Enhancing Landslide Susceptibility Analysis through Citizen Science, Geospatial Analysis, and Precipitation Thresholds in Urbanizing Environments

by

Tyler Rohan

Bachelor of Science in Earth Sciences: Geophysics, Ohio State University, 2016

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Tyler Rohan

It was defended on

May 15, 2023

and approved by

Michael S. Ramsey, Professor, Geology and Environmental Science

Daniel Bain, Associate Professor, Geology and Environmental Science

Josef Werne, Professor, Geology and Environmental Science

Anthony Iannacchione, Professor, Swanson School of Engineering

Thesis Advisor/Dissertation Director: Eitan Shelef, Assistant Professor, Geology and Environmental Science

Copyright © by Tyler Rohan

2023

Enhancing Landslide Susceptibility Analysis through Citizen Science, Geospatial Analysis, and Precipitation Thresholds in Urbanizing Environments

Tyler Rohan, PhD

University of Pittsburgh, 2023

Landslides pose a significant threat to human life and critical infrastructure, with increasing occurrences and severity attributed to climatic and anthropogenic factors, particularly urbanization. This doctoral thesis introduces an approach to landslide susceptibility analysis that combines citizen science, geospatial analysis, and precipitation thresholds to create a framework for landslide risk assessment in urban environments. The research explores the potential of using citizen science data to develop reliable landslide susceptibility models, addressing the pressing challenge of scarce landslide location data, and offering an alternative to conventional field and remote sensing work. The research also investigates the utility of citizen science data for identifying precipitation conditions that trigger landslides, emphasizing the importance of accurate information about landslide timing and preceding precipitation conditions. Despite the inherent uncertainty in citizen science data, the research demonstrates its value for approximating triggering precipitation conditions, as it aligns with local and global thresholds based on field-validated data. Furthermore, the research examines the long-term effects of urbanization on landslide susceptibility, using digitized United States Geological Survey (USGS) maps of pre-historic and active landslides in southwest Pennsylvania. This research confirms that urbanization has a lasting impact on geophysical and hydrological conditions, increasing an area's landslide susceptibility. The study provides valuable insights into the temporal dynamics of landslide risk, which are critical for effective risk assessment and land-use planning. By addressing the potential advantages and challenges of integrating non-expert data, this approach enhances the understanding of urbanization's complex interactions with landslide susceptibility. It also provides decision-makers with a tool for implementing targeted risk reduction measures, ultimately contributing to more effective risk management and land-use planning. This doctoral thesis thus paves the way for future research on the integration of citizen science data in natural hazard assessment and highlights the importance of interdisciplinary approaches for addressing the increasing challenges posed by natural hazards in a rapidly urbanizing world.

Table of Contents

Preface xi

1.0 Introduction 1

2.0 Landslide Susceptibility Analysis Based on Citizen Reports 4

2.1 Introduction 4

2.2 Study Area and Data 7

2.3 Methods 11

2.3.1 Field Validation of Reported 311 Landslide Locations 11

2.3.2 Conditional Probability Analysis 11

2.3.2 Filtration of Factor Maps 18

2.4 Results 22

2.4.1 Field Validation of Reported 311 Landslide Locations 22

2.4.2 Conditional Probability Analysis 22

2.4.3 Filtering of the Digital Elevation and Factor Maps 23

2.5 Discussion 27

2.5.1 Uncertainty in Reported Locations of 311 Landslides 27

2.5.2 Comparison to Previous Landslide Studies 27

2.5.3 Filtering of Factor Maps to Overcome Uncertainty 30

2.5.4 Methodological Limitations 31

2.5.5 Potential Use of 311 Data in the Future Study of Landslides 32

2.6 Conclusion 32

3.0 Exploring the Influence of Precipitation on Landslide Occurrence 34

3.1 Introduction **34**

3.2 Methods **39**

2.3.1 Study Area **39**

2.3.2 Data **39**

2.3.3 Data Processing **41**

3.3 Results **43**

3.4.1 Rainfall Intensity-Duration Curve **43**

3.4.2 Factor Analysis **44**

3.4 Discussion **47**

3.5.1 Rainfall Intensity Duration Pattern and Threshold **47**

3.5.2 Association of Landslide Characteristics with Rainfall Intesnity-Duration 48

3.5.3 Implications and Limitations **50**

3.5 Conclusion **51**

4.0 Prolonged Influence of Urbanization on Landslide Susceptibility **53**

4.1 Introduction **53**

4.2 Methods and Study Area **56**

4.2.1 Study Area **56**

4.2.2 Road Density **57**

4.2.3 Digital Elevation Data **58**

4.2.4 Landslide Inventory **60**

4.2.5 Landslide Related Factors **61**

4.2.6 Random Forest and ROC Validation **63**

4.2.7 Landslide Susceptibility Map Comparision 66

4.3 Results..... 67

4.3.1 Ranking of Landslide-Related Factors 67

4.3.2 Landslide Susceptibility Maps 69

4.3.3 Partial Dependence Analysis 72

4.3.4 Model Validation 74

4.4 Discussion 74

4.4.1 Differences Between Models based on Different Inventories 74

4.4.2 Comparison to Previous Landslide Studies 79

4.4.3 Relations between Factors and Landslide Suscpetibility 80

4.4.3 Data Limitations.....82

4.5 Conclusion 85

5.0 Conclusions and Future Directions 86

Appendix A: Landslide Susceptibility Analysis Based on Citizen Reports 89

Appendix B: Exploring the Influence of Precipitation on Landsldie Occurence..... 97

Appendix C Prolonged Influence of Urbanization on Landslide Susceptibility 100

Bibliography 106

List of Tables

Table 2.1 Landslide Inventories 10

Table 4.1 Cross Correlation Between Suscpetibility Maps 72

Appendix Table A.1 Weighted Contrast 311 Total 88

Appendix Table A.2 Weighted Contrast USGS 89

Appendix Table A.3 Weighted Contrast 311 (Field Validated - Total) 90

Appendix Table A.4 Weighted Contrast 311 (Field Validated - True) 91

Appendix Table A.5 Validated Locations, Displacement, and Size 92

Appendix Table A.6 Area Under the Curve Statistics 93

Appendix Table C.1 Model AUC Validation 104

Appendix Table C.2 Landslide Related Factors 105

List of Figures

Figure 2.1 Location Map and Photos of 311 Landslides 8

Figure 2.2 311 Field-Validated and Total Landslide Suscpetibility Maps 13

Figure 2.3 Landslide Susceptibility Maps from each Inventory..... 14

Figure 2.4 Ranking of Field-Validated Landslide-Related Factors. 17

Figure 2.5 Correlation Values for Filtered Maps. 20

Figure 2.6 Filtering Landslide Susceptibility Maps..... 21

Figure 2.7 Field-Validated versus 311 Filtered Maps..... 24

Figure 2.8 Filtering and Factor Influence..... 25

Figure 2.9 Filter Radius Correlation..... 26

Figure 3.1 Landslide Location Map with Overlain Validated Locations. 38

Figure 3.2 Precipitation Threshold. 45

Figure 3.3 Precipitation Factor Analysis. 46

Figure 4.1 Location and Road Density Map..... 59

Figure 4.2 Random Forest Factor Ranking..... 68

Figure 4.3 Landslide Susceptibility Map Comparison. 71

Figure 4.4 Partial Dependence Plots. 73

Appendix Figure A.1 Filtered Map Cross Correlation 94

Appendix Figure A.2 Susceptibility Map Correlation..... 95

Appendix Figure B.1 48 Hour Threshold Rainfall Threshold..... 96

Appendix Figure B.2 48 Hour Factor Analysis..... 97

Appendix Figure B.3 PennDOT Disctrict 10 Rainfall Threshold 98

Appendix Figure C.1 Landslide Factor Ranking (Conditional Probability) 101

Appendix Figure C.2 Landslide Factor Ranking (Out of Bag Permutation)..... 102

Appendix Figure C.3 Landslide Suscpetibility Maps 103

Preface

I would like to express my sincere gratitude and appreciation to my entire doctoral committee: Eitan Shelef, Michael Ramsey, Daniel Bain, Josef Werne, and Anthony Iannacchione for their unwavering support and guidance throughout my doctoral journey. Their contributions have been invaluable in helping me achieve my goal of obtaining a doctoral degree. I would like to extend my heartfelt thanks to my doctoral advisor, Eitan Shelef, who not only taught me the skills necessary to become a successful researcher but also inspired me to pursue my passion for natural hazard science. His mentorship and guidance have been instrumental in shaping my research and career. I would also like to thank Ben Mirus for his invaluable support in publishing my work and for introducing me to the United States Geological Survey landslide community, which has been an incredible resource for me. Additionally, I would like to express my gratitude to Emily Elliot and the Pittsburgh Collaboratory for water research, education, and outreach for their support in my research and for teaching me the importance of conducting research that has a positive impact on the local community. Furthermore, I would like to highlight that the Impact Resilient Infrastructure Science and Engineering Consortium (IRISE) not only provided me with various opportunities, but also gave me real-world experience that allowed me to translate my research into practical skills. Special thanks also go out to the members of the geomorphology lab group (Nicolas Wondolowski, Emrah Ozpolat, and Tianyue Qu) for their support and guidance throughout my research. Lastly, I would like to express my deepest appreciation to my family and friends for their unwavering support and encouragement throughout this journey. Their love and support have been my motivation and inspiration to strive for excellence. Thank you all.

1.0 Introduction

This Ph.D. thesis, titled " Enhancing Landslide Susceptibility Analysis through Citizen Science, Geospatial Analysis, and Precipitation Thresholds in Urbanizing Environments," utilizes citizen science, geospatial analysis, and precipitation thresholds to assess landslide susceptibility in urban regions. Landslides, being a prominent natural hazard, pose considerable threats to human life and infrastructure, and can be influenced by climate change and anthropogenic activities, with urbanization potentially acting as a key exacerbating factor. Considering the ongoing expansion of urban areas, an accurate estimation and prediction of landslide susceptibility is critical for protecting life and property. The thesis consists of three chapters, each focused on a different facet of landslide susceptibility analysis.

Chapter 1, "Landslide Susceptibility Analysis Based on Citizen Reports," explores the potential of utilizing crowdsourced data to enhance landslide susceptibility models. Addressing the acute need for extensive landslide inventory data, this chapter explores the utility of landslides reported by citizens for conducting susceptibility estimates, an alternative to conventional fieldwork and remote sensing. The study explores landslides reported by citizens through a 311-phone and online system, a nationwide system that updates real-time and records reported landslide's location and timing. Main findings show that: (a) approximately 70% of the 311-reported landslides are associated with an identifiable landslide in the field; (b) the spatial uncertainty of the 311-reported landslides is 104 ± 25 m; (c) 311-reported landslides differ from other inventories in that they are primarily associated with proximity to roads, however, field-correction of 311-reported landslide locations rectifies this anomaly; (d) a spatial filter,

scaled by the uncertainty in location as determined from a subset of the 311-data, can increase the consistency between the 311-reported inventory and field validated inventories. Potential benefits and inherent limitations of incorporating citizen data into landslide research are explored and discussed.

Chapter 2, " Using Citizen Science Data to Explore the Influence of Precipitation on Landslide Occurrence," probes into the utility of citizen science data for identifying precipitation conditions that induce landslides. Emphasizing the critical role of precise landslide timing and preceding precipitation conditions, this chapter explores empirical precipitation thresholds and underlines the challenges in determining landslide timing. By employing a landslide inventory rooted in citizen science data obtained from a 311-municipal system, the chapter demonstrates that citizen-reported precipitation thresholds generally align with physical expectations and with a threshold based on an independent landslide inventory based on field validated data. This establishes the value of citizen-reported data for estimating precipitation conditions that trigger landslides.

Chapter 3, "Urbanization and Its Prolonged Influence on Landslide Susceptibility," concentrates on the long-lasting impact of urbanization on landslide susceptibility. Using digitized USGS maps of pre-historic and active landslides in southwest Pennsylvania, this chapter reveals that urbanization has persistently altered the geophysical and hydrological conditions, thereby increasing an area's proneness to landslides for decades after peak urbanization. Through examination of the interplay between urbanization and landslide susceptibility over time, this chapter brings forth insights into the temporal dynamics of landslide risk, a crucial factor for effective risk management and land-use planning.

This doctoral thesis explores different approaches for landslide susceptibility analysis in urban environments. The subsequent chapters delve into the practical application of crowdsourced data for constructing trustworthy landslide susceptibility models, understanding precipitation thresholds through a citizen science perspective, and investigating the enduring effects of urbanization on landslide susceptibility. By shedding light on the prospective advantages and addressing the challenges tied to the integration of non-expert data, this approach aims to both add to the understanding of the complex interaction between urbanization and landslide susceptibility and provide useful information for policymakers that aids in designing effective risk mitigation strategies.

2.0 Landslide Susceptibility Analysis Based on Citizen Reports

2.1 Introduction

Landslide inventories of high quality can improve landslide susceptibility maps and understanding of landslide mechanisms and thus advance the implementation of safe land-use planning and prioritization of preventative efforts (Fell et al., 2008; Leventhal et al., 2008). Landslides typically occur over steep terrain where gravitational forces translate soil and rock downslope along weak planes of low frictional resistance (Vardoulakis, 2000; Miao et al., 2001). Their occurrence depends on factors such as the magnitude of topographic slope, soil and rock properties, the inclination of layered rock units, reinforcement due to roots, and hydrologic factors that may reduce the frictional-resistance of soil and rock by increasing pore-pressure (Iverson, 2000; Wang et al., 2003; Pfeil-McCullough et al., 2015; Bogaard et al., 2016). Such factors may covary in linear and non-linear ways that influence the magnitude and likelihood of landslide occurrence and make the production of reliable landslides susceptibility maps a challenging task (Guzzetti, 2000; Marjanovic, 2011).

Landslide susceptibility maps typically utilize landslide inventories that require extensive field mapping efforts and/or analysis of high-resolution remote sensing data. These inventories help identify factors that are associated with landslide susceptibility (Guzzetti, 2000; Yilmaz, 2010; Harp et al., 2011). Landslides are often influenced by temporally variable factors (e.g., precipitation, urban expansion, deforestation, fires) (Meusburger et al., 2008; Crozier, 2010; Huggel et al., 2012) and repeated mapping efforts are thus needed to study the impact of temporal changes in these factors on landslide occurrence. However, the extensive cost and effort associated

with such repeated mapping impede the progressive quantification of landslide susceptibility under changing environmental conditions.

Citizen science enables the public to participate in data collection to help provide solutions to scientific problems and has become increasingly popular over the last two decades (Franzoni, 2014; Can et al., 2019; Cieslik et al., 2019; Paul et al., 2019). Citizen science allows for inexpensive collection of large amounts of data at a rapid rate. However, to obtain data of sufficient quantity and quality, citizen science data must be of both interest to the public and have a standard procedure for which the data is collected (Can et al., 2019; Juang et al., 2019; Kocaman and Gokceoglu, 2019; Paul et al., 2019). Landslide related citizen science has been applied through university and government led programs (Juang et al., 2019; Kocaman and Gokceoglu, 2019; Mirus et al., 2020) and indicated that the primary improvements needed to this approach are increased citizen participation and validation of their reports (Cieslik et al., 2019; Kocaman and Gokceoglu, 2019). With such improvements, citizen science may be an optimal path for improving local landslide susceptibility estimates as well as augmenting global landslide catalogs (Juang et al., 2019).

A new and publicly available citizen science data source that is based on a 311-municipal service citizens' reports system, has the potential to provide a low-cost and progressively updated landslide inventory, that will enable progressive evaluation of landslide susceptibility. The 311 data combines a non-emergency phone and online reporting system where citizens report issues that warrant a response from county officials. The dataset is updated hourly and is accessible in over 300 cities in the United States and Canada (Schellong et al., 2007; Schwester, 2009; O'Brien, 2016; Choi et al., 2018). The 311 system reports a multitude of different categories and includes landslide locations and the time of their reporting. This data is thus gathered with a minimal cost

and effort and has the potential to create a progressively updated landslide inventory. This inventory has not yet been utilized to study landslide occurrence and can potentially improve landslide susceptibility estimates at the national level.

Data-driven approaches for mapping landslide susceptibility typically rely on landslide inventory data (e.g., location, size, timing, degree of activity), as well as maps of factors that are related to landslide occurrence (hereafter termed landslide-related factors) such as topography, geology, soil, vegetation, hydrologic properties, etc. (Kamp et al., 2008; Bălteanu et al., 2010; Santoso et al., 2011; Arabameri et al., 2019; Zhao and Chen, 2020). These approaches are divided into (Yilmaz, 2009; Huang and Zhao, 2018; Reichenbach et al., 2018): (a) Heuristic methods, where expert opinion determines the weighting of different factors on the relative likelihood (i.e., susceptibility) of landslide occurrence; (b) Physical methods that compute the relative magnitude of the physical forces that drive and resist landslides, and identify locations where the conditions for landslide occurrence are likely attained; and (c) Statistical methods rely on a large dataset and utilize the covariance between landslide-related factors and the occurrence of landslides to weight these factors and predict the relative likelihood of landslides. Heuristic methods are highly subjective and difficult to reproduce, and physical methods are accurate and reproducible, but require detailed information about soil and hydrologic properties that can rarely be attained over large areas (Francipane et al., 2014). Statistical methods, such as conditional probability and machine learning approaches (Yilmaz, 2010; Pourghasemi et al., 2012; Do et al., 2018), often produce consistent results over large areas at a high spatial resolution (Komac, 2006). In general, conditional probability enables simpler interpretation compared to machine learning methods and produces landslide susceptibility maps of comparable quality (Pradhan et al., 2010; Yilmaz, 2010; Goetz et al., 2015).

This research combines field work, statistical analyses, and comparison between different landslide inventories to explore the accuracy and applicability of the 311-dataset for landslide susceptibility estimates. More specifically, we use landslide inventories from Pittsburgh, PA, where landslide risk is among the highest in the nation (Highland, 2006; Gray, 2011), to: (a) Quantify the spatial accuracy of landslides reported via the 311 system; (b) Compare landslide susceptibility estimates based on a 311-based landslide inventory to those based on field validated landslide inventories; and (c) Use these comparisons to explore procedures for producing reliable landslide susceptibility maps from a 311-based landslide inventory. We first present the field area and datasets used in this study, and the method used to produce landslide susceptibility maps and ranking of landslide-related factors. We then use the spatial uncertainty associated with 311-based landslide reports to design a simple filter and test if it increases the consistency between 311-based susceptibility maps and those produced from established, field-based landslide inventories. Our results suggest that 311-based landslide inventories can guide landslide mapping at a low cost and effort and thus can improve landslide susceptibility estimates.

2.2 Study Area and Data

Given the large datasets required to study landslide susceptibility we focus on the city of Pittsburgh in Allegheny County, PA (Figure 2.1), where multiple landslides are recorded in various datasets. The area has a history of landslide occurrence and is located next to the Allegheny, Ohio, and Monongahela rivers (Monongahela means "falling banks", in a native language (Staats, 1942), which likely refers to the geological instability of the surrounding slopes). In this area, the lithologic, climatic, and topographic characteristics, as well as anthropogenic modifications, cause

a generally high susceptibility for landslides and increases the risks associated with their effects (Pomeroy, 1982). The high susceptibility and social awareness for landslide risk makes Pittsburgh a data rich location for studying landslide occurrence and reporting.

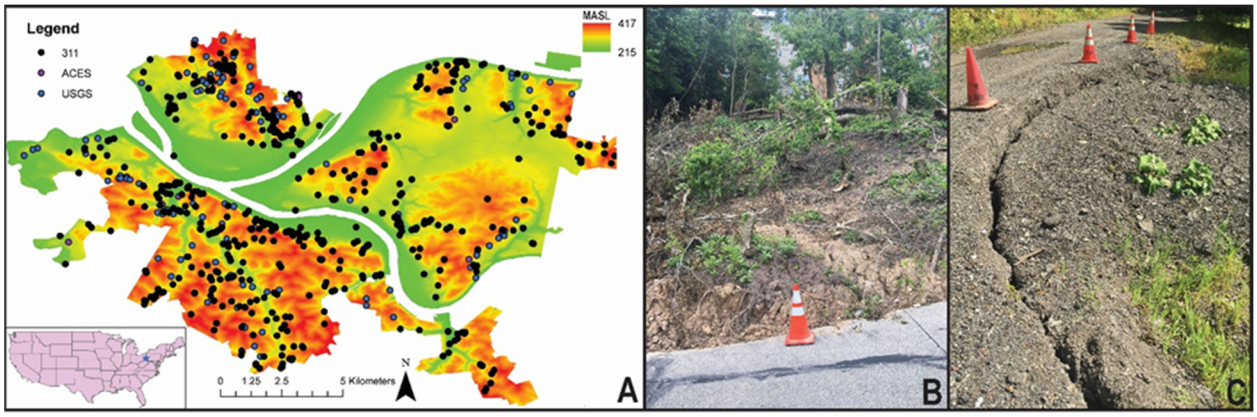


Figure 2.1: (A) Location of 311 reported landslides overlain on a DEM of the City of Pittsburgh. (B-C) Field Photos of validated 311 reported landslides. Inset map shows study area over map of the United States.

There are two established landslide inventories of high spatial accuracy in Pittsburgh, (Table 2.1); (1) Maps produced by the United States Geological Survey (USGS) where landslides are mapped as polygons based on field mapping efforts conducted in the 1970s-1980s (Pomeroy, 1977; Bridges et al., 1975; Pomeroy, 1982, Southwestern Pennsylvania Commission, 2017). (2) A landslide map produced by Allegheny County Emergency Services (ACES) for locations that are being monitored by the county's department of public works (Allegheny County Landslide Task Force, 2019). In addition to these established datasets, Pittsburgh has a publicly available 311-based dataset, where a relatively large number of landslides is being reported. However, the uncertainty associated with this data is yet to be determined, and thus it remains unclear whether it can be utilized to progressively improve landslide susceptibility estimates. Because landslides are mapped and digitized as polygons in the USGS dataset, and as point locations in the ACES and 311-based inventories, we converted polygons to point locations so that the datasets are consistent. This conversion selects the highest point within a landslide polygon as the representative location of this landslide, assuming that it most closely represents the location of slope failure.

To produce and compare landslide susceptibility maps based on different landslide inventories we analyzed landslide locations in each inventory in the context of nine topographic and environmental factors at these locations. We computed topographic factors (slope, curvature, drainage area, relative location on hillslope, distance from nearest channel, and aspect) from a 10-meter resolution Digital Elevation Model (DEM) from the National Elevation Database (NED), and the environmental factors (distance from nearest road, land-use, and stratigraphic group) from Pennsylvania Spatial Data Access (PASDA). Slope was calculated from the DEM as the magnitude of the gradient vector and expressed in degrees. Profile curvature was calculated as the

second numerical derivative of the DEM through the MATLAB based software TopoToolBox (Schmidt et al., 2003; Schwanghart and Kuhn, 2010). The relative location of the landslide on the hillslope is the fraction of the landslide elevation relative to the hillslope relief, as estimated from local relief value over a circular disk with a 200-meter radius (i.e., similar to the length scale of local hillslopes). Lithological information for Pittsburgh was acquired from a categorical digital dataset that is based on the map of Berg, 1980, and includes five different lithologic groups (Dunkard, Monongahela, Casselman, Glenshaw, and Allegheny).

Inventory	USGS (1973 –1982)	ACES (2019)	311 (2015–2020)
Number of landslides	110	24	720
Collection method	Field mapping	Field mapping	Citizen reports

Table 2.1: Inventories of landslides and the dates of collection that are located in Pittsburgh, PA, USA provided by, the United States Geological Survey (USGS), and Allegheny County Emergency Services (ACES).

2.3 Methods

2.3.1 Field Validation of Reported 311 Landslide Locations

To quantify uncertainty in the location and reliability of the 311-landslide inventory, we validated 311 reported landslides (Figure 2.1) in May-August 2019. At each site, we recorded the coordinates of the landslide in the field (if such a landslide was identified) and compared them to the reported coordinates to define the spatial uncertainty in landslide locations. To cast this spatial uncertainty in the context of landslide dimensions, we also recorded the spatial dimensions of each landslide. We then compiled the field validated landslides into a new inventory and used these datasets to produce and compare susceptibility maps based on the: (1) original, non-field validated landslide locations, (2) field corrected landslide locations, and (3) USGS and ACES inventories (Table 2.1).

2.3.2 Conditional Probability Analysis

We used a conditional probability approach to produce landslide susceptibility maps and rank the influence of the different landslide-related factors (e.g., topography, land-use, lithology) on landslide occurrence (e.g., Chung, 2006; Ozdemir, 2009; Yilmaz et al., 2010; Regmi et al., 2014; Costanzo & Irigaray, 2020). To do so, we divided each of the m landslide-related factor to n classes that span the range of values for this factor in the maps of the study area. For each of the resulting nm factor-class combinations, we then computed the conditional probability C_p:

$$C_{p\text{-}j}=N_{l\text{-}j}/N_{p\text{-}j}\,, \qquad (1)$$

Where subscript j is the index of the factor-class combination, and $N_{l\text{-}j}$ and $N_{p\text{-}j}$ are the number of landslides locations and map pixels within this combination, respectively. A factor-class combination that produces a relatively high C_p indicates that spatial locations that are characterized by this combination tend to generate a relatively large number of landslides. To create a landslide susceptibility map, we assigned the computed values of C_p for each factor-class combination to all the map pixels associated with this combination. Percentile maps were produced by normalizing each C_p value by the range of C_p values and multiplying by 100 (e.g., Figures 2.2, 2.3).

The analysis requires a small number of factor class combinations relative to the number of landslides so that the number of landslides in each such combination suffices to minimize the effect of outliers. We determine the number of factors (m=5), and of classes in each factor (n=5) based on prior studies with similarly sized datasets (Chung, 2006; Pradhan et al., 2010; Pourghasemi et al., 2012).

To define the 5 classes in each factor while accounting for the distribution of values in each continuous factor (i.e., non-categorical factors such as slope, curvature, drainage area), the lower class for each factor is defined between the minimum to the 5th percentile of the factor map values for which landslides occur, and the upper class between the maxima and the 95th percentile of the factor map values for which landslides occur. The factor values between the 5th and 95th percentiles were divided into three equally spaced classes, thus resulting in a total of five classes for a factor. We classified categorical factors (i.e., lithology, land-use) according to their mapped categories. To standardize the comparison between landslide inventories this classification is based on landslide locations from all the aforementioned inventories (Table 2.1).

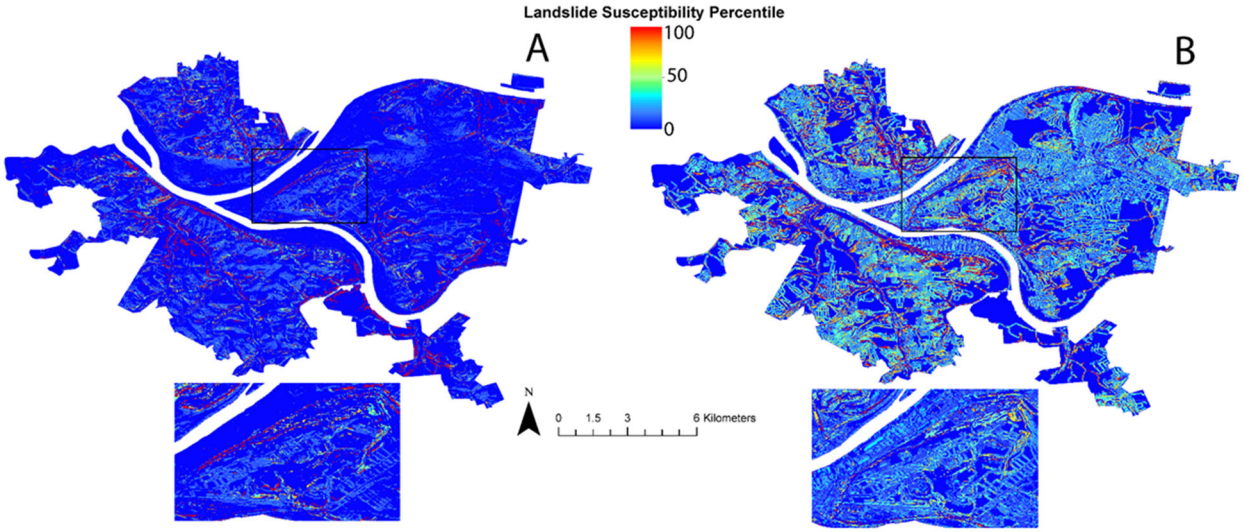


Figure 2.2: Landslide susceptibility maps (10-meter resolution) for Pittsburgh, PA: (A) Map based on 55 field corrected 311 landslide locations. (B) Map-based on 77 originally reported 311 landslide locations. Color bar based on the percentile of computed conditional probability values. Black rectangles mark the location of inset maps that show more details. Note that this figure has a different color-scheme than other maps to help identify differences between these maps.

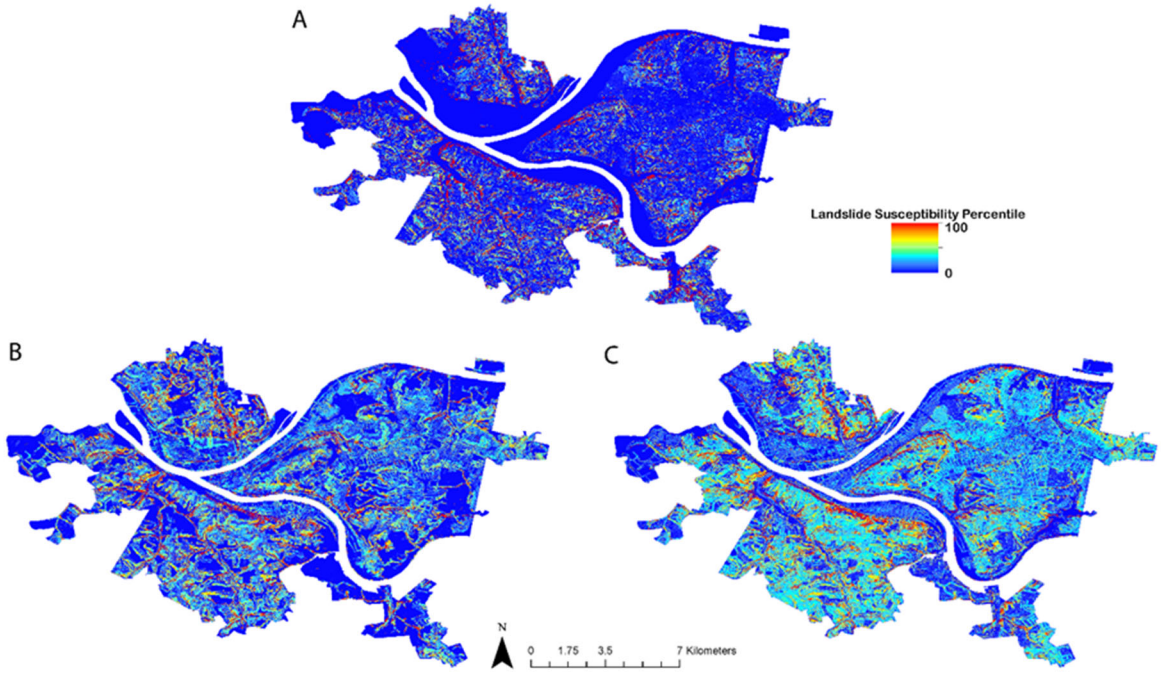


Figure 2.3: 10-meter resolution nonfiltered landslide susceptibility maps created from: (A) ACES-USGS inventories (N=134) (B) The entire 311 landslide inventory (N=720) (C) Total combined landslide inventories (N=834).

To identify the five-primary landslide-related factors (m=5) out of the nine total factors used, we calculated Weighted Contrast (Wc) values (Supplementary Information) for each class in each factor (Schicker 2012; Guo 2015).

$$W_p = \frac{\frac{A_1}{A_1+A_2}}{\frac{A_3}{A_3+A_4}}, \tag{2}$$

$$W_n = \frac{\frac{A_2}{A_1+A_2}}{\frac{A_4}{A_3+A_4}}, \tag{3}$$

$$W_c = W_p - W_n \tag{4}$$

Where W_p = Weighted Positives, W_n = Weighted Negatives, A_1 = Number of Landslides that fell inside a class, A_2 = Number of Landslides that fall outside a class, A_3 = Number of map pixels that fell inside a class, and A_4 = Number of map pixels that fell outside of a class. Weighted contrast values between 0.5-1, 1-2, and >2 is indicative of moderate, good, and extreme predictability, respectively. Negative values indicate the inverse predication of a factor class. We used the maximal weighted contrast for each landslide-related factor to rank the top five factors to be used in the conditional probability analysis. To test if this ranking is dependent on ranking methodology, we also ranked the factors with an alternate method that is based on a probabilistic parameter called frequency-ratio (Lee et al., 2005; Pradhan, 2010; Yilmaz, 2010). This produced similar ranking to that produced with the weighted contrast approach.

To quantify the predictive power of the conditional probability (C_p) - based landslide susceptibility maps, we computed receiver operating characteristic curves (ROC, Fawcett, 2006, Gorsevski et al., 2006). These curves evaluate the performance of a binary classifier system, (i.e., yes/no landslide occurrence), such as the conditional probability method, by analyzing true-positive and false positive rates for different discrimination thresholds (i.e. C_p values). In this context, a map pixel is considered a true positive if it contains a mapped landslide

and is also predicted to contain a landslide for a given C_p threshold. A pixel is considered a false positive when it does not contain a landslide but is predicted to contain one for a given C_p threshold. Similarly, a true negative occurs when a pixel that does not contain a landslide is predicted not to contain one, and a false negative occurs when a pixel that contains a landslide is predicted not to contain one.

ROC analysis is used to evaluate model performance through calculation of the area under the ROC curve (AUC) (Gorsevski, 2006; Cantarino et al., 2019; Pham 2020). ROC-AUC analysis can be used to rank landslide-related factors by their influence on model performance, and thus point at commonalities and differences between landslide inventories and guide further analyses (Pham et al., 2020). For each landslide inventory, we explored the relative influence of each landslide-related factor on a model prediction by excluding one factor at a time from the ROC-AUC analysis (Gorsevski, 2006; Marjanović, 2013; Cantarino et al., 2019; Pham 2020) and calculating the relative difference (dAUC= 100*(AUCa-AUCe)/AUCa) between the AUCs for a model with excluded factor (AUCe) and that with all five factors (AUCa). We then rank the factors based on their relative influence on the AUC. To quantify the uncertainty associated with this procedure, we run a bootstrap analysis (n = 1000) where in each iteration we run the aforementioned procedure while excluding a random subset of 25% of the landslide locations. We use the 5 and 95 percentiles from these iterations to define the uncertainty in the AUC difference (Figure 2.4).

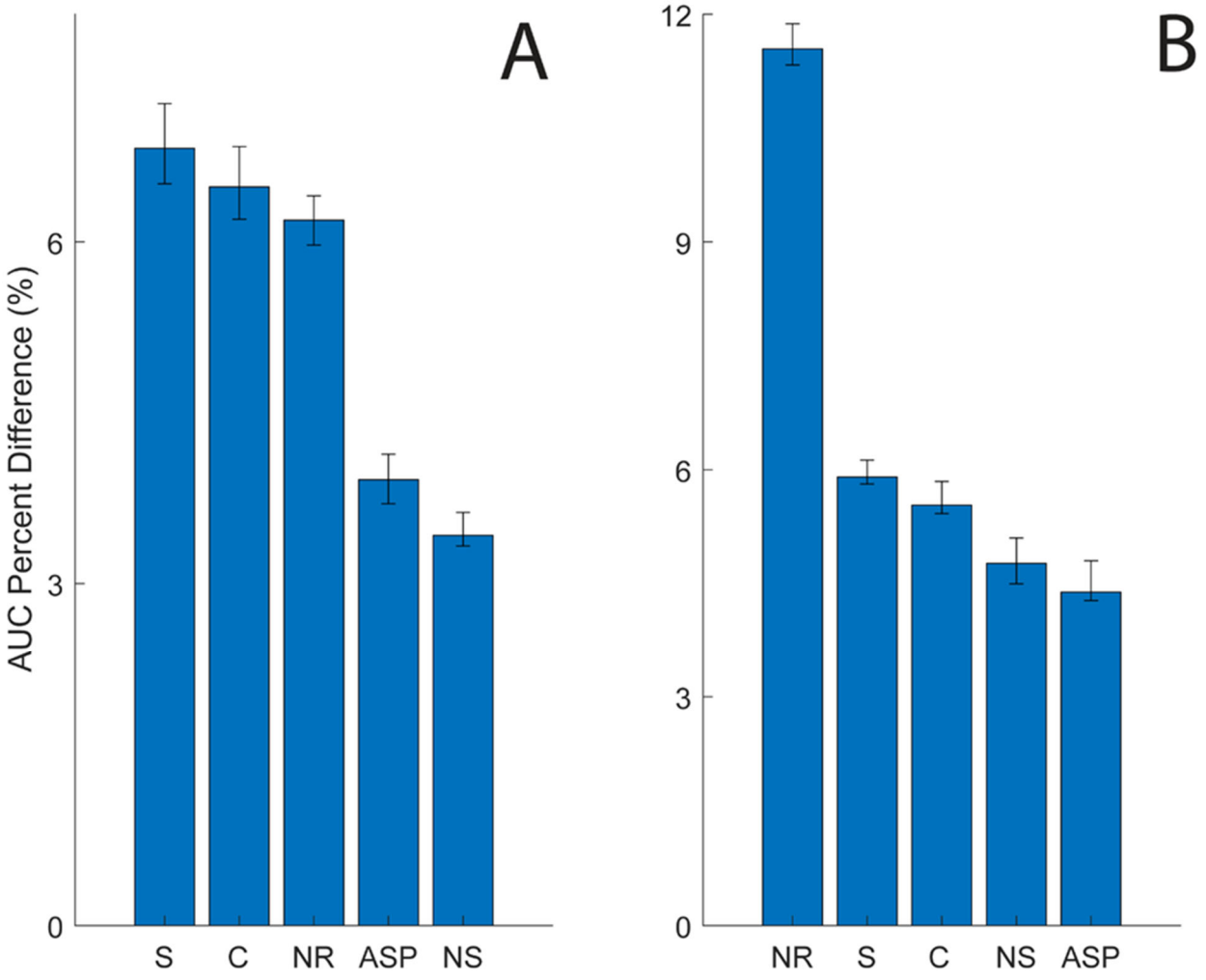


Figure 2.4: Ranking of landslide-related factors by their influence on the AUC for: (A) the field corrected 311-based inventory (N=55), and (B) the original 311-reported landslide locations (N=77). Nearest Road (NR), Slope (S), Aspect (Asp), Profile Curvature (C), and Nearest Stream (NS).

2.3.3 Filtration of Factor Maps

The spatial uncertainty in 311-reported landslide location can cause erroneous association between landslide-related factors and landslide occurrence that may cause inaccuracies in landslide susceptibility maps. To ameliorate this problem, we use a two-dimensional circular average filter. The radius of this filter defines a spatial scale over which each landslide-related factor is averaged to compute a representative value that accounts for the uncertainty in landslide location. We use the filtered factor maps in the conditional probability procedure to test if filtering increases the similarity (measured through two-dimensional correlation) between the landslide susceptibility maps that are produced from the original 77 landslides reported through the 311 system, and the 55 field corrected locations of these landslides (Figures 2.5, 2.6). We then use a similar procedure to explore the influence of filtering on the similarity between the entire inventory that is based on the 311-reports (N=720, non-field corrected) and the field-based inventory produced by combining the USGS and ACES inventories (N=134). We further explore this approach by applying a range of filter radii to identify the scale that maximizes the similarity between maps from field-validated and non-field-validated landslide inventories. Factor classes were generated based on the distribution of the combined landslide datasets used in each correlation experiment.

To explore the generality of the filtration approach given the uncertainty in landslide location, we generated 100 different quasi-random landslide inventories and used filtration to explore the correlation between the susceptibility maps they produce. The quasi-random landslide locations (N=55) were selected within a ring, centered at pixels that contain landslides, whose dimensions are based on the distribution of measured distances between the 311-reported and field corrected landslide locations (approximately 60 to 120 meters). Each of the 100 landslide inventories is then filtered using the filtration procedure described above to generate 100 landslide

susceptibility maps at each filter diameter. These susceptibility maps are then compared spatially through correlation with the susceptibility map that is based on the 311-field corrected landslide inventory, and the mean and standard deviation of these 100 correlations are recorded. For consistency, factor classes were generated based on the distribution of the combined landslide datasets for each experiment (i.e., the 311-field corrected and the randomly generated landslide locations). We conducted this experiment with both 6 and 5 classes per factor to further explore the sensitivity of these results to the number of classes.

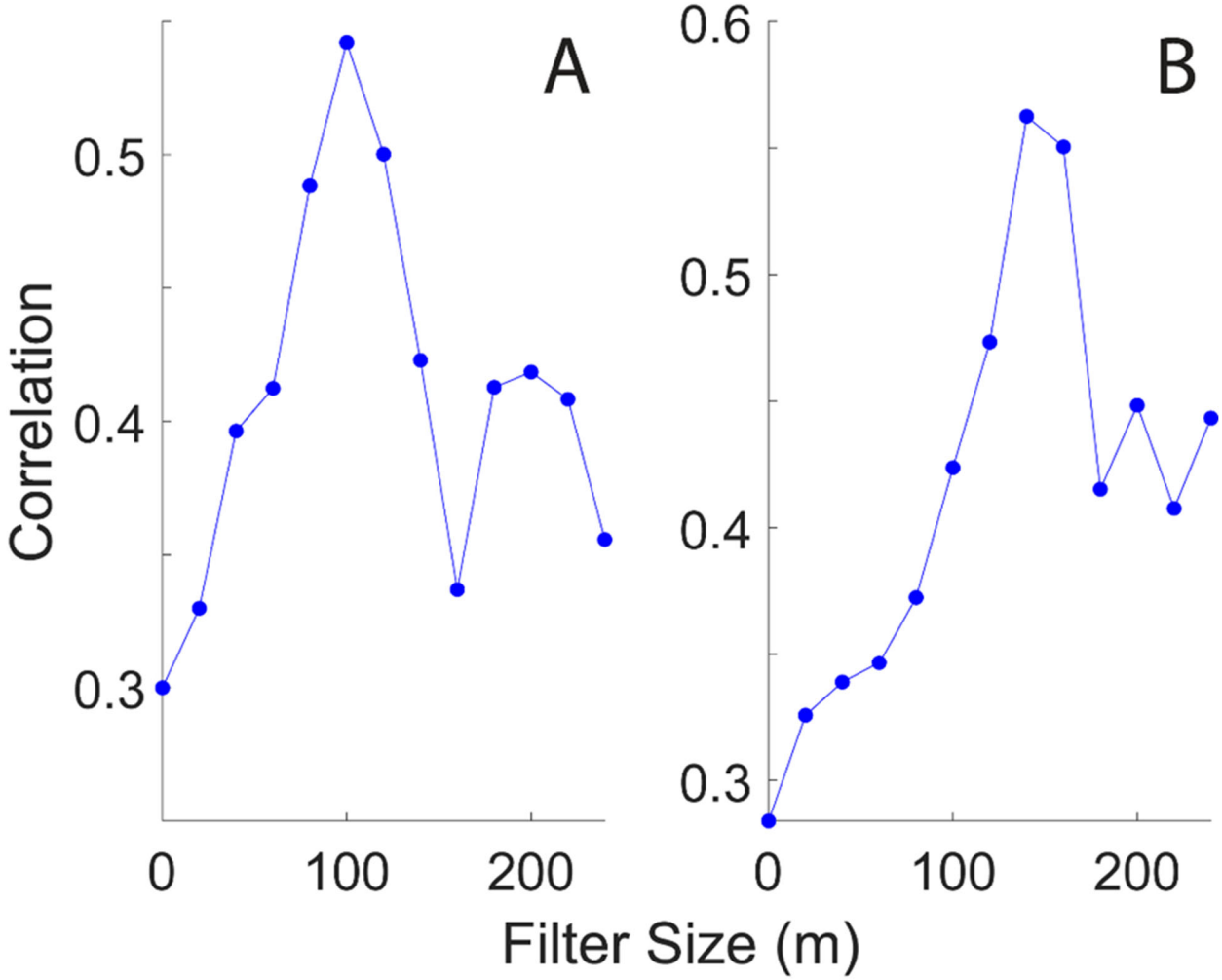


Figure 2.5: 2D Correlation Values between maps produced with filter radii from 20-240 in steps of 20 meters: (A) 2D Correlation between susceptibility maps based on entire 311 (N=720) and combined (USGS and ACES, N=134) landslide inventories. (B) 2D Correlation between susceptibility maps based on the original (N=77), and field-corrected (N=55) landslide inventories.

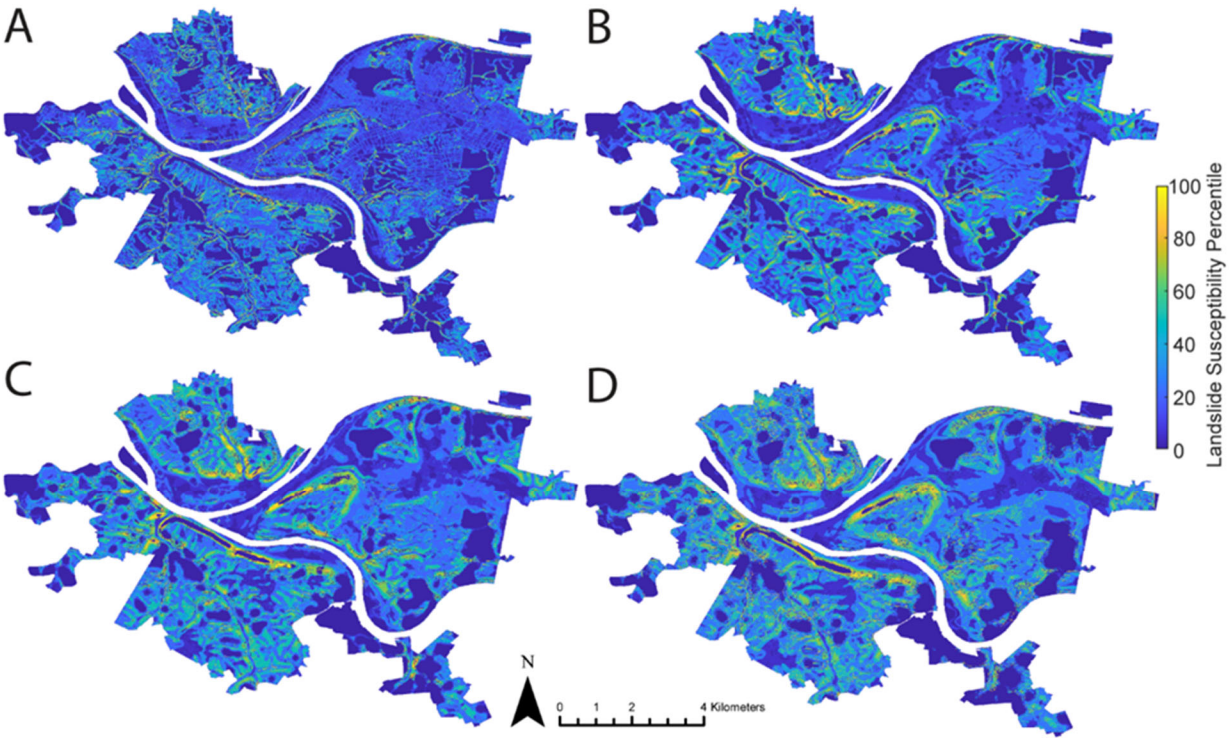


Figure 2.6: Landslide susceptibility maps based on the total 311 landslide inventory (N=720) demonstrate the effect of filtration with different filter radius: (A) Non-Filtered, (B) 80m, (C) 140m, (D) 200m. The color bar is based on the percentile of computed conditional probability values.

2.4 Results

2.4.1 Field Validation of 311 Landslide Locations

Field validation of 311-based citizen reports of landslide locations quantifies the uncertainty in the reported location of landslides. Out of 77 field validated locations, 55 were associated with an identifiable landslide in the proximity of the reported location. Out of the 22 locations discarded, 7 are duplicate reports of the same landslides. The mean distance between the reported and field validated locations is 104±25 meters (uncertainty is one standard deviation, most landslides occur between 60 to 120 meters away from the reported landslide location), and the typical size of a field validated landslide is approximately 5x13 meters (Supplementary Information).

2.4.2 Conditional Probability Analysis

The five highest ranked factors calculated for both the original and field-adjusted 311 data were similar for both inventories, and include nearest road, slope, profile curvature, distance to nearest stream, and aspect. Landslide susceptibility analysis based on the original locations of the 77 311- reported landslide differ from that based on the 55 field corrected landslide locations. The differences between the field corrected and original inventories are reflected in the spatial pattern of landslide susceptibility (Figure 2.2, Figure 2.3) and the ranking of landslide-related factors (Figure 2.4). The two-dimensional correlation between the susceptibility maps (0.3775, Figures 2.2, 2.5A) reflects the different susceptibility estimates from these two datasets. The two inventories also differ in the factor class combination that produces the highest conditional

probability (Cp). The highest Cp based on the field corrected 311 inventory, occurs at the following factor-class combination: slope (25 – 35°), nearest road (0 – 9 meters), profile curvature (0.016 - 0.1 meters-1) nearest stream (144 – 277 meters), and aspect (10 – 123°). In contrast, the highest Cp for the original, non-field corrected inventory occurs at the following factor-class combination: slope (14 – 25°), profile curvature (-0.011 - -0.002 meters-1), nearest stream (10 – 144 meters), and similar aspect and nearest road to that of the field corrected inventory. Similarly, the original and field-corrected inventories also differ in the magnitude and ranking of the landslide-related factors (Figure 2.4).

This is particularly apparent in the magnitude and ranking of the distance to the nearest road factor (NR, Figure 2.4), whose influence on the AUC is meaningfully larger for the original 311-reported locations compared to the field corrected ones (Figure 2.4A vs. 2.4B). This influence of roads on susceptibility estimates is apparent in Figure 2.22B compared to 2.22A. The range of AUC values for our analyses was 0.82 to 0.94 (Supplementary Information)

2.4.3 Filtering of Digital Elevation and Factor Maps

The correlation between the susceptibility maps produced from the field corrected and original inventories is sensitive to the scale of the averaging filter applied to the maps of landslide-related factors. Conditional probability analysis based on different filter radii (0 to 240 meters, in intervals of 20 m, Figure 2.5), shows that the spatial correlation between the susceptibility maps produced from the field corrected and original landslide inventories increases with filter radius (Figure 2.5B) up to a maxima at a radius of 140 meters, which is slightly larger than the uncertainty in landslide locations (104±25 meters). This trend repeats when comparing the landslide susceptibility map produced from the entire 311 inventory (N=720 non-field-validated landslides,

Table 2.1, Figure 2.1) to that produced from a combination of the ACES and USGS inventories (N=134 field-validated landslides, Figures 2.5A, 2.6). For these inventories, the maximal correlation is attained at a filter radius of 100 meters (Figure 2.7). Likewise, the filtering also increases the similarity in ranking of landslide-related factors (Figure 2.8). Compared to these inventories, the randomly generated landslide inventories (Figure 2.5B vs. Figure 2.9) produces a peak correlation at a similar filter radius as well as a similar decline in correlation for larger filter radii. Experiments with six factor classes rather than five show more ambiguous relations between filter radii and correlation between susceptibility maps (Supplementary Information).

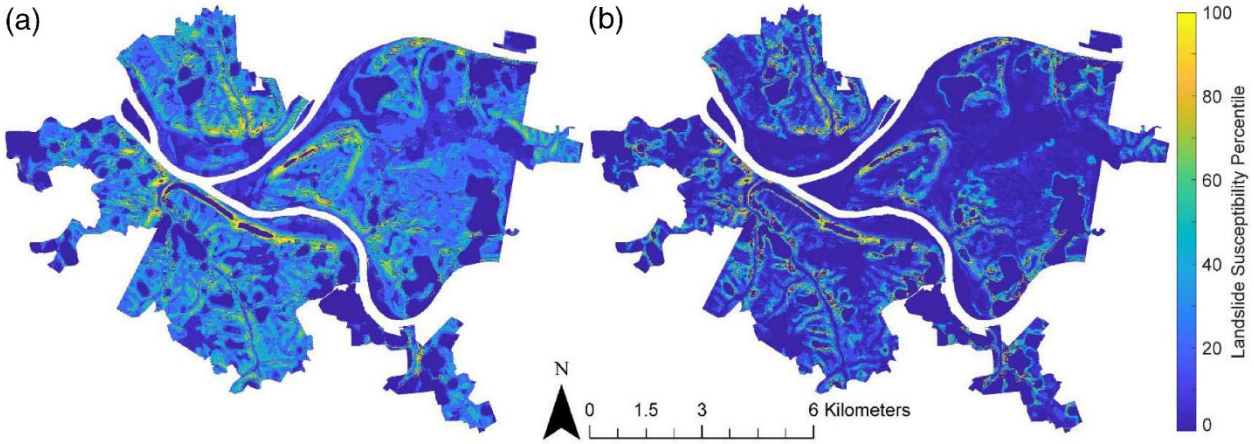


Figure 2.7: Landslide susceptibility maps that have been filtered using a 100-meter filter radius that produced a maximal spatial correlation. (A) Map based on the entire 311 landslide inventory and (B) Map based on combined (ACES/USGS) inventory. Lower susceptibility in ACES/USGS map due to less landslide coverage in the study area.

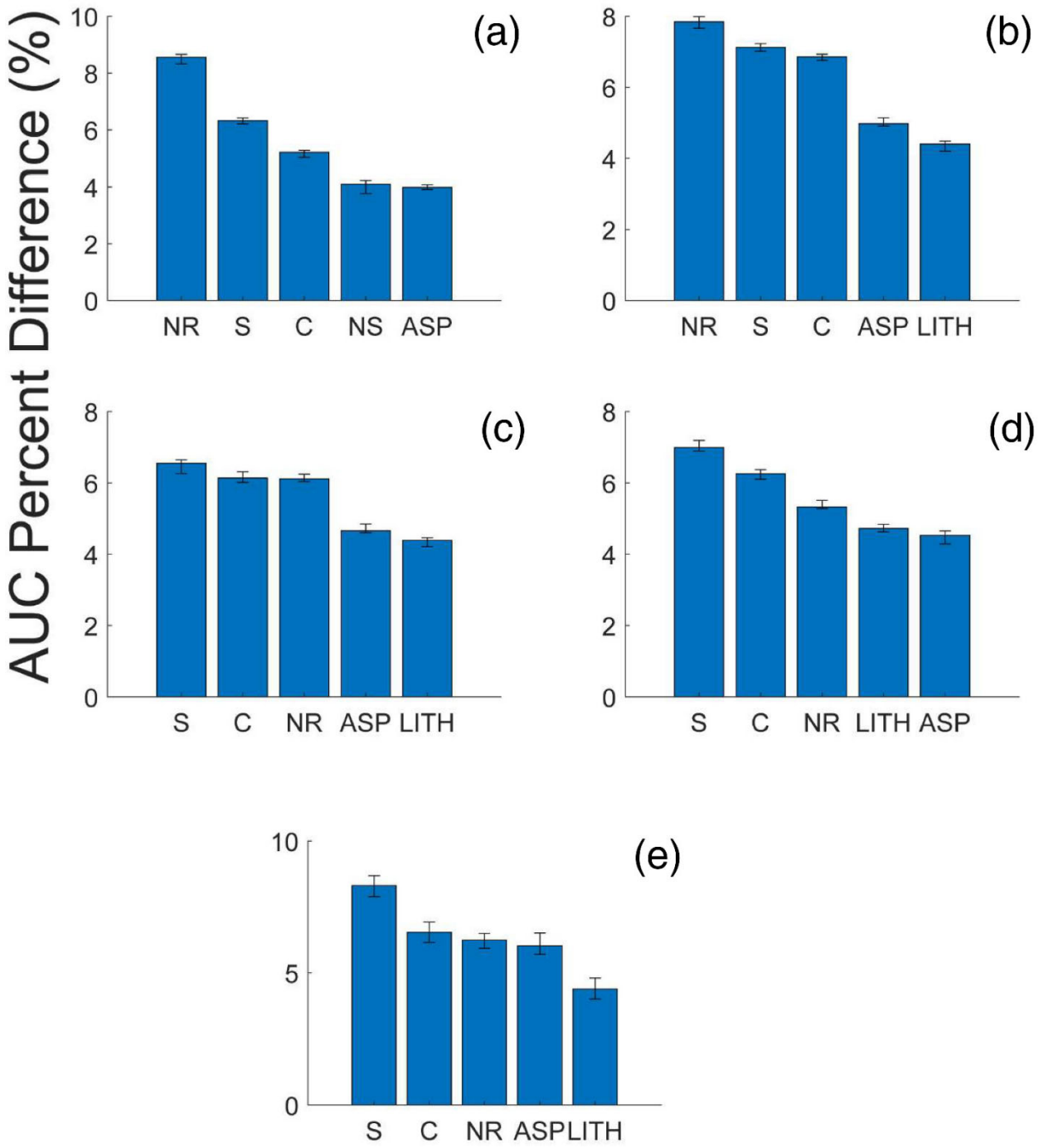


Figure 2.8: Ranking of landslide-related based on the total 311 landslide inventory (N=720) at filtering steps: (A) 20 m, (B) 80 m, (C) 140 m, (D) 200 m, (E) Influential factors of USGS/ACES inventory with no filtering. Note that higher filter radii for this non-filed validated 311-based landslide inventory (panels a-d), increases the similarity in factor ranking with the non-filtered

field based USGS/ACES inventory. Nearest Road (NR), Profile Curvature (C), Aspect (Asp), Lithology (Lith), and Nearest Stream (NS).

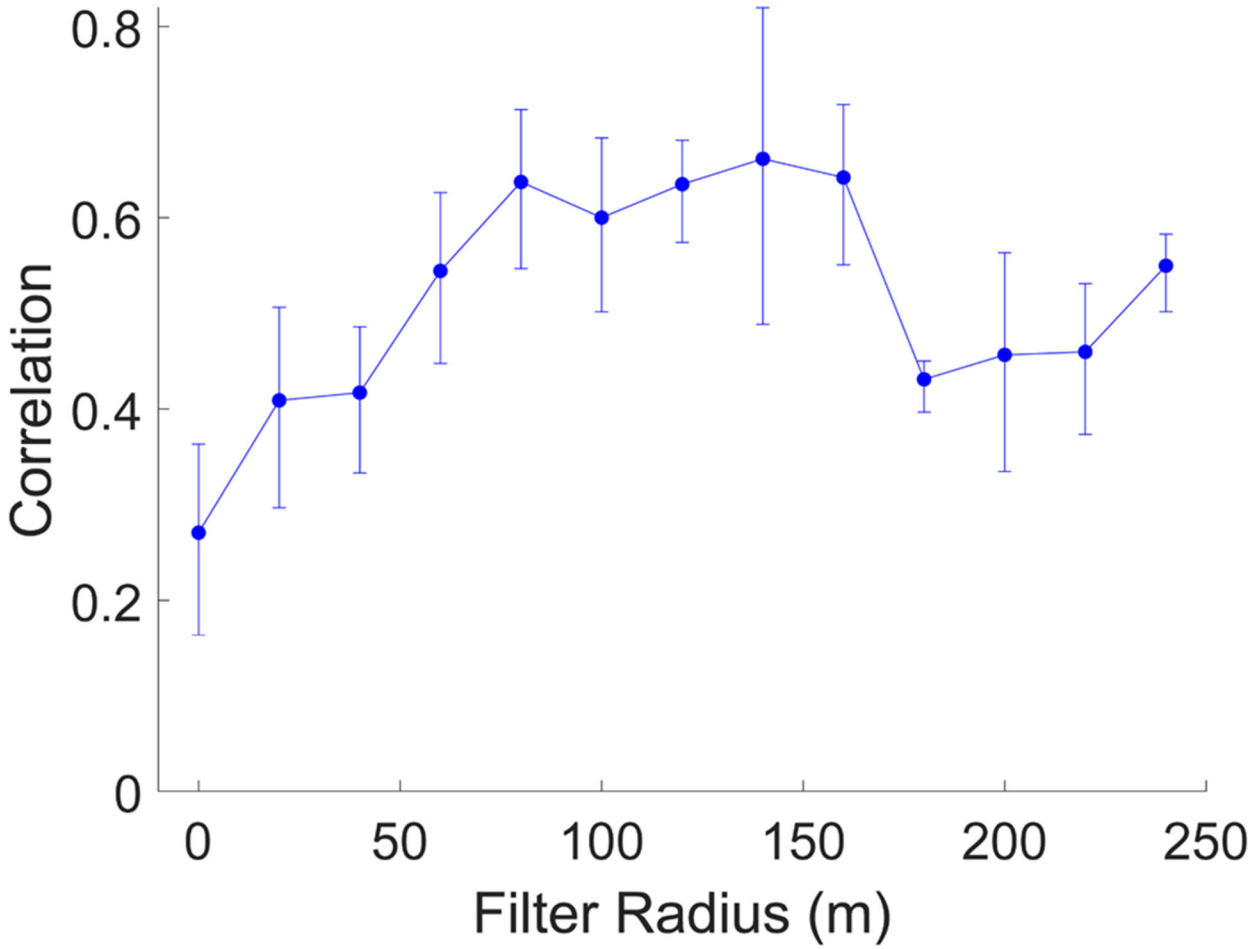


Figure 2.9: Correlation values between susceptibility maps produced from quasi- random landslide inventories versus the field corrected 311 landslide location. Error bars represent the 95th and 5th percentiles.

2.5 Discussion

2.5.1 Uncertainty in the reported locations of 311 landslides

The field validation quantifies the uncertainty in the reported 311-landslide locations. This uncertainty either reflects inaccurate locations taken by citizens or inaccuracies in how the 311 system interprets and outputs reported locations. Given that the spatial inaccuracy (104 ± 25 m) is meaningfully larger than the typical landslide size (5 m × 13 m), as well as the DEM resolution used for the susceptibility analysis, this inaccuracy may cause erroneous evaluation of landslide-related factors and susceptibility estimates (Guzzetti et al., 2000; Steger et al., 2016). Thus, validation with field or remote sensing products is essential for properly utilizing a 311-based landslide inventory in such settings. The validation of 55 out of 77,311-reported landslides (71%) suggests that 311-based reports can efficiently guide landslide mapping efforts that rely on field or remote sensing techniques, and that this dataset can help produce a progressively updating landslide inventory at a relatively low effort and cost.

2.5.2 Comparison to previous landslide studies

Our analyses of a 311-based landslide inventory in Pittsburgh are generally comparable to a previous study (Pomeroy, 1982) of landslides in Pittsburgh and surrounding counties. For example, Pomeroy (1982) found that 90% of landslides in the Pittsburgh West quadrangle map (Pomeroy, 1977) occur on slopes greater than 14°, similar to our results, where landslides are most likely to occur over slopes of 25° to 35° and 14° to 25°, for the field-corrected and original 311 inventories, respectively. Similarly, our results suggest that landslides are most like to occur on

slopes that primarily face to the northeast, similar to the findings of Pomeroy (1982). This likely stems from higher soil saturation and pore pressure on these slopes. North-facing slopes are exposed to comparably less sunlight and east-facing slopes experience sunlight in the early mornings, when temperatures are low so that overall, the drying effect from evapotranspiration is minimized on northeast-facing slopes (Pomeroy, 1982). Depending on climate, slope aspect can also influence vegetation density and the associated root strength of a hillslope (McGuire et al., 2016). Landslides analyzed by Pomeroy (1977, 1982) are more likely to occur on locations of concave upward profile curvature which is consistent with our analysis of the entire 311 inventory (i.e., non-field corrected, N = 720). This likely stems from convergence of water into these concave portions of the landscape, resulting in increased pore pressure. In contrast, analysis of the field-corrected 311 and the combined ACES and USGS landslide inventories, indicate that landslides are more likely in areas of concave downward profile curvature (Appendix Tables A.1–A.4). For the USGS inventory, this difference may be associated with the conversion from landslide-polygons to the point of highest elevation within each polygon. For the 311 inventory, where landslides often occur adjacent to roads, this may reflect the influence of roads on the profile curvature. The lithologies that are most likely to be associated with landslides are different between the two studies: the Dunkard Group in Pomeroy's study and the Monongahela Group in our study. This difference is likely due to the Dunkard Group being exposed mostly outside of Pittsburgh, in neighboring counties that were included in the USGS landslide inventory but not included in our study. This is supported by the similarity in ranking of landslide related factors when comparing the 311 vs. the ACES and USGS inventories (i.e., as mentioned earlier, the USGS inventory is a digitized version of the map produced by Pomeroy [1977]), where the USGS data is clipped to the

city limits (Figure 2.8). Overall, the similarities between our findings and the results of prior work (Pomeroy, 1977, 1982) support the value of the 311-based landslide inventory.

Comparisons between the analysis of the 311-landslide inventory to studies that use similar statistical methods in areas with different environmental characteristics (Clerici et al., 2002; Dahal et al., 2008; Pradhan & Lee, 2010; Yilmaz, 2010) reveal both similarities and differences. Unlike our study, where 311-reported landslides are mapped as point-locations, these studies rely on field verified landslides that are mapped as polygon. The studies are similar in that slope, curvature, and lithology are among the most influential landslide related factors. Studies that do examine distance to the nearest road (Pradhan & Lee, 2010; Yilmaz, 2010) also indicate that it is an important factor; however, the distance from road at which landslides are most likely is larger (i.e., 100+ m) than that computed for the total and field-corrected 311 landslide inventories (0–9 m). This difference is likely due to the higher road density in Pittsburgh compared to the other study areas and a potential bias in reporting to the 311 system where landslides next to roads are more likely to be observed and reported. Roads can be associated with modification of topography, changes to near surface hydrology, and formation of groundwater dams, and thus influence slope stability (Mirus et al., 2007). Aspect was another meaningful landslide-related factor in all studies, but whereas Pradhan et al. (2010) indicate that landslides were most likely on north and north-eastern hillslopes, similar to our findings, other studies (Dahal et al., 2008; Yilmaz, 2010) indicate that landslide likelihood is higher on south-eastern hillslopes. These differences may reflect variations in lithology, land use, or climate between study areas as well as the influence of faults and seismicity (i.e., Yilmaz, 2010). The AUC values of all studies range from 0.85 to 0.95, pointing at the high predictability of the conditional probability model. The similarities between these studies

generally supports the usability of a 311-based landslide inventory for landslide susceptibility mapping.

2.5.3 Filtering of factor maps to overcome uncertainty

Our experiment with a two-dimensional averaging filter generally suggests that the uncertainty in landslide location, as computed through field validation, can help improve susceptibility maps that are based on the 311 data. The two-dimensional correlation between the landslide susceptibility maps that rely on the original (N = 77) and field validated (N = 55) 311-based inventories, peaks at a filter radius of 140 m (Figure 2.5B), which is somewhat larger than the scale of the spatial uncertainty in landslide locations (104 ± 25 m). This filter radius also produces a similar ranking of landslide-related factors between these two 311-based landslide inventories (Figure 8). This improvement is similar to that shown in Figure 2.5A, where a filter radius of 100 m maximizes the two-dimensional correlation between the susceptibility map that is based on the entire 311-based inventory (N = 720) and the map based on the combined ACES and USGS inventories (N = 134) (Figure 5A). This suggests that the magnitude of spatial uncertainty in landslide locations, as measured from a subset of the 311-based inventory, may help scale a filter that reduce the influence of this uncertainty on susceptibility estimates from the entire dataset. Our experiment with the quasi-random landslide inventory (Figure 2.9) further demonstrates that high correlation values are attained at filter radii that are similar to the uncertainty in landslide locations. Whereas larger filter radii can also produce relatively high correlation values between these maps (Figure 2.9), they also reduce the effective map resolution (i.e., they increase the spatial extent of areas with the same landslide susceptibility, Figure 2.6), and are thus less preferable.

The filtration results are sensitive to the number of factor classes. Our experiments with six rather than five factor-classes (Supporting Information, Figures A1-A2), show ambiguous relation between filter radius and correlation between susceptibility maps. This may reflect a larger influence of outliers on susceptibility estimates, that becomes more pronounced as the number of factor classes increases.

2.5.4 Methodological limitations

The results we present are influenced by the limitations of the data sources and methodology. For example, in the USGS inventory, landslides are mapped as polygons whereas the 311 data is provided as a point location. Although we overcome this mismatch in landslide localization by converting the USGS data to point data (by using the pixel of highest elevation in the polygon), this conversion is somewhat arbitrary and may influence results. Additionally, lithology and soil characteristics play an important role in landslide susceptibility (Hamel, 1972), The analysis relies on geologic maps at the group and formation level (Berg, 1980), a coarse stratigraphic resolution that cannot resolve individual stratigraphic layers and/or soil cover that are particularly prone to landslides. This may influence the ranking of lithologic factor and decrease the accuracy of the susceptibility maps. There are also temporal and spatial difference between inventories, with the 311 and USGS inventories being collected 40–50 years apart, and the USGS data having larger landslides (at the scale of 10s–100s of meters) compared to the size of the typical landslides reported by the 311-based inventory (at the scale of 10 m). Also, limitations associated with the conditional probability model (Yilmaz, 2010), where the results are sensitive to the number of factor classes being used. In general, the more factors and/or classes that are added to the analysis, the lesser the statistical significance of the conditional probability estimates. Previous

work has shown (Li & Chen, 2020; Pham et al., 2018; Yilmaz, 2010) that the usage of machine learning methods may overcome this obstacle, however, the robustness of such results for datasets of that scale is yet to be evaluated and the interpretation of the results is not as straight forward compared to conditional probability.

2.5.5 Potential use of 311 data in the future study of landslides

Changes in weather, in particular precipitation, can meaningfully influence landslide occurrence (Hsu et al., 2018; Kumar et al., 2019; Ray & Jacobs, 2007). Whereas this is not addressed in this study. The temporal information in 311-based landslide inventories may be suitable for examining the influence of precipitation on landslide occurrence. Similarly, the association between distance to roads (NR, Figures 2.4 and 2.7) and landslide occurrence points at the potential influence of urban development on landslides. The progressive updates to the 311 data can therefore be used to compare between landslide occurrence and precipitation data over various timescales, as well as between landslide occurrence and urban development. Overall, the public availability of progressively updated 311-data enables such explorations in different cities, climates, and environmental conditions across the United States and Canada.

2.6 Conclusion

Our analysis of landslide inventories in Pittsburgh, PA, USA, suggests that a landslide inventory that relies on citizen reports to the 311 system can be used to create landslide

susceptibility maps that are consistent with field-validated inventories. Whereas the 311-based inventory is associated with a substantial spatial uncertainty, it can guide targeted field-validation efforts. Our comparison with field-validated landslide inventories suggests that the spatial uncertainty computed from a field-validated subset of the 311-inventory can help scale a simple two-dimensional filter that can reduce the influence of this uncertainty on susceptibility estimates. Future work can likely utilize the progressive updates to the 311-dataset to explore the temporal covariance between landslide, precipitation, and urban development, as well as differences in landslide patterns across climatic, lithologic, and topographic gradients in the United States and Canada.

3.0 Exploring the Influence of Precipitation on Landslide Occurrence

3.1 Introduction

Landslides are a natural hazard that occurs frequently and have a significant impact on social and economic well-being worldwide (Froude and Petley, 2018). Precipitation is widely recognized as the most common trigger of landslide occurrence, and significant efforts have been devoted to characterizing the precipitation conditions that cause landslides (Ashland, 2021; Segoni et al., 2018; Brunetti et al., 2010; Guzzetti et al., 2007; Wieczorek and Glade, 2005; Aleotii, 2004; Crosta and Frattini, 2001; Corominas, 2000; Caine, 1980). Precipitation can affect slope stability through three main mechanisms (Terzaghi,1950, Wilson, 1989): (1) by weakening the bonds between soil particles through mineral dissolution or washout, (2) by increasing pore pressure along potential sliding planes, reducing the soil's resistance to shear and making it more susceptible to failure, and (3) by eroding the base of the slope (Tucker and Whipple, 2002). Regions that have experienced landslides can be vulnerable to future landslides during heavy rainfall, especially in areas where the soil is shallow or unconsolidated (Handwerger et al., 2019). For most precipitation-triggered landslides, various climatic, surface, and subsurface factors also contribute to predisposing the slope to failure by increasing the effects of downgradient forces and/or reducing the strength of the underlying substance (Sidle and Ochiai, 2006; Terzaghi, 1950). Given the complexity of these factors and their interactions, empirical data are typically used to identify the precipitation conditions that trigger landslides.

To mitigate the damage caused by rainfall induced landslides, empirical models have been proposed that define threshold precipitation conditions for landslide induction (Crosta and Frattini,

2001; Innes, 1983; Caine, 1980). Such thresholds can guide early warning systems that reduce potential landslide damage (Guzzetti et al., 2007; Wieczorek and Glade, 2005). Typically, such threshold conditions are defined based on statistical analysis of multiple landslides in the context of the duration [T] and intensity [L/T] of the rainfall events that predate them (Caine, 1980), as both duration and intensity can affect soil saturation and erosion that trigger landslides (Posner and Georgakakos, 2015). Hence, reliable information about the timing of landslide occurrence and the precipitation condition prior to this time is vital for computing empirical precipitation thresholds.

Difficulties in determining the time of landslide occurrence hamper attempts to robustly define threshold precipitation conditions for different areas (Kirschbaum et al., 2010, Benz and Blum, 2009). The need for reliable up-to-date information on the timing of landslide occurrence is further heightened by the potential effects of climate change and landscape modification (e.g., urban expansion, Rohan et al., 2023), as changing conditions may modify landslide susceptibility. Ideally, the timing of landslide occurrence can be determined using remotely sensed data acquired at high temporal frequency (Dussaillant and Hovius, 2016). However, the spatial resolution of high frequency airborne remotely sensed data, as well as masking by vegetation canopy or clouds, may obstruct detection of small or slow-moving landslides. Citizen-science, where the public participates in data collection (Paul et al., 2019; Cieslik et al., 2019; Can et al., 2019; Franzoni, 2014), has the potential to address this gap as it allows for inexpensive collection of large amounts of data at real time. Such large datasets can potentially help define location-specific precipitation thresholds for areas with unique environmental characteristics (e.g., soil, vegetation, lithology, topography, climate). However, to obtain data of sufficient quantity and quality, citizen science data must be of interest to the public and have a standard, unbiased, and reliable procedure for data

collection (Can et al., 2019; Juang et al., 2019; Kocaman and Gokceoglu, 2019; Paul et al., 2019). Because of these limitations the scientific value of citizen-produced data is questionable, and its validation has been shown to be crucial for robust landslide research (Rohan et al., 2021; Can et al., 2019).

A publicly available source of citizen science data is based on a 311-municipal system that includes citizens' reports about landslide location and time of reporting. The 311 data combines a non-emergency phone and online reporting system for issues that warrant a response from municipal officials. The dataset is updated hourly and is accessible in over 300 cities in the United States and Canada (Choi et al., 2018; O'Brien, 2016; Schwester, 2009; Schellong et al., 2007). 311-based landslide reports have been previously field-validated in Pittsburgh PA, revealing that approximately 70% of the reported landslides coincide with landslides observed in the field (Rohan et al., 2021). However, the temporal accuracy of 311-based landslide reports, and their potential for identifying precipitation conditions that trigger landslides was never evaluated.

In this study, we explored whether a landslide inventory based on citizen science-data can be used to approximate precipitation conditions that trigger landslides. To do so, we analyzed the association between the intensity and duration of precipitation and the occurrence of landslides reported through a 311 system in Pittsburgh, PA, where landslide risk is among the highest in the USA (USGS, 2020). We examined whether the general trend of these relations is consistent with the trend expected based on prior studies and physical rationale, and whether the threshold precipitation produced by the 311-based dataset aligns with a local threshold based on an independent dataset of field validated landslides. We also examined whether the 311-based landslide dataset is consistent with expectations for the association between precipitation conditions and other landslide characteristics (reactivation of pre-existing landslides, topographic

slope, and vegetation cover). Overall, our results show that 311-based landslide inventories produce patterns that are consistent with expectations and with field validated landslides and suggest that they can be used to approximate precipitation conditions that trigger landslides.

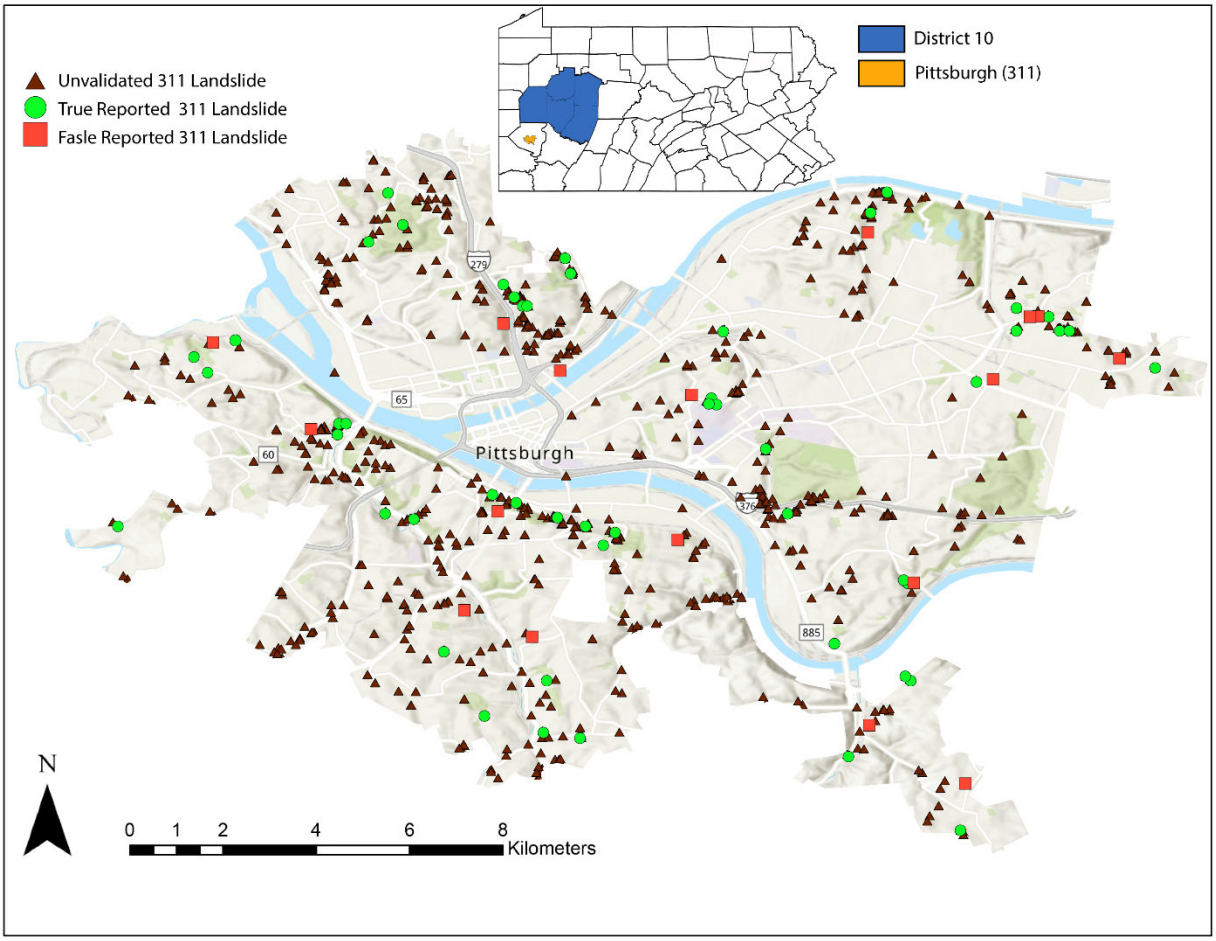


Figure 3.1: Map of the study area with locations of 311 landslide reports (2015-2022) and inset map showing location of PennDOT District 10 and city of Pittsburgh within the state of Pennsylvania. The 77 reports validated in Spring 2019 are marked by red squares (false reports) and green circles (true reports). Brown triangles mark all other reports.

3.2 Methods

3.2.1 Study Area

The study focuses on citizen reports to the 311-system in Pittsburgh, Pennsylvania (Figure 3.1). Pittsburgh is highly susceptible to landslides due to a combination of hilly topography, humid climate, clay-rich lithology, and anthropogenic disturbances to the landscape (Ashland, 2021; Rohan et al., 2021; Mirus et al., 2020; Pomeroy, 1982). The study area is primarily drained by the Allegheny and Monongahela rivers, which merge in Pittsburgh to form the Ohio River. The surficial deposits of the area are primarily composed of sand, shale, colluvium, alluvium, and gravel. Stream and river valleys are typically flanked by steep hillslopes, with a local relief of tens to a few hundred meters. A large number of prehistoric slope failures formed landslide deposits that are commonly reactivated throughout the study area (Pomeroy, 1982). The range of annual precipitation for Pittsburgh between 2015-2022 was 89 - 146 cm (National Oceanic and Atmospheric Administration, 2023), and high intensity rainstorms, which often trigger landslides, typically occur between the months of June to October (Ashland, 2021). Urbanization decreased vegetation cover and increased impervious cover across the area, with potential implications for landslide occurrence (Rohan et al., 2023; Hussain et al., 2017; Guzzetti et al., 2008).

3.2.2 Data

The landslide inventory used in this study is based on online and phone-based landslide reports to the 311 non-emergency system between 2015-2022 (Figure 1, N=1,033). Rohan et al. (2021) field validated 77 landslides reported in the spring of 2019 and found that 55 of them (71%)

were associated with an actual landslide (hereafter true reports), whereas the remaining reports were false or duplicate (hereafter false reports, Figure 1). The average area of these landslides was 65 m2, and the average distance between the landslide observed in the field and the 311-reported location was 105 ± 21 m. To examine which of the 311-reported landslides occurred over previously existing landslides (i.e., reactivation) we used an inventory of prehistoric landslides that were mapped in the 1970s by the United States Geological Survey (Pomeroy, 1982). For comparison with a 311 based rainfall intensity-duration threshold, an inventory of landslides (N=138) collected by District 10 of the Pennsylvania Department of Transportation (PennDOT), approximately 33 km north of Pittsburgh (Figure 3.1), was used for the creation of an additional rainfall intensity-duration threshold. This dataset was selected because of its close proximity to Pittsburgh, similarity in environmental characteristics, and abundant information on the timing of landslide reporting. The timing of landslides in the dataset of District 10 is based on the date when landslides were reported to PennDOT. Unlike the 311 data, all the landslides in District 10 have been identified in the field.

3.2.3 Data Processing

We used rainfall intensity-duration (ID) analysis (Segoni et al., 2018; Brunetti et al., 2010; Guzzetti et al., 2007; Wieczorek and Glade, 2005; Aleotii, 2004; Caine, 1980) in conjunction with the 311-based reporting time to explore the association between precipitation and landslide occurrence. Rainfall duration is defined as the period between the time of landslide reporting to the 311 system and the beginning of the rainfall event that predated it. The beginning of the rainfall was determined at the end of the closest period prior to the landslide report with 24 hours or more of zero precipitation. The weighted cumulative rainfall (P_t) is computed using a dimensionless

decay factor (k = 0.84, Melillo et al., 2018; Crozier, 1999) that accounts for the effects of soil saturation and evapotranspiration:

$$P_t = \sum_{i=0}^{N} k^i [E_L(i)] \ . \quad (1)$$

The index i marks the number of days (i.e., 24-hour periods) before the landslide report, so that the value of k^i decreases non-linearly with time (for k<1), the parameter $E_L\ (i)$ is the cumulative rainfall in the i'th day and N is the duration of the rainfall event in days. To calculate rainfall intensity [mm hr^{-1}], the weighted cumulated rainfall (P_t) is divided by the duration of the rainfall event. To account for the uncertainty that stems from potential deviation between the timing of landslide occurrence and that of 311 reporting, a 48-hour threshold was also explored (supplementary information).

An intensity-duration exceedance threshold identifies the intensity-duration conditions above which landslides become more likely (Guzzetti et al., 2007) and is typically expressed as:

$$I = \alpha \cdot D^{\gamma}, \quad (2)$$

Where I is the rainfall intensity (mm hr^{-1}), D is the duration of the rainfall event (hr), α is a dimensionless scaling constant, and γ is a dimensionless exponent. To constrain the variables α and γ (Equation 2) we used the Frequentist approach (Ashland, 2021; Brunetti et al., 2010; Guzzetti et al., 2007) which relies on the frequency of rainfall conditions associated with reported landslides (Silverman, 2018; Scott, 2015; Venables and Ripley, 2013). These parameters are computed through least square regression over log transformed intensity and duration values (Brunetti et al., 2010), such that the value of γ is the slope of the regression line. The distribution of regression residuals was used to compute a precipitation intensity threshold corresponding to 5% exceedance probability (T_5) which was used to compute the value of α (Equation 2) (Brunetti et al., 2015; Guzzetti et al. 2008; Iverson, 2000).

Mean values of α and γ and their associated uncertainties Δα and Δγ, were estimated using a bootstrap statistical technique. Bootstrapping is a non-parametric technique for assessing the mean sample distribution of a population from an empirical data set (Peruccacci et al., 2011; Efron and Tibshirani, 1994). The technique generates k sets of m randomly selected events from an empirical data set of n events. Analysis of the k sets allows calculating the representative (mean) value of α and γ (Equation 1) at the 5% exceedance level. The uncertainties Δα and Δγ (Figure 2) are computed from the standard deviation of the distribution of α and γ values in the k sets. we used k=5000, with m=403 and m=104 (75% of each dataset) with repetitions for the 311 and District 10 data sets, respectively.

To account for the uncertainty associated with false 311 landslide reports (i.e., citizen reports that do not correspond with an actual landslide at the reported location) we relied on the rainfall intensity associated with false 311-landslide reports identified by Rohan et al., (2021) through field validation (Figure 1). We used the precipitation intensity associated with the 80th percentile of these false reports (0.21 mm hr^{-1}) to exclude from the analysis landslides associated with precipitation intensity that is lower than this threshold.

To explore the relationship between precipitation, landslide occurrence, and the aforementioned landslide related factors (i.e., reactivation, slope, and vegetation cover), we computed threshold curves for 311-reported landslide associated with different categories of these variables. Reactivated landslides were identified as those within prehistoric landslides mapped by the USGS (Pomeroy, 1982). Other factors were interpolated from the reported point location of the 311-landslide report.

3.3 Results

3.3.1 Rainfall Intensity Duration Curve

Out of all 1032 311-reported landslides between 2015-2023, only 537 (52%) were associated with precipitation within the 24 hours prior to the landslide report. The mean intensity and duration of precipitation events associated with a 311 reported landslide were 2.3 mm hr-1 and 19.1 hours, respectively. The upper limit of rainfall intensity that was associated with a landslide was 22.5 mm hr-1. A significant portion of rainfall events associated with a reported landslide (36%) occurred in 2018 - 2019, with 88% of all landslides related precipitation events with intensities greater than 19 mm hr-1 occurring during this period.

The analysis (Figure 3.2) shows that 311-based landslide reports generally produce a similar pattern to that of published data, with a negative correlation between precipitating intensity and duration (Brunetti et al., 2010; Guzzetti et al., 2007), and that the 311-data show some unique patterns. First, whereas in prior studies the entire landslide data follows a negative intensity-duration correlation, the 311 data displays such trend mainly for the lower bound of the main cluster of landslides. Second, a small cluster of landslides is associated with very low intensity and/or duration (Figure 3.2). Analysis of this cluster in conjunction with a field validation effort (Figure 1 and 2, Rohan et al., 2021) shows that it is characterized by a relatively high occurrence of false reports (82% of the false reports detected by Rohan et al., 2021, occur within this cluster). The threshold curve calculated from the 311-data, $I = (22.43 \pm 0.51)\ D^{-0.51 \pm 0.039}$, is similar within error (Figure 2) to the curve produced from the PennDOT District 10 data, , $I = (23.27 \pm 1.91)\ D^{-0.64 \pm 0.053}$.

3.3.2 Factor Analysis

Comparing precipitation intensity and duration with different landslide-related factors revealed associations with reactivation, slope, and vegetation cover (Figure 3.3A-C). The threshold curve for reactivated landslides (N=66) is associated with lower precipitation intensity and duration compared to 311-reported landslides that are not located over old landslides (Figure 3.3A). Plotting for three topographic slope categories (0-15°, 15-30°, 30-45°) suggests that landslides that occur over lower topographic slopes are typically associated with storms of higher intensity and/or duration (Figure 3.3B). Landslides in areas with slope >45° are also associated with a higher intensity-duration rainfall compared to areas with slopes of 30-45°. Heavily vegetated areas are associated with the highest rainfall intensity-duration for landslide occurrence during low duration storm and areas of bare earth for high duration storm events. Loosely vegetated and impermeable areas are associated with lower rainfall intensity/duration (Figure 3.3C) that closely overlap and have similar slopes. The slope of the threshold curve for landslides in bare earth areas is lower than that of other areas, implying lower sensitivity to precipitation. Other factors, such as curvature, lithology, and distance to nearest stream and road showed little to no distinctive rainfall intensity-duration patterns (supplementary information).

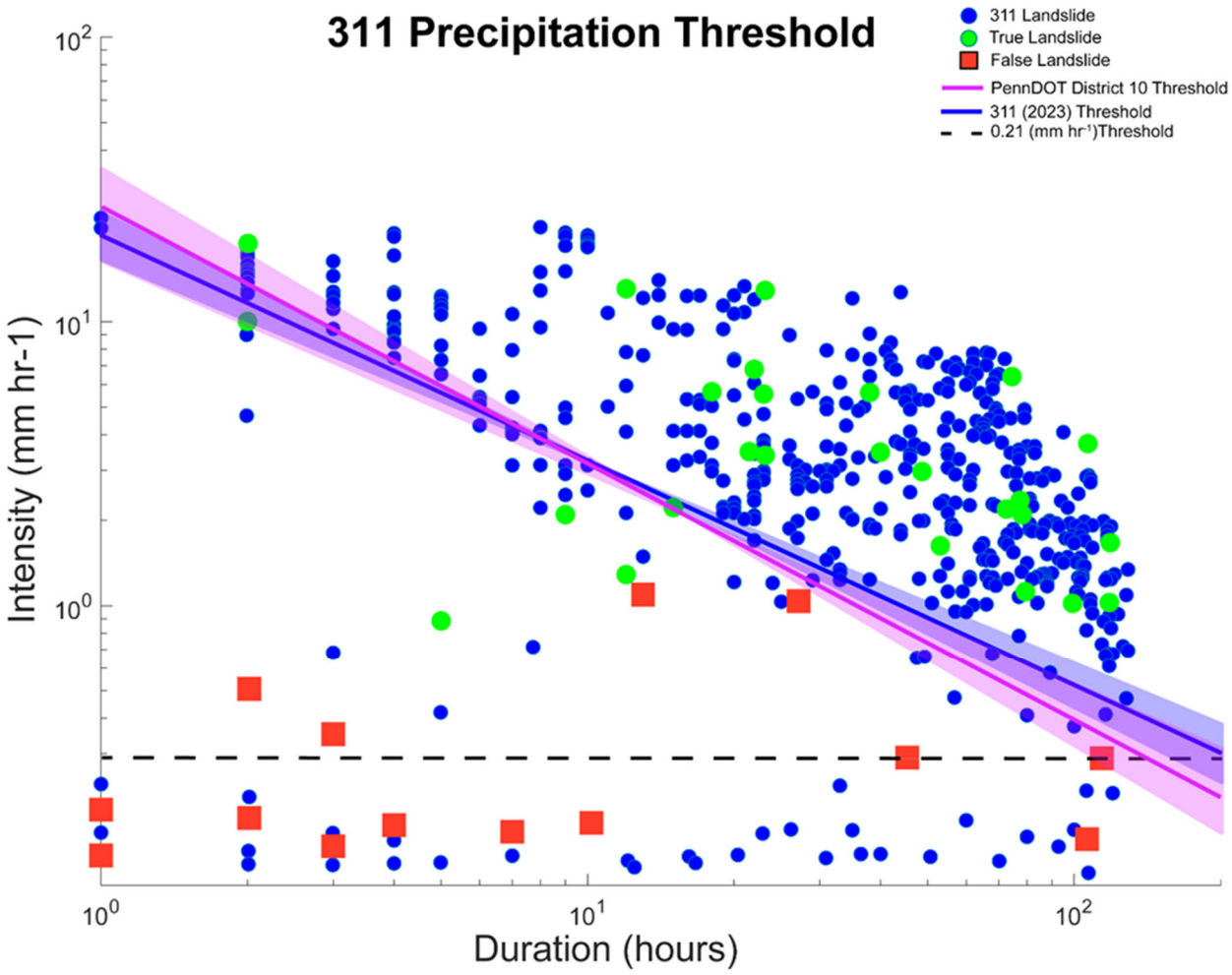


Figure 3.2: (A) Rainfall Intensity Duration for a 311 based landslide inventory. Each datapoint represents a 311-reported landslide and the intensity and duration of the rainfall that preceded it. Green circles and red squares mark true and false reports, respectively, following field validation of 77 311-based landslide reports in 2019 by Rohan et al., (2021). Lines mark intensity-duration threshold curves based on the 311-data (this study), as well as the curve by PennDOT District 10. Uncertainty in the 311 and District 10 threshold curves is marked by the shaded area around the line. Note that the threshold curve for the 311-data is based on landslides associated with rainfall

intensity > 0.21 mm hr-1, such that it excludes the cluster of landslides at low intensity values that likely represents false reports.

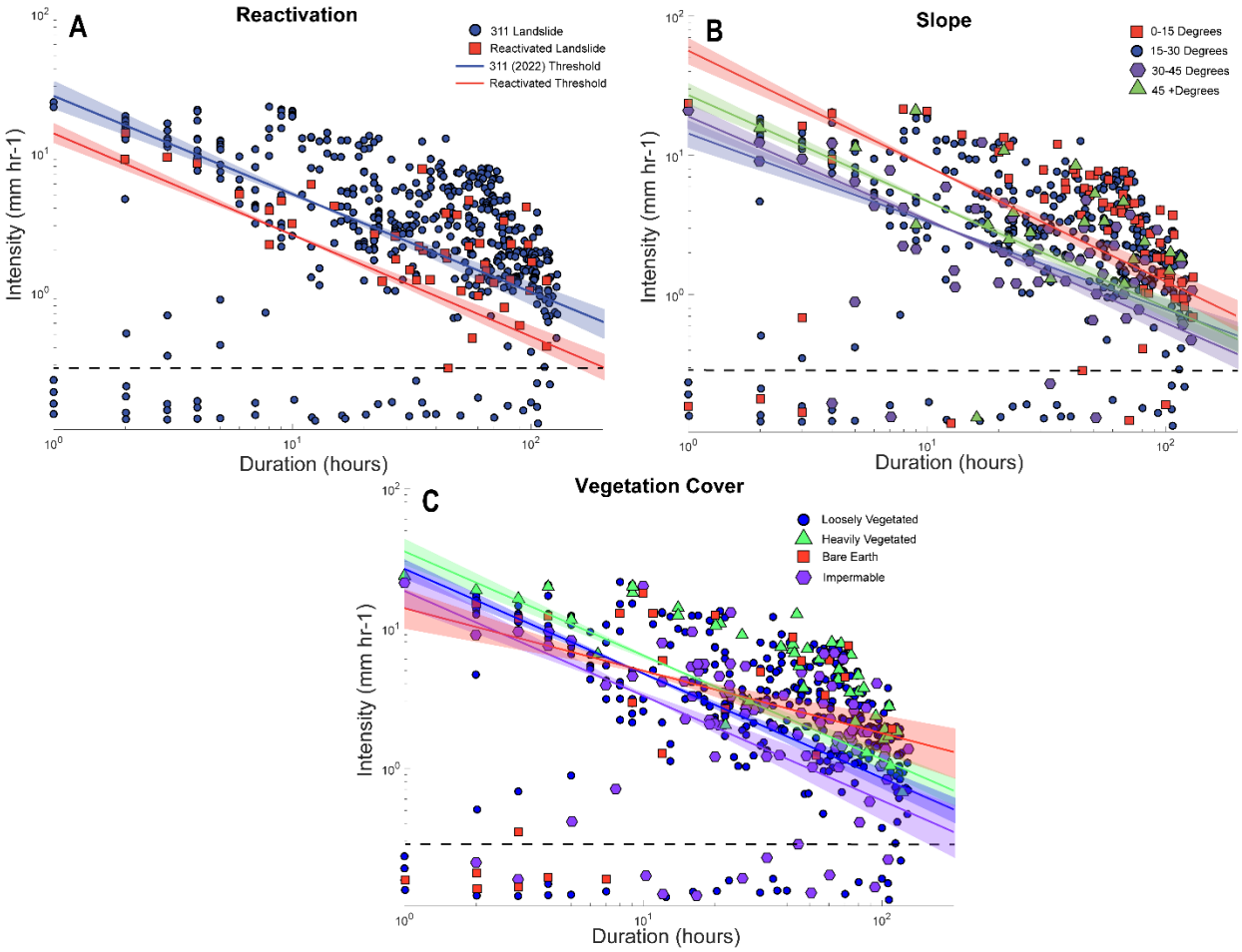


Figure 3.3: Rainfall Intensity Duration plots for different categories: (A) landslide reactivation, (B), topographic slope, and (C), vegetation cover. The colors of threshold curves match the category-based coloring of the datapoints used to compute each curve with uncertainties represented by shaded areas. The lower cluster (intensity < 0.21 mm hr-1) is excluded from threshold calculation.

3.4 Discussion

3.4.1 Rainfall Intensity Duration Pattern and Threshold

The general pattern of rainfall intensity duration for 311-reported landslides is consistent with that observed in multiple studies of field validated landslides (Segoni et al., 2018; Brunetti et al., 2010; Guzzetti et al., 2007). This pattern, where the rainfall intensity required to generate landslides decreases as the duration of rainfall increases (i.e., negative correlation), is consistent with expectations based on the role of rainfall induced saturation and erosion in generating landslides (Caine, 1980). One exception to this consistency is the cluster of 311 reported landslides at low rainfall intensity duration conditions (Figure 3.1). However, the observation that this cluster coincides with most (82 %) of the false reports identified through the field validation conducted in 2019 (Rohan et al., 2021), suggests that this entire cluster reflects false reports rather than true conditions for landslide generation. The distinction between this cluster and most of the data enables its exclusion from the analysis to minimize the effect of false reports. A second exception is that the negative correlation between intensity and duration is reflected primarily by the lower bound of the data (Figure 3.2), whereas the upper bound is defined by the extent of the intensity values. This pattern may reflect the time represented by the 311 inventory's (2015-2022), that includes two years of record precipitation (2018, 2019) for the Pittsburgh area. Seventy-six percent of landslides associated with >13 mm hr-1 rainfall intensity were reported in either 2018 or 2019 and removal of these events significantly reduces the number of points (62% decrease) in the upper right portion of Figure 2. Despite the outliers related to the high amount of precipitation in 2018-2019 the overall trend is dominated by landslides that follow a negative association between intensity and duration (Brunetti et al., 2010; Guzzetti et al., 2007).

The specific threshold values computed based on the 311-reported landslides are similar (yet not identical) to those computed based on landslides reported to PennDOT District 10 (Figure 2). The small differences between the threshold curves may reflect the difference in locations and extent of urbanization between District 10 and Pittsburgh (Figure 3.1), a difference in landslide size (the 311-reported landslides are relatively small (65 m2, Rohan et al, 2021), or statistical effects of the difference in dataset size (N=1,033, N=138, for the 311 and District 10 datasets, respectively). Despite these small differences, the similarity between the 311 and District 10 threshold curves, suggests that 311-data can be used to approximate precipitation conditions that trigger landslides at the local scale.

3.4.2 Association of Landslide Characteristics with Rain Intensity Duration

Reactivated landslides (i.e., located within areas mapped as prehistoric landslides, Pomeroy, 1982) are characterized by a lower intensity-duration threshold compared to that for all other 311-reported landslides (Figure 3.3A). Past landslides can indeed form soil and rock debris of modified hydrology and reduced cohesion, which can make a slope more susceptible to further landslide occurrences (Floris and Bozzano, 2008; Rochetti, 2007, Doglioni et al., 2012) and lower the rainfall intensity-duration conditions that trigger landslides.

Our analysis suggests that higher rainfall intensity and duration are needed to generate 311-reported landslides over lower topographic slopes (Figure 3.3B, for the 3 slope categories between 0 to 45°). From a physical perspective, higher topographic slope increases the gravitationally induced shear force that drives landslide compared to the forces that resist it (cohesion, friction) (Pourghasemi et al., 2018). The resisting forces can be also reduced by increased saturation and

pore pressure and thus when the rainfall intensity-duration is sufficiently high, landslides may occur even on low topographic slopes (Fall et al., 2006, Lee and Talib, 2005). Interestingly, our results also show that the intensity-duration threshold decreases for slopes higher than 45°. The confidence in this result is relatively low because of the small number of landslides on slopes greater than 45° (16 landslides, compared to 195, 255, and 101 in the 0-15°, 15-30°, 30-45° slope categories, respectively), yet the higher threshold for slopes >45° may reflect that steep slopes are primarily composed of bedrock, which has higher strength compared to soil.

The analysis also suggests that heavily vegetated areas require comparably higher rainfall intensity and duration to generate 311-reported landslides (Figure 3C), which is consistent with expectations. Heavy vegetation can stabilize hillslopes through soil reinforcement by roots (McGuire et al., 2016), decreased soil saturation due to increased evapotranspiration (Gonzalez-Ollauri and Mickovski, 2017), and decreased infiltration due to canopy and leaf cover. Additionally, vegetation can slow down surface runoff and reduce the erosive power of water runoff (Gonzalez-Ollauri and Mickovski, 2017). Therefore, landslide triggering over densely vegetated slopes is expected to require higher rainfall intensity and duration compared to sparsely vegetated slopes. The relatively low rainfall threshold for impermeable areas is surprising, given that such areas are expected to be less affected by rainfall induced infiltration and erosion. However, examination of landslide locations in these areas shows that they mostly occur next to roads, where "cut and fill" road construction, as well as drainage from impermeable road surface to nearby areas, may induce landslides. The relatively low coefficient (α, intercept in Figure 3) for bare earth areas is consistent with the expectation that such areas will require a lower rainfall threshold compared to vegetated areas. However, the low exponent (γ, slope in figure 3), suggest that, somewhat unexpectedly, landslides in bare earth areas are relatively less sensitive to changes

in precipitation. Examination of landslide locations in the bare earth areas (N=17) indicates that 81% were associated with sites of new land development, mostly in the suburbs of Pittsburgh. This suggests that these slides were primarily triggered by anthropogenic landscape modifications and that their association with precipitation may be coincidental, resulting in low sensitivity to precipitation. This may align with the observation that only 52% of 311 landslide reports in Pittsburgh are associated with a rainfall event, which hints at the important role of other landslide triggers.

3.4.3 Implications and Limitations

Our analysis suggests that landslide reports by citizens can help identify precipitation conditions that trigger landslides, and hence help mitigate landslide damage. The abundance and availability of information on the location and timing of landslides at a local scale through the 311-system enable evaluation of local, rather than regional or global precipitation threshold for landslide occurrence. Such threshold reflects local environmental factors and can therefore be used by local authorities to implement effective mitigation strategies. Highlighting the benefits of citizen reports and their outcomes through outreach efforts can increase public awareness and community engagement in data collection (Haque et al., 2019). Overall, our results suggest that landslide research can meaningfully benefit from a citizen science approach that engages a large group of individuals to collect data at a high pace and a lower cost compared to traditional centralized data collection approaches.

Our analysis also reveals that despite its advantages, citizen-science landslide data collected through the 311-system has some inherent flaws and biases that should be considered when analyzing this data. The occurrence of false reports, as well as the spatial uncertainty in

landslide location (Rohan et al., 2021), should be considered, potentially through validation of a subset of the reported data or cross-validation with other sources and/or remote sensing products (Ponti et al., 2018). The lag between the time of landslide occurrence and time of reporting to the 311-system can directly influence rainfall intensity-duration analysis, and therefore a validation study that focuses on this lag time is needed. However, a similar problem exists for the District 10 reports as well as for other data collection methods such as remote sensing or news reports (Hardmann et al., 2019). We also note that the landslides reported through the 311-system may represent a biased subset of all landslides (e.g., those that disturb residents through damaging driveways and property). This bias is disadvantageous in that findings cannot be generalized broadly; however, it is advantageous in that resulting analyses can be directly applied to mitigation of damage that disturbs residents.

3.5 Conclusion

This study explored the potential of landslide data reported by citizens to identify precipitation-threshold for landslide occurrence. We used data on the timing and location of landslides, from a 311-municipal service system in Pittsburgh, Pennsylvania, where landslide risk is among the highest in the USA. Analysis of 311-reported landslides produce a precipitation threshold whose trend is consistent with physical expectations, and whose slope and intercept are generally similar to those computed from an independent dataset of field-validated landslides in this area. Precipitation threshold computed for landslides that occur on different topographic slopes, vegetation cover, and within areas of old landslides, are consistent with physical expectations. These results suggest that citizen science data, obtained at a relatively low cost and

effort, can be used to compute reliable precipitation thresholds. Although field validation of citizen-science landslide data is crucial to ensure its accuracy, the time-progressive reporting by citizens may make this data particularly useful in accounting for the effect of climate change and anthropogenic landscape modifications on landslide occurrence.

4.0 Prolonged Influence of Urbanization on Landslide Susceptibility

4.1 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 et al., 2004; Wasowski et al., 2010; Chen and Huang, 2013; Reichenbach et al., 2014; Pisano et al., 2014; 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; Dragicevic 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).

4.2 Methods and Study Area

4.2.1 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 (Figure 4.1A). 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 C2). 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 1900-2020; NOAA, 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 was 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.

4.2.2 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 (Figure 4.1B). 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 (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 1947-1977 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.

4.2.3 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 U.S. Geologic Survey (USGS) and used in this study. The resolution of the DEM used for this 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 (USGS, 1999) is a seamless mosaic of best-available bare earth elevation data for the conterminous United States.

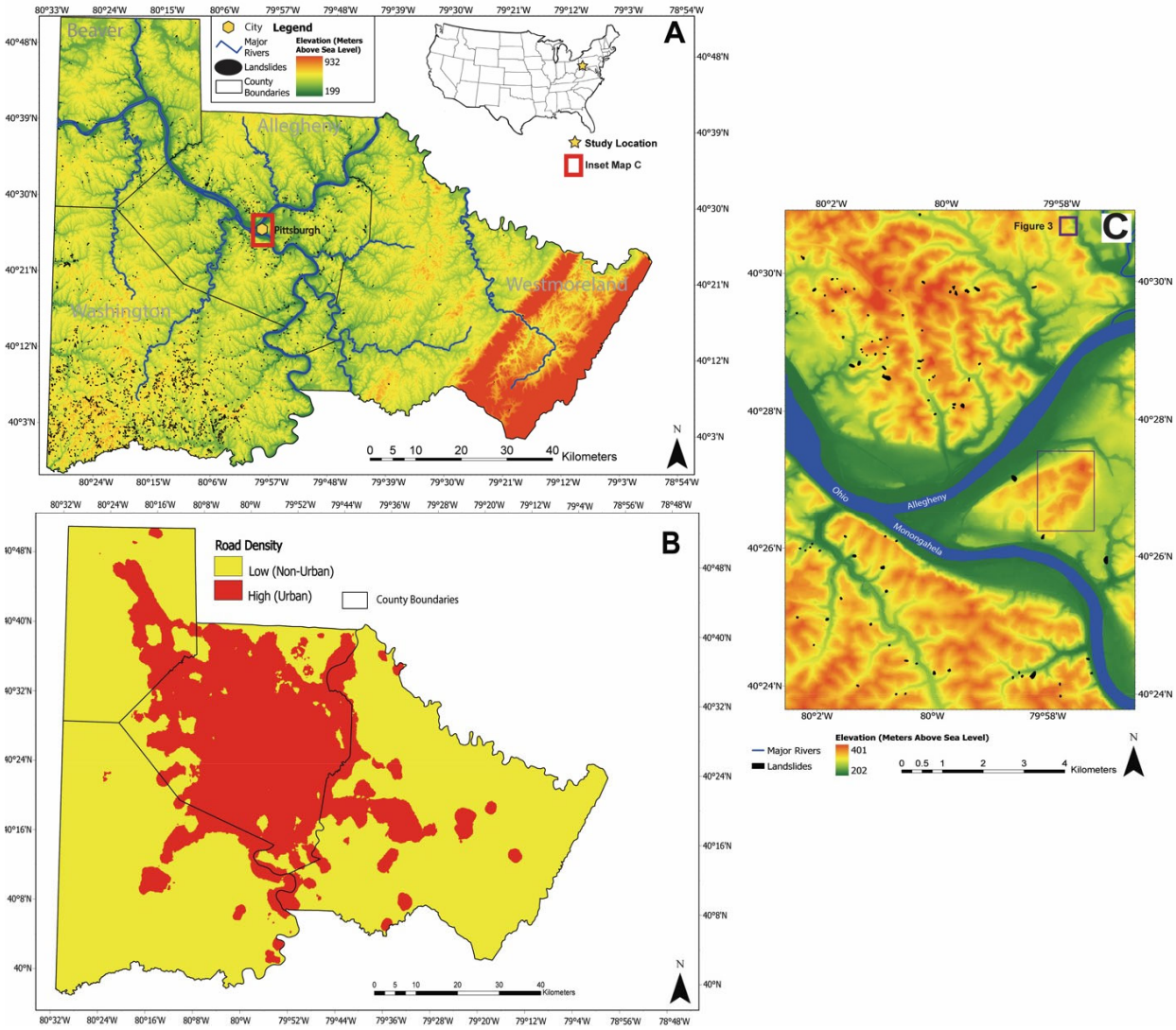


Figure 4.1: (A) Elevation map of the study area overlain by county lines, major rivers, cities, and active landslides. (B) Map of urban (red) and non-urban (yellow) areas as separated by road density (threshold 3.3 km/km2). (C) An example zoomed in map of a subset of the study area showing major rivers and mapped landslides

4.2.4 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 1973-1975 (scale 1:24,000) (Pomeroy, 1977; Figure 4.1C). 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 post-date the urbanization of the study area (Pomeroy, 1977; Briggs et al., 1975; Pomeroy, 1982). 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 km2 with maximum and minimum areas of 0.178 km2 and 0.0001 km2, 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 14C 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 km2 , 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 (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=1,762), (2) Active landslides in non-urbanized areas (i.e., areas of low road density, hereafter Active-Non-Urbanized) (N=581), (3) Old landslides in urbanized areas (hereafter Old-Urbanized) (N=319), and (4) Old landslides in non-urbanized areas (hereafter Old-Non-Urbanized) (N=348).

4.2.5 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 km2. 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.

4.2.6 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: Figures C1-C2) 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).

4.2.7 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., $100\times(|a-b|)/((a+b)/2)$, 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.

4.3 Results

4.3.1 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 (Figure 4.2) are similar for random forest, out of bag permutation, and conditional probability computations (Supplementary Information: Figure C1). 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).

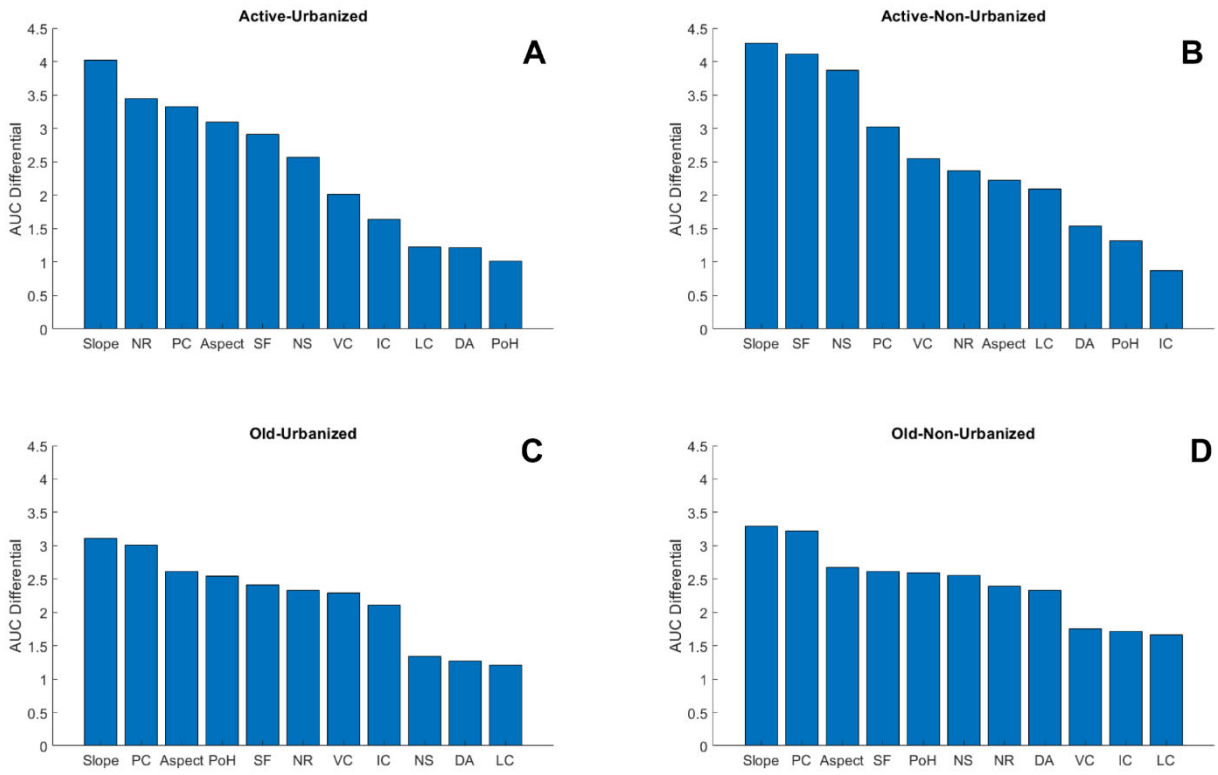


Figure 4.2: Ranking of landslide-related factors by a random forest-based AUC differential method for models based on landslide inventories of (A) Active-Urbanized Landslides, (B) Old-Urbanized Landslides, (C) Active-Non-Urbanized Landslides, and (D) Old-Non-Urbanized Landslides. Factors: Slope, Nearest 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)

4.3.2 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: Figure C3). Figure 4.3 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 the active-urbanized (Figure 4.3A) and active-non-urbanized (Figure 4.3B) 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 (Figure 4.3C) and non-urbanized (Figure 4.3D) 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 (Figures 4.3C and 4.3D), the association between high susceptibility to topographic slope is less distinct compared to models based on inventories of active landslides (Figures 4.3A and 4.3B).

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 (α=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 4.1). 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 Figure 4.3A and 4.3B, 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.

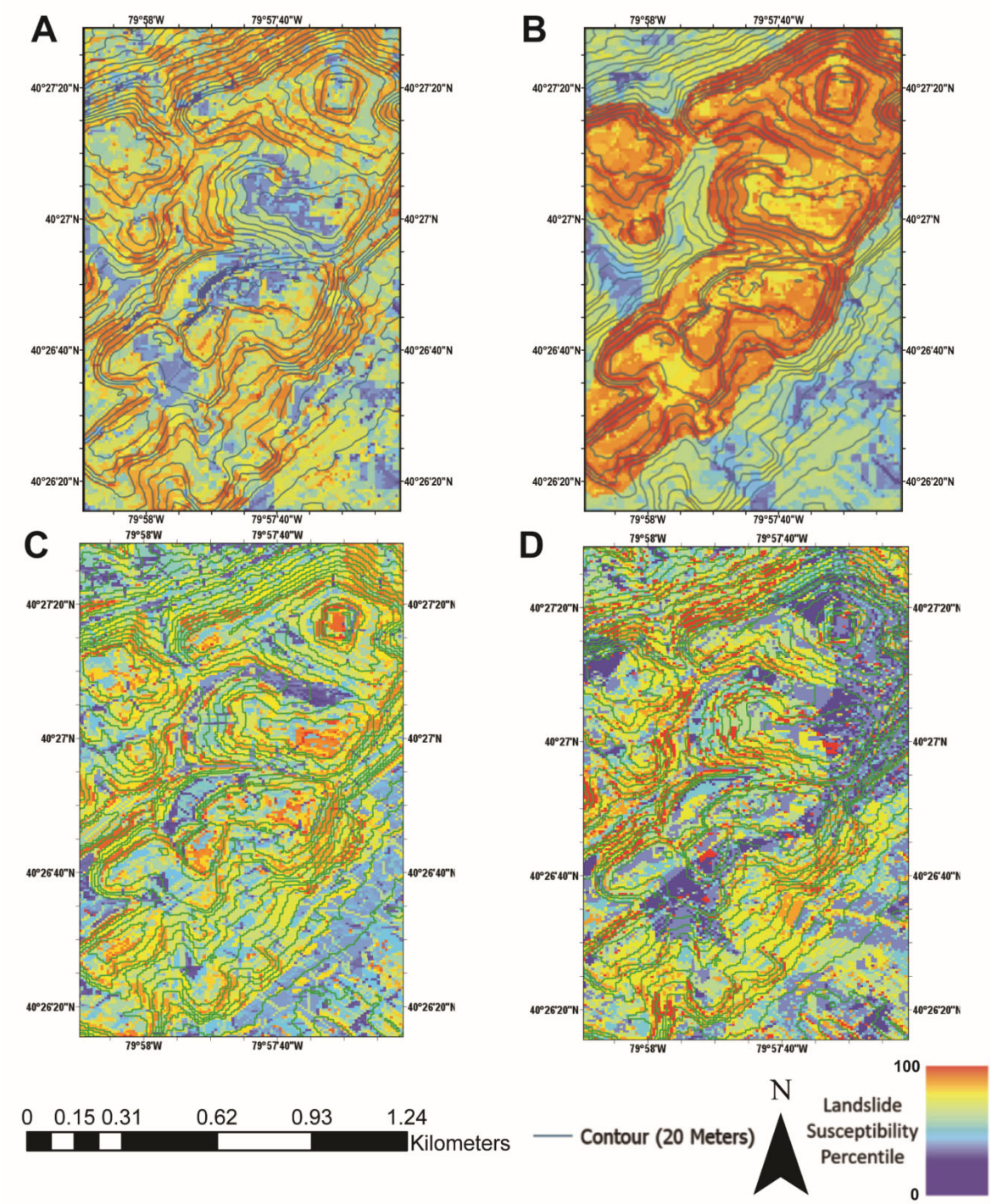


Figure 4.3: Landslide susceptibility maps of the same (urbanized) area, produced with a random forest models based on: (A) Active landslides in urban areas, B) Active landslides in non-urban areas, C) Old landslides in urban areas, and D) Old landslides in non-urban areas. Note the difference in susceptibility patterns between these models.

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

Table 4.1: Table of 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.

4.3.3 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 (Figure 4.4) 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 (Figure 4.4), and similar values for the aspect factor.

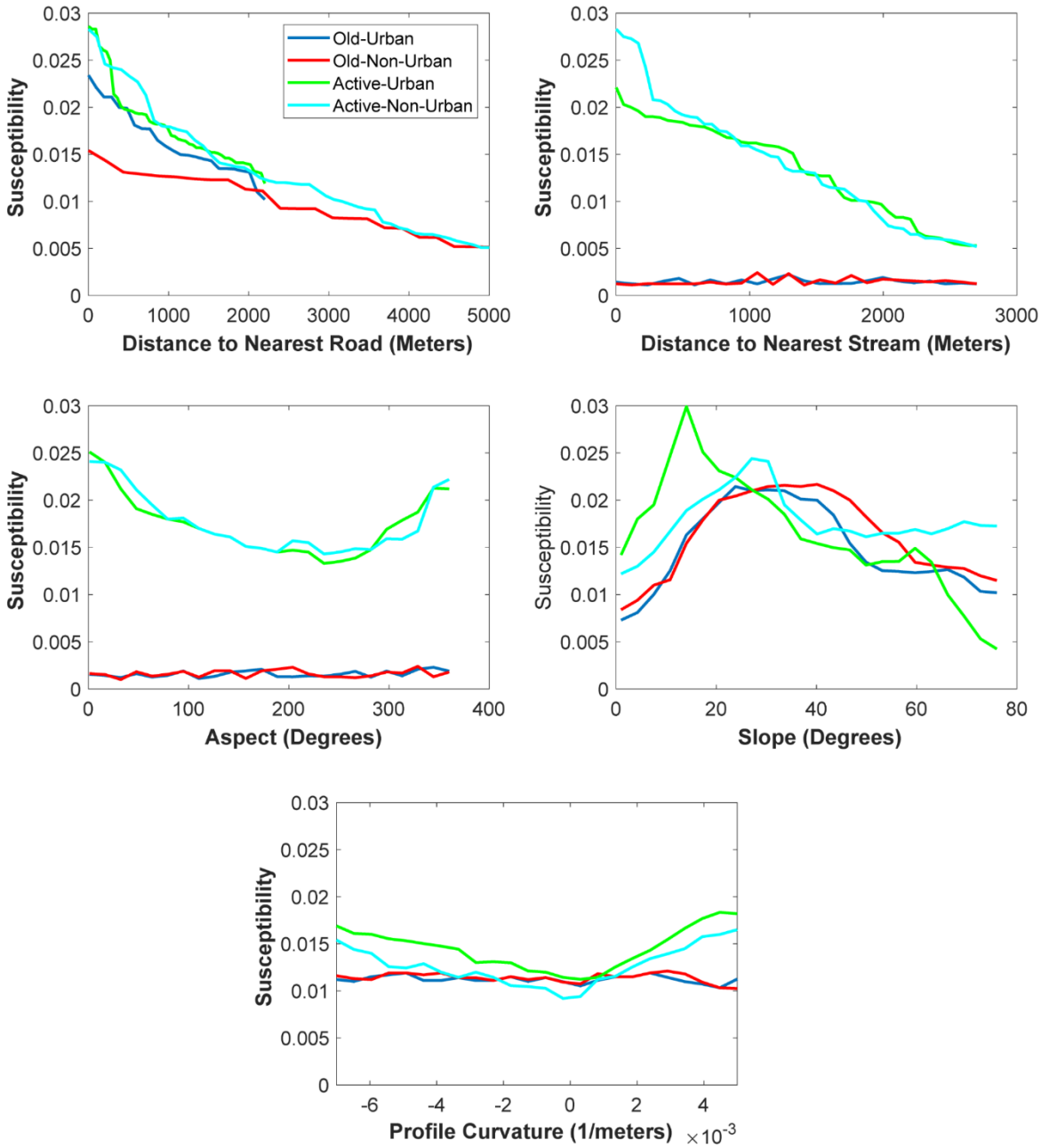


Figure 4.4: Partial dependence plots based on model results for the active and old landslide inventories in urban (green/red) and non-urban (blue/light blue) areas for key factors.

4.3.4 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 C1).

4.4 Discussion

4.4.1 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 C1), (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 (Figure 4.2), and (d) differences in partial dependence relation for some factors (e.g., slope, Figure 44.4). 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 C1), factor ranking (Figure 4.2), partial dependence relations (Figure 4.4) 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., Figure 4.3) are generally aligned with the ranking of factor. A visual comparison of Figure 4.3A to Figure 4.3B 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 on the active-non-urbanized model (Figure 4.3B), 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 Figure 4.3C and Figure 4.3D) likely reflect the influence of profile curvature, which indeed varies over short length scales. The strip of low susceptibility values in Figure 4.3D 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 (Figure 4.2D).

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 "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 section 4.3 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 (Figure 4.4), 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 1920-1954, 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 1920-1954 were still detectable at the time of their mapping in the 1970s.

4.4.2 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 (Dragicevic 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 (Dragicevic 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.

4.4.3 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 (Figure 4.4). 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 to landslides. As discussed in section 4.1, the association between susceptibility and distance to roads in the model based on the inventory of old landslides (Figure 4.4) 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 (Figure 4.4). 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 (Figure 4.4). 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 (Figure 4.4). 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 (Figure 4.4), show that the slope values associated with the highest landslide susceptibility are ~15 and ~30 degrees 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 (Figure 4.4) 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.

4.4.4 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 (Figure 4.4). Such a bias may stem from the exposure and access enabled by roads, that 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.

Another important limitation is the scale of available lithologic data. The analysis we present relies on geologic maps at the group and formation level (Miles et al., 2001), and therefore cannot explore association between landslides and specific lithologic horizons within a group or formation. It is therefore possible that the effect of lithology compared to other factors is larger than that implied by our analysis. Further, the geologic maps we use describe the underlying bedrock, and do not include soil composition and properties, such as cohesion, porosity, and permeability, that may directly affect landslide susceptibility (Hamel, 1972).

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.

4.5 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.

5.0 Conclusions and Future Directions

This doctoral thesis contributed to the understanding of landslide susceptibility and hazard assessment. And focused on leveraging citizen-reported landslide data through the 311 system in Pittsburgh, PA, USA, and on the effect of urbanization of landslide occurrence.

The thesis demonstrated the potential of utilizing citizen-reported data to create landslide susceptibility maps and to identify precipitation thresholds for landslide occurrence. It showed that while the 311-based inventory has inherent spatial uncertainty, it can be effectively used to guide targeted field-validation efforts and scale a two-dimensional filter to improve susceptibility estimates. Furthermore, the study showed that citizen science data can be used to compute reliable precipitation thresholds, emphasizing the importance of field validation to ensure accuracy.

The thesis also revealed the prolonged effects of urbanization on landslide susceptibility estimates. By comparing urbanized and non-urbanized areas, as well as pre and post urbanization landslides, it showed that urbanization has a lasting impact on landslide susceptibility, resulting in differences in the ranking of landslide-related factors and in susceptibility values. This indicates that separate susceptibility estimates should be made for urbanized and non-urbanized environments. The consistent differences in susceptibility patterns between these two types of environments further highlight the importance of understanding the complex interplay between urbanization and landslide susceptibility.

Future work can benefit from focusing on utilizing the progressive updates to the 311-dataset to investigate temporal changes in the covariance between landslides, precipitation, and urban development. This could lead to a better understanding of the relationships between these factors, and how they may change over time. Additionally, examining differences in landslide

patterns across climatic, lithologic, and topographic gradients in the United States and Canada could provide valuable insights. Such research could reveal regional variations in landslide susceptibility and help inform more targeted and effective mitigation strategies. Further research can also explore the potential biases in landslide mapping due to proximity to roads and investigate methods to account for these biases in landslide susceptibility studies. This will ultimately lead to more accurate and reliable landslide susceptibility assessments.

The time-progressive reporting by citizens can also be leveraged to account for the effect of climate change on landslide occurrence. As the frequency and intensity of precipitation events are expected to change, understanding how these changes may impact landslide hazards is essential for effective disaster risk management. Citizen science data can play a crucial role in this regard by providing valuable information on landslide occurrences and their relationship with precipitation patterns.

By harnessing the power of citizen science data and incorporating field validation, this research has provided new insights into the complex relationships between landslides, precipitation, urban development, and other factors that influence landslide occurrence. The results of this study offer promising avenues for future research, with the potential to improve our ability to predict and mitigate landslide hazards in a rapidly changing world.

Appendix A Landslide Susceptibility Analysis Based on Citizen Reports

Landslide Inventory	Factor	Bin Edge 1	Bin Edge 2	Weighted Contrast	Frequency Ratio
Total 311 (N=720)	Slope (Degrees)	0	1.5	-0.125267246	0.423094988
		1.5	11.28	0.781631438	0.758820968
		11.28	21.05	2.27246993	1.456468926
		21.05	30.83	3.697871014	2.180400402
		30.83	77.49	2.295959303	1.565799749
	Aspect (Degrees)	0	7.82	2.206447462	1.520115907
		7.82	122.24	1.147474579	0.957396588
		122.24	236.66	1.285113503	0.989511418
		236.66	351.08	1.079217554	0.922112005
		351.08	360	2.522870618	1.668582754
	Profile Curvature (Meters^{-1})	-0.16423841	-0.016018	4.766464946	2.739054596
		-0.01601837	-0.006342	2.897386811	1.809943988
		-0.006341523	0.0033353	0.999791175	0.80207483
		0.003335325	0.0130122	1.425355719	1.101592438
		0.013012173	0.0988957	1.523562302	1.198566879
	NearRoad (Meters)	0	1.96	5.822015563	3.228523455
		1.96	25.21	2.754991765	1.43203102
		25.21	48.47	-0.256881477	0.365918053
		48.47	71.72	-0.89966577	0.064894678
		71.72	758	-1.015689591	0.02515976
	NearStream (Meters)	0	10	2.597170694	1.693349347
		10	128.67	1.660105575	1.120113271
		128.67	247.35	0.713524606	0.76658148
		247.35	366.02	0.989544663	0.915423926
		366.02	1049.57	1.31441646	1.094082909
	Drainage Area (Square Meters)	0	100	-1.868171753	0.137893877
		100	4146.667	0.884188343	0.835788069
		4146.667	8193.334	1.615294377	1.126917819
		8193.334	12240	1.708667267	1.265281228
		12240	24287900	1.838157648	1.340689899
	Position on Hillslope	0	0.077	2.081403185	1.455825527
		0.077	0.33	1.594406983	1.155087799
		0.33	0.59	1.565597165	1.110293575
		0.59	0.84	0.912366073	0.854732935
		0.84	1	-0.108573251	0.429371636
	Lithology	Glenshaw Fromation		0.0306282	0.492263957
		Casselman Formation		2.5198924	1.673326727
		Monongehela Group		2.095954221	1.111339123
	Land Cover	Water		0.61774521	0.769747827
		Transportation		1.471355805	1.098762699
		Forest		-0.336627058	0.321167353
		Agriculture		1.675743365	1.392079396
		Low-Density Residential		1.902211142	1.648418413
		Medium-Density Residential		0.380974383	0.644048864
		High-Density Residential		0.178558329	0.564433805
		Identified Malls		-0.822941684	0.095299998
		Commercial		-0.578831685	0.204047291
		Light Industrial		-0.702739292	0.146284388
		Heavy Industrial		-1.001346424	0
		Strip Mine		-1.003258399	0

Appendix Table A.1: Calculated weighted contrast values and frequency ratios for each factor of the total 311 (N=720) landslide inventory.

Landslide Inventory	Factor	Bin Edge 1	Bin Edge 2	Weighted Contrast	Frequency Ratio
ACES/USGS (N = 134)	Slope (Degrees)	0	1.5	-0.954910754	0.05265471
		1.5	11.28	-0.468893488	0.311181957
		11.28	21.05	3.526633159	1.995913298
		21.05	30.83	7.019137321	3.708499278
		30.83	77.49	7.259039801	3.915030513
	Aspect (Degrees)	0	7.82	3.023889177	1.905816958
		7.82	122.24	1.627346192	1.157449606
		122.24	236.66	1.007567339	0.879262124
		236.66	351.08	1.148195734	0.950431748
		351.08	360	0.916176141	0.911012442
	Profile Curvature (Meters^{-1})	-0.16423841	-0.016018	0.471135155	0.702416986
		-0.01601837	-0.006342	1.609596825	1.218396827
		-0.006341523	0.0033353	0.523002736	0.64801742
		0.003335325	0.0130122	2.986396396	1.783421683
		0.013012173	0.0988957	7.039623174	3.803660937
	NearRoad (Meters)	0	1.96	-0.478649299	0.252981316
		1.96	25.21	0.43755206	0.630384983
		25.21	48.47	2.428468536	1.548324895
		48.47	71.72	2.140824974	1.46448885
		71.72	758	2.858982607	1.756583094
	NearStream (Meters)	0	10	-1.02453574	0
		10	128.67	1.577143461	1.08807565
		128.67	247.35	1.606051007	1.126274096
		247.35	366.02	0.956401723	0.900807646
		366.02	1049.57	0.62519107	0.771573395
	Drainage Area (Square Meters)	0	100	-2.02222736	0.109762874
		100	4146.667	1.397282106	1.04281594
		4146.667	8193.334	2.09280236	1.318445017
		8193.334	12240	0.536593544	0.725944302
		12240	24287900	-0.665957323	0.166850565
	Position on Hillslope	0	0.077	-1.020732322	0
		0.077	0.33	-0.101865381	0.435834535
		0.33	0.59	2.172992033	1.355201902
		0.59	0.84	1.805541492	1.222535819
		0.84	1	1.025208627	0.947905454
	Lithology	Glenshaw Fromation		-0.770204722	0.146171073
		Casselman Formation		1.834848919	1.348651392
		Monongehela Group		2.445040631	1.216652206
	Land Cover	Water		-0.152766123	0.407133412
		Transportation		1.627822073	0.428266608
		Forest		2.238661896	1.528840488
		Agriculture		0.762877296	0.810623523
		Low-Density Residential		-0.263299442	0.362841034
		Medium-Density Residential		-0.502607998	0.261319765
		High-Density Residential		-1.001499046	0
		Identified Malls		-0.739054246	0.134415669
		Commercial		-1.008350026	0
		Light Industrial		-1.011685707	0
		Heavy Industrial		-1.001346424	0
		Strip Mine		2.21771009	1.402751428

Appendix Table A.2: Calculated weighted contrast values and frequency ratios for each factor of the ACES/USGS (N=134) landslide inventory.

Landslide Inventory	Factor	Bin Edge 1	Bin Edge 2	Weighted Contrast	Frequency Ratio
311 Original (N = 77)	Slope (Degrees)	0	1	0.116128335	0.53008611
		1.84	14.22	0.707900792	0.725020105
		14.22	24.69	3.853167175	2.160001029
		24.69	35.16	2.234419378	1.523145393
		35.16	77.49	2.631275964	1.734961324
	Aspect (Degrees)	0	10.06	3.597703499	2.177811035
		10.06	122.98	1.365490028	1.047685025
		122.98	237.44	1.25957501	0.978861242
		237.44	352.45	0.829622248	0.819089425
		352.45	360	2.949722794	1.873255342
	Profile Curvature (Meters^{-1})	-0.16423841	-0.011234	2.995627762	1.742648591
		-0.011234201	-0.00228	2.588133932	1.5929236
		-0.002279647	0.0066749	0.936869873	0.789979283
		0.006674907	0.0156295	0.635041639	0.770956267
		0.015629459	0.0988957	0.516130391	0.72414408
	NearRoad (Meters)	0	8.65	5.355243263	3.001804084
		8.65	27.36	2.866083835	1.475763048
		27.36	48.47	-0.382562142	0.310458577
		48.47	71.72	-1.03719107	0
		71.72	758	-1.078081219	0
	NearStream (Meters)	0	10	1.008705565	0.943100781
		10	143.66	2.015320961	1.246947571
		143.66	277.31	0.195259889	0.557961874
		277.31	366.02	1.342100034	1.081240504
		366.02	1049.57	3.237228805	2.005699469
	Drainage Area (Square Meters)	0	100	-2.182877162	0.080427752
		100	2043.3334	0.084032226	0.514413081
		2043.3334	3986.6667	2.954146794	1.648484598
		3986.6667	5930	0.989284449	0.927492842
		5930	24287900	1.000663095	0.942553503
	Position on Hillslope	0	0.077	1.338871237	1.094874055
		0.077	0.33	1.107087908	0.95033503
		0.33	0.59	2.275854671	1.410552805
		0.59	0.84	0.588427411	0.724599587
		0.84	1	0.222543415	0.579913802
	Lithology	Glenshaw Fromation		-0.372675978	0.317969542
		Casselman Formation		0.227289343	0.58675093
		Monongehela Group		2.366893391	1.193076591
	Land Cover	Water		0.487639066	0.708517886
		Transportation		1.043838434	1.33989244
		Forest		0.239286098	0.5912399
		Agriculture		2.607056005	1.612223407
		Low-Density Residential		1.597471468	1.183945582
		Medium-Density Residential		-0.355982212	0.324831619
		High-Density Residential		2.688014352	4.156285291
		Identified Malls		-1.02732199	0
		Commercial		0.573058414	0.751265026
		Light Industrial		0.12579882	0.538592518
		Heavy Industrial		-1.001346424	0
		Strip Mine		-1.003258399	0

Appendix Table A.3: Calculated weighted contrast values and frequency ratios for each factor of original 311 (N=77) landslide inventory.

Landslide Inventory	Factor	Bin Edge 1	Bin Edge 2	Weighted Contrast	Frequency Ratio
311 Field Corrected (N = 55)	Slope (Degrees)	0	1	-1.089505731	0
		1.84	14.22	-1.104670879	0.095158889
		14.22	24.69	1.808798549	1.268129637
		24.69	35.16	19.52934813	11.72821953
		35.16	77.49	14.07334698	14.57367512
	Aspect (Degrees)	0	10.06	1.568338376	1.21957418
		10.06	122.98	1.251137496	1.000062978
		122.98	237.44	1.862550636	1.218138434
		237.44	352.45	0.451003094	0.663893534
		352.45	360	3.147882082	1.966918109
	Profile Curvature (Meters^{-1})	-0.16423841	-0.0112342	-0.154467986	0.406618005
		-0.011234201	-0.0022796	3.072130825	1.801228993
		-0.002279647	0.00667491	0.521871676	0.652239305
		0.006674907	0.01562946	0.836291819	0.863471019
		0.015629459	0.09889568	11.79961973	6.08281027
	NearRoad (Meters)	0	8.65	1.016859248	0.547814
		8.65	27.36	2.730127967	2.080712632
		27.36	48.47	1.132277419	0.977944518
		48.47	71.72	1.592387635	1.214472927
		71.72	758	0.816406211	0.841784027
	NearStream (Meters)	0	10	0.393730938	0.660170547
		10	143.66	2.595137433	1.577360926
		143.66	277.31	0.471305794	0.669554249
		277.31	366.02	1.385912163	1.100899422
		366.02	1049.57	1.365916832	1.123191703
	Drainage Area (Square Meters)	46	46.05	-2.392091687	0.04222457
		46.05	62.07	1.598824392	1.152285301
		62.07	78.09	2.465304702	1.454966841
		78.09	94.11	0.733536086	0.811556236
		94.11	170.06	0.105355115	0.527829961
	Position on Hillslope	0	0.077	-1.041421927	0
		0.077	0.33	1.81617453	1.252206157
		0.33	0.59	1.899906677	1.256674317
		0.59	0.84	1.444976803	1.082068717
		0.84	1	-0.340314276	0.324751729
	Lithology	Glenshaw Fromation		-0.284385063	0.356125886
		Casselman Formation		0.722491603	0.821451302
		Monongehela Group		2.311192698	1.176272695
	Land Cover	Water		-1.017874886	0
		Transportation		1.658873862	1.020145524
		Forest		0.743599241	0.82773586
		Agriculture		1.14341633	0.410695481
		Low-Density Residential		0.416936466	0.663009526
		Medium-Density Residential		-0.475969115	0.27285856
		High-Density Residential		0.163819712	0.818799407
		Identified Malls		-1.02732199	0
		Commercial		-1.008350026	0
		Light Industrial		-1.011685707	0
		Heavy Industrial		-1.001346424	0
		Strip Mine		-1.003258399	0

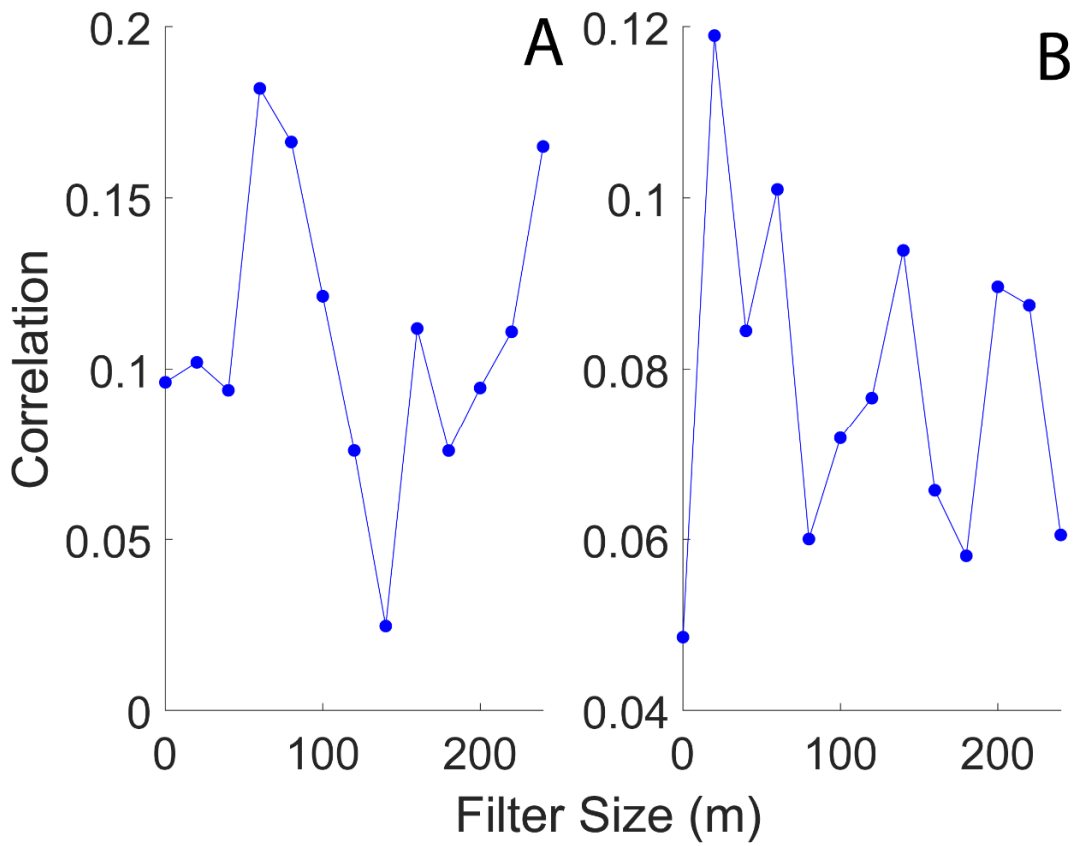
Appendix Table A.4: Calculated weighted contrast values and frequency ratios for each factor of field corrected 311 (N=55) landslide inventory.

Report Date	Validation Date	Distance Away from Reported Location (m)	Size (m)
5/2/2019	5/14/2019	58.44549986	1x4
5/7/2019	5/14/2019	88.05737656	4x8
5/20/2019	6/10/2019	138.0847507	2x5
5/21/2019	6/10/2019	82.75485942	1x3
5/21/2019	6/10/2019	129.1097482	1x4
5/22/2019	6/10/2019	110.0283354	13x25
5/24/2019	6/11/2019	139.956683	10x17
5/24/2019	6/11/2019	62.49148001	4x10
5/27/2019	6/11/2019	102.6924804	7x16
5/28/2019	6/11/2019	79.66112294	8x17
5/29/2019	6/12/2019	173.5025831	2x6
5/29/2019	6/12/2019	126.0149908	2x5
5/29/2019	6/12/2019	138.6405611	11x25
5/30/2019	6/13/2019	83.07349012	5x13
5/30/2019	6/13/2019	96.51556475	6x16
5/30/2019	6/14/2019	100.9876975	5x14
5/30/2019	6/14/2019	132.2440813	5x14
5/31/2019	6/18/2019	100.8645199	2x6
6/3/2019	6/18/2019	123.1951452	15x29
6/4/2019	6/18/2019	60.41858836	16x33
6/4/2019	6/18/2019	99.13222005	4x12
6/5/2019	6/24/2019	88.42222723	9x19
6/6/2019	6/24/2019	71.24309988	5x14
6/7/2019	6/26/2019	118.781822	4x11
6/7/2019	6/26/2019	113.6292367	5x14
6/8/2019	6/26/2019	107.9633413	4x10
6/11/2019	6/26/2019	76.79214289	6x16
6/11/2019	6/26/2019	132.9068239	3x5
6/13/2019	6/26/2019	115.1840906	12x14
6/14/2019	6/28/2019	100.3812945	10x14
6/14/2019	6/28/2019	107.7218873	3x7
6/20/2019	6/28/2019	101.2265041	5x10
6/20/2019	7/2/2019	67.295158	7x14
6/21/2019	7/2/2019	100.6871578	5x10
6/24/2019	7/2/2019	88.24484354	9x19
7/2/2019	7/10/2019	84.87409624	4x8
7/2/2019	7/10/2019	80.83404225	12x26
7/2/2019	7/10/2019	95.03489791	8x16
7/5/2019	7/11/2019	61.53990522	7x14
7/5/2019	7/12/2019	129.1844308	5x11
7/5/2019	7/12/2019	119.0573703	3x8
7/8/2019	7/12/2019	106.743365	1x5
7/8/2019	7/12/2019	106.4072192	9x22
7/8/2019	7/14/2019	89.00187027	8x14
7/9/2019	7/14/2019	130.4260244	5x13
7/9/2019	7/14/2019	104.1626415	5x14
7/10/2019	7/14/2019	90.90871227	4x7
7/10/2019	7/14/2019	138.0115955	3x7
7/11/2019	7/14/2019	102.0752199	14x26
7/11/2019	7/15/2019	93.7701381	6x14
7/11/2019	7/15/2019	100.502418	4x9
7/12/2019	7/15/2019	87.86728165	3x7
7/12/2019	7/15/2019	81.66107473	3x6
7/12/2019	7/15/2019	164.792793	5x13
7/12/2019	7/15/2019	144.9453451	8x18

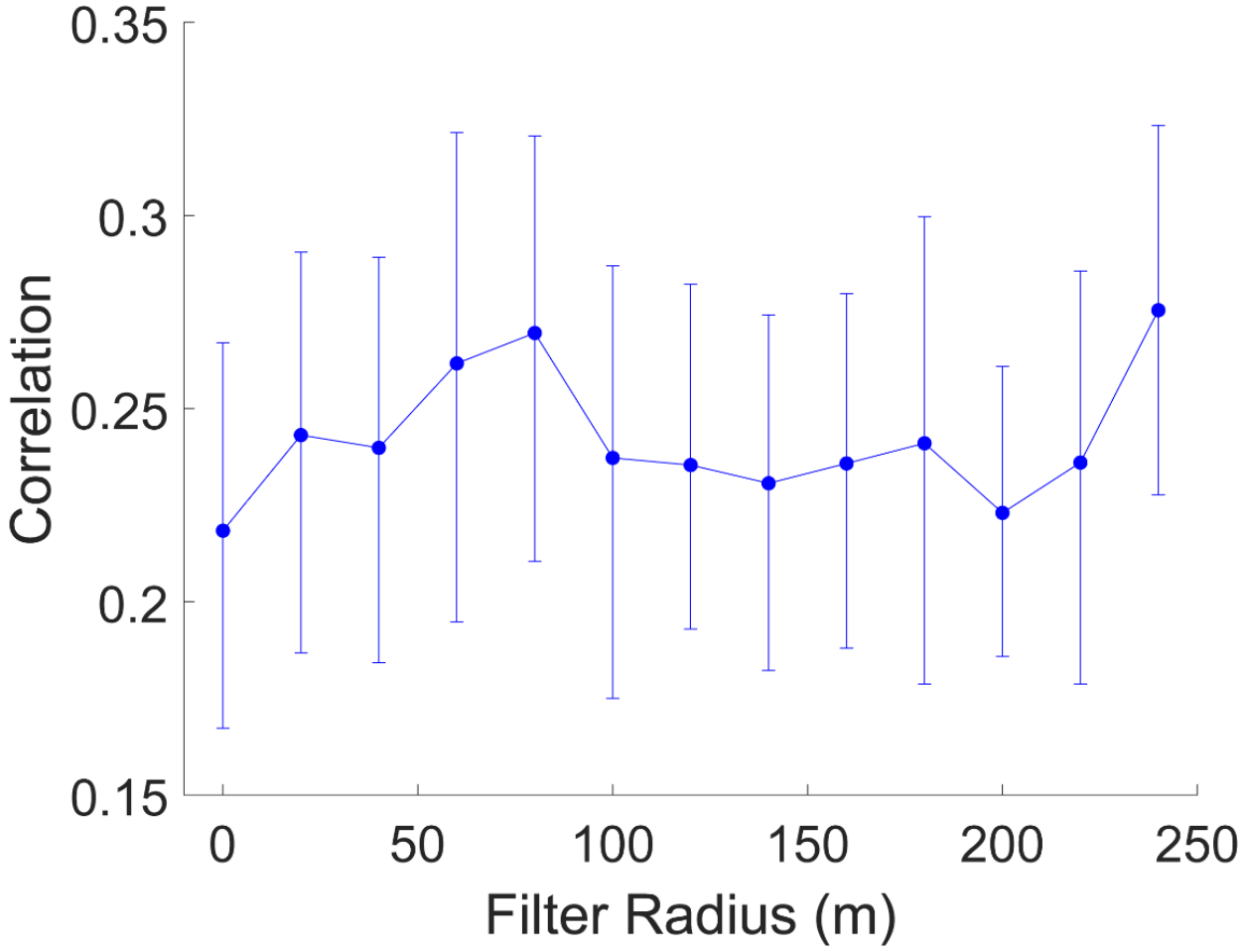
Appendix Table A.5: Calculated weighted contrast values and frequency ratios for each factor of field corrected 311 (N=55) landslide inventory.

Inventory	Filter Radius (m)	AUC	Inventory	Filter Radius (m)	AUC
311 Total	0	0.8203	311 Original	0	0.8925
	20	0.8211		20	0.8936
	40	0.8215		40	0.8957
	60	0.8345		60	0.8967
	80	0.8378		80	0.9084
	100	0.8391		100	0.9097
	120	0.8411		120	0.9001
	140	0.8395		140	0.8917
	160	0.8312		160	0.9057
	180	0.8341		180	0.8963
	200	0.8357		200	0.8977
	220	0.8369		220	0.8943
	240	0.8304		240	0.8957
ACES/USGS	0	0.9382	311 Field Corrected	0	0.9507
	20	0.9411		20	0.9514
	40	0.9427		40	0.9527
	60	0.9483		60	0.9452
	80	0.9441		80	0.9412
	100	0.9472		100	0.941
	120	0.9451		120	0.9405
	140	0.9439		140	0.9412
	160	0.9431		160	0.9394
	180	0.9407		180	0.9387
	200	0.9432		200	0.9395
	220	0.9449		220	0.9388
	240	0.9428		240	0.9373
Total	0	0.8269			

Appendix Table A.6: AUC values for all conditional probability analyses that were used to generate landslide susceptibility maps, 311 total (N=720), ACES/USGS (N=134), 311 Original (N=77), and 311 Field Corrected (N=55). Total is the combination of all landslide inventories for Pittsburgh, PA.

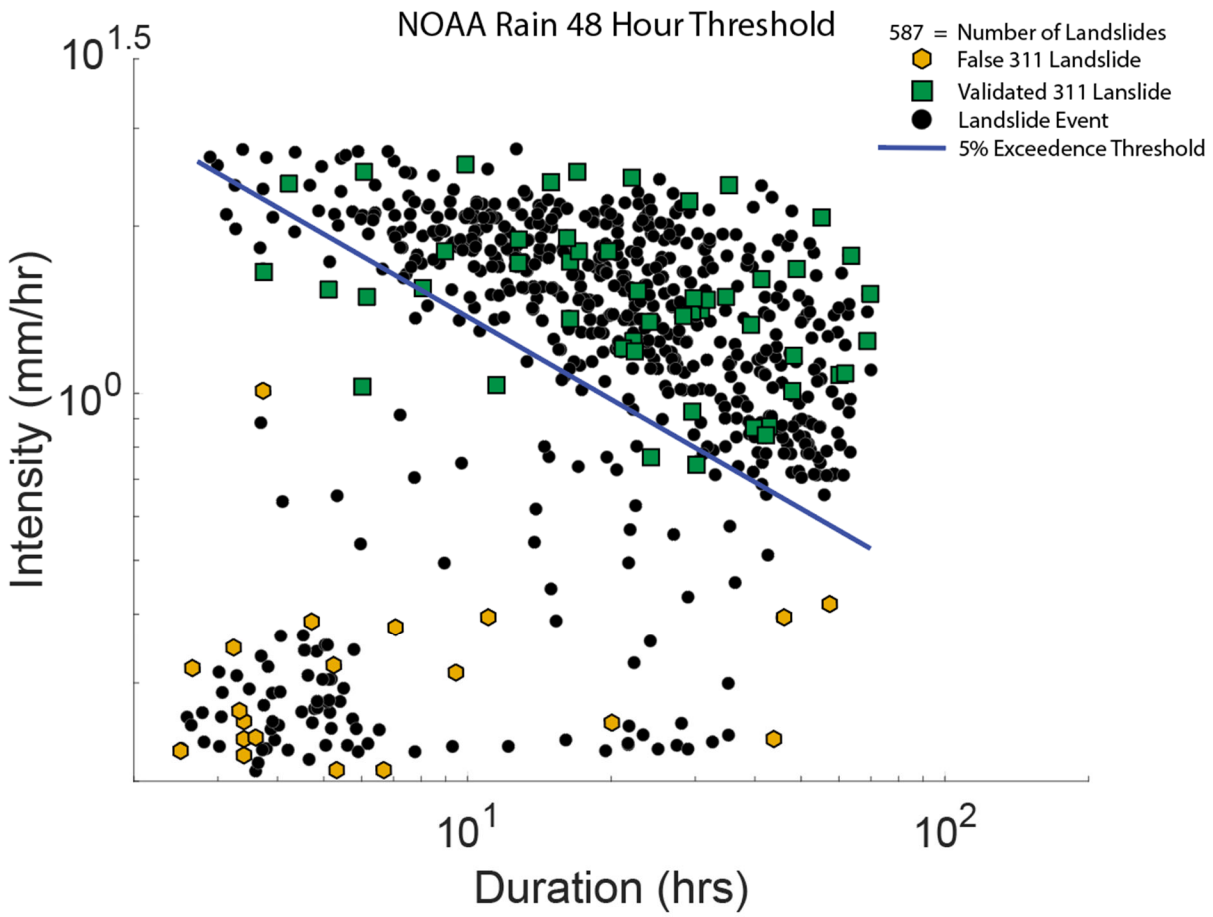


Appendix Figure A.1: Cross-Correlation Plots for the filtered maps with six factor class bins for (A) the original (N=77) and field corrected 311 (N=55) locations and (B) the combined (N=134) USGS/ACES inventory and the total (N =720) 311 inventory.

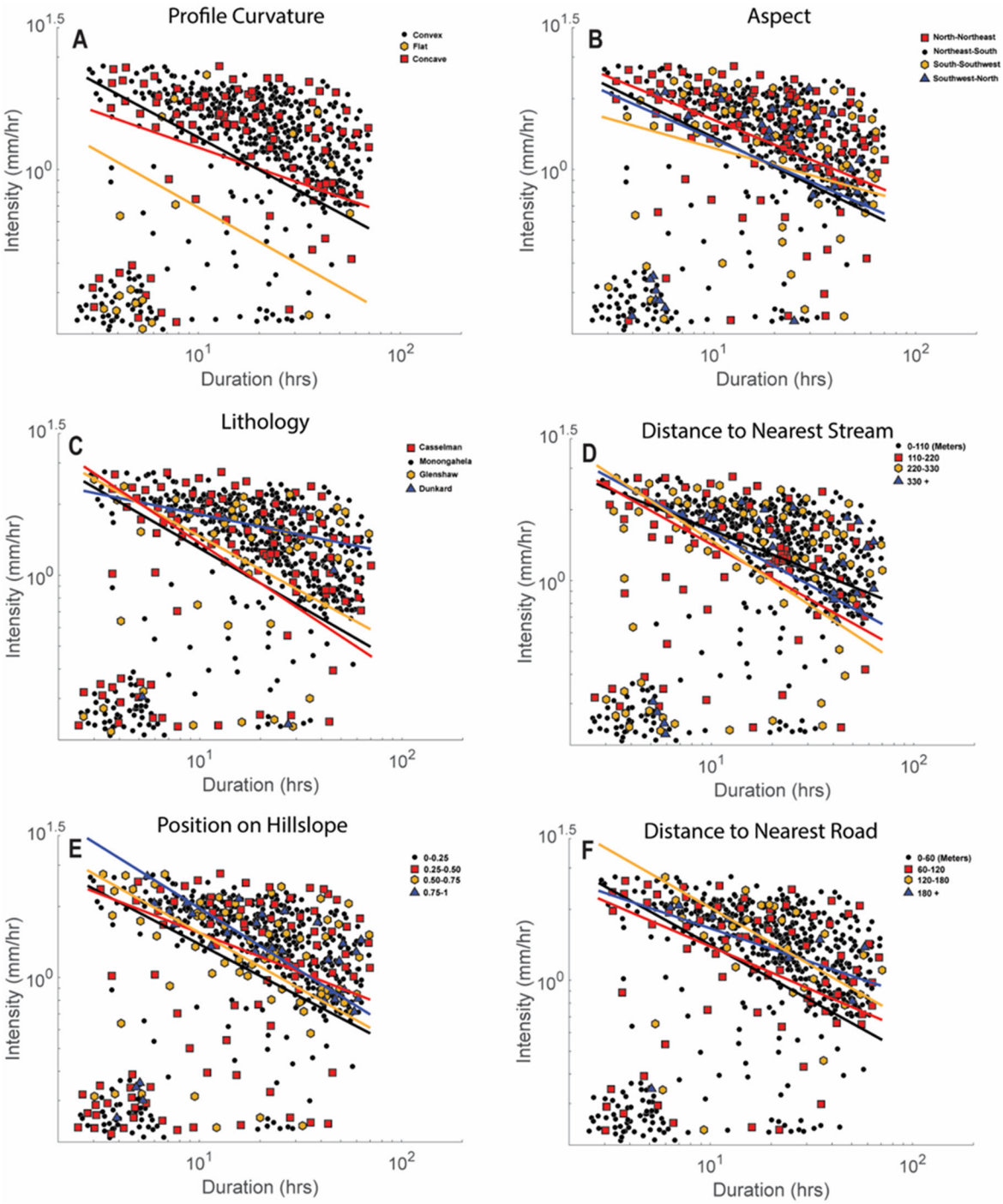


Appendix Figure A.2: Correlation value between susceptibility maps produced from quasi-random landslide inventories versus the field corrected 311 landslide location. This experiment based on conditional probability with six classes per factor rather than five. Error bars represent the 95th and 5th percentiles.

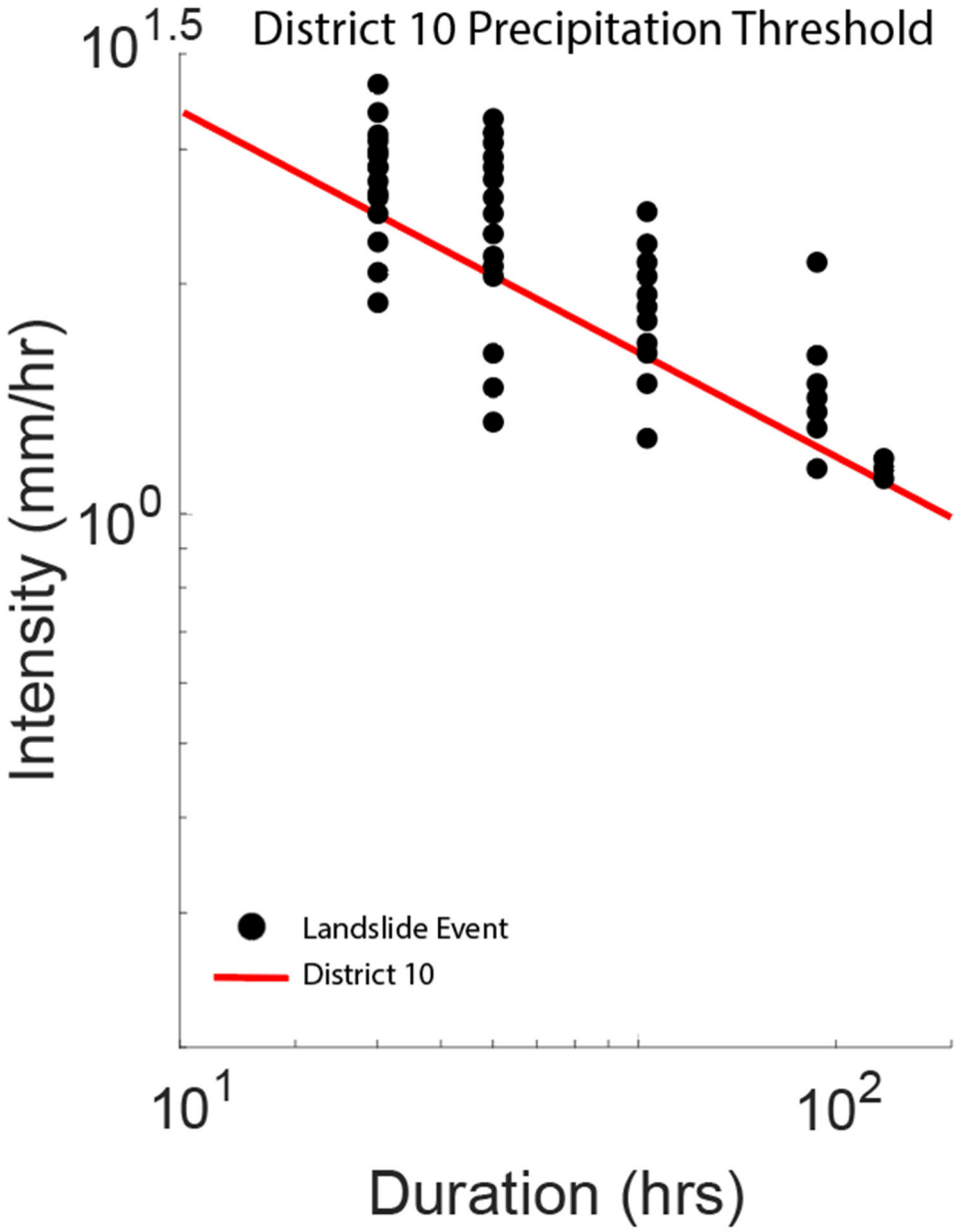
Appendix B Using Citizen Science to Explore the Influence of Precipitation



Appendix Figure B.1: Rainfall Intensity Duration plot for 48-hour threshold using 311 citizen science data.



Appendix Figure B.2: Rainfall Intensity Duration plots with overlain landslide-related factors: reactivation of old landslides (A), Profile Curvature (B), Aspect (C) Lithology, (D) Distance to Nearest Stream, (E) Position on Hillslope, and (F) Distance to Nearest Road. Threshold lines calculated for each binned category.



Appendix Figure B.3: Rainfall Intensity Duration plot with overlain threshold for PennDOT district 10 data.

Appendix C Prolonged Influence of Urbanization on Landslide Susceptibility

Appendix C.1 Conditional Probability

To corroborate the random forest analysis, we used a conditional probability approach to produce landslide susceptibility maps and rank the influence of the different landslide-related factors (e.g., topography, land use, lithology) on landslide occurrence (e.g., Chung, 2006; Ozdemir, 2009; Yilmaz, 2010; Regmi et al., 2014; Costanzo and Irigaray, 2020). To do so, we divided each of the 11 landslide-related factor to five classes that span the range of values for this factor in the maps of the study area. For each of the resulting factor-class combinations, we then computed the conditional probability C_p:

$$C_p\text{-}j=N_l\text{-}j/N_p\text{-}j, \quad (1)$$

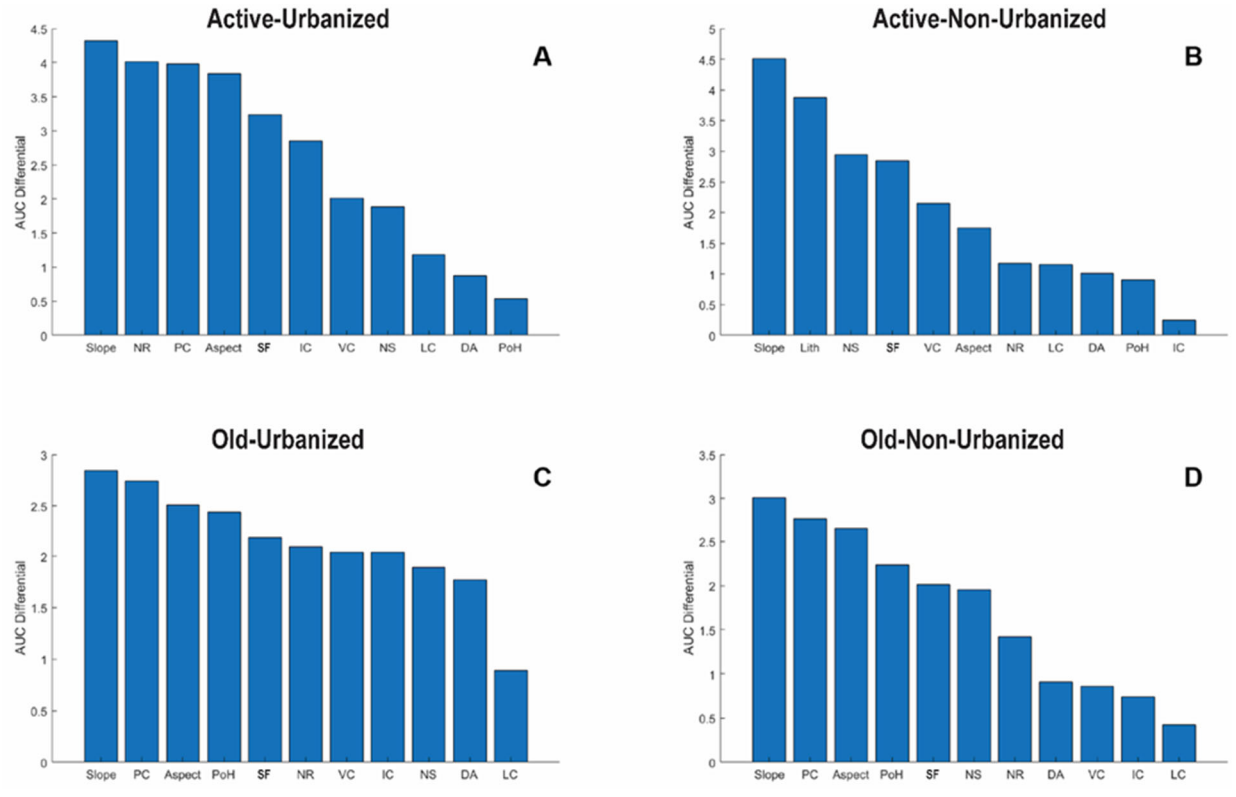
where subscript j is the index of the factor-class combination, and Nl-j and Np-j are the number of landslides locations and map pixels within this combination, respectively. A factor-class combination that produces a relatively high Cp indicates that spatial locations characterized by this combination tend to generate a relatively large number of landslides.

Appendix C.2 AUC Factor Rankings

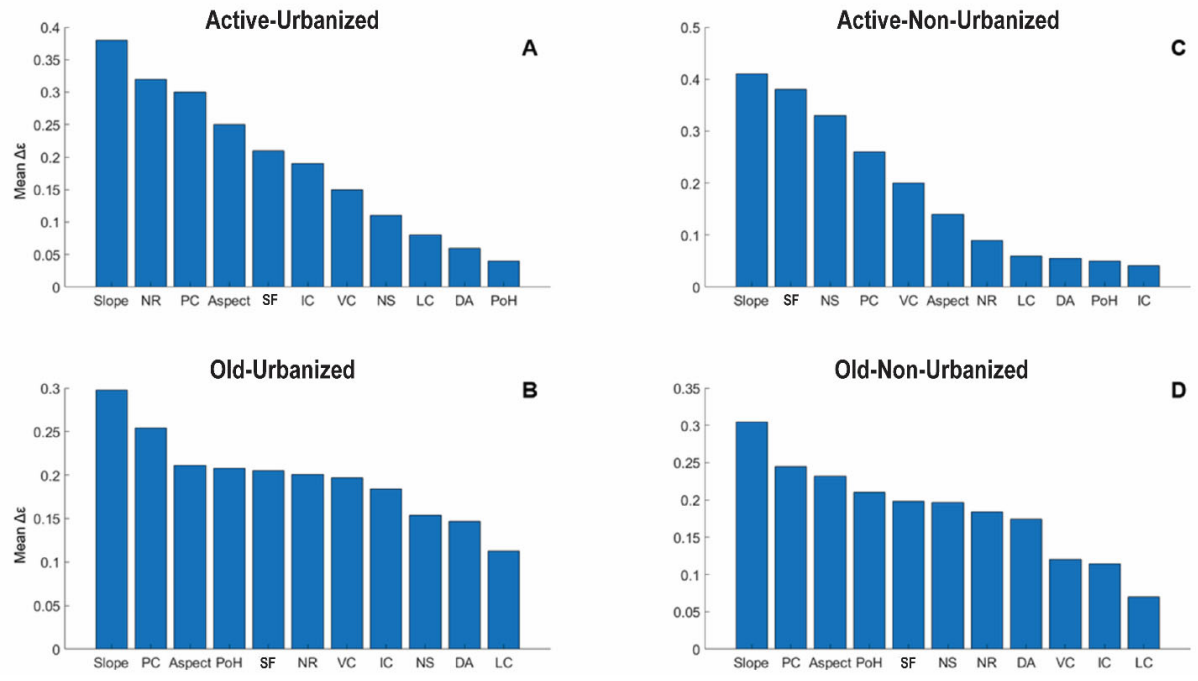
For each landslide inventory, we also explored the relative influence of each landslide-related factor on a model prediction by excluding one factor at a time from a Receiver Operating Characteristic (ROC)-area under the ROC curve (AUC) analysis (Gorsevski et al., 2006; Marjanović, 2013; Cantarino et al., 2019; Pham et al., 2020) and calculating the relative difference (dAUC= 100*(AUCa-AUCe)/AUCa) between the AUC values for a model with excluded factor (AUCe) and that with all 11 factors (AUCa). We then rank the factors based on their relative influence on the AUC.

Appendix C.3 Out of Bag Permutation Factor Ranking

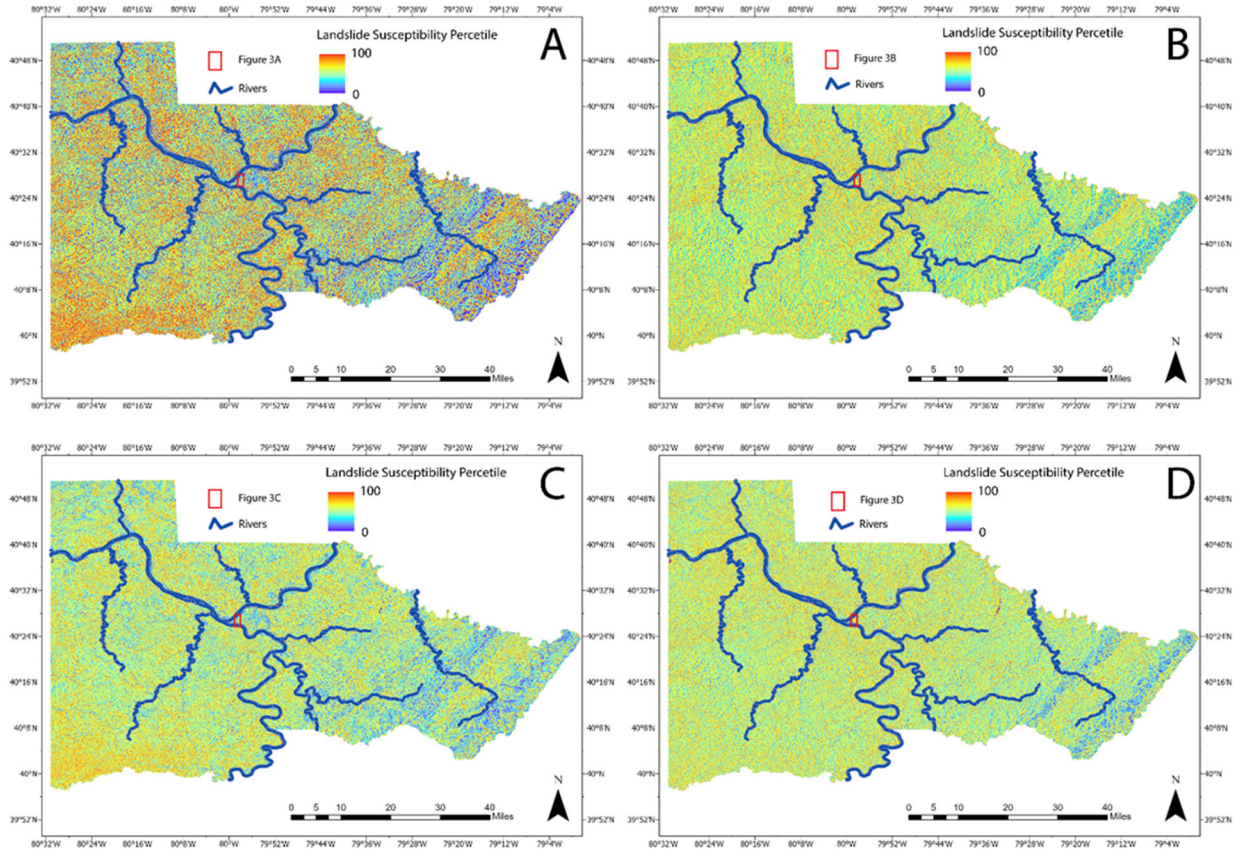
Landslide related factors were also ranked using a permutation out-of-bag method (Taalab et al., 2018). The method randomly permutes the data for each factor at a time for the out of bag samples of each tree and generates predictions based on the permuted values. It then evaluates the quality of this new prediction (εn) compared to that based on the non-permuted data (εj). The difference in prediction quality ($\Delta\varepsilon = \varepsilon n - \varepsilon i$: D'Amato et al., 2021) is used to rank the importance of the factors, where highly ranked factors are those associated with high $\Delta\varepsilon$, and random permutation of a factor meaningfully degrades the prediction.



Appendix Figure C.1: Ranking of landslide-related factors by the conditional probability AUC-Differential approach. (A) Active-Urbanized, (B) Active-Non-Urbanized, (C) Old-Urbanized, and (D) Old-Non-Urbanized. 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).



Appendix Figure C.2: Ranking of landslide-related factors by the random forest out of bag permutation approach. (A) Active-Urbanized, (B) Active-Non-Urbanized, (C) Old-Urbanized, and (D) Old-Non-Urbanized. 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).



Appendix Figure C.3: Landslide susceptibility maps for the entire study area based on landslide inventories from: (A) Active-Urbanized, (B) Active-Non-Urbanized, (C) Old-Urbanized, and (D) Old-Non-Urbanized.

Landslide Inventory Modeled AUC Values	Urban Areas	Non-Urban Areas
Urban (Active)	0.7942	0.6861
Non-Urban (Active)	0.6103	0.7854
Urban (Old)	0.6793	0.6222
Non-Urban (Old)	0.6142	0.6966

Appendix Table C.1: Table of AUC values for all models in Urban, Non-Urban, and the total study area.

Landslide Related Factors	Range	Mean	Median
Slope (degrees)	0-88.91	8.1	7.5
Aspect (degrees)	0-360	94	84
Profile Curvature (1/meters)	-198.87-350.25	-0.132	-0.124
Distance to Nearest Road (meters)	0-4990	143	213
Distance to Nearest Stream (meters)	0-4673	344	387
Elevation (meters)	202-909	416	488

Appendix Table C.2: Table of topographic landslide related factors and their range, mean, and median values.

Bibliography

Alexander, D. (1986). Landslide damage to buildings. Environmental Geology and Water Sciences, 8(3), 147-151. https://doi.org/10.1007/BF02509902

Allegheny County Landslide Task Force (2019), Allegheny County Landslide Portal. <https://landslide-portal-alcogis.opendata.arcgis.com>

Arabameri, A., Pradhan, B., Rezaei, K., Sohrabi, M., & Kalantari, Z. (2019). GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. Journal of Mountain Science, 16(3), 595-618.

Archer, K. J. (2010). rpartOrdinal: An R package for deriving a classification tree for predicting an ordinal response. Journal of Statistical Software, 34, 7. https://doi.org/10.18637/jss.v034.i07

Ashland, F. X. (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

Ashland, F. X. (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.

Ávila, F. F., Alvalá, R. C., Mendes, R. M., & Amore, D. J. (2021). The influence of land use/land cover variability and rainfall intensity in triggering landslides: A back-analysis study via

physically based models. Natural Hazards, 105(1), 1139-1161. https://doi.org/10.1007/s11069-020-04324-x

Bălteanu, D., Chendeş, V., Sima, M., & Enciu, P. (2010). A country-wide spatial assessment of landslide susceptibility in Romania. Geomorphology, 124(3-4), 102-112.

Barla, M., & Antolini, F. (2016). An integrated methodology for landslides' early warning systems. Landslides, 13(2), 215-228.

Benz, S. A., & Blum, P. (2019). Global detection of rainfall-triggered landslide clusters. Natural Hazards and Earth System Sciences, 19(7), 1433-1444.

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. Natural Hazards and Earth System Sciences, 21(1), 147-169. https://doi.org/10.5194/nhess-21-147-2021

BeVille, S. H., Mirus, B. B., Ebel, B. A., Mader, G. G., & Loague, K. (2010). Using simulated hydrologic response to revisit the 1973 Lerida Court landslide. Environmental Earth Sciences, 61(6), 1249-1257. https://doi.org/10.1007/s12665-010-0448-z

Bogaard, T. A., & Greco, R. (2016). Landslide hydrology: From hydrology to pore pressure. Wiley Interdisciplinary Reviews: Water, 3(3), 439-459. https://doi.org/10.1002/wat2.1126

Bogaard, T. A., & Greco, R. (2016). Landslide hydrology: from hydrology to pore pressure. Wiley Interdisciplinary Reviews: Water, 3(3), 439-459.

Bousta, M., & Brahim, L. A. (2018). Weights of evidence method for landslide susceptibility mapping in Tangier, Morocco. In MATEC Web of Conferences (Vol. 149, p. 02042). EDP Sciences.

Braun, A., Urquia, E. L. G., Lopez, R. M., & 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, R. P., Pomeroy, J. S., & Davies, W. E. (1975). Landsliding in Allegheny County, Pennsylvania. U.S. Geological Survey Circular 728, 18 pp. https://doi.org/10.3133/cir728

Briggs, R. P., Pomeroy, J. S., & Davies, W. E. (1975). Landsliding in Allegheny County, Pennsylvania (Vol. 728). US Geological Survey.

Brunetti, M. T., Melillo, M., Peruccacci, S., Ciabatta, L., & Brocca, L. (2018). How far are we from the use of satellite rainfall products in landslide forecasting?. Remote Sensing of Environment, 210, 65-75.

Brunetti, M. T., Peruccacci, S., Antronico, L., Bartolini, D., Deganutti, A. M., Gariano, S. L., ... & Guzzetti, F. (2015). Catalogue of rainfall events with shallow landslides and new rainfall thresholds in Italy. In Engineering Geology for Society and Territory-Volume 2 (pp. 1575-1579). Springer, Cham.

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in Italy. Natural Hazards and Earth System Sciences, 10(3), 447-458.

Caine N. (1980) The rainfall intensity-duration control of shallow landslides and debris flows. Geogr Ann A 62:23–27

Can, R., Kocaman, S., & Gokceoglu, C. (2019). A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. ISPRS International Journal of Geo-Information, 8(7), 300.

Can, R., Kocaman, S., & Gokceoglu, C. (2019). A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. ISPRS International Journal of Geo-Information, 8(7), 300.

Cantarino, I., Carrion, M. A., Goerlich, F., & Ibañez, V. M. (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

Cantarino, I., Carrion, M. A., Goerlich, F., & Ibañez, V. M. (2019). A ROC analysis-based classification method for landslide susceptibility maps. Landslides, 16(2), 265-282.Cascini, L., Bonnard, C., Corominas, J., Jibson, R., & Montero-Olarte, J. (2005). Landslide hazard and risk zoning for urban planning and development. In Landslide risk management (pp. 209-246). CRC Press.

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., and Eberhardt, E. (eds), Landslide Risk Management (pp. 209-246), CRC Press, London.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. Natural Hazards and Earth System Sciences, 13(11), 2815-2831. https://doi.org/10.5194/nhess-13-2815-2013

Chae, B.-G.; Park, H.-J.; Catani, F.; Simoni, A.; Berti, M. Landslide prediction, monitoring, and early warning: A concise review of state-of-the-art. Geosci. J. 2017, 21, 1033–1070

Chaparro, J. C. M. (2020). Identifying and Mapping the Risk of Rockfall and Landslide on Roads and Urban Areas. In Mapping the Risk of Flood, Mass Movement and Local Subsidence (pp. 23-39). Springer, Cham.

Chen, C. Y., & Huang, W. L. (2013). Land use change and landslide characteristics analysis for community-based disaster mitigation. Environmental Monitoring and Assessment, 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, B. B. (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

Choi, B. Y., Al-Mansoori, M. K., Zaman, R., & Albishri, A. A. (2018, January). Understanding what residents ask cities: open data 311 call analysis and future directions. In Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking (pp. 1-6).

Choi, B. Y., Al-Mansoori, M. K., Zaman, R., & Albishri, A. A. (2018, January). Understanding what residents ask cities: open data 311 call analysis and future directions. In Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking (pp. 1-6).

Chung, C. J. (2006). Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment. Computers & Geosciences, 32(8), 1052-1068.

Cieslik, K., Shakya, P., Uprety, M., Dewulf, A., Russell, C., Clark, J., ... & Dhakal, A. (2019). Building Resilience to Chronic Landslide Hazard Through Citizen Science. Frontiers in Earth Science, 7, 278.

Cieslik, K., Shakya, P., Uprety, M., Dewulf, A., Russell, C., Clark, J., ... & Dhakal, A. (2019). Building Resilience to Chronic Landslide Hazard Through Citizen Science. Frontiers in Earth Science, 7, 278.

Clerici, A., Perego, S., Tellini, C., & Vescovi, P. (2002). A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphology, 48(4), 349-364.

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

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.

Crosta GB, Frattini P. (2001) Rainfall thresholds for triggering soil slips and debris flow. In: Proc. 2nd EGS Plinius Conf. on Mediterranean Storms (Mugnai A, Guzzetti F,Roth G, eds). Siena, pp 463–487

Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. Geomorphology, 124(3-4), 260-267.

Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. Geomorphology, 124, 260–267. https://doi.org/10.1016/j.geomorph.2010.04.009

Crozier, M.J., (1999). Prediction of rainfall-triggered landslides: a test of the antecedent water status model. Earth Surf. Process. Landforms 24, 825-833.

Cruden, D.M., Varnes, D.J. (1996). Landslide types and processes. Transportation Research Board, U.S. National Academy of Sciences, Special Report, 247: 36-75.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783-2792. https://doi.org/10.1890/07-0539.1

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., & Paudyal, P. (2008). Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence. Geomorphology, 102(3-4), 496-510.

Dai, F., & Lee, C. F. (2002). Landslides on natural terrain. Mountain Research and Development, 22(1), 40-47. https://doi.org/10.1659/0276-4741(2002)022[0040:LONT]2.0.CO;2

Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. Proceedings of the National Academy of Sciences, 118(4), 1–7.

David-Novak, H. B., Morin, E., & Enzel, Y. (2004). Modern extreme storms and the rainfall thresholds for initiating debris flows on the hyperarid western escarpment of the Dead Sea, Israel. Geological Society of America Bulletin, 116(5-6), 718-728.

Davis, J. C., Chung, C. J., & Ohlmacher, G. C. (2006). Two models for evaluating landslide hazards. Computers & Geosciences, 32(8), 1120-1127. https://doi.org/10.1016/j.cageo.2006.02.006

Do, H. M., & Yin, K. L. (2018). Rainfall Threshold Analysis and Bayesian Probability Method for Landslide Initiation Based on Landslides and Rainfall Events in the Past. Open Journal of Geology, 8(7), 674-696.

Doglioni, A., Fiorillo, F., Guadagno, F. M., & Simeone, V. (2012). Evolutionary polynomial regression to alert rainfall-triggered landslide reactivation. Landslides, 9(1), 53-62.

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

East, A. E., & Sankey, J. B. (2020). Geomorphic and sedimentary effects of modern climate change: Current and anticipated future conditions in the western United States. Reviews of Geophysics, 58(4), e2019RG000692.

Ebel, B. A., Rengers, F. K., & Tucker, G. E. (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

Fan, L., Lehmann, P., & Or, D. (2016). Effects of soil spatial variability at the hillslope and catchment scales on characteristics of rainfall-induced landslides. Water Resources Research, 52(3), 1781-1799.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Engineering Geology, 102(3-4), 99-111.

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

Floris, M., & Bozzano, F. (2008). Evaluation of landslide reactivation: a modified rainfall threshold model based on historical records of rainfall and landslides. Geomorphology, 94(1-2), 40-57.

Francipane, A., Arnone, E., Lo Conti, F., Puglisi, C., & Noto, L. V. (2014). A comparison between heuristic, statistical, and data-driven methods in landslide susceptibility assessment: an application to the briga and giampilieri catchments.

Franzoni C, Sauermann H (2014). Crowd science: The organization of scientific research in open collaborative projects. Res Policy. Elsevier B.V. 43:1–20.

Franzoni C, Sauermann H (2014). Crowd science: The organization of scientific research in open collaborative projects. Res Policy. Elsevier B.V. 43:1–20.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 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

Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Nat Hazards Earth Syst Sci 18:2161–2181.

Giarratani, F., & Houston, D. B. (1989). Structural change and economic policy in a declining metropolitan region: Implications of the Pittsburgh experience. Urban Studies, 26(6), 549-558. https://doi.org/10.1080/00420988920080661

Gibbons, J. D., & 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. Die Erde, 134(2), 123-146.

Glade, T., Crozier, M., & Smith, P. (2000). Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “Antecedent Daily Rainfall Model”. Pure and Applied Geophysics, 157(6), 1059-1079.

Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & geosciences, 81, 1-11.

Gonzalez-Ollauri, A., & Mickovski, S. B. (2017). Hydrological effect of vegetation against rainfall-induced landslides. Journal of Hydrology, 549, 374-387.

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(3), 395-415. https://doi.org/10.1111/j.1467-9671.2006.01004.x

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(3), 395-415.

Gray, R. E., Hamel, J. V., & Adams, W. R. (2011). Landslides in the vicinity of Pittsburgh, Pennsylvania. In: Ruffolo, R. M., & Ciampaglio, C. N. (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)

Gray, R. E., Hamel, J. V., & Adams, W. R. (2011). Landslides in the vicinity of Pittsburgh, Pennsylvania. Field Guides, 20, 61-85.

Guo, C., Montgomery, D. R., Zhang, Y., Wang, K., & Yang, Z. (2015). Quantitative assessment of landslide susceptibility along the Xianshuihe fault zone, Tibetan Plateau, China. Geomorphology, 248, 93-110.

Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing Landslide Maps: A Case Study in the Upper Tiber River Basin, Central Italy. Environmental management, 25(3).

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and atmospheric physics, 98(3), 239-267.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: an update. Landslides, 5(1), 3-17.

Habib, E., Henschke, A., & Adler, R. F. (2009). Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. Atmospheric Research, 94(3), 373-388.

Hamel, J. V., & Flint, N. K. (1972). Failure of colluvial slope. Journal of the Soil Mechanics and Foundations Division, 98(2), 167-180. https://doi.org/10.1061/JSFEAQ.0001736

Handwerger, A. L., Fielding, E. J., Huang, M. H., Bennett, G. L., Liang, C., & Schulz, W. H. (2019). Widespread initiation, reactivation, and acceleration of landslides in the northern California Coast Ranges due to extreme rainfall. Journal of Geophysical Research: Earth Surface, 124(7), 1782-1797.

Harp, E. L., Keefer, D. K., Sato, H. P., & Yagi, H. (2011). Landslide inventories: the essential part of seismic landslide hazard analyses. Engineering Geology, 122(1-2), 9-21.

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, T. A. (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 Sensing of Environment, 199, 201-217. https://doi.org/10.1016/j.rse.2017.07.010

Highland, L. M. (2006). Estimating landslide losses-preliminary results of a seven-state pilot project (No. 2006-1032).

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

Hsu, Y. C., Chang, Y. L., Chang, C. H., Yang, J. C., & Tung, Y. K. (2018). Physical-based rainfall-triggered shallow landslide forecasting. Smart Water, 3(1), 3.

Huang, Y., & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. Catena, 165, 520-529.

Huggel, C., Clague, J. J., & Korup, O. (2012). Is climate change responsible for changing landslide activity in high mountains?. Earth Surface Processes and Landforms, 37(1), 77-91.

Innes JL (1983) Debris flows. Prog Phys Geog 7: 469–501

Istanbulluoglu, E., & Bras, R. L. (2005). Vegetation-modulated landscape evolution: Effects of vegetation on landscape processes, drainage density, and topography. Journal of Geophysical Research: Earth Surface, 110(F2), F02012. https://doi.org/10.1029/2004JF000249

Iverson, R. M. (2000). Landslide triggering by rain infiltration. Water Resources Research, 36(7), 1897-1910. https://doi.org/10.1029/2000WR900090

Johnston, E. C., Davenport, F. V., Wang, L., Caers, J. K., Muthukrishnan, S., Burke, M., & Diffenbaugh, N. S. (2021). Quantifying the effect of precipitation on landslide hazard in urbanized and non-urbanized areas. Geophysical Research Letters, 48(16), e2021GL094038. https://doi.org/10.1029/2021GL094038

Johnston, E. C., Davenport, F. V., Wang, L., Caers, J. K., Muthukrishnan, S., Burke, M., & Diffenbaugh, N. S. (2021). Quantifying the Effect of Precipitation on Landslide Hazard in Urbanized and Non-Urbanized Areas. Geophysical Research Letters, 48(16), e2021GL094038.

Juang, C. S., Stanley, T. A., & Kirschbaum, D. B. (2019). Using citizen science to expand the global map of landslides: Introducing the Cooperative Open Online Landslide Repository (COOLR). PloS one, 14(7), e0218657.

Juang, C. S., Stanley, T. A., & Kirschbaum, D. B. (2019). Using citizen science to expand the global map of landslides: Introducing the Cooperative Open Online Landslide Repository (COOLR). PloS one, 14(7), e0218657.

Juang, C., Stanley, T., & Kirschbaum, D. (2017). Citizen science, GIS, and the global hunt for landslides. AGUFM, 2017, IN24B-03.

Kafy, A. A., Rahman, M. S., & 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).

Kamp, U., Growley, B. J., Khattak, G. A., & Owen, L. A. (2008). GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology, 101(4), 631-642.

Kim, J. C., Lee, S., Jung, H. S., & Lee, S. (2018). Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocarto International, 33(9), 1000-1015. https://doi.org/10.1080/10106049.2017.1323964

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: method, results, and limitations. Natural Hazards, 52(3), 561-575.

Kocaman, S., & Gokceoglu, C. (2019). A CitSci app for landslide data collection. Landslides, 16(3), 611-615.

Kocaman, S., & Gokceoglu, C. (2019). A CitSci app for landslide data collection. Landslides, 16(3), 611-615.

Komac, M. (2006). A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine Slovenia. Geomorphology, 74(1-4), 17-28.

Kumar, S. V., & Bhagavanulu, D. V. S. (2008). Effect of deforestation on landslides in Nilgiris district—A case study. Journal of the Indian Society of Remote Sensing, 36(1), 105-108. https://doi.org/10.1007/s12524-008-0011-5

Kumar, V., Gupta, V., & Sundriyal, Y. P. (2019). Spatial interrelationship of landslides, litho-tectonics, and climate regime, Satluj valley, Northwest Himalaya. Geological Journal, 54(1), 537-551.

Lagomarsino, D., Segoni, S., Rosi, A., Rossi, G., Battistini, A., Catani, F., & Casagli, N. (2015). Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting. Natural Hazards and Earth System Sciences, 15(10), 2413-2423.

Lee, C. F., Huang, W. K., Chang, Y. L., Chi, S. Y., & Liao, W. C. (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. Journal of the Korean GEO-environmental Society, 17(4), 17-31. https://doi.org/10.14481/jkges.2016.17.4.17

Lee, S., & Dan, N. T. (2005). Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. Environmental Geology, 48(6), 778-787.

Lessing, Peter, Kulander, B. R., Wilson, B. D., Dean, S. L., and Woodring, S. M. (1976). West Virginia landslides and slide-prone areas. West Virginia Geological and Economic Survey Environmental Geology Bulletin EGB-15a, 64 pp. (1:24,000 scale, 28 maps on 27 sheets).

Leventhal, A. R., & Kotze, G. P. (2008). Landslide susceptibility and hazard mapping in Australia for land-use planning—with reference to challenges in metropolitan suburbia. Engineering Geology, 102(3-4), 238-250.

Li, Y., & Chen, W. (2020). Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. Water, 12(1), 113.

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

Marjanović, M. (2013). Comparing the performance of different landslide susceptibility models in ROC space. In Landslide science and practice (pp. 579-584). Springer, Berlin, Heidelberg.

Marjanović, M., Kovačević, M., Bajat, B., & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology, 123(3), 225-234.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784-2817. https://doi.org/10.1080/01431161.2018.1433343

McAdoo, B. G., Quak, M., Gnyawali, K. R., Adhikari, B. R., Devkota, S., Rajbhandari, P. L., & Sudmeier-Rieux, K. (2018). Roads and landslides in Nepal: How development affects environmental risk. Natural Hazards and Earth System Sciences, 18(12), 3203-3210. https://doi.org/10.5194/nhess-18-3203-2018

McGuire, L. A., Rengers, F. K., Kean, J. W., Coe, J. A., Mirus, B. B., Baum, R. L., & Godt, J. W. (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. Geophysical Research Letters, 43(17), 9084-9092. https://doi.org/10.1002/2016GL070741

McGuire, L. A., Rengers, F. K., Kean, J. W., Coe, J. A., Mirus, B. B., Baum, R. L., & Godt, J. W. (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. Geophysical Research Letters, 43(17), 9084-9092.

McGuire, L. A., Rengers, F. K., Kean, J. W., Coe, J. A., Mirus, B. B., Baum, R. L., & Godt, J. W. (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. Geophysical Research Letters, 43(17), 9084-9092.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., Roccati, A., & Guzzetti, F. (2018). A tool for the automatic calculation of rainfall thresholds for landslide occurrence. Environmental Modelling & Software, 105, 230-243.

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

Meusburger, K., & Alewell, C. (2008). Impacts of anthropogenic and environmental factors on the occurrence of shallow landslides in an alpine catchment (Urseren Valley, Switzerland). Natural Hazards and Earth System Sciences, 8, 509-520.

Miao, T., Liu, Z., Niu, Y., & Ma, C. (2001). A sliding block model for the runout prediction of high-speed landslides. Canadian geotechnical journal, 38(2), 217-226.

Miles, C. E., Whitfield, G. T., and others (2001). Bedrock geologic units of Pennsylvania, scale 1:250,000. based on: Berg, T. M., Edmunds, W. E., Geyer, A. R., Glover, A. D., Hoskins, D. M., MacLachlan, D. B., Root, S. I., Sevon, W. D., & Socolow, A. A. (1980). Geologic map of Pennsylvania, scale 1:250000. Pennsylvania Geological Survey, Map 1.

Mirus, B. B., Ebel, B. A., Loague, K., & Wemple, B. C. (2007). Simulated effect of a forest road on near-surface hydrologic response: Redux. Earth Surface Processes and Landforms, 32(1), 126-142. https://doi.org/10.1002/esp.1387

Mirus, B. B., Ebel, B. A., Loague, K., & Wemple, B. C. (2007). Simulated effect of a forest road on near-surface hydrologic response: Redux. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 32(1), 126-142.

Mirus, B. B., Jones, E. S., Baum, R. L., Godt, J. W., Slaughter, S., Crawford, M. M., Lancaster, J., Stanley, T., Kirschbaum, D. B., Burns, W. J., Schmitt, R. G., Lindsey, K. O., & McCoy, K. M. (2020). Landslides across the USA: Occurrence, susceptibility, and data limitations. Landslides, 17, 2271-2285. https://doi.org/10.1007/s10346-020-01424-4

Mirus, B. B., Jones, E. S., Baum, R. L., Godt, J. W., Slaughter, S., Crawford, M. M., ... & Schmitt, R. G. (2020). Landslides across the USA: occurrence, susceptibility, and data limitations.

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).

O'Brien, D. T. (2016). Using small data to interpret big data: 311 reports as individual contributions to informal social control in urban neighborhoods. Social science research, 59, 83-96.

O'Brien, D. T. (2016). Using small data to interpret big data: 311 reports as individual contributions to informal social control in urban neighborhoods. Social science research, 59, 83-96.

Okagbue, C. O. (1986). An investigation of landslide problems in spoil piles in a strip coal mining area, West Virginia (USA). Engineering Geology, 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. Environmental Geology, 57(7), 1675-1686. https://doi.org/10.1007/s00254-008-1449-z

Ozdemir, A. (2009). Landslide susceptibility mapping of vicinity of Yaka Landslide (Gelendost, Turkey) using conditional probability approach in GIS. Environmental geology, 57(7), 1675-1686.

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 (pp. 293-320). UNU-Press, Tokyo.

Paul, J. D., Buytaert, W., Paudel, S., Sah, N. K., Parajuli, B., Shakya, P., ... & Nayaval, J. L. (2019). Landslide EVO: Monitoring landslides in remote western Nepal by leveraging technological advances and citizen science. AGUFM, 2019, H14A-08.

Paul, J. D., Buytaert, W., Paudel, S., Sah, N. K., Parajuli, B., Shakya, P., ... & Nayaval, J. L. (2019). Landslide EVO: Monitoring landslides in remote western Nepal by leveraging technological advances and citizen science. AGUFM, 2019, H14A-08.

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) (2007) PAMAP Program – Roads. Available at https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=8 (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).

Pennsylvania. Bureau of Topographic and Geologic Survey, & Berg, T. M. (1980). Geologic map of Pennsylvania. The Survey.

Pfeil-McCullough, E., Bain, D. J., 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. Science of the Total Environment, 536, 538-545. https://doi.org/10.1016/j.scitotenv.2015.06.145

Pfeil-McCullough, E., Bain, D. J., 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. Science of the Total Environment, 536, 538-545.

Pfeil-McCullough, E., Bain, D. J., 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. Science of the Total Environment, 536, 538-545.

Pham, B. T., Nguyen-Thoi, T., Qi, C., Van Phong, T., Dou, J., Ho, L. S., 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

Pham, B. T., Prakash, I., & Bui, D. T. (2018). Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. Geomorphology, 303, 256-270.

Pham, B. T., Prakash, I., Dou, J., Singh, S. K., Trinh, P. T., Tran, H. T., ... & Bui, D. T. (2020). A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto International, 35(12), 1267-1292.

Pisano, L., Zumpano, V., Malek, Ž., Rosskopf, C. M., & 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. Science of The Total Environment, 601, 1147-1159. https://doi.org/10.1016/j.scitotenv.2017.05.231

Pomeroy, J. S. (1977). Landslide Susceptibility Map of the Pittsburgh West Quadrangle, Allegheny County, Pennsylvania (No. 1035).

Pomeroy, J. S. (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, J. S. (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

Pomeroy, J. S. (1982). Landslides in the greater Pittsburgh region, Pennsylvania (p. 48). US Government Printing Office.

Pourghasemi, H. R., Mohammady, M., & Pradhan, B. (2012). Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena, 97, 71-84.

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. Natural Hazards, 63(2), 965-996. https://doi.org/10.1007/s11069-012-0217-2

Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling & Software, 25(6), 747-759.

Rahardjo, H., Li, X. W., Toll, D. G., & Leong, E. C. (2001). The effect of antecedent rainfall on slope stability. In Unsaturated soil concepts and their application in geotechnical practice (pp. 371-399). Springer, Dordrecht.

Rappaport, J. (2003). U.S. urban decline and growth, 1950 to 2000. Economic Review, 88(3), 15-44.

Ray, R. L., & Jacobs, J. M. (2007). Relationships among remotely sensed soil moisture, precipitation and landslide events. Natural Hazards, 43(2), 211-222.

References

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., 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. Arabian Journal of Geosciences, 7(2), 725-742. https://doi.org/10.1007/s12517-012-0807-z

Regmi, A. D., Yoshida, K., Pourghasemi, H. R., DhitaL, M. R., & Pradhan, B. (2014). Landslide susceptibility mapping along Bhalubang—Shiwapur area of mid-Western Nepal using frequency ratio and conditional probability models. Journal of Mountain Science, 11(5), 1266-1285.

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

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

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180, 60-91.

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

Rohan, T. J., Wondolowski, N., & Shelef, E. (2021). Landslide susceptibility analysis based on citizen reports. Earth Surface Processes and Landforms, 46(4), 791-803. https://doi.org/10.1002/esp.5064

Rohan, T. J., Wondolowski, N., & Shelef, E. (2021). Landslide susceptibility analysis based on citizen reports. Earth Surface Processes and Landforms, 46(4), 791-803.

Running, S. W., Nemani, R. R., & Hungerford, R. D. (1987). Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. Canadian Journal of Forest Research, 17(6), 472-483. https://doi.org/10.1139/x87-081

Samodra, G., Chen, G., Sartohadi, J., & Kasama, K. (2017). Comparing data-driven landslide susceptibility models based on participatory landslide inventory mapping in Purwosari area, Yogyakarta, Java. Environmental Earth Sciences, 76(4), 184.

Santoso, A. M., Phoon, K. K., & Quek, S. T. (2011). Effects of soil spatial variability on rainfall-induced landslides. Computers & Structures, 89(11-12), 893-900.

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

Schellong, A., & Langenberg, T. (2007, January). Managing citizen relationships in disasters: Hurricane wilma, 311 and miami-dade county. In 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07) (pp. 96-96). IEEE.

Schellong, A., & Langenberg, T. (2007, January). Managing citizen relationships in disasters: Hurricane wilma, 311 and miami-dade county. In 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07) (pp. 96-96). IEEE.

Schicker, R., & Moon, V. (2012). Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. Geomorphology, 161, 40-57.

Schmidt, J., Evans, I.S., Brinkmann, J., 2003. Comparison of polynomial models for land surface curvature calculation. International Journal of Geographical Information Science 17,797-814.

Schwanghart, W., & Kuhn, N. J. (2010). TopoToolbox: A set of Matlab functions for topographic analysis. Environmental Modelling & Software, 25(6), 770-781.

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

Schwester, R. W., Carrizales, T., & Holzer, M. (2009). An examination of the municipal 311 system. International Journal of Organization Theory & Behavior, 12(2), 218-236.

Schwester, R. W., Carrizales, T., & Holzer, M. (2009). An examination of the municipal 311 system. International Journal of Organization Theory & Behavior, 12(2), 218-236.

Scott, D. W. (2015). Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons.

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 Sensing, 12(9), 1483. https://doi.org/10.3390/rs12091483

Seneviratne, S. I., Nicholls, D., Easterling, C. M., Goodess, S., Kanae, J., Kossin, Y., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change

(IPCC). In C. B. Field, V. Barros, T. F. Stocker, & Q. Dahe (Eds.), Managing the risks of extreme events and disasters to advance climate change adaptation (pp. 109– 230). Cambridge University Press.

Shelef, E., & Hilley, G. E. (2014). Symmetry, randomness, and process in the structure of branched channel networks. Geophysical Research Letters, 41(10), 3485-3493. https://doi.org/10.1002/2014GL059816

Sidle, R. C., & Ochiai, H. (2006). Natural factors influencing landslides. In Landslides: Prcoesses, prediction, and land use (Water Resources, pp. 41– 119). American Geophysical Union.

Sidle, R. C., Ziegler, A. D., Negishi, J. N., Nik, A. R., Siew, R., & Turkelboom, F. (2006). Erosion processes in steep terrain—Truths, myths, and uncertainties related to forest management in Southeast Asia. Forest Ecology and Management, 224(1-2), 199-225. https://doi.org/10.1016/j.foreco.2005.12.019

Silverman, B. W. (2018). Density estimation for statistics and data analysis. Routledge.

Simon, N., Crozier, M., de Roiste, M., Rafek, A. G., & Roslee, R. (2015). Time series assessment on landslide occurrences in an area undergoing development. Singapore Journal of Tropical Geography, 36(1), 98-111. https://doi.org/10.1111/sjtg.12096

Smyth, C. G., & Royle, S. A. (2000). Urban landslide hazards: Incidence and causative factors in Niterói, Rio de Janeiro State, Brazil. Applied Geography, 20(2), 95-118. https://doi.org/10.1016/S0143-6228(00)00004-7

Soma, A. S., & Kubota, T. (2017). The performance of land use change causative factor on landslide susceptibility map in Upper Ujung-Loe Watersheds South Sulawesi, Indonesia. Geoplanning: Journal of Geomatics and Planning, 4(2), 157-170. https://doi.org/10.14710/geoplanning.4.2.157-170

Southwestern Pennsylvania Commission. (2017). Previously Active Documented Landslides in Southwestern Pennsylvania <https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=1622>

Staats, J. R. (1942). The Monongahela River and its Stategic Importance. Journal of Geography, 41(8), 281-288.

Steger, S., Brenning, A., Bell, R., & Glade, T. (2016). The propagation of inventory-based positional errors into statistical landslide susceptibility models. Natural Hazards & Earth System Sciences, 16(12).

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

Terzaghi, K. (1950). Mechanism of landslides. In S. Paige (Ed.), Application of geology to engineering practice. Geological Society of America.

Theodorou, P., Baltz, L. M., Paxton, R. J., & Soro, A. (2021). Urbanization is associated with shifts in bumblebee body size, with cascading effects on pollination. Evolutionary Applications, 14(1), 53-68. https://doi.org/10.1111/eva.13087

Tubbs, D. W. (1974). Landslides in Seattle. Washington Department of Natural Resources, Information Circular 52, 15 pp., 1 plate.

U.S. Geological Survey EROS Data Center, 1999, 7.5-minute Digital Elevation Model (10-meter resolution): U. S. Geological Survey, Sioux Falls, SD.

Van Beek, L. P. H., & Van Asch, T. W. (2004). Regional assessment of the effects of land-use change on landslide hazard by means of physically based modelling. Natural 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 and Planetary Science Letters, 256(3-4), 588-603. https://doi.org/10.1016/j.epsl.2007.01.040

Vardoulakis, I. (2000). Catastrophic landslides due to frictional heating of the failure plane. Mechanics of Cohesive-frictional Materials: An International Journal on Experiments, Modelling and Computation of Materials and Structures, 5(6), 443-467.

Venables, W. N., & Ripley, B. D. (2013). Modern applied statistics with S-PLUS. Springer Science & Business Media.

Vieira, B. C., & Fernandes, N. F. (2004). Landslides in Rio de Janeiro: The role played by variations in soil hydraulic conductivity. Hydrological Processes, 18(4), 791-805. https://doi.org/10.1002/hyp.1363

Waltz, J. P. (1971). An analysis of selected landslides in Alameda and Contra Costa Counties, California. Bulletin of the Association of Engineering Geologists, 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, G., & Sassa, K. (2003). Pore-pressure generation and movement of rainfall-induced landslides: effects of grain size and fine-particle content. Engineering geology, 69(1-2), 109-125.

Wang, H., Zhang, L., Yin, K., Luo, H., & Li, J. (2021). Landslide identification using machine learning. Geoscience Frontiers, 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. Environmental Geology, 35(2), 197-209. https://doi.org/10.1007/s00254005030

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

Wasowski, J., Lamanna, C., & Casarano, D. (2010). Influence of land-use change and precipitation patterns on landslide activity in the Daunia Apennines, Italy. Quarterly Journal of Engineering Geology and Hydrogeology, 43(4), 387-401. https://doi.org/10.1144/1470-9236/08-101

White, Ariel, and Kris-Stella Trump. "The promises and pitfalls of 311 data." Urban affairs review 54.4 (2018): 794-823.

Wilson, R. C. (1989). Rainstorms, pore pressures, and debris flows: a theoretical framework. Landslides in a semi-arid environment (Morton DM, Sadler PM, eds), 2, 101-117.

Wilson, R. C., & Wieczorek, G. F. (1995). Rainfall thresholds for the initiation of debris flows at La Honda, California. Environmental & Engineering Geoscience, 1(1), 11-27.

Winant, G. (2021). The Next Shift. Harvard University Press, Boston, 368 pp. (Map: The Pittsburgh Area, pp. xi-xii).

Xu, C., Dai, F., Xu, X., & Lee, Y. H. (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, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., & Xian, G. (2018). A new

generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing, 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. Environmental Earth Sciences, 61(4), 821-836. https://doi.org/10.1007/s12665-009-0394-9

Yilmaz, I. (2010). Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences, 61(4), 821-836.

Zêzere, J. L., Ferreira, A. B., & Rodrigues, M. L. (1999). Landslides in the North of Lisbon Region (Portugal): Conditioning and triggering factors. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, 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. Environmental Earth Sciences, 76(11), 405. https://doi.org/10.1007/s12665-017-6731-5

Zhao, X., & Chen, W. (2020). GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. Applied Sciences, 10(1), 16.

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 Sensing, 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, P. E., Eldho, T. I., & 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