Received: 15 July 2020 | Revised: 17 November 2020 | Accepted: 9 December 2020

DOI: 10.1002/esp.5064

RESEARCH ARTICLE

ESPL WILEY

Landslide susceptibility analysis based on citizen reports

Tyler J. Rohan | Nicholas Wondolowski | Eitan Shelef

Department of Geology and Environmental Science, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence
Tyler Rohan, Geology and Environmental Science, University of Pittsburgh, 4107 O'Hara Street, Pittsburgh, PA 15260-3332, USA.
Email: tjr68@pitt.edu

Abstract

Landslide susceptibility estimates are essential for reducing the risk posed by landslides to social and economic well-being. However, estimates of landslide susceptibility depend on reliable landslide inventories whose production requires extensive field or remote sensing efforts. Further, most inventories are not updated through time and thus may not capture the influence of changes in climate and/or land use. Inventories based on citizen reports of landslide occurrence, have the potential to overcome these limitations. Such an inventory can be produced from citizen reports to a 311-phone and online system, a nationwide database that updates real-time and records reported landslides location and timing. Whereas this landslide inventory is promising, it has not used for landslide susceptibility analyses and may be associated with spatial uncertainties and reporting biases. In this study we explore the use of 311-based landslide inventory for landslide susceptibility estimates in Pittsburgh, PA, USA, where landslide risk is among the highest in the nation. We compare the 311-based inventory to field-validated inventories through a multi-pronged approach that combines field validation of 311-reported landslides, probabilistic analysis of the association between landslides and the underlying topographic and geologic factors, and spatial filtering. Our results 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 simple 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. These results suggest that 311-based landslide inventories can improve susceptibility estimates at a relatively low cost and high temporal resolution.

KEYWORDS
311 citizen science, conditional probability, landslide inventory, landslides, susceptibility mapping

1 | INTRODUCTION

Landslides pose a substantial risk to social and economic well-being that is escalated in areas of high population and infrastructure density (Cascini et al., 2005; Chaparro, 2020). 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 & Kotze, 2008). Landslides typically occur over steep terrain where gravitational forces translate soil and rock downslope along weak planes of low frictional resistance (Miao et al., 2001; Vardoulakis, 2000). 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 (Bogaard & Greco, 2016; Iverson, 2000; Pfeil-McCullough et al., 2015; Wang & Sassa, 2003). 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 et al., 2000; Marjanović et al., 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 et al., 2000; Harp et al., 2011; Yilmaz, 2010). Landslides are often influenced by temporally variable factors (e.g., precipitation, urban expansion, deforestation, fires) (Crozier, 2010; Huggel et al., 2012; Meusburger & Alewell, 2008) 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 (Can et al., 2019; Cieslik et al., 2019; Franzoni & Sauermann, 2014; Juang *et al.*, 2017; 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 & Gokceoglu, 2019; Paul et al., 2019). Landslide related citizen science has been applied through university and government led programs (Juang et al., 2019; Kocaman & 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 & 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 updating 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 in real time and is accessible in over 300 cities in the United States and Canada (Choi et al., 2018; O'Brien, 2016; Schellong & Langenberg, 2007; Schwester et al., 2009). 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, so forth. (Arabameri et al., 2019; Bălteanu et al., 2010; Kamp et al., 2008; Santoso et al., 2011; Zhao & Chen, 2020). These approaches are divided into (Yilmaz, 2010; Huang & 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 (Do & Yin, 2018; Pourghasemi et al., 2012; Yilmaz, 2010), 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 (Goetz et al., 2015; Pradhan & Lee, 2010; Yilmaz, 2010).

This research combines fieldwork, 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, USA, where landslide risk is among the highest in the nation (Gray et al., 2011; Highland, 2006), 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 | STUDY AREA AND DATA

Given the large datasets required to study landslide susceptibility we focus on the city of Pittsburgh in Allegheny County, Pennsylvania (Figure 1), where multiple landslides are recorded in various datasets. Pittsburgh 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

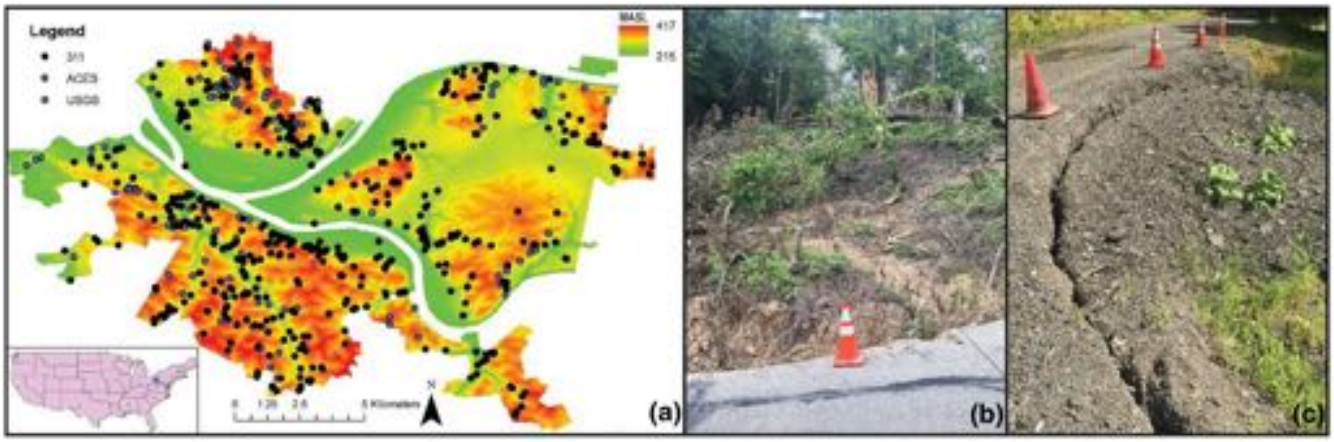

FIGURE 1 (a) Location of 311 reported landslides overlain on a DEM of the City of Pittsburgh. Inset map shows study area over map of the United States. (b, c) Field photographs of validated 311 reported landslides [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 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)

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

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.

There are two established landslide inventories of high spatial accuracy in Pittsburgh (Table 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 (Briggs et al., 1975; Pomeroy, 1977, 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-m 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 & 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-m 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).

3 | METHODS

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 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 1).

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; Costanzo & Irigaray, 2020; Ozdemir, 2009; Regmi et al., 2014; Yilmaz et al., 2010). 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 n^m factor-class combinations, we then computed the conditional probability C_p:

$$C_{p-j} = N_{l-j}/N_{p-j} \quad (1)$$

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 and 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 & Lee, 2010; Pourghasemi et al., 2012).

To define the five 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 1).

To identify the five-primary landslide-related factors (m = 5) out of the nine total factors used, we calculated weighted contrast (W_c) values (Supporting Information, Tables S1–S4) for each class in each factor (Schicker & Moon, 2012; Guo et al., 2015).

$$W_p = \frac{\frac{A_1}{A_1+A_2}}{\frac{A_3}{A_3+A_4}}, \quad (2)$$

$$W_n = \frac{\frac{A_2}{A_1+A_2}}{\frac{A_4}{A_3+A_4}}, \quad (3)$$

$$W_c = W_p - W_n \quad (4)$$

where W_p represents the weighted positives, W_n the weighted negatives, A_1 the number of landslides that fell inside a class, A_2 the number of landslides that fall outside a class, A_3 the number of map pixels that fell inside a class, and A_4 the number of map pixels that fell outside of a class. Weighted contrast values between 0.5–1, 1–2, and > 2 are 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 & Dan, 2005; Pradhan & Lee, 2010; Yilmaz,

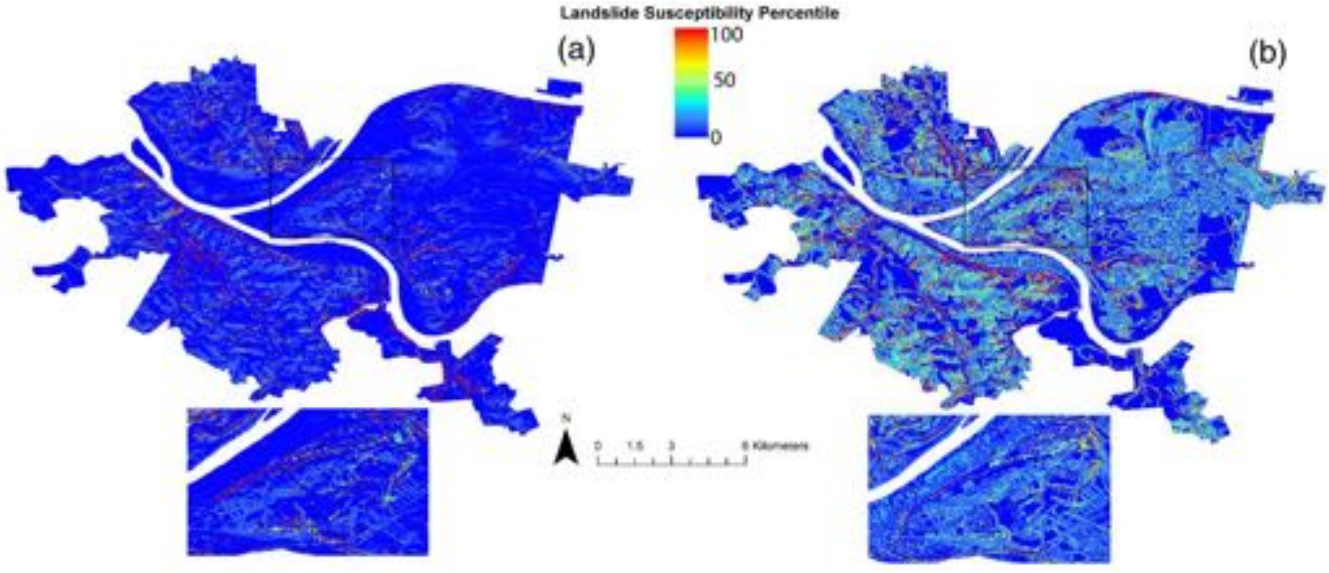

FIGURE 2 Landslide susceptibility maps (10-m resolution) for Pittsburgh, PA, USA: (a) map based on 55 field-corrected 311 landslide locations; (b) map-based on 77 originally reported 311 landslide locations. The Color bar is based on the percentile of computed conditional probability values. Black rectangles mark the location of inset maps that show more details [Colour figure can be viewed at wileyonlinelibrary.com]

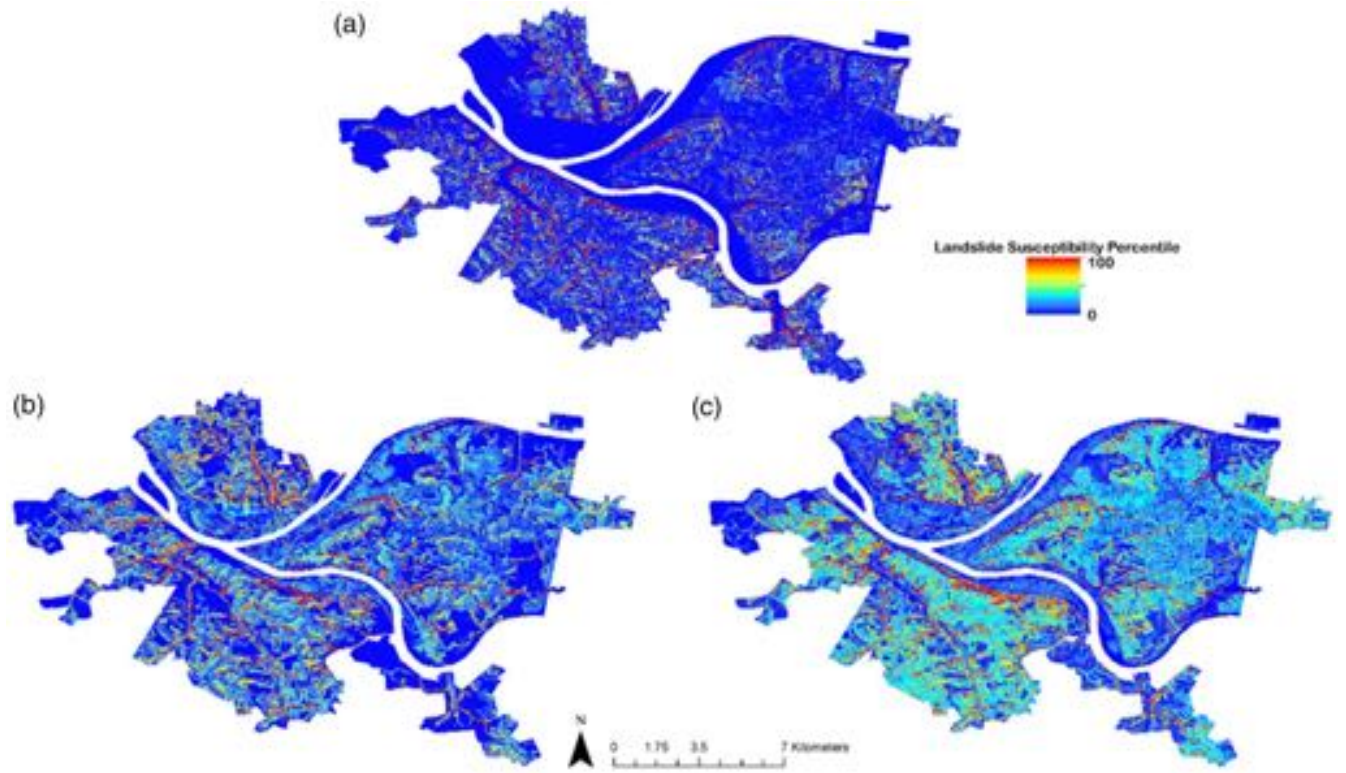

FIGURE 3 A non-filtered landslide susceptibility maps (10-m resolution) created from: (a) ACES and USGS inventories (N = 134); (b) the entire 311 landslide inventory (N = 720); (c) total combined landslide inventories (N = 834) [Colour figure can be viewed at wileyonlinelibrary.com]

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 (ROC) curves (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) (Cantarino et al., 2019; Gorsevski et al., 2006; Pham et al., 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 (Cantarino et al., 2019; Gorsevski et al., 2006; Marjanović, 2013; Pham et al., 2020) and calculating the relative difference [$dAUC = 100 \times (AUC_a - AUC_e)/AUC_a$] between the AUCs for a model with excluded factor (AUC_e) and that with all five factors (AUC_a). 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 5th and 95th percentiles from these iterations to define the uncertainty in the AUC difference (Figure 4).

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 averaging 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 5 and 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

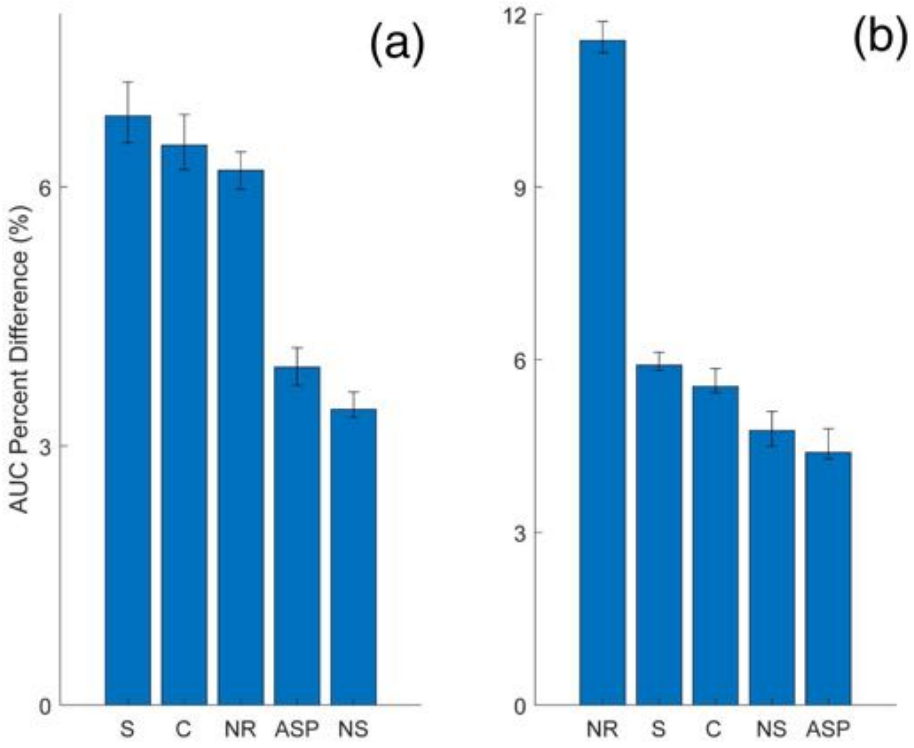

FIGURE 4 Ranking of landslide-related factors by their influence on the AUC for: (a) the field-corrected 311-based inventory (N = 55); (b) the original 311-reported landslide locations for the field validated sites (N = 77). NR, nearest road; S, slope; ASP, aspect; C, profile curvature; NS, nearest stream [Colour figure can be viewed at wileyonlinelibrary.com]

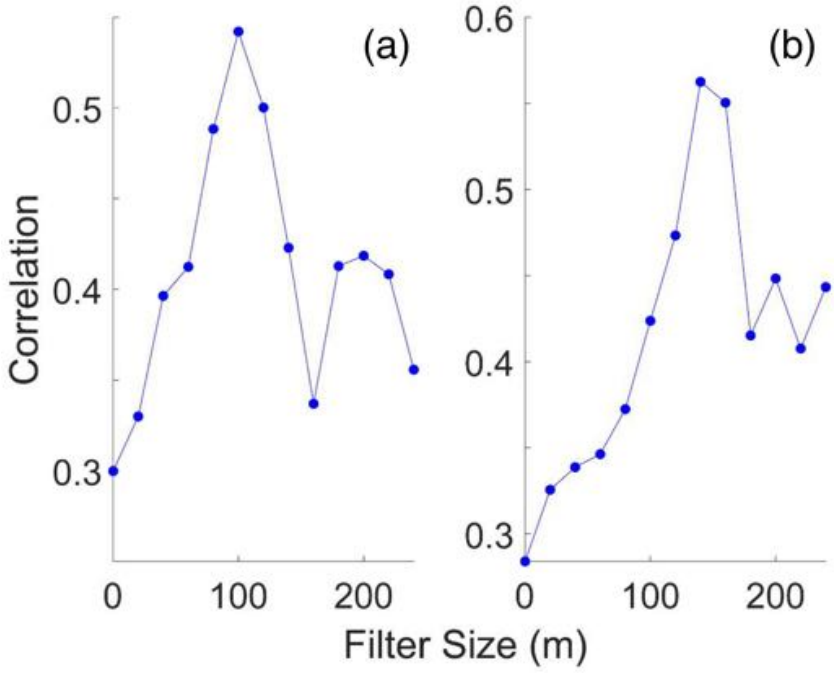

FIGURE 5 Two-dimensional correlation values between maps produced with filter radii from 20 to 240 m in steps of 20 m: (a) Correlation between susceptibility maps based on the entire 311 (N = 720) and the combined (USGS and ACES, N = 134) landslide inventories; (b) Correlation between susceptibility maps based on the original (N = 77), and the field-corrected (N = 55) landslide inventories [Colour figure can be viewed at wileyonlinelibrary.com]

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 each pixel that contains a landslide, whose dimensions are based on the distribution of measured distances between the 311-reported and field-corrected landslide locations (approximately 60 to 120 m). Each of the 100 landslide inventories is then filtered, using the filtration procedure described earlier, 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 six and five classes per factor to further explore the sensitivity of these results to the number of classes.

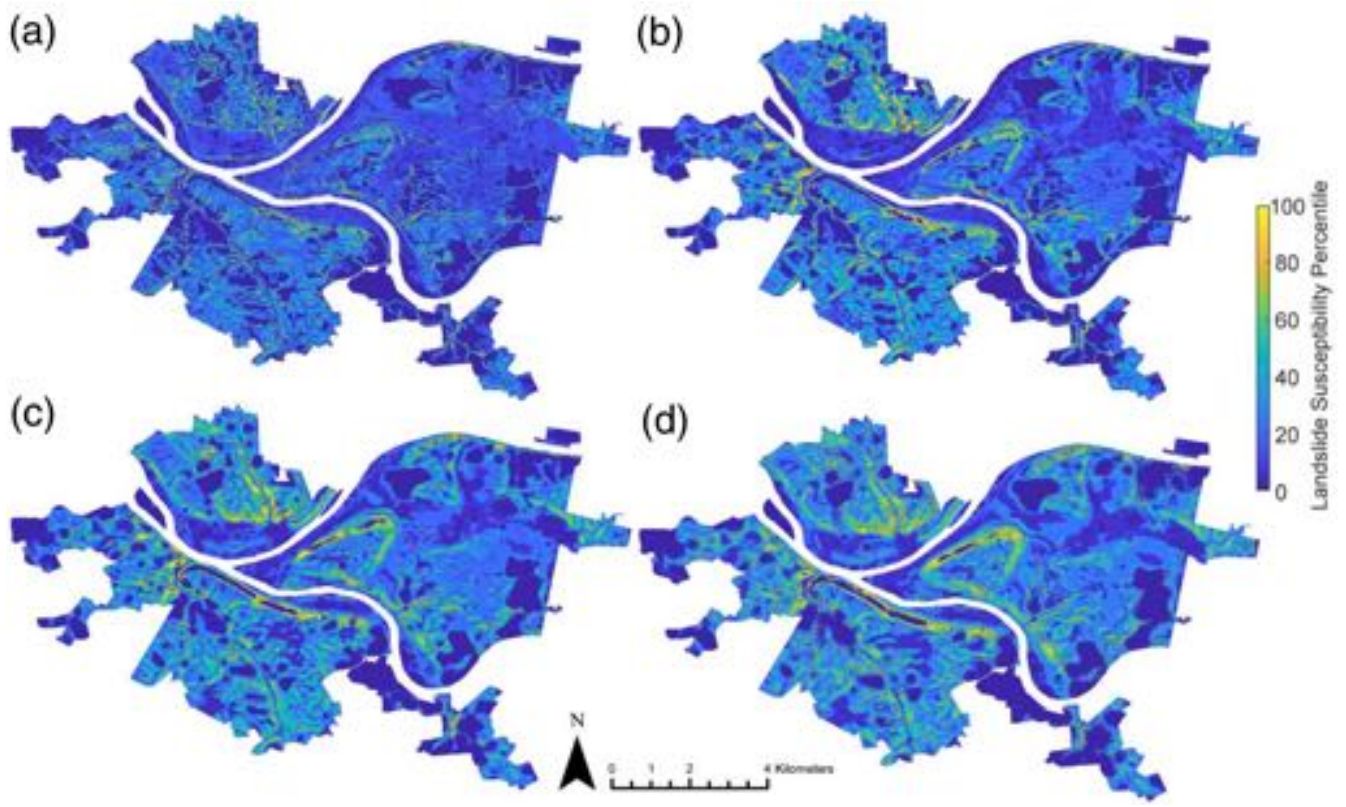

FIGURE 6 Landslide susceptibility maps based on the total 311 landslide inventory (N = 720) demonstrate the effect of filtration with different filter radii: (a) non-filtered; (b) 80 m; (c) 140 m; (d) 200 m. The color bar is based on the percentile of computed conditional probability values [Colour figure can be viewed at wileyonlinelibrary.com]

4 | RESULTS

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, seven are duplicate reports of the same landslides. The mean distance between the reported and field-validated locations is 104 ± 25 m (uncertainty is one standard deviation, most landslides occur between 60 to 120 m away from the reported landslide location), and the typical size of a field-validated landslide is approximately 5 m × 13 m (Supporting Information, Table S5).

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 (Figures 2 and 3) and the ranking of landslide-related factors (Figure 4). The two-dimensional correlation between the susceptibility maps (0.3775, Figures 2 and 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 (C_p). The highest C_p based on the field-corrected 311 inventory, occurs at the following factor-class combination: slope (25°–35°), nearest road (0–9 m), profile curvature (0.016–0.1 m^{-1}) nearest stream (144–277 m), and aspect (10°–123°). In contrast, the highest C_p for the original, non-field-corrected inventory occurs at the following factor-class combination: slope (14°–25°), profile curvature (–0.011 to –0.002 m^{-1}), nearest stream (10–144 m), 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 4).

This is particularly apparent in the magnitude and ranking of the distance to the nearest road factor (NR, Figure 4), whose influence on the AUC is meaningfully larger for the original 311-reported locations compared to the field corrected ones (Figure 4a vs. Figure 4b). This influence of roads on susceptibility estimates is apparent in Figure 2(b) compared to Figure 2(a). The range of AUC values for our analyses was 0.82 to 0.94 (Supporting Information, Table S6).

4.3 | Filtering of DEM 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 (0–240 m, in intervals of 20 m, Figure 5), shows that the spatial correlation

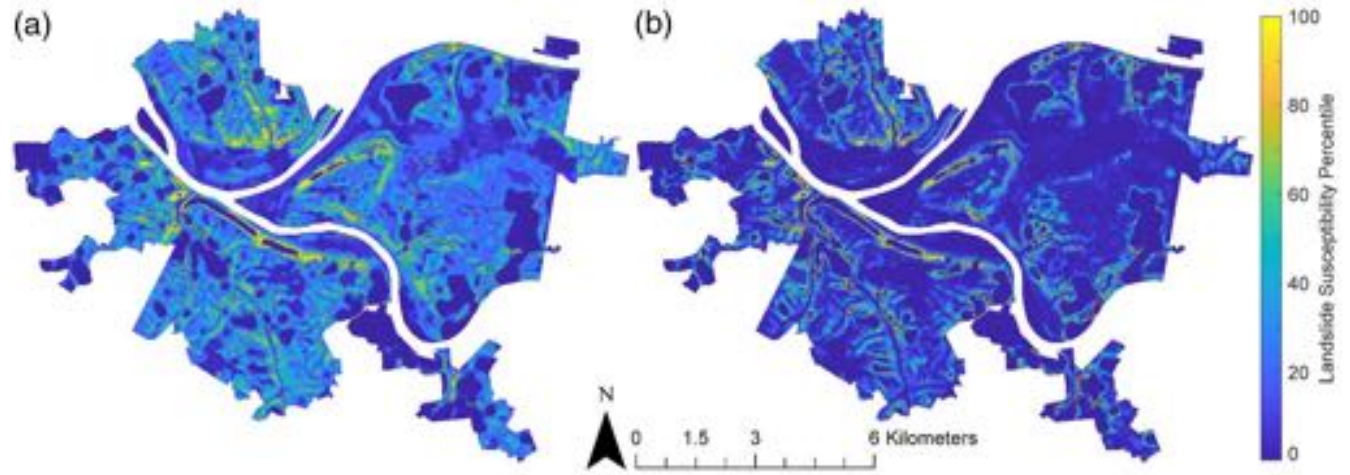

FIGURE 7 Landslide susceptibility maps that have been filtered using a 100-m filter radius that produced a maximal spatial correlation: (a) map based on the entire 311 landslide inventory; (b) map based on combined (ACES and USGS) inventory. The Lower susceptibility values in the ACES and USGS map are due to a lesser number of landslides in these datasets compared to the 311-based dataset [Colour figure can be viewed at wileyonlinelibrary.com]

between the susceptibility maps produced from the field-corrected and original landslide inventories increases with filter radius (Figure 5b) up to a maxima at a radius of 140 m, which is slightly larger than the uncertainty in landslide locations (104 ± 25 m). This trend repeats when comparing the landslide susceptibility map produced from the entire 311 inventory (N = 720 non-field-validated landslides, Table 1, Figure 1) to that produced from a combination of the ACES and USGS inventories (N = 134 field-validated landslides, Figures 5A and 6). For these inventories, the maximal correlation is attained at a filter radius of 100 m (Figure 7). Likewise, the filtering also increases the similarity in ranking of landslide-related factors (Figure 8). Compared to these inventories, the randomly generated landslide inventories (Figure 5b vs. Figure 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 (Supporting Information, Figures S1 and S2).

5 | DISCUSSION

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.

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 (Supporting Information, Tables S1–S4). 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.

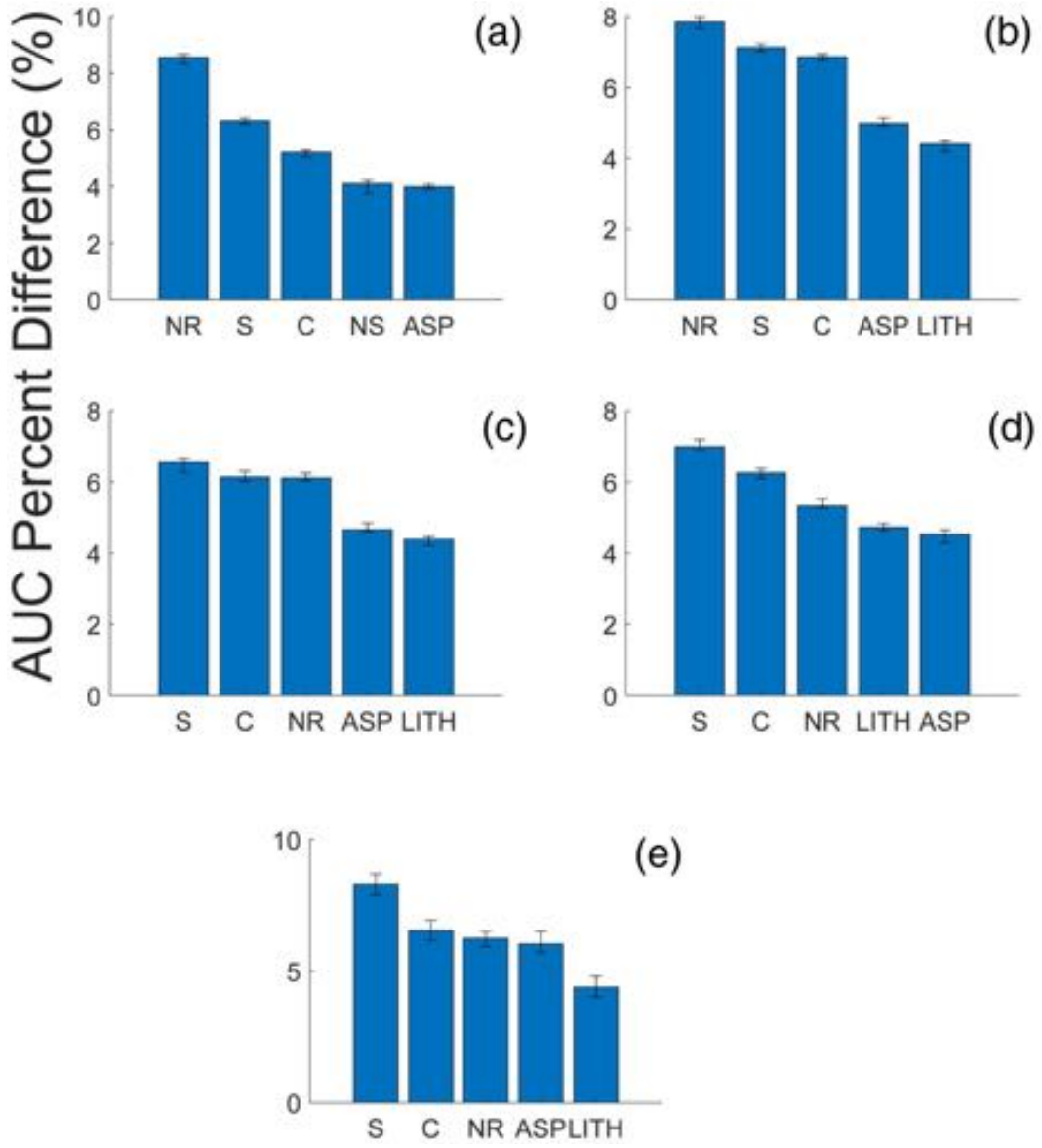

FIGURE 8 Ranking of landslide-related factors based on the total 311 landslide inventory (N = 720) at filtering steps: (a) 20 m; (b) 80 m; (c) 140 m; (d) 200 m; (e) influential factors of the combined USGS and ACES inventory with no filtering. Note that higher filter radius for this non-field validated 311-based landslide inventory (panels A–D), increases the similarity in factor ranking with the non-filtered field based ACES and USGS combined inventory. NR, nearest road; S, slope; ASP, aspect; C, profile curvature; NS, nearest stream; LITH, lithology [Colour figure can be viewed at wileyonlinelibrary.com]

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 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.

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 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 5(A), 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 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 9), they also reduce the effective map resolution (i.e., they increase the spatial extent of areas with the same landslide susceptibility, Figure 6), and are thus less preferable.

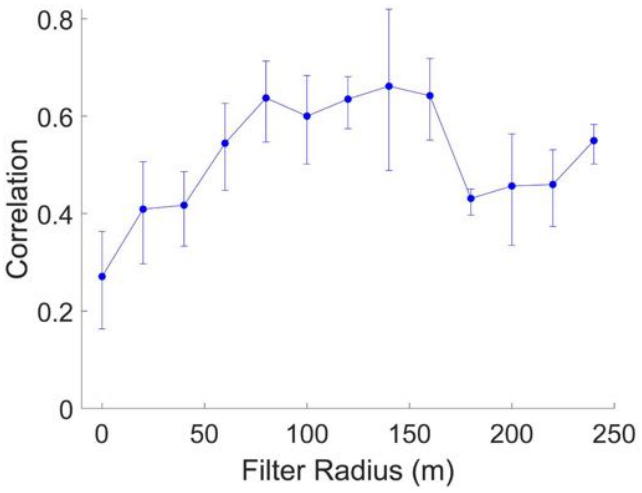

FIGURE 9 Correlation values between susceptibility maps produced from quasi-random landslide inventories versus the field-corrected 311 landslide locations. The error bars represent the 95th and 5th percentiles [Colour figure can be viewed at wileyonlinelibrary.com]

The filtration results are sensitive to the number of factor classes. Our experiments with six rather than five factor-classes (Supporting Information, Figures S1-S2), 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.

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. 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). There are 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.

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 4 and 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.

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.

ACKNOWLEDGEMENTS

The authors thank the Heinz Foundation for funding this study, the Pittsburgh Department of Public Works and Allegheny County Emergency Services for providing data and field guidance. The authors also thank Dr. Dan Bain for pointing at the potential value of a 311-based landslide dataset and to Ian Pamerleau for comments on an earlier version of this manuscript.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest for this study.

DATA AVAILABILITY STATEMENT

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

ORCID

Tyler J. Rohan https://orcid.org/0000-0001-9242-7252

REFERENCES

Allegheny County Landslide Task Force. (2019) Allegheny County Landslide Portal. Available from: 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. https://doi.org/10.1007/s11629-018-5168-y

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. https://doi.org/10.1016/j.geomorph.2010.03.005

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

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

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

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. https://doi.org/10.3390/ijgi8070300

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*. Boca Raton, FL: CRC Press, pp. 209–246.

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*. Cham: Springer, pp. 23–39.

Choi, B.Y., Al-Mansoori, M.K., Zaman, R. & Albishri, A.A. (2018) 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. https://doi.org/10.1016/j.cageo.2006.02.003

Cieslik, K., Shakya, P., Uprety, M., Dewulf, A., Russell, C., Clark, J., *et al.* (2019) Building resilience to chronic landslide Hazard through citizen science. *Frontiers in Earth Science*, 7, 278. https://doi.org/10.3389/feart.2019.00278

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. https://doi.org/10.1016/S0169-555X(02)00079-X

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

Crozier, M.J. (2010) Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, 124(3–4), 260–267. https://doi.org/10.1016/j.geomorph.2010.04.009

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. https://doi.org/10.1016/j.geomorph.2008.05.041

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. https://doi.org/10.4236/ojg.2018.87040

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

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. https://doi.org/10.1016/j.enggeo.2008.03.014

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*, (43), 1–20.

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. https://doi.org/10.1016/j.cageo.2015.04.007

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

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. https://doi.org/10.1016/j.geomorph.2015.07.012

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), 247–263. https://doi.org/10.1007/s002679910020

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. https://doi.org/10.1016/j.enggeo.2010.06.013

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

Hosmer, D.W., Jr., Lemeshow, S. & Sturdivant, R.X. (2013) *Applied logistic regression*, Vol. 398. Chichester: 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. https://doi.org/10.1186/s40713-018-0011-8

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

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. https://doi.org/10.1002/esp.2223

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

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

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. https://doi.org/10.1371/journal.pone.0218657

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. https://doi.org/10.1016/j.geomorph.2008.03.003

Kocaman, S. & Gokceoglu, C. (2019) A CitSci app for landslide data collection. *Landslides*, 16(3), 611–615. https://doi.org/10.1007/s10346-018-1101-2

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. https://doi.org/10.1016/j.geomorph.2005.07.005

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. https://doi.org/10.1002/gj.3204

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. https://doi.org/10.1007/s00254-005-0019-x

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. https://doi.org/10.1016/j.enggeo.2008.03.021

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

Marjanović, M. (2013) Comparing the performance of different landslide susceptibility models in ROC space. In: *Landslide Science and Practice*. Berlin: Springer, pp. 579–584.

Marjanović, M., Kovačević, M., Bajat, B. & Voženílek, V. (2011) Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225–234. https://doi.org/10.1016/j.enggeo.2011.09.006

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

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(3), 509–520. https://doi.org/10.5194/nhess-8-509-2008

Miao, T., Liu, Z., Niu, Y. & Ma, C. (2001) A sliding block model for the run-out prediction of high-speed landslides. *Canadian Geotechnical Journal*, 38(2), 217–226. https://doi.org/10.1139/t00-092

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. *et al.* (2020) Landslides across the USA: Occurrence, susceptibility, and data limitations.

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. https://doi.org/10.1016/j.ssresearch.2016.04.009

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

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

Pennsylvania. Bureau of Topographic and Geologic Survey, Berg TM. (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

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. https://doi.org/10.1016/j.geomorph.2017.12.008

Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., *et al.* (2020) A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto International*, 35(12), 1267–1292. https://doi.org/10.1080/10106049.2018.1559885

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

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

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. https://doi.org/10.1016/j.catena.2012.05.005

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. https://doi.org/10.1016/j.envsoft.2009.10.016

Ray, R.L. & Jacobs, J.M. (2007) Relationships among remotely sensed soil moisture, precipitation and landslide events. *Natural Hazards*, 43(2), 211–222. https://doi.org/10.1007/s11069-006-9095-9

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. https://doi.org/10.1007/s11629-013-2847-6

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

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. https://doi.org/10.1007/s12665-017-6475-2

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. https://doi.org/10.1016/j.compstruc.2011.02.016

Schellong, A. & Langenberg, T. (2007) Managing citizen relationships in disasters: Hurricane Wilma, 311 and Miami-Dade county. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. Piscataway, NJ: IEEE, p. 96.

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(8), 797–814. https://doi.org/10.1080/1365881031000159605 8

Schwanghart, W. & Kuhn, N.J. (2010) TopoToolbox: A set of Matlab functions for topographic analysis. *Environmental Modelling & Software*, 25(6), 770–781. https://doi.org/10.1016/j.envsoft.2009.12.002

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. https://doi.org/10.1108/IJOTB-12-02-2009-B003

Southwestern Pennsylvania Commission. (2017) Previously Active Documented Landslides in Southwestern Pennsylvania. Available from: 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. https://doi.org/10.1080/00221344208986578

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), 2729–2745. https://doi.org/10.5194/nhess-16-2729-2016

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.

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. https://doi.org/10.1016/S0013-7952(02)00268-5

White, A. & Trump, K.S. (2018) The promises and pitfalls of 311 data. *Urban Affairs Review*, 54(4), 794–823. https://doi.org/10.1177/1078087416673202

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

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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Rohan TJ, Wondolowski N, Shelef E. Landslide susceptibility analysis based on citizen reports. *Earth Surf. Process. Landforms*. 2021;46:791–803. https://doi.org/10.1002/esp.5064