. 2018). Each decision tree divides the data into more homogenous subsets based on a recursive procedure that identifies the factors (and associated thresholds) that best divide a target variable data (Catani et al. 2013). In the random forest model, every tree is trained using a subset of training samples and factors. The randomness introduced by subsampling observations and restricting the factors available at each node leads to less accurate individual trees that are less correlated with each other, reducing the variance of the ensemble forest (Culter et al. 2007; He et al. 2017), and can be used to quantify the probabilistic prediction of each factor. We used the randomForest package in R-CRAN with a bagging ensemble method (Liaw and Wiener 2002). The length of the input dataset is the number of DEM pixels ($N = 23{,}262{,}133$ and $N = 33{,}473{,}337$ for urbanized and non-urbanized areas, respectively), the target variable is landslide occurrence (i.e., binary, based on the USGS dataset), and the input variables are the previously mentioned 11 landslide-related factors. The factors used in the analysis randomly vary in each decision tree, with six factors used for each node. To handle the imbalance in the input data, each tree was structured to subsample 75% of the overall pixels for training, while the remaining 25% (i.e., the out of bag samples) were used for validation. The model was trained using 500 trees.

To produce a landslide susceptibility map, the model computes landslide susceptibility in each pixel. Each trained tree in the random forest model uses the landslide-related factors associated with each pixel to make a binary prediction regarding whether the pixel is associated with landslide occurrence. The susceptibility of the pixel to landslides is computed as the fraction of positive decisions (i.e., landslide occurrence) that a pixel receives from all trees. The susceptibility values can be converted into percentiles based on the range of all the computed susceptibility values in the map; hence, pixels of higher percentile signify a relatively higher risk of landslide occurrence.

We used partial dependence analysis to further explore the influence of high-ranking continuous landslide-related factors on landslide susceptibility. Partial dependence plots show the marginal effect of landslide-related factors on the predicted outcome of the model (Friedman 2001) and thus can display the relations between landslide susceptibility and related factors. The partial dependence at a particular factor value is computed by forcing all data points (i.e., map pixels) to assume that factor value while holding the other factors at their original value for each pixel and computing the mean susceptibility prediction from all trees. This reveals the relations between different factor values and landslide susceptibility, and thus provides valuable insights into the random forest model predictions.

The performance of the random forest model was evaluated through a Receiver Operating Characteristic (ROC) curve and calculation of the area under the ROC curve (AUC) (Gorsevski et al. 2006; Cantarino et al. 2019; Pham et al. 2020). An ROC curve shows the true positive rate against the false positive rate at various thresholds. For a given susceptibility threshold, a true positive is defined as when the model correctly predicts landslide occurrence for a location with a landslide, and a false positive is where the model predicts landslide occurrence at a location without a landslide. Because landslides are relatively isolated phenomena, determination of the rate of true and false positives for the ROC curves, rather than their absolute values, helps provide a more balanced assessment of the model's capability to correctly distinguish between areas of high and low landslide susceptibility. The larger the area under this curve (AUC), the better the model is at distinguishing between areas of high and low landslide susceptibility. Values over 0.8 generally indicate good model prediction (Robin et al. 2011); values between 0.7 and 0.8 are considered fair; values < 0.7 are considered poor predictions; and values of 0.5 would be the result of random guessing (Zhu et al. 2010).

Landslide-related factors were ranked using ROC-AUC analysis by excluding one factor at a time from the random forest analysis and computing the relative difference in AUC between a model with excluded factor and that with all factors (Gorsevski et al. 2006; Marjanović, 2013; Cantarino et al. 2019; Pham et al. 2020). Important factors are expected to be associated with a larger difference. For robustness, to explore whether the results are sensitive to the analysis method, we also used a conditional probability approach and an out of bag permutation approach (Supplementary information: Figs. S1-S2) with identical input data to produce landslide susceptibility maps and factor ranking (e.g., Davis et al. 2006; Ozdemir 2009; Yilmaz 2010; Regmi et al. 2014; Costanzo and Irigaray 2020; Rohan et al. 2021).

Landslide susceptibility map comparison

To statistically quantify the effect of urbanization on modeled landslide susceptibility estimates, we calculated median susceptibility for the total, urbanized, and non-urbanized area in each of the modeled susceptibility maps. The susceptibility difference between maps was calculated from the percentage difference between these median values (i.e., $100 \times \frac{|a-b|}{(a+b)/2}$, where a and b are the median susceptibilities for the areas of interest in each of the modeled susceptibility maps). To test whether the differences in median susceptibility estimates are statistically significant ($\alpha = 0.01$), we used the non-parametric Wilcoxon rank sum test that compares susceptibility values between matched-paired pixels and modeled susceptibility maps over an area of interest (Gibbons and Chakraborti 2014).

To quantify spatial similarities or differences in the distribution of landslide susceptibility estimates generated from different landslide inventories, we calculated cross correlation between the modeled susceptibility maps (e.g., Shelef and Hilley 2014; Rohan et al. 2021). End member values of 1, −1, and 0 are indicative of perfect, inverse, and no correlation, respectively. This analysis focuses on correlation between large-scale susceptibility patterns by smoothing each map, before computing the correlation, with a circular filter with a radius of 35 m (based on the radius of a circle whose area is the same as the average area of all the active landslides in the inventory). To further evaluate the difference between models, we also compared AUC values when applying a model based on an urbanized landslide inventory to that based on non-urbanized inventory when applied on the same area.

Results

Ranking of landslide-related factors

The ranking of landslide-related factors had differences between the models trained on the four different landslide inventory classes: (1) Active-urbanized, (2) Active-non-urbanized, (3) Old-urbanized, and (4) Old-non-urbanized. The rankings computed through the AUC approach (Fig. 2) are similar for random forest, out of bag permutation, and conditional probability computations (Supplementary Information: Figure S1). Aside from slope, which is ranked as the factor that is most strongly associated with landslides, the rankings of the top five factors, as well as factors of lower ranking, had variations between models based on the different inventories. Comparison between models based on the two inventory classes of active landslides (i.e., in urban and non-urban areas), whose occurrence likely postdate urbanization, shows that in the urbanized (high road density) areas, active landslides are strongly associated with distance from roads and aspect. In models based on the non-urbanized inventory, active landslides are strongly associated with stratigraphic formation, distance to streams, and vegetation cover. In contrast, the ranking of factors is generally similar for the models based on the two inventory classes of old landslides (i.e., in urbanized and non-urbanized areas) that likely predate urbanization (Slope, Profile Curvature, Aspect, and Position on Hillslope being ranked as the most important factors).

Landslide susceptibility maps

To explore the effect of urbanization on modeled susceptibility estimates, we compared the susceptibility maps produced by each landslide inventory for the entire study area (supplementary information: Fig. S3). Figure 3 contrasts such maps for an urbanized section of the map both in terms of the spatial pattern and magnitude of susceptibility. The contrasting maps produced by models based on

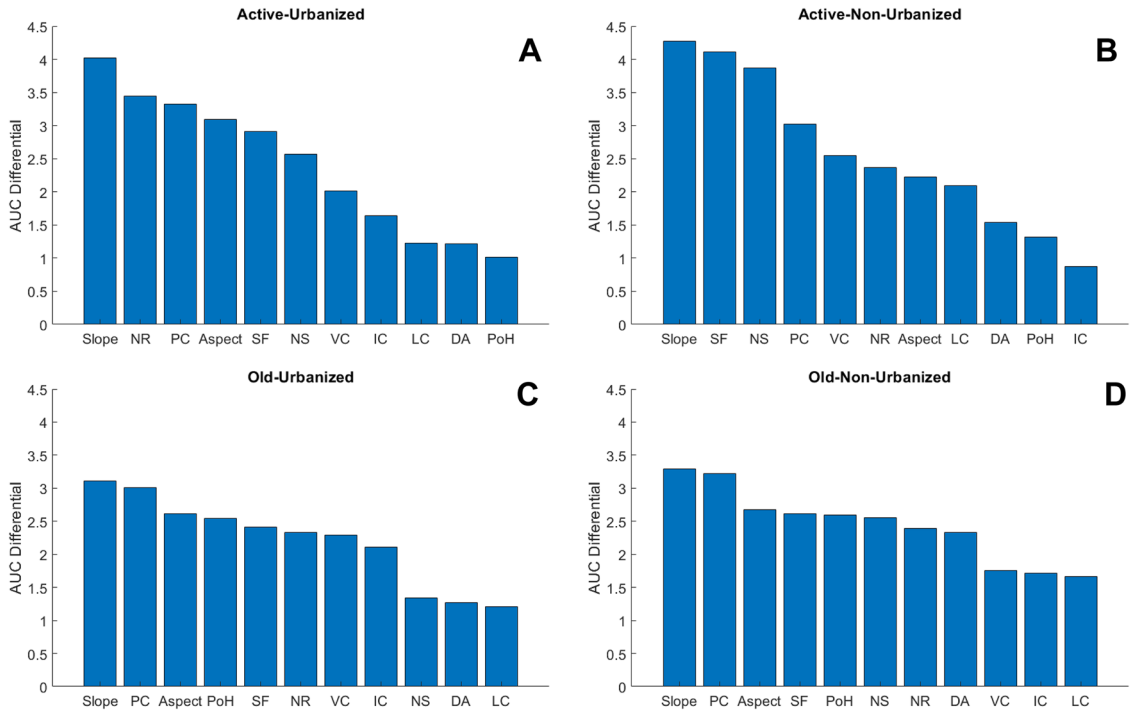

Fig. 2 Ranking of landslide-related factors by a random forest-based AUC differential method for models based on landslide inventories of **A** Active-Urbanized Landslides, **B** Old-Urbanized Landslides, **C** Active-Non-Urbanized Landslides, and **D** Old-Non-Urbanized Landslides. Factors: Slope, Nearest Road (NR), Profile Curvature (PC), Aspect, Stratigraphic Formation (SF), Nearest Stream (NS), Land Cover (LC), Vegetation Cover (VC), Drainage Area (DA), Impervious Cover (IC), and Position on Hillslope (PoH)

the active-urbanized (Fig. 3A) and active-non-urbanized (Fig. 3B) inventories show that the susceptibility model that is based on the active-non-urbanized inventory produces a smoother susceptibility pattern and a relatively high susceptibility at the upper portion of the hillslopes compared to the model based on the active-urbanized inventory. To explore the effect of time on modeled susceptibility estimates, we also compare modeled susceptibility maps based on inventories of old landslides in urbanized (Fig. 3C) and non-urbanized (Fig. 3D) areas. The modeled susceptibility based on these inventories is characterized by a noisy pattern that reflects fluctuation in susceptibility over short spatial distances. In these models (Fig. 3C, D), the association between high susceptibility to topographic slope is less distinct compared to models based on inventories of active landslides (Fig. 3A, B).

Statistical analysis of susceptibility values reveals differences between susceptibility maps modeled based on different landslide inventories. Comparison of median landslide susceptibility values between maps based on urbanized and non-urbanized inventories of active landslides can help quantify the effect of urbanization on landslide susceptibility. Comparison shows that the median susceptibility value modeled over urbanized areas with the model based on the active-urbanized inventory is 16% larger than the median susceptibility calculated by the model based on the active-non-urbanized inventory for the same area. For non-urbanized areas, the active-non-urbanized model median susceptibility is 8% larger than that modeled by the active-urbanized inventory. Comparison between susceptibility values modeled with the active-non-urbanized model to those produced by the old-urbanized and old-non-urbanized models can help quantify whether active-non-urbanized landslide susceptibility patterns differ from pre-urbanization patterns. The difference in median susceptibility value for such comparisons shows a somewhat higher (< 5%) susceptibility in both urbanized and non-urbanized areas for the models based on the old landslide inventories. Comparison between susceptibility values produced with the old-non-urbanized model to those produced with the old-urbanized model can help quantify whether areas that are now urbanized were more susceptible to landslides even prior to urbanization. Such comparison shows that the median susceptibility produced by the model based on the old-urbanized landslide inventory areas is 4% higher in urbanized areas compared to that produced by the old-non-urbanized landslide inventory. For each of the comparisons of medians mentioned above, a Wilcoxon rank sum test rejects the null hypothesis ($\alpha = 0.01$) that the compared susceptibility estimates come from continuous distributions with equal medians, indicating that the differences between these susceptibility maps are statistically significant.

Differences in susceptibility mapping are also apparent in the correlations between the smoothed susceptibility maps for the entire study area (Table 1). The correlation is 0.2991 between the two susceptibility maps produced from the inventories of active landslides in urbanized and non-urbanized areas (a subsection of these maps is shown Fig. 3A, B, respectively). This value is substantially lower than the correlation value of 0.6552 between the two susceptibility maps based on the inventories of old landslides in urbanized and non-urbanized areas.

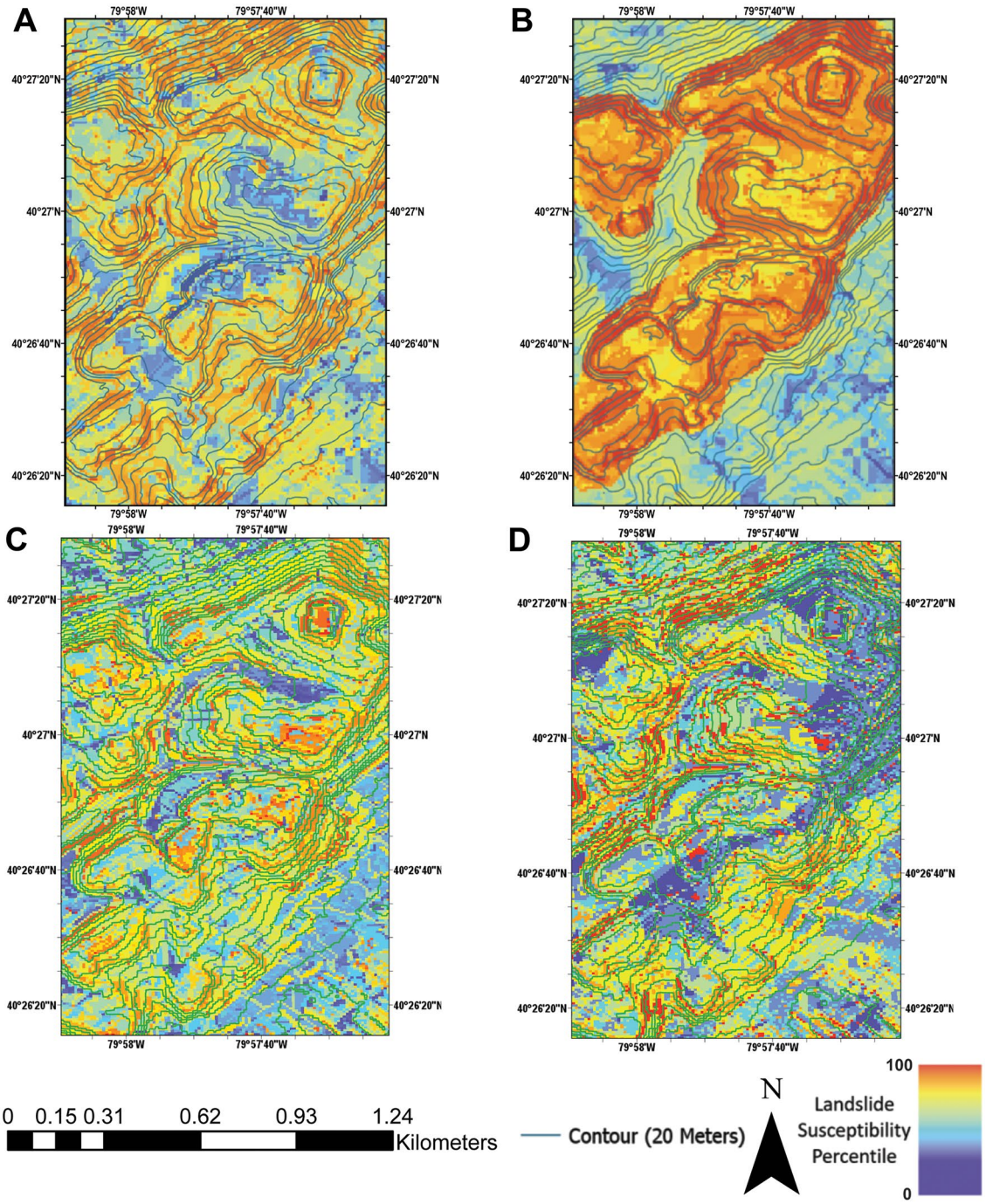

Fig. 3 Landslide susceptibility maps of the same (urbanized) area, produced with a random forest models based on: **A** Active landslides in urban areas, **B** Active landslides in non-urban areas, **C** Old landslides in urban areas, and **D** Old landslides in non-urban areas. Note the difference in susceptibility patterns between these models

Partial dependence analysis

To examine the marginal effect of a specific landslide-related factor on landslide susceptibility, we used a partial dependence analysis based on random forest. Results (Fig. 4) show that the models based on the old landslides inventory classes for (a) urbanized and (b) non-urbanized areas produced generally similar partial dependence relations (apart from the distance to nearest road factor). In contrast, the models trained with the two active landslide inventory classes for (a) urbanized and (b) non-urbanized areas produced different partial dependence relations for the slope, profile curvature, distance to nearest road, and nearest stream factors (Fig. 4), and similar values for the aspect factor.

Table 1 Cross correlation values between smoothed susceptibility maps for the entire study area, as produced from random-forest models based on each of the landslide inventories used in this study

Cross-correlation	Active-urban	Active-non-urban	Old-urban	Old-non-urban
Active-urban	1			
Active-non-urban	0.2991	1		
Old-urban	0.3146	0.4215	1	
Old-non-urban	0.3222	0.4032	0.6552	1

Model validation

The results from the ROC-AUC validation point at differences in landslide susceptibility between urbanized and non-urbanized areas. Applying the model based on inventory of active landslides in urbanized areas to estimate landslide susceptibility in urbanized areas produces the highest AUC (0.7942). In contrast, applying this model to estimate landslide susceptibility in non-urbanized areas produces a much lower AUC (0.6103). Similarly, the model based on inventory of non-urban-active landslides produces AUC values of 0.7942 for non-urbanized areas and 0.6966 for urbanized areas. For models based on the old landslide inventories, the AUC values for urbanized and non-urbanized areas are all within a narrower range of 0.6142–0.6973 (supplementary information: Table S1).

Discussion

Differences between models based on different inventories

Comparison of factor ranking, susceptibility maps, and partial dependence relations between models that are based on different landslide inventories quantifies the prolonged effect of urbanization on landslide susceptibility. The influence of urbanization on landslide occurrence is illustrated by the differences between the two landslide susceptibility models based on inventories of active landslides (i.e., post-urbanization) from urbanized and non-urbanized areas. These differences are quantified by (a) the relatively low spatial correlation between the smoothed landslide susceptibility maps based on these two models (correlation of 0.2991, Table 1), (b) the large (16%) and statistically significant difference in median susceptibility value between the models based on these inventories, (c) the difference in factor ranking between the two models (Fig. 2), and (d) differences in partial dependence relation for some factors (e.g., slope, Fig. 4). The prolonged effect of urbanization on landslides is further supported by the models based on the inventories of old landslides (i.e., pre-urbanization) that produce relatively similar susceptibility maps (i.e., correlation of 0.6552, Table 1), factor ranking (Fig. 2), partial dependence relations (Fig. 4) for urban and non-urban areas, and a relatively small difference in median susceptibility (4%) between these areas. These similarities indicate that prior to urbanization, landslide susceptibility patterns over the entire study area were more similar compared to the patterns portrayed by the inventories of active landslides, and therefore indicate that the differences in landslide susceptibility between urban and non-urban areas reflect the influence of urbanization.

Although topographic slope is widely recognized as a primary driver of slope failure, regardless of setting, the difference in factor ranking between models of active landsides in urbanized and non-urbanized areas hints at the contrasting influences on landslide susceptibility in these areas. Whereas in non-urbanized areas stratigraphic formation, distance to nearest stream, and vegetation cover are ranked as important factors, in urbanized areas the distance to nearest road, profile curvature, and aspect were more influential. The influence of distance to streams (dominant in non-urbanized areas) and roads (dominant in urbanized areas) on landslide occurrence may be akin in that both can induce landslides in their proximity by steepening the lower part of hillslopes through stream erosion and the construction of road cuts, respectively. Roads can also disrupt subsurface and surface hydrology and induce landslides by changing surface and subsurface flow pathways and increasing pore pressure (Mirus et al. 2007). Although foundation soils and embankments provide adequate support for the initial construction of roads, overstressing the embankment or foundation soil with additional fill can also result in rotational displacement and hillslope failure (Fell et al. 2005). The influence of vegetation cover on landslide occurrence likely reflects the effect of root reinforcement (Istanbulluoglu and Bras 2005). In urbanized settings, areas of removed vegetation are often covered by impermeable surfaces that may reduce infiltration and pore pressure in the underlying soil, and thus mitigate the loss of root reinforcement (Lee and Kim 2016) compared to areas of removed vegetation in non-urbanized settings (Glade 2003). This may explain the relative dominance of the vegetation cover factor in non-urbanized areas. The difference in the influence of stratigraphic formation may be linked to the prevalence of a sedimentary unit with numerous "red beds" that is particularly prone to landslides (Hamel and Flint 1972; Pomeroy 1982; Okagbue 1986) in non-urbanized areas (31%) compared to urbanized areas (8%). However, stratigraphic formation is ranked relatively high (fifth most important factor) even in the active and old urbanized landslide inventories, indicating that its influence is meaningful even in urbanized settings. Together with topographic slope, profile curvature is the only factor that is ranked within the four most influential factors in all four inventories, which is aligned with the findings of Pomeroy (1982) who noted that the majority (~ 60%) of active landslides mapped in the USGS inventory occur on concave slopes. Slope aspect is dominant in models based on both urbanized and non-urbanized inventories, likely because north-facing slopes are exposed to comparably less sunlight (Pomeroy 1982). Thus, north-facing slopes may sustain higher soil saturation and pore pressure that can increase the likelihood of landslide occurrence in both urbanized and non-urbanized areas.

The differences between the calculated susceptibility maps (e.g., Fig. 3) are generally aligned with the ranking of factor. A visual comparison of Fig. 3A, B together with inspection of the digital geologic map used in the analysis shows a strong association between susceptibility and stratigraphic formation in the susceptibility map based

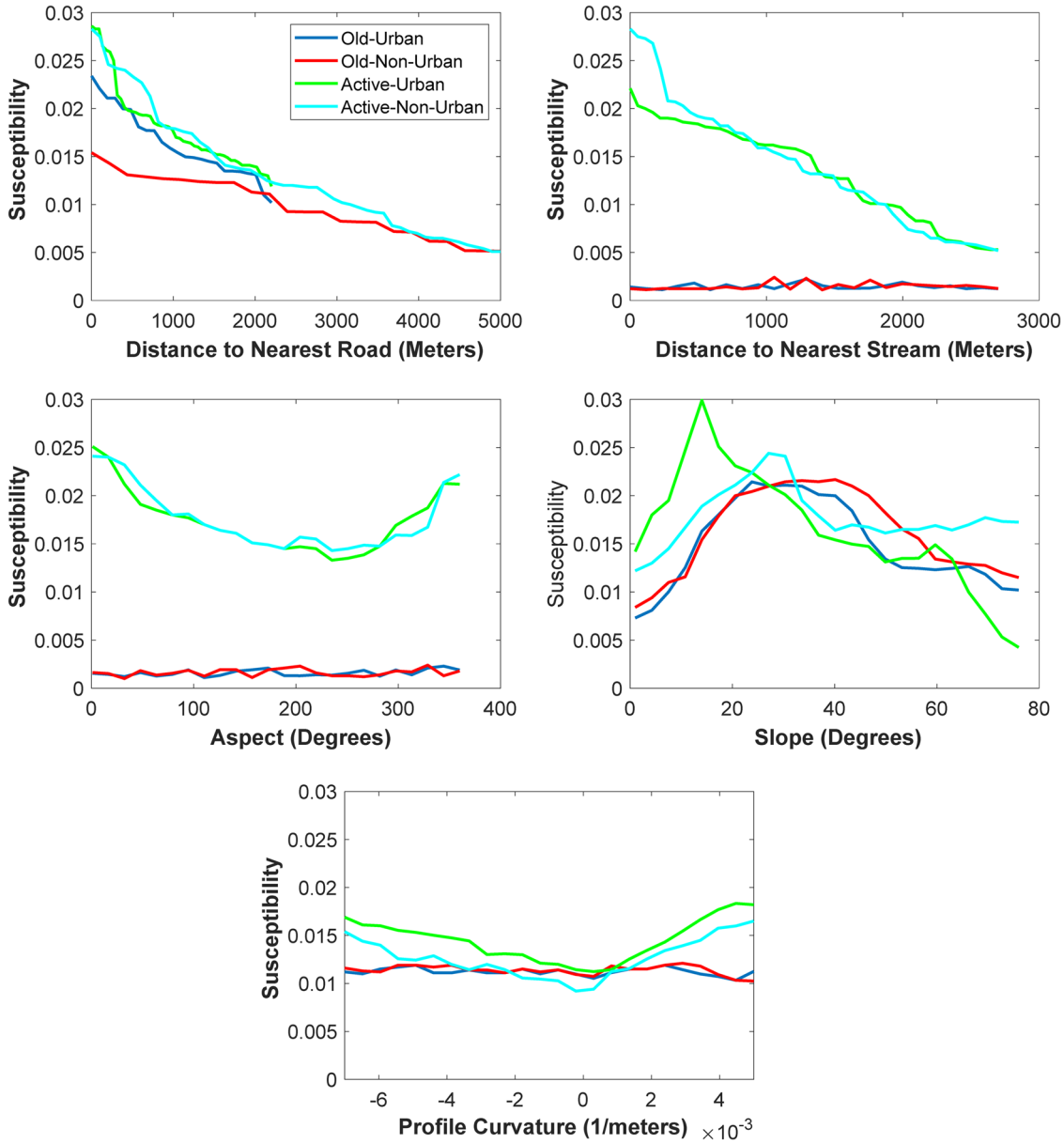

Fig. 4 Partial dependence plots based on model results for the active and old landslide inventories in urban (green/red) and nonurban (blue/light blue) areas for key factors

on the active-non-urbanized model (Fig. 3B), where higher susceptibility values occur within the "red bed" clay formation that has previously been highlighted as highly susceptible to landslide (Briggs et al. 1975; Gray et al. 2011). After heavy rains, the soil above these "red beds" becomes heavy and saturated with water causing the softened clay layer to break apart and slide (Pomeroy 1982). The localized fluctuations in susceptibility values in the maps modeled based on the old landslide inventories (i.e., the rough patterns in Fig. 3C, D) likely reflect the influence of profile curvature, which indeed varies over short length scales. The strip of low susceptibility values in Fig. 3D aligns with the combined effect of large distance to streams, stratigraphic formation, and southeastern and eastern aspects, all factors that rank highly in the corresponding susceptibility model (Fig. 2D).

The results indicate that urbanization may also influence the absolute susceptibility for landslides. Overall, the model based on the active-urbanized landslide inventory produced higher median landslide susceptibility estimates than the model based on the active-non-urbanized inventory (16% difference). The similarity in median landslide susceptibility values (< 5%) calculated by models based on the active-non-urbanized, old-non-urbanized, and old-urbanized landslide inventories indicates that, while the urbanized areas experienced increased susceptibility over time, the susceptibility remained similar in the non-urban area. These differences may reflect an increased susceptibility that stems from landscape change associated with urbanization. Given the time span between urbanization and landslide mapping, this indicates that landslide

inventories from non-urbanized areas may underestimate landslide susceptibility in urbanized areas even decades after urbanization. This inference has major implications for using such inventories to plan new urban developments in terrain previously considered of "moderate" susceptibility.

The low cross correlation between maps of similar median susceptibility (e.g., < 5% difference between median susceptibility values calculated by models based on the active-non-urbanized, old-non-urbanized, and old-urbanized landslide inventories) and rejection of the null hypothesis of the Wilcoxon rank sum test for comparisons between these maps indicate that although the median susceptibility may be similar, the susceptibility patterns differ between models. The higher median susceptibility values in urbanized compared to non-urbanized areas when susceptibility is calculated using the old-urbanized landslide inventory is an interesting result because it indicates that urbanization preferentially occurred in areas that are more susceptible to landslides. Although this can be a true pattern that reflects preferential urbanization of high slopes surrounding the industrial hubs along major river valleys, it can also reflect the bias in landslide mapping or road construction (see Sect. "Relations between factors and landslide susceptibility" for more discussion).

Roads may introduce a bias in landslide susceptibility mapping. In general, the high modeled susceptibility in urbanized areas is influenced by the higher number of landslide pixels in the active landslide inventory in urbanized areas (2.5%) compared to that in non-urbanized areas (0.7%). Although this may reflect a true pattern, it may be biased to some extent by the high density of roads in urbanized areas, where roads produce accessibility and expose outcrops that can increase the number of mapped landslides. The analysis of the old landslide inventory indeed hints at such a bias, as it shows a clear association between landslide susceptibility and distance to roads (Fig. 4), although the old landslides occurred prior to road construction.

The USGS landslide inventory may also be biased by exclusion of small landslides that occurred during peak urbanization (1930–1950). Comparing Pomeroy's old and active landslide polygons with landslide locations and volumes reported by Ackenheil (1955), based on landslides that occurred between 1920 and 1954, we find that the largest landslides reported by Ackenheil (≥ 1530 cubic meters) fall within active landslide polygons mapped by Pomeroy (1977). Smaller reported landslides (< 1530 cubic meters) either fall outside of the USGS landslide polygons or within polygons of old landslides. This difference likely reflects the resolution of the USGS mapping effort, as well as the slower healing (i.e., topographic smoothing and cover via vegetation growth and sediment transport) of large landslides scarps and deposits compared to small ones, such that only large landslides occurring between 1920 and 1954 were still detectable at the time of their mapping in the 1970s.

Comparison to previous landslide studies

Comparison of our results to studies that examined the influence of recent urbanization on landslide susceptibility reveals both similarities and differences. Studies that focused on recent urbanization (Dragićević et al. 2015; Simon et al. 2015; Frodella et al. 2018; Braun et al. 2019) report similar results to ours in that landslide estimates between urbanized and non-urbanized areas have low spatial correlation. Our results indicate that, in both urbanized and non-urbanized areas within our study area, the influence of aspect and stratigraphic formation are higher compared to other study areas. These differences do not necessarily reflect a difference between prolonged versus immediate influences of urbanization on landslides, and they may stem from differences in the analysis methodology or from regional differences in climate and stratigraphic formation that distinguished southwestern Pennsylvania from the warmer tropical climates of previous studies. A major difference between this study and previous studies is the low ranking of the land cover factor in all four models, whereas in prior studies landcover is typically highly ranked (Kumar and Bhagavanulu 2008; Kafy et al. 2017; Pisano et al. 2017; Avila et al. 2021). This may reflect the difference between the effect of decades-old urbanization (this study) versus recent urbanization (Dragićević et al. 2015; Simon et al. 2015; Frodella et al. 2018; Braun et al. 2019) and indicates that the relative influence of land cover on landslide susceptibility may decrease through time. However, this may also reflect differences in environmental and/or climatic conditions and in land-cover categories between this and previous studies.

Relations between factors and landslide susceptibility

The partial dependence plots provide insights into the relations between landslide-related factors and landslide susceptibility. In all four models, the susceptibility decreases gradually with increased distance to nearest road (Fig. 4). This likely reflects the effect of road construction on slope undercutting, soil weight, fill-induced loading, and changes to the natural drainage system (Sidle et al. 2006). This correlation may also reflect the abundance of roads along narrow river valleys, where nearby slopes are highly susceptible for landslides. As discussed in Sect. "Differences between models based on different Inventories", the association between susceptibility and distance to roads in the model based on the inventory of old landslides (Fig. 4) is somewhat unexpected because these old landslides likely predate road construction. This may be due to a preference for road construction over slopes covered with unconsolidated landslide deposits compared to bedrock. This association may also stem from a landslide mapping bias, where landslides and their deposits are more likely to be mapped along road cuts due to ease of access and good exposure. It is also possible that another factor that was not investigated in this study, but correlates with distance to road, influences the occurrence and/or mapping of old landslides. Increased distance to the nearest stream is also associated with a gradual decrease in susceptibility and increased similarity for the models based on the inventories of active landslide (Fig. 4). However, inversely to distance to nearest roads, streams are more strongly associated with landslide occurrence in the non-urbanized model. Given the low density of roads in these non-urbanized areas, streams likely play a more dominant role in undercutting the hillslopes compared to urbanized areas where undercutting is caused by construction activities. In contrast, in the models based on the inventories of old landslides, there is no association between landslide susceptibility and distance to stream (Fig. 4). This may reflect the difference between the mapped pattern of old versus active landslides, where old landslides are often mapped as broad features that cover entire hillslopes (so their locations are

not influenced by proximity to streams), and active landslides are mapped as small-scale, locally defined features whose location is influenced by such proximity. Partial dependence plots also reflect the association of aspect with landslide susceptibility. The models based on the inventories of active landslides show that susceptibility values increase in north-facing slopes in both urbanized and non-urbanized areas (Fig. 4). This likely reflects the influence of comparably lower solar radiation on north-facing slopes (Pomeroy 1982), which may sustain higher soil saturation, and pore pressure that overall can increase the likelihood of landslide occurrence in both urbanized and non-urbanized areas (Running et al. 1987; McGuire et al. 2016). In Colorado, landslide mapping demonstrated an entirely different relation, where landslides occurred preferentially on south-facing slopes (Ebel et al. 2015), which can be attributed to the higher level of root-reinforcement on north-facing slopes (McGuire et al. 2016).

The partial dependence plots for slope, based on the inventories of active landslides (Fig. 4), show that the slope values associated with the highest landslide susceptibility are ~ 15 and ~ 30° in urbanized and non-urbanized areas, respectively. This difference is consistent with a recent study by Johnston et al. (2021), showing that landslides in urbanized areas occur over lower slope and precipitation conditions, likely due to modification of the natural drainage system, loss of vegetation, and increased impermeable cover. This is supported by models based on the inventory of old landslides (i.e., pre-urbanization), where the slope – landslide susceptibility relations are similar between urbanized and non-urbanized areas. For profile curvature, the partial dependence plots based on the inventories of active landslides show (Fig. 4) an increased susceptibility in areas of both concave (positive) and convex (negative) curvatures and a lower susceptibility in flat areas (i.e., where concavity is approximately zero). The highest susceptibility values are associated with the largest concave values. This result is aligned with the aforementioned findings of Pomeroy (1982) who noted that the majority (~ 60%) of active landslides mapped in the USGS inventory occur on concave slopes. It also agrees with other studies that have found that landslide occurrence is less common where the slope is convex (Waltz 1971; Lessing et al. 1976) and likely reflect water convergence toward concave slopes that result in increased pore pressure and erosion, which may trigger landslides (Vieira and Fernandes 2004; Xu et al. 2012). The rise in susceptibility for localities of convex profile is most pronounced for the active-urbanized inventory and may reflect an association between urbanization and increased instability of convex profiles.

Data limitations

Although this study systematically quantifies the prolonged effects of urbanization on landslide occurrence, its findings are limited by the datasets and methods used. Our findings relied on a relatively small area in the Appalachian Plateau using limited landslide inventories that span one major urbanization phase. Larger multi-temporal landslide inventories with more detailed information on landslide occurrence during multiple distinct phases of urbanization may help better quantify the role of urbanization on landslide susceptibility. Further, the landslides explored in this study were mapped in the 1970s, several decades after the major urbanization phase in the study area. Due to the age of old landslide deposits, weathering, sediment transport, and vegetation growth may smooth and obscure their features, thus hampering identification and mapping of old landslides through field and remote sensing techniques (Pomeroy 1982). Mapped old landslide polygons likely reflect an amalgamation of smaller landslides through time leading to overestimation of the size of individual old landslides in our inventory. Systematic mapping of recent landslides in this area may produce an updated inventory that can quantify additional temporal patterns. The spatial and temporal resolution of the inventories and datasets used in this study also precludes investigation of landslide responses to local construction efforts within a previously urbanized area (e.g., road improvement, new neighborhoods, best-practices in landslide mitigation). Landslide inventories of higher spatial and temporal resolutions may enable exploration of such responses. Also, higher resolution maps of factors that were available at a relatively low resolution (e.g., stratigraphic formation, land use, and vegetation cover) may improve the robustness of the statistical models. We note that the patterns we report may be unique to the conditions of this specific study area (e.g., climate, stratigraphic formation, topography, construction practices), and studies that explore similar questions in different conditions may reveal different patterns. There is also an apparent bias of the distance to nearest roads factor, particularly in the old landslide susceptibility calculations, where landslide events occurred prior to road construction, are associated with distance to nearest roads (Fig. 4). Such a bias may stem from the exposure and access enabled by roads, which may have resulted in preferred mapping of old landslides next to roads. Alternatively, it may reflect preferred road construction within the deposits of old landslides. Accounting for this bias may help improve susceptibility maps. Finally, our analysis does not directly account for the effect of precipitation on landslide occurrence (Wasowski 1998; Wasowski and Pisano 2020; Ashland 2021). Although precipitation is recognized as an important trigger for landslides in southwestern Pennsylvania (Pomeroy 1982; Gray et al. 2011; Ashland 2021), its association with the occurrence of landslides reported in the USGS inventories we investigate here cannot be explored due to lack information about the timing of landslide occurrence. However, other data sources such as information reported by citizens to the 311 system do include such temporal data (Rohan et al. 2021), and future studies using these inventories may provide new insights into the role of precipitation in landslide occurrence in southwestern Pennsylvania. Despite these limitations, our results point at systematic and persistent differences in landslide patterns between urbanized and non-urbanized areas, even decades after urbanization, and indicate that distinguishing between these two environments would be beneficial for landslide susceptibility estimates.

Conclusion

Analysis of the prolonged effect of urbanization on landslide susceptibility estimates indicates long-term effects of urbanization on landslide susceptibility. This relies on susceptibility estimates based on inventories of active landslides in urbanized and non-urbanized areas, which show differences in ranking of landslide-related factors between these areas, as well as low spatial correlation of mapped susceptibility values between them. This is corroborated by analysis of old, pre-urbanization landslide inventories, used as a control dataset, which shows a general similarity between urbanized and non-urbanized areas in the ranking of landslide-related factors, as well as a relatively high correlation between mapped susceptibility patterns, hence indicating that the analysis of active landslide inventories indeed captures the influence of urbanization on landslide

susceptibility. Analysis of inventories of active landslides further shows that compared to non-urbanized areas, urbanized areas are associated with higher susceptibility values, stronger association between landslide occurrence and proximity to roads, and more likely occurrence of landslides over lower slopes compared to non-urbanized areas. Our analysis of old, pre-urbanization landslides indicates that the mapping of landslides might be biased by proximity to roads and that accounting for this bias in landslide susceptibility studies would be beneficial. Despite this bias, the consistent differences in susceptibility patterns between urbanized and non-urbanized areas indicate that urbanization has a decades-lasting effect on landslide susceptibility and that landslide susceptibility estimates should be made separately for these two different environments.

Acknowledgements

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Funding

The authors would like to thank both the Heinz and Andrew Mellon Foundations for providing the funding to accomplish this project.

Data availability

Data that support the findings of this study are available from the corresponding author (Tyler Rohan), upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Ackenheil AC (1955) A soil mechanics and engineering geology analysis of landslides in the area of Pittsburgh, Pennskylvania (Doctoral dissertation, University of Pittsburgh). Available from ProQuest Dissertations & Theses Global database. (UMI No. 301997212)

Alexander D (1986) Landslide damage to buildings. Environ Geol Water Sci 8(3):147–151. https://doi.org/10.1007/BF02509902

Archer KJ (2010) rpartOrdinal: an R package for deriving a classification tree for predicting an ordinal response. J Stat Softw 34:7. https://doi.org/10.18637/jss.v034.i07

Ashland FX (2021) Critical shallow and deep hydrologic conditions associated with widespread landslides during a series of storms between February and April 2018 in Pittsburgh and vicinity, western Pennsylvania, USA. Landslides 18(6):2159–2174. https://doi.org/10.1007/s10346-021-01665-x

Ávila FF, Alvalá RC, Mendes RM, Amore DJ (2021) The influence of land use/land cover variability and rainfall intensity in triggering landslides: a back-analysis study via physically based models. Nat Hazards 105(1):1139–1161. https://doi.org/10.1007/s11069-020-04324-x

Bernardie S, Vandromme R, Thiery Y, Houet T, Grémont M, Masson F, Grandjean G, Bouroullec I (2021) Modelling landslide hazards under global changes: the case of a Pyrenean valley. Nat Hazard 21(1):147–169. https://doi.org/10.5194/nhess-21-147-2021

BeVille SH, Mirus BB, Ebel BA, Mader GG, Loague K (2010) Using simulated hydrologic response to revisit the 1973 Lerida Court landslide. Environ Earth Sci 61(6):1249–1257. https://doi.org/10.1007/s12665-010-0448-z

Bogaard TA, Greco R (2016) Landslide hydrology: from hydrology to pore pressure. Wiley Interdiscip Rev Water 3(3):439–459. https://doi.org/10.1002/wat2.1126

Braun A, Urquia ELG, Lopez RM, Yamagishi H (2019) Landslide susceptibility mapping in Tegucigalpa, Honduras, using data mining methods. In IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018-Volume 1 (pp. 207–215). Springer, Cham. https://doi.org/10.1007/978-3-319-93124-1_25

Briggs RP, Pomeroy JS, Davies WE (1975) Landsliding in Allegheny County, Pennsylvania. U.S. Geological Survey Circular 728:18. https://doi.org/10.3133/cir728

Cantarino I, Carrion MA, Goerlich F, Ibañez VM (2019) A ROC analysis-based classification method for landslide susceptibility maps. Landslides 16(2):265–282. https://doi.org/10.1007/s10346-018-1063-4

Cascini L, Bonnard C, Corominas J, Jibson R, Montero-Olarte J (2005) Landslide hazard and risk zoning for urban planning and development. In: Hungr O, Fell R, Couture R, Eberhardt E (eds) Landslide risk management. CRC Press, London, pp 209–246

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazard 13(11):2815–2831. https://doi.org/10.5194/nhess-13-2815-2013

Chen CY, Huang WL (2013) Land use change and landslide characteristics analysis for community-based disaster mitigation. Environ Monit Assess 185(5):4125–4139. https://doi.org/10.1007/s10661-012-2855-y

Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S, Li S, Jaafari A, Ahmad BB (2019) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 172:212–231. https://doi.org/10.1016/j.catena.2018.08.025

Costanzo D, Irigaray C (2020) Comparing forward conditional analysis and forward logistic regression methods in a landslide susceptibility assessment: a case study in Sicily. Hydrology 7(3):37. https://doi.org/10.3390/hydrology7030037

Cruden DM, Varnes DJ (1996) Landslide types and processes. Transportation Research Board, U.S. National Academy of Sciences, Special Report 247:36–75

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

Dai F, Lee CF (2002) Landslides on natural terrain. Mt Res Dev 22(1):40–47. https://doi.org/10.1659/0276-4741(2002)022[0040:LONT]2.0.CO;2

Davis JC, Chung CJ, Ohlmacher GC (2006) Two models for evaluating landslide hazards. Comput Geosci 32(8):1120–1127. https://doi.org/10.1016/j.cageo.2006.02.006

Dragićević S, Lai T, Balram S (2015) GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. Habitat Int 45(2):114–125. https://doi.org/10.1016/j.habitatint.2014.06.031

Ebel BA, Rengers FK, Tucker GE (2015) Aspect-dependent soil saturation and insight into debris-flow initiation during extreme rainfall in the Colorado Front Range. Geology 43(8):659–662. https://doi.org/10.1130/G36741.1

Fell R, Ho KK, Lacasse S, Leroi E (2005) A framework for landslide risk assessment and management. In: Hungr O, Fell R, Couture R, Eberhardt E (eds) Landslide risk management. CRC Press, London, pp 13–36

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451

Frodella W, Ciampalini A, Bardi F, Salvatici T, Di Traglia F, Basile G, Casagli N (2018) A method for assessing and managing landslide residual hazard in urban areas. Landslides 15(2):183–197. https://doi.org/10.1007/s10346-017-0875-y

Giarratani F, Houston DB (1989) Structural change and economic policy in a declining metropolitan region: implications of the Pittsburgh experience. Urban Stud 26(6):549–558. https://doi.org/10.1080/00420988920080661
Gibbons JD, Chakraborti S (2014) Nonparametric statistical inference. In: Lovric M (ed) International encyclopedia of statistical science (pp. 977–979), Springer, Berlin. https://doi.org/10.1007/978-3-642-04898-2_420
Glade T (2003) Vulnerability assessment in landslide risk analysis. Erde 134(2):123–146
Gorsevski PV, Gessler PE, Foltz RB, Elliot WJ (2006) Spatial prediction of landslide hazard using logistic regression and ROC analysis. Trans GIS 10(3):395–415. https://doi.org/10.1111/j.1467-9671.2006.01004.x
Gray RE, Hamel JV, Adams WR (2011) Landslides in the vicinity of Pittsburgh, Pennsylvania. In: Ruffolo RM, Ciampaglio CN (eds) GSA field guide 20: from the shield to the sea (pp. 61–85), Geological Society of America, Boulder, CO. https://doi.org/10.1130/2011.0020(04)
Hamel JV, Flint NK (1972) Failure of colluvial slope. J Soil Mech Found Div 98(2):167–180. https://doi.org/10.1061/JSFEAQ.0001736
Hastie T, Tibshirani R, Friedman J (2009) Random forests. In: Hastie T, Tibshirani R, Friedman J (eds) The elements of statistical learning, 2nd edition (pp. 587–604). Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7_15
He Y, Lee E, Warner TA (2017) A time series of annual land use and land cover maps of China from 1982 to 2013 generated using AVHRR GIMMS NDVI3g data. Remote Sens Environ 199:201–217. https://doi.org/10.1016/j.rse.2017.07.010
Istanbulluoglu E, Bras RL (2005) Vegetation-modulated landscape evolution: effects of vegetation on landscape processes, drainage density, and topography. J Geophys Res Earth Surf 110(F2):F02012. https://doi.org/10.1029/2004JF000249
Iverson RM (2000) Landslide triggering by rain infiltration. Water Resour Res 36(7):1897–1910. https://doi.org/10.1029/2000WR900090
Johnston EC, Davenport FV, Wang L, Caers JK, Muthukrishnan S, Burke M, Diffenbaugh NS (2021) Quantifying the effect of precipitation on landslide hazard in urbanized and non-urbanized areas. Geophys Res Lett 48(16):e2021GL094038. https://doi.org/10.1029/2021GL094038
Kafy AA, Rahman MS, Ferdous L (2017) Exploring the association of land cover change and landslides in the Chittagong hill tracts (CHT): a remote sensing perspective. In Proceedings of the International Conference on Disaster Risk Management, Dhaka, Bangladesh (Vol. 23)
Kim JC, Lee S, Jung HS, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang. Korea Geocarto Int 33(9):1000–1015. https://doi.org/10.1080/10106049.2017.1323964
Kumar SV, Bhagavanulu DVS (2008) Effect of deforestation on landslides in Nilgiris district—a case study. J Indian Soc Remote Sens 36(1):105–108. https://doi.org/10.1007/s12524-008-0011-5
Lee CF, Huang WK, Chang YL, Chi SY, Liao WC (2018) Regional landslide susceptibility assessment using multi-stage remote sensing data along the coastal range highway in northeastern Taiwan. Geomorphology 300:113–127. https://doi.org/10.1016/j.geomorph.2017.10.019
Lee G, Kim M (2016) Shallow landslide assessment considering the influence of vegetation cover. J Korean GEO-environmen Soc 17(4):17–31. https://doi.org/10.14481/jkges.2016.17.4.17
Lessing P, Kulander BR, Wilson BD, Dean SL, Woodring SM (1976) West Virginia landslides and slide-prone areas. West Virginia Geol Economic Survey Environ Geol Bull EGB-15a, 64 pp. (1:24,000 scale, 28 maps on 27 sheets)
Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2(3):18–22
Marjanović M (2013) Comparing the performance of different landslide susceptibility models in ROC space. In: Margottini C, Canuti P, Sassa K (eds) Landslide science and practice (pp. 579–584). Springer, Berlin. https://doi.org/10.1007/978-3-642-31325-7_76
Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. Int J Remote Sens 39(9):2784–2817. https://doi.org/10.1080/01431161.2018.1433343
McAdoo BG, Quak M, Gnyawali KR, Adhikari BR, Devkota S, Rajbhandari PL, Sudmeier-Rieux K (2018) Roads and landslides in Nepal: how development affects environmental risk. Nat Hazard 18(12):3203–3210. https://doi.org/10.5194/nhess-18-3203-2018
McGuire LA, Rengers FK, Kean JW, Coe JA, Mirus BB, Baum RL, Godt JW (2016) Elucidating the role of vegetation in the initiation of rainfall-induced shallow landslides: insights from an extreme rainfall event in the Colorado Front Range. Geophys Res Lett 43(17):9084–9092. https://doi.org/10.1002/2016GL070741
Merghadi A, Yunus AP, Dou J, Whiteley J, Thai Pham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. Earth-Sci Rev 207:103225. https://doi.org/10.1016/j.earscirev.2020.103225
Miles CE, Whitfield GT, and others (2001) Bedrock geologic units of Pennsylvania, scale 1:250,000. based on: Berg TM, Edmunds WE, Geyer AR, Glover AD, Hoskins DM, MacLachlan DB, Root SI, Sevon WD, Socolow AA (1980). Geologic map of Pennsylvania, scale 1:250000. Pennsylvania Geological Survey, Map 1
Mirus BB, Ebel BA, Loague K, Wemple BC (2007) Simulated effect of a forest road on near-surface hydrologic response: redux. Earth Surf Proc Land 32(1):126–142. https://doi.org/10.1002/esp.1387
Mirus BB, Jones ES, Baum RL, Godt JW, Slaughter S, Crawford MM, Lancaster J, Stanley T, Kirschbaum DB, Burns WJ, Schmitt RG, Lindsey KO, McCoy KM (2020) Landslides across the USA: occurrence, susceptibility, and data limitations. Landslides 17:2271–2285. https://doi.org/10.1007/s10346-020-01424-4
National Oceanic and Atmospheric Administration, National Centers for Environmental Information (2021) Precipitation Frequency Data Server. Available at https://hdsc.nws.noaa.gov/hdsc/pfds/. (Accessed April 2021)
Okagbue CO (1986) An investigation of landslide problems in spoil piles in a strip coal mining area, West Virginia (USA). Eng Geol 22(4):317–333. https://doi.org/10.1016/0013-7952(86)90002-5
Ozdemir A (2009) Landslide susceptibility mapping of vicinity of Yaka Landslide (Gelendost, Turkey) using conditional probability approach in GIS. Environ Geol 57(7):1675–1686. https://doi.org/10.1007/s00254-008-1449-z
Papathoma-Köhle M, Glade T (2013) The role of vegetation cover change for landslide hazard and risk. In: Renaud G, Sudmeier-Rieux K, Estrella M (eds) The role of ecosystems in disaster risk reduction. UNU-Press, Tokyo, pp 293–320
Pennsylvania Spatial Data Access (PASDA) (2007) PAMAP program—roads. Available at https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=8. (Accessed April 2021)
Pennsylvania Spatial Data Access (PASDA) (2000) Fractional vegetation cover for southwest Pennsylvania, 2000 Available at https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=357. (Accessed April 2021)
Pennsylvania Spatial Data Access (PASDA) (2017) Previously active documented landslides in southwestern Pennsylvania. Available at https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=1622. (Accessed April 2021)
Pfeil-McCullough E, Bain DJ, Bergman J, Crumrine D (2015) Emerald ash borer and the urban forest: changes in landslide potential due to canopy loss scenarios in the City of Pittsburgh, PA. Sci Total Environ 536:538–545. https://doi.org/10.1016/j.scitotenv.2015.06.145
Pham BT, Nguyen-Thoi T, Qi C, Van Phong T, Dou J, Ho LS, Van Le H, Prakash I (2020) Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping. Catena 195:104805. https://doi.org/10.1016/j.catena.2020.104805

Pisano L, Zumpano V, Malek Ž, Rosskopf CM, Parise M (2017) Variations in the susceptibility to landslides, as a consequence of land cover changes: a look to the past, and another towards the future. Sci Total Environ 601:1147–1159. https://doi.org/10.1016/j.scitotenv.2017.05.231

Pomeroy JS (1977) Preliminary reconnaissance map showing landslides in Butler County, Pennsylvania. U.S. Geological Survey Open-File Report 77–246:3 pp. https://doi.org/10.3133/ofr77246

Pomeroy JS (1982) Landslides in the greater Pittsburgh region, Pennsylvania. U.S. Geological Survey Professional Paper 1229:48 pp., 12 plates. https://doi.org/10.3133/pp1229

Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed. Iran Nat Hazards 63(2):965–996. https://doi.org/10.1007/s11069-012-0217-2

Rappaport J (2003) U.S. urban decline and growth, 1950 to 2000. Econ Rev 88(3):15–44

Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A (2014) Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arab J Geosci 7(2):725–742. https://doi.org/10.1007/s12517-012-0807-z

Reichenbach P, Mondini AC, Rossi M (2014) The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). Environ Manage 54(6):1372–1384. https://doi.org/10.1007/s00267-014-0357-0

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically based landslide susceptibility models. Earth Sci Rev 180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 12:77. https://doi.org/10.1186/1471-2105-12-77

Rohan TJ, Wondolowski N, Shelef E (2021) Landslide susceptibility analysis based on citizen reports. Earth Surf Proc Land 46(4):791–803. https://doi.org/10.1002/esp.5064

Running SW, Nemani RR, Hungerford RD (1987) Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. Can J Res 17(6):472–483. https://doi.org/10.1139/x87-081

Sassa K, Wang G, Fukuoka H, Wang F, Ochiai T, Sugiyama M, Sekiguchi T (2004) Landslide risk evaluation and hazard zoning for rapid and long-travel landslides in urban development areas. Landslides 1(3):221–235. https://doi.org/10.1007/s10346-004-0028-y

Schwanghart W, Scherler D (2014) TopoToolbox 2–MATLAB-based software for topographic analysis and modeling in Earth surface sciences. Earth Surf Dyn 2(1):1–7. https://doi.org/10.5194/esurf-2-1-2014

Senanayake S, Pradhan B, Huete A, Brennan J (2020) Assessing soil erosion hazards using land-use change and landslide frequency ratio method: a case study of Sabaragamuwa Province. Sri Lanka Remote Sens 12(9):1483. https://doi.org/10.3390/rs12091483

Shelef E, Hilley GE (2014) Symmetry, randomness, and process in the structure of branched channel networks. Geophys Res Lett 41(10):3485–3493. https://doi.org/10.1002/2014GL059816

Sidle RC, Ziegler AD, Negishi JN, Nik AR, Siew R, Turkelboom F (2006) Erosion processes in steep terrain—truths, myths, and uncertainties related to forest management in Southeast Asia. For Ecol Manage 224(1–2):199–225. https://doi.org/10.1016/j.foreco.2005.12.019

Simon N, Crozier M, de Roiste M, Rafek AG, Roslee R (2015) Time series assessment on landslide occurrences in an area undergoing development. Singap J Trop Geogr 36(1):98–111. https://doi.org/10.1111/sjtg.12096

Smyth CG, Royle SA (2000) Urban landslide hazards: incidence and causative factors in Niterói, Rio de Janeiro State. Brazil Appl Geogr 20(2):95–118. https://doi.org/10.1016/S0143-6228(00)00004-7

Soma AS, Kubota T (2017) The performance of land use change causative factor on landslide susceptibility map in Upper Ujung-Loe Watersheds South Sulawesi, Indonesia. Geoplan: J Geomatics Plan 4(2):157–170. https://doi.org/10.14710/geoplanning.4.2.157-170

Tarolli P, Sofia G (2016) Human topographic signatures and derived geomorphic processes across landscapes. Geomorphology 255:140–161. https://doi.org/10.1016/j.geomorph.2015.12.007

Theodorou P, Baltz LM, Paxton RJ, Soro A (2021) Urbanization is associated with shifts in bumblebee body size, with cascading effects on pollination. Evol Appl 14(1):53–68. https://doi.org/10.1111/eva.13087

Tubbs DW (1974) Landslides in Seattle. Department Natural Resources

U. S. Geological Survey EROS Data Center (1999) National elevation dataset 10 meter 7.5x7.5 minute quadrangle for Pennsylvnia. National Cartography & Geospatial Center

Van Beek LPH, Van Asch TW (2004) Regional assessment of the effects of land-use change on landslide hazard by means of physically based modelling. Nat Hazards 31(1):289–304. https://doi.org/10.1023/B:NHAZ.0000020267.39691.39

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(3–4):588–603. https://doi.org/10.1016/j.epsl.2007.01.040

Vieira BC, Fernandes NF (2004) Landslides in Rio de Janeiro: the role played by variations in soil hydraulic conductivity. Hydrol Process 18(4):791–805. https://doi.org/10.1002/hyp.1363

Waltz JP (1971) An analysis of selected landslides in Alameda and Contra Costa Counties, California. Bull Assoc Eng Geol 8(2):153–163

Wang F, Yin Y, Huo Z, Zhang Y, Wang G, Ding R (2013) Slope deformation caused by water-level variation in the Three Gorges Reservoir, China. In: Sassa K, Rouhban B, Briceño S, McSaveney M, He B (eds) Landslides: global risk preparedness (pp. 227–237). Springer, Berlin. https://doi.org/10.1007/978-3-642-22087-6_15

Wang H, Zhang L, Yin K, Luo H, Li J (2021) Landslide identification using machine learning. Geosci Front 12(1):351–364. https://doi.org/10.1016/j.gsf.2020.02.012

Wasowski J (1998) Understanding rainfall-landslide relationships in man-modified environments: a case-history from Caramanico Terme. Italy Environ Geol 35(2):197–209. https://doi.org/10.1007/s00254005030

Wasowski J, Lamanna C, Casarano D (2010) Influence of land-use change and precipitation patterns on landslide activity in the Daunia Apennines, Italy. Q J Eng Geol Hydrogeol 43(4):387–401. https://doi.org/10.1144/1470-9236/08-101

Wasowski J, Pisano L (2020) Long-term InSAR, borehole inclinometer, and rainfall records provide insight into the mechanism and activity patterns of an extremely slow urbanized landslide. Landslides 17(2):445–457. https://doi.org/10.1007/s10346-019-01276-7

Winant G (2021) The next shift. In The next shift. Harvard University Press

Xu C, Dai F, Xu X, Lee YH (2012) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology 145–146:70–80. https://doi.org/10.1016/j.geomorph.2011.12.040

Yang L, Jin S, Danielson P, Homer C, Gass L, Bender SM, Case A, Costello C, Dewitz J, Fry J, Funk M, Granneman B, Liknes GC, Rigge M, Xian G (2018) A new generation of the united states national land cover database: requirements, research priorities, design, and implementation strategies. ISPRS J Photogramm Remote Sens 146:108–123. https://doi.org/10.1016/j.isprsjprs.2018.09.006

Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci 61(4):821–836. https://doi.org/10.1007/s12665-009-0394-9

Zêzere JL, Ferreira AB, Rodrigues ML (1999) Landslides in the North of Lisbon Region (Portugal): conditioning and triggering factors. Phys Chem Earth Part A 24(10):925–934. https://doi.org/10.1016/S1464-1895(99)00137-4

Zhang K, Wu X, Niu R, Yang K, Zhao L (2017) The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area. China Environ Earth Sci 76(11):405. https://doi.org/10.1007/s12665-017-6731-5

Zhou C, Cao Y, Yin K, Wang Y, Shi X, Catani F, Ahmed B (2020) Landslide characterization applying Sentinel-1 images and InSAR technique: the Muyubao landslide in the three gorges reservoir area. China Remote Sens 12(20):3385. https://doi.org/10.3390/rs12203385

Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland 9 pp

Zope PE, Eldho TI, Jothiprakash V (2016) Impacts of land use–land cover change and urbanization on flooding: a case study of Oshiwara River Basin in Mumbai, India. Catena 145:142–154. https://doi.org/10.1016/j.catena.2016.06.009

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10346-023-02050-6.

Tyler Rohan (✉) · **Eitan Shelef**
Department of Geology and Environmental Science, University of Pittsburgh, 4107 O'Hara Street, Pittsburgh, PA 15260, USA
Email: tjr68@pitt.edu

Ben Mirus
Geologic Hazards Science Center, US Geological Survey, PO Box 25046, Denver, CO 80225, USA

Tim Coleman
Department of Data Sciences and Operations, University of Southern California, Los Angeles, CA 90089, USA