Landslide Hazard Assessment for Fayzabad District, Badakhshan Province, Afghanistan

By

Nathan Schlagel
B.S., University of Nebraska – Omaha, 2015

Submitted to the graduate degree program in Department of Geology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Arts.

Chair: William C. Johnson

Mike Taylor

Mike Blum

Date Defended: 6/8/2018

The thesis committee for Nathan Schlagel certifies that this is the approved version of the following thesis:

Landslide Hazard Assessment for Fayzabad District, Badakhshan Province, Afghanistan

_____

Chair: William C. Johnson

Date Approved: 6/11/2018

Abstract:

Fayzabad District is one of those most impacted by landslide hazards in Afghanistan, accounting for 71% of all national landslide fatalities reported between 2012 and 2017. Necessary elevation data did not cover the very south of Fayzabad District; consequently, this study focuses on the northern two thirds of the district, where data were available. A landslide inventory was developed by mapping landslides using DEMs and high-resolution satellite imagery to aid in development and assessment of both Heuristic and bivariate statistical models of landslide susceptibility. Landslide statistics, including length, area, width, and pertinent relationships to geology, elevation, aspect, slope, and proximity to faults and streams were quantitatively calculated using geoprocessing tools. Hazard maps were produced using landslide susceptibility and proximity of villages to mapped landslides.

Mapped susceptibility results indicate that in this part of Afghanistan landslides occur primarily on north to northwest aspects in loess or soil media over gneiss bedrock. Landslides are concentrated between 1500 m and 2000 m elevation and on 18° to 45° slopes within 60 m of a stream channel and or within 1 km of a fault. Landslide dimensions plot linearly on log-log scales, simplifying the development of predictive associations.

Model results encapsulate a high proportion of landslide pixels within areas of high susceptibility, although there were significant variations between Heuristic and bivariate methods. Bivariate methods performed better universally, but may be over trained when the entire dataset is used to produce statistical weights. Use of subset of data to develop weights results in a more even distribution of landslides between low- to high-susceptibility zones. Findings in both the landslide inventory and susceptibility models are supported by prior studies of landslide behavior in Afghanistan.

Programmatic workflows allowed for rapid production of many model components after initial reclassification and will facilitate further research in Afghanistan, and application of the methodology elsewhere. Map products potentially provide a new tool for hazard planners and aid groups in

northeastern Afghanistan, and supplemental code will allow for rapid incorporation of new datasets as they are developed.

Acknowledgements:

I extend my appreciation to my advisor, William Johnson, for all his support, encouragement, and expertise shared; to Emlyn Hagen and the ASDC for hosting past research and for providing encouragement to continuing this work; to my committee members for their valuable feedback and constructive criticisms; and to the University of Kansas and ESRI for providing academic computing facilities and software licenses for ArcMap. No external funding was required or received for this project.

Table of Contents

1 Introduction

1.1 Landslide Hazard

The people of Afghanistan are subject to an array of natural hazards, including floods, earthquakes, landslides, avalanches, and extreme seasonal temperatures (Boyd et al.,2007; Shroder, 1989; Shroder et al., 2011; OCHA, 2017). Floods and earthquakes have the greatest national impact in terms of frequency, extent, affected individuals and property damage (Table 1; OCHA, 2017). Landslide hazards, though typically localized, are complicated in that they have a propensity to be triggered and amplified by floods, heavy rains, and earthquakes (Schuster & Krizek, 1978; Shroder et al., 2011; Shroder, 2014; Turner & Schuster, 1996).

Landslides can be disproportionately devastating relative to their relatively low frequency of occurrence (Schuster & Krizek, 1978; Turner & Schuster, 1996). This is the case in Afghanistan, where landslides account for only 3.6% of reported natural disasters affecting the populace as recorded by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) from 2012 to 2017, but they account for 16.8% of natural disaster related deaths. Landslides contribute more proportionally, relative to number of occurrences, to injuries, deaths, displacement of inhabitants, and damage to infrastructure and personal property.

Nearly 90 landslides were documented in Afghanistan between 2012 and 2017 (OCHA, 2017), though the impact of these landslides varied by event. Many small landslides may have little or no impact on even proximal inhabitants or structures. Landslides can occur very slowly, harming only infrastructure over months or years, or be devastatingly quick, resulting in massive harm. This trend extends globally, with individual historic landslides resulting in tens, to tens-of-thousands killed by single landslides or sequence of multiple failures triggered by a common event (Schuster & Krizek, 1978; Turner & Schuster, 1996; Zhang et al., 2015). However, landslides pose the greatest risk, in terms of lives lost, in developing countries where there is a paucity of research and few mitigation efforts have been

weighed against landslide issues (Nadim et al., 2006; Petley, 2012). In Afghanistan, the mudslide at Aab Barik village, Badakhshan, on May 2nd, 2014 (Vergano, 2014; Zhang et al., 2015) alone resulted in 71% of all casualties reported by OCHA between 2012 and 2017. OCHA estimates differ, however, from other reports for the Aab Barik slide, suggesting inconsistent reporting between organizations, and, in general, tracking any population related statistic in Afghanistan is difficult due to incomplete and outdated census data (Thompson, 2013).

1.2 Landslide literature overview

This paper will focus on modeling landslide susceptibility at lower elevations in the northern portion of the Hindu Kush where loess mantles gneiss and granite bedrock, although Afghanistan experiences all types of mass wasting due to the high relief, varying geology, and harsh climate. In the highest elevations of the Hindu Kush mountains, rock and snow avalanches are the primary forms of slope failure, with rock falls and slides occurring at median elevations. Earth falls, slides, and flows in loess or soil dominate lower elevations such as those examined in this study (Shroder et al., 2011a/b). 'Landslide' is used henceforth to refer to all forms of slope failure, including flows, falls, slides, slumps, and avalanches.

Mechanisms of slope failure in the region are primarily ground shaking during earthquakes and increases in soil moisture or ground water pressure during the rainy season. Surface run off, increase in ground water pressure, dissolution of cement, and subsurface piping can cause debris flows or slumps along shear planes. These often occur where water acts at the soil-bedrock contact, i.e., where the bedrock surface serves as a plane of failure. Moisture-related failures typically occur on higher-angle slopes, whereas liquefaction and flows in earthquake shaking are more commonly triggered on low-angle slopes (Naseri & Kang, 2016; Schuster & Krizek, 1978; Shroder et al., 2011a; Turner & Schuster, 1996; Zhang et al., 2015).

Landslide inventory maps are a key component in assessing susceptibility and risk. Conventional mapping techniques involve on-site field investigations and examination of aerial photographs or satellite imagery. Landslides can be identified in imagery by discolorations over areas of failure where a landslide has removed vegetation, dips in elevation where material has been removed, or formation of mounds where mobilized material comes to rest (Galli et al., 2008; Guzzetti et al., 2000, 2012; Schulz, 2004, 2005, 2007; Schuster & Krizek, 1978; Turner & Schuster, 1996). Incorporation of digital elevation models (DEMs) allows for specific examination of change in elevation and slope where landslides have occurred. With high-resolution DEMs (~1 m) it is possible to achieve a semi-automated scheme for identifying landslides (Booth et al., 2008; McKean and Roering, 2004).

With an event inventory it is easier to assess susceptibility and risk via modeling or statistical analysis. Temporal data from successive mapping or dating of landslides combined with calculations or estimates of material volume for landslides allows estimation of recurrence intervals for given slide sizes by frequency-magnitude relationships (Shroder, 2014). Multi-temporal data also allow assessment of specific triggering mechanisms, such as storms, floods, or earthquakes (Kamp et al., 2010), and can help determine relationships between trigger and landslide magnitudes. Regardless of temporality in inventory maps, they primarily enable the validation of susceptibility or predictive models by comparing results to actual occurrences.

Statistical assessment of landslide susceptibility makes use of inventories to develop statistical weights of variables, where susceptibility of a given pixel equals the sum of all input layers at that pixel's location (Kayastha et al., 2013; Pardeshi et al., 2013). Susceptibility can be modeled without an inventory by means of multi-criteria analysis. In this approach, variable weights are assigned based on researcher knowledge of landslide principles and local conditions. Each variable is ranked by significance of contribution as a percentile, and layers are combined in a GIS to produce a susceptibility map (Effat & Hegazy, 2014; Kayastha et al., 2013; Pardeshi et al., 2013; Pradhan & Kim, 2016). Reichenbach et al.

(2018) provide further, comprehensive presentation of peer reviewed, statistical landslide susceptibility studies including a variety of modeling and assessment methods, and scale of study area.

Previous landslide modeling in Afghanistan has been done at the national scale (Schlagel, 2015), and provincial scale for Badakhshan Province (Zhang et al., 2015) and Takhar Province (Naseri & Kang, 2016). Schlagel (2015) and Zhang et al. (2015) both used 90 m SRTM elevation data and DEM derivatives, but both lacked a landslide inventory; low-resolution data and lack of inventory are significant limitations in landslide studies. Naseri and Kang (2016) developed a landslide susceptibility map for Takhar Province using higher-resolution DEMs (30 m) and a landslide inventory map created using Google Earth Imagery. These were significant improvements, but considering the scale of many landslides this does not represent small magnitude events well, which are the majority of landslide occurrences.

For this study in Fayzabad District, higher-resolution DEMs from the Department of Defense (DOD; HRTE, 2008) were available than SRTM derived DEMs used in prior studies (Naseri & Kang, 2016; Schlagel, 2016; Zhang et al., 2015). At 5 m horizontal resolution, the DOD DEMs provided a significant improvement in ability to discern and attribute landslide characteristics, but they were still insufficient to see smaller landslides for mapping purposes, which would require resolution on the order of 1 m, or even sub-meter (Booth et al., 2009; McKean and Roering, 2004; Schulz, 2004, 2005, 2007). Satellite images at 2 m resolution or higher were also available through ESRI's ArcGIS basemaps, which were an improvement over Google Earth imagery of Afghanistan and greatly aided development of a landslide inventory. Use of higher-resolution DEMs and imagery allows for inclusion of smaller, more frequent landslides that often go overlooked even though they pose risk of damage to proximal roads or buildings (Shroder, 2014). The goal of this project is to take advantage of access to high-resolution DOD DEMs to develop a landslide hazard package for Fayzabad District for use in aid organizations, hazard mitigation,

and development planning to be accessible through the Afghan Spatial Data Center (ASDC), an affiliate of the USAID iMMAP agency, and establish a foundation for future work.

2 Study Area

Fayzabad, in northeast Afghanistan, is one of the country's most landslide-affected districts. For the purposes of this study, the newly organized districts of Arghanj Khaw, Argo, Darayim, and Taftal-e-Sufla are considered as part of Fayzabad. During the period of 2012 to 2017, 88 landslides were reported by OCHA, and, of the reported landslide events, 35 (39.7%) occurred in Badakhshan Province; 16 of the landslides in Badakhshan (18.2% of national landslides, and 45.7% of those in Badakhshan) were within Fayzabad District. Of national landslides, those in Fayzabad account for over one third of affected individuals and damaged or destroyed structures, and 71.3% of fatalities (Figure 1; OCHA, 2017).

Coverage of DEM data was available for the northern two thirds of Fayzabad District, limited in the south either by incomplete flight paths or exclusion from the dataset. Developed susceptibility models were applied to three areas in the district. First, to two watersheds totaling an area of 167 km^2 in which model variations were developed and tested. Second, on half of the district for which detailed surface mapping identified not only landslides, but likely landslide terrains and unfailed terrain that was used to determine material critical acceleration; total area for this second model was 1,263 km^2. Finally, the full district was modeled, with a total area of 2,168 km^2 (Figure 2).

Elevation and slope in the district range from 813 m to 3,580 m and from flat to 81°, respectively. Geology in the district is primarily gneiss overlain by loess deposits and associated soils on northwest aspects. Intense faulting occurs throughout Badakhshan province, including the study area, as well as intense seismicity, which can initiate landslides and increase the frequency and magnitude of landslide occurrences (Boyd et al., 2007; Ruleman et al., 2007). Landslides in this region typically occur in soils or loess and are initiated by heavy rainfall or earthquakes (Shroder et al., 2011a). Exposed bedrock in Fayzabad appears to be the result of primarily successive reactivation of past landslides that stripped

away overlying soils; the remaining rock exhibits characteristics of debris flows in satellite images and is later identified as 'landslide terrains' in mapped surface classification.

3 Methods

3.1 Data

High-resolution Terrain Elevation data (HRTE) from airborne Interferometric Synthetic Aperture Radar (IFSAR) are used to produce digital elevation derivatives. DEM tiles are 15 minute by 15 minute at 5 m horizontal and 0.01 m vertical resolution. Horizontal accuracy is 2.5 m Root Mean Square Error (RMSE) for slopes under 20°. Vertical accuracy varies by slope angle, with ± 3 m RMSE for slopes under 20°, ± 5 m for slopes 20° to 30°, ± 9 m for slopes greater than 30° (HRTE, 2008). ESRI streamed basemap satellite imagery was used extensively for mapping and visual assessment. Basemap imagery includes 2.5 m SPOT, 1 m GeoEye IKONOS, and AeroGRID where coverage available for Afghanistan. Specific image metadata were not available. LandSAT8 data for November 7, 2016 was used to calculate Normalized Difference Vegetation Index (NDVI).

Fault maps (Ruleman et al., 2007) and critical acceleration for common lithologies (MRI, 2003) were used as tectonic components in this study. Geologic maps compiled by the USGS from mid- to late-20th century Soviet-sourced material, refined during United States presence in Afghanistan following 2001 were used herein for modeling (Steinshouer et al., 2006), and, though the resolution and quality are low for the scale of this project, this was the only geologic mapping available for the entire study area.

3.2 DEM Processing

All processing and modeling were done in ESRI's ArcMap (ver 10.4.1), with exception of some workflows being implemented in Python (ver 2.7) using the ArcPy library, and some data being exported to Excel for calculation and then re-imported to ArcMap. Reference hereafter to a tool, function, process, or creation of a map or layer refers to an ESRI ArcMap tool unless specified otherwise. DEMs

were used to produce slope, aspect, curvature, topographic wetness index (TWI) (Equation 1), drainage basin, and drainage channel maps (Cooley, 2016; Sorensen et al., 2006). Drainage density and buffer maps were created from the produced channel map using line density and Euclidean distance, respectively. Fault density and buffer maps were similarly produced from source feature class (Ruleman et al., 2007). LandSAT8 bands for red and near infrared (NIR), originally 30 m resolution, were pansharpened using the 15 m panchromatic band. NDVI was calculated by at both 30 m and 15 m cell sizes using Equation 2 (Johnson, 2014).

$$TWI = Ln[(FlowAccumulation + 1) * cellsize \,/\, \tan(Slope * \left(\frac{pi}{2}\right) / 90] \qquad \text{Equation 1}$$

$$NDVI = (NIR - RED)/(NIR + RED) \qquad \text{Equation 2}$$

Critical Acceleration (Ac) was determined by adapting procedures in FEMA's Multi-hazard Loss Estimation Methodology: Earthquake Model Hazus - MH 2.1 Technical Manual (MRl, 2003). Linear equations were developed to describe lines from Figure 4.12 in the Hazus manual (Figure 3 herein; MRl, 2003), which depicts the relationship between the Ac of a given material and its topographic slope following the format of Equation 3, where m is the slope of the line defining the relationship between Ac and topographic slope for a lithology (M), *Slope* is a surface slope raster in degrees for the study area, and b is the y-intercept for the line describing lithology (M) (Figure 3). Lithologies were reclassified according to three requisite classes: crystalline and strongly cemented rock; weakly cemented rock and sandy soil; and shale, clayey soil and existing landslides. Equation 3 was applied as raster algebra. The y-intercept used for Ac is theoretical, and minimum bounding slope angles should be used (Table 4.16 in MRl, 2003) when compared to peak ground acceleration (PGA) to determine coseismic landslide susceptibility or for implementation in the Hazus - MH 2.1 software. In this study spatial differences in Ac are used to represent relative susceptibility using weighted values, rather than comparing Ac to PGA.

Weights ascribed to given Ac values in this manner make adherence to minima and maxima less important in this context.

$$Ac = mM * Slope + bM \quad \text{Equation 3}$$

3.3 Mapping

Landslides were mapped by heads-up digitizing in ArcMap using ESRI streamed base maps with elevation, slope, and hillshade DEMs. Because the satellite images were of higher resolution (2.5 m or better) than available DEMs (5 m), they proved most useful of available data for mapping. Methods for identifying and mapping landslides were followed as detailed in existing literature, i.e., looking for bare earth stripped of vegetation, headscarps, extension cracks, hummocks, internal drainage development, and flow or lobate deposits (Figure 4; e.g. Galli et al., 2008; Guzzetti et al., 2000; Schulz 2004, 2005, 2007; Schuster & Krizek, 1978; Turner & Schuster, 1996). Land surface was mapped in four categories: 1) high-confidence landslides distinguishable by removed vegetation cover and character clearly depicting flow of material with definite scarps; 2) lower confidence landslides distinguished from highly confident landslides by post event erosion of the landslide area, return of vegetation cover, or complexing of multiple landslides together with poor boundaries defining each; 3) eroded or stripped terrain not clearly discernable as a landslide, but possibly formed by repeated landslides or mass wasting; and 4) unfailed loess or soil cover (Schulz, 2004, 2005; Schuster & Krizek, 1978). The fourth category was appended to the lithology map, as no loess was on the original geologic map, but this region of the N-NW Hindu Kush is characterized by a loess cover (Shroder et al., 2011). Distinction between landslides in category 1 and 2 are subjective, and in statistical analyses these groups are merged into one.

Landslide attributes, e.g. length, area, and DEM related statistics, were calculated with geoprocessing workflows applied to the watershed subset of data containing 837 landslides to save time over manual measurement or processing time to run on the entire dataset. Length was calculated using a landslide region group raster to mask elevation for specific landslides or landslide complexes and

produce a flow length map using hydrology tools. The maximum length value for each landslide was extracted using zonal statistics and considered the length of the landslide. Landslide width was geometrically estimated in Excel using the area of respective polygons and calculated length.

3.4 LSI Modeling: Heuristic model

Heuristic, or multi-criteria models engage researcher knowledge qualitatively to assign weights to environmental variables expected to contribute to landslide susceptibility (Effat & Hegazy, 2014; Kayastha et al., 2013). Methods used were primarily those of Effat and Hegazy (2014), and as was done in Schlagel (2015). Data layers were classified into 9 bins using the Natural Breaks Jenks method and reassigned standard weights between 0 and 9 based on author knowledge and information gleaned from the literature (Shroder et al., 2011a) for use in the weighted overlay tool to create a landslide susceptibility index (LSI). Fewer bins were used where appropriate or necessary for categorical data with fewer than 9 types or for data with fewer desired classes, such as curvature being concave, convex, or straight (Buckley, 2010), or compass directions with similar empirical susceptibility to landsliding. Some lithologies with expected similar behavior were merged to reduce the class count to 9 in accordance with limitations in the weighted overlay tool. Weight assignments were informed by observations in Shroder et al. (2011a).

3.5 LSI Modeling: Bivariate statistical analysis

Previously reclassified layers were compared to the landslide inventory using the tabulate area tool. Results were in table format with each bin in the layer classification ascribed the landslide area in square meters corresponding to those elevation ranges. Table values from tabulate area results were used to calculate statistical weights following methods by Kayastha et al. (2013) (Equation 4), where LS refers to landslide area. Respective layers were reclassified from class values ranging 0-9 to respective statistical weights using conditional operators in raster algebra; values were reassigned in this manner because the statistical weights have decimals and the reclassify tool forces an integer output. The sum

of all statistical layers represents relative susceptibility to land sliding for a given pixel. This bivariate method is not limited to 9 classes like the weighted overlay, so lithology was expanded to allow all classes as originally mapped by Steinshouer et al. (2006). Lithologies in the study area that lacked distinct, mapped landslides were still excluded from the calculation and assigned a weight of 0 to not influence the outcome of the summation.

$$Statistical\ weight = \ln\left(\frac{LS\ Area\ in\ class}{Class\ Area} * \frac{Map\ Area}{LS\ Area\ in\ map}\right) \quad \text{Equation 4}$$

3.6 LSI Modeling: Informed Heuristic model

After calculating statistical weights in the bivariate method resulting values were used to inform and reassign weights in the Heuristic method. As the weighted overlay is limited to integers between 0 and 9, statistical values were sorted low to high and class bin weights reassigned according to that order. For example, with 1 being low and 9 being high, if bin 7 is shown to be more strongly correlated to mapped landslides through a higher statistical weight than bin 8, the weight of those bins is switched in the model. Percentages of each layer in the weighted overlay are still calculated using the straight rank sum method of the normal Heuristic method, but significance or order of each layer is changed based on the sorted absolute value of the sum of weights in each layer. This was done using two methods, both utilizing the sum of all statistical weights in a layer. The variation was calculating the absolute value of the sum for one method to force a positive sign that represented the magnitude of influence that layer would exert relative to other layers. The importance of this is that the sum of weights alone allowed negative values and would change how the weights were sorted. For example, a weight sum of 3 would be ranked higher than a sum of -4, even though the latter had a stronger total impact on the model output. The method used and uncommon inputs for all models are listed in Table 2.

3.7 Hazard Modeling

Villages were mapped by heads up digitizing following edges of the outermost buildings or apparent associated outdoor space. Focus was placed on structure groupings rather than lone or isolated buildings, which were not mapped. Buffer and density maps were produced for both villages and landslides. Buffers were calculated by Euclidean distance up to 5000 m and classified into bins 0-10 using Natural Breaks Jenks, with 0 being greater than 5000 m. To calculate densities, polygons were converted to points, restricted to within polygons where centroid points fell outside of irregular shapes. The point density tool was used with a search radius of 307 cells, output units in km^2, and a cell size of 5 m to match DEM resolution.

Landslide and village buffer layers in combination with landslide density and results of susceptibility models were used to describe hazard. The intention was for assessed pixels to scale hazard output up for villages near landslides, and scale down for villages removed from landslide events, or landslide areas with no proximal villages. Input layers were classified with low numbers reflecting high distance, or low hazard, and high numbers reflecting close distances, or high hazard. The final district scale hazard calculation followed Equation 5, where LSI is landslide susceptibility index, and LS refers to landslide features.

$$Hazard = LSI + VillageBuffer + \left(\frac{LSBuffer}{2}\right) + \left(\frac{LSDensity}{2}\right) \qquad \text{Equation 5}$$

3.8 Analysis

Models were quantitatively assessed by two measures: 1) percent distribution of landslides within each susceptibility class, where greater proportion of landslide pixels falling in high susceptibility values is considered most accurate; and 2) percent shared area, or agreement, between models (Kayastha et al., 2013). The distributions of landslides by LSI was calculated for all models, but shared area was only determined for all combinations of test models at watershed scale, and for one pair of larger models, between the full district and the half district. Distributions of landslides within each LSI

zone were calculated using the tabulate area tool between model output and the rasterized landslide inventory. Shared area was calculated by subtraction of model layers, where a result of 0 reflects cells of equal value and area shared between the two models; a result of 1 reflects a difference of one class between the models, for example, a pixel of "medium" LSI (value 3) minus a pixel of "low-medium" (value 2) or "medium-high" (value 4) LSI would result in a deviation of ± 1. Results were in pixel count for each value and converted to area using cell size. Percent shared area between models was calculated for both perfect shared area (0 deviation), and for a deviation of ± 1 LSI class as in the preceding example.

4 Results

4.1 Mapping

Mapping results included 3486 landslides in the district (Figure 5). The half of the district with completed surface classification, including landslides, eroded or landslide terrains, and unfailed soil and loess had 2419 landslides, and 903 landslides were in the watershed areas used to develop methodologies (Figure 6). Approximately 1.7% of the district surface area is a distinct landslide, not considering the mapped 'eroded or landslide terrain' class, which comprises significantly more area.

Mapped landslides show association with specific environmental factors. Of all lithologies, 36% of landslide pixels occur over gneiss bedrock, and 23.8% in conglomerate and sandstone presumably of weak cementation (MRI, 2003). This follows the approximate proportion of rock types in the area (Figure 7) and follows landslide trends in the literature (Schuster & Krizek, 1978; Shroder et al., 2011a, b). Notably, no distinct landslides were mapped in shale units, which is contrary to that expected (Schuster & Krizek, 1978; Shroder, 1968; Turner & Schuster, 1996). This is perhaps because shale makes up a minute portion of outcrops in the study area and consists primarily of eroded or landslide terrains without distinct observable landslides in imagery.

North and northwest aspects host a combined 42% of landslide pixels, and 57.8% occur between 1494 m and 1937 m elevation. Concentration of landslides at similar elevations in Kashmir were noted

by Kamp et al. (2010). Slopes show an increasing proportion of landslides from low- to high-angle with 25.8% between 27° and 36°, 25.6% from 36° to 45° followed by a sharp drop to 8.6% at 45° presumably due to angle of repose in local media (Figure 8). Faults and stream channels also seem to exert a strong control on landslide location, with 40% of landslides occurring within 1 km of a fault, 35.8% within 30 m of a stream channel, and 54.8% on low (dry) TWI values. That landslide pixels occur primarily with low TWI values, yet also in close proximity to streams suggests association with drainage slopes near channels, and perhaps runoff as a more significant initiation factor than channel erosion (Figure 9). Such association may be more related to shallow debris flows, whereas deeper-seated slumps and earthflows would likely be initiated by erosion and undercutting or over-steepening at channel cut-banks.

Maximum dimensions calculated for landslides show lengths up to 824 m, widths up to 360 m, relief from top of scar to bottom of deposit up to 446 m, and maximum area of 212,359 m^2 (Table 3), all of which are positively correlated. Plots of landslide length and width versus area, length-to-width ratios, and height plot linearly with area on log-log scales and may allow for discovery of predictive relationships (Figure 10).

The completed village map has 720 villages in the entire district, of which 440 are in the southern region represented by the mapped extent in Figure 6, and 68 are in the watersheds used to develop methods; to note, no distinction was made between villages, towns, or cities. Of village pixels, including individual pixels within the same village, 58% are within 601 m, and 27% are within 275 m of a landslide pixel. Most villages are relatively small, with mean area of 41,216 m^2, and median area of 11,923 m^2. The smallest village mapped was 230 m^2, and largest 3,591,935 m^2.

4.2 LSI model results

Statistical models of this study performed universally better than Heuristic variants (Figure 11). Model performance increases with addition of critical acceleration data (per Table 2), but decreases with addition of NDVI, which is likely due to resolutions being reduced from 5 m to 30 m, and 15 m for

the pansharpened NDVI. The initial Heuristic model did not perform well, placing only 9% of landslide pixels in the highly susceptible category. Figure 12 shows the best statistical model at watershed scale, with 83% of landslides portioned in the high susceptibility zone. FayzMod1, the first model applied to the full district, is the poorest of all statistically-based models in terms of capturing landslides in the highest susceptibility category, but had the most linear landslide-to-LSI distribution of all models (Figure 7), which occurred by inadvertently using the weights from FayzMod_1/2 (Figure 13), the half district area, for the model of the entire district. This is significant because the weights are dependent on the total area and landslide area in the specific region being examined. However, the linear distribution may indicate that the statistical models with a more exponential distribution are over trained. This oversite was corrected in FayzMod2, and dedicated weights were calculated for the full district, which increased landslide inclusion in the high susceptibility zone from 48% to 66% (Figure 14). This is still 17% lower than the highest test model at watershed scale, which may be explained in part by exclusion of critical acceleration due to an incomplete component for calculating Ac at the district scale. Based on the change from StatMod1 to StatMod2, however, it would not likely improve more than a few percent. An increase in performance from the first to second Heuristic model is due to both improved weighting and inclusion of Ac with input data.

Using statistical calculations to reorder integer weights of the Heuristic model increased Heuristic accuracy (Figure 11). Shared area, and distribution of landslides in high susceptibility zones were highest in bivariate models. StatMod1 and StatMod2, both 5 m resolution, with critical acceleration added in the second model, had 82% perfect agreement, i.e., 82% of pixels in each have the same value. StatMod2 and StatMod3, with addition of NDVI at 15 m resolution shared 81% area; outside of this pairing, adding NDVI to other model variations or changing from 15 m NDVI to 30 m resolution reduced performance by at least 6%. All bivariate models and Heuristic_2 share over 95% area, when allowing for a deviation of ± 1 LSI class, thereby showing that models herein do not in principle

contradict each other. Deviation may be explained in part by different ranges of values in model output being grouped in different final bins by the chosen classification scheme, i.e., bin ranges, and Natural Breaks Jenks, etc. Expanding area analysis to allow for ± 1 susceptibility class between models shows high agreement between all models, suggesting agreement of trend, i.e., no model drastically disagrees with any other. All statistical variations have shared area between 62% and 82%; Heuristic to statistical model shared area ranges 37.2-38.9%, and informed-Heuristic to statistical range 51.2-57%. Allowing for ± 1 deviation in LSI all models exceed 88.5% agreement, with a maximum agreement of 99.48% with ± 1 deviation (Table 4).

4.3 Hazard results

As noted in mapping results, 58% of village pixels are within 601 m of an existing landslide, and as a result, most villages show relatively high hazard potential in conducted models. Model methods effectively reduce LSI representation in unpopulated areas and accentuate populated areas, but may over-estimate risk relative to comparing village locations to LSI models. Using the final susceptibility model for the district, 15.35% of village pixels are in the high landslide susceptibility zone with 34.44% being medium-high. The final hazard model followed Equation 5, with output classified to 5 categories by Natural Breaks Jenks (Figure 15). No village pixels were in the low hazard classification, and less than 1% were in the low-medium hazard zone, as distance to villages was heavily weighted; 15% were in medium hazard, 36% in medium-high, and 48% in high hazard.

5 Discussion

Mapping of landslides and villages in the study area was facilitated by high-resolution images, but landslide mapping by image interpretation is extremely time consuming and took several months to develop a landslide inventory for use in this study. Similar mapping periods are reported in literature, with average mapping rate of 788 in km^2 per month per mapper, but individual rates varied drastically depending on the scale of production (Guzzetti et al., 2000, 2012). In the future automated mapping

techniques utilizing DEMs may be applicable in Afghanistan, but new datasets are required. Some automated methods (e.g. Booth et al., 2009; McKean & Roering 2004) were attempted for the study area, but DEM resolutions were too low for detecting many of the mapped features, and in places the terrain was rugged enough to limit accuracy of wavelet techniques. Image analysis on satellite photos or LandSAT8 may be viable alternative. However, high-resolution multi-temporal satellite catalogs may be limited by proprietary access or leases. LandSAT8, while having the advantage of an ongoing collection of multi-temporal, freely accessible images in an analysis-friendly format lacks resolution to see many but the largest landslides. LandSAT8 is limited due to its recency, having launched in 2013, and LandSAT7, although having a longer catalog, has data gaps over Afghanistan.

Expanding on temporality of data, datasets herein are several years removed from each other. Geologic mapping (lithology, faults, etc.), while modified in the early 2000's, is based heavily on mid-late 20th century sources. DEMs were developed from data created in 2006 and do not reflect some of the larger landslides visible in newer satellite photos, such as the 2014 Aab Barik mudslide (Vergano, 2014; Zhang et al., 2015). Although it is clear from apparent freshness of some landslides in high-resolution imagery that much of the imagery is relatively new, the respective dates, tile boundaries, and specific image sensor metadata were not available for streamed high-resolution satellite imagery in ArcMap. ESRI's source sensors, satellite projects, and companies that supplied imagery for the basemaps were included together in a single citation in both ArcMap and on their website.

Issues between differing dates in data may impact map results because of a known, mapped feature being associated with pre-failure elevation data in GIS analyses. Mapping bias resulting from differing resolutions, cloud cover, and shadows in basemap imagery, as well as varied time-labor spent mapping each area used in modeling may bias susceptibility and hazard results. This is particularly so for the watershed areas, where greater time was spent during development and may have influenced quantity and quality of identified landslides relative to the rest of the district. The north and northeast

portions of the district were particularly problematic for mapping, relative to the center and south, due to the many shadows that made identifying landslides difficult. In addition, some imagery tiles had lower-resolution images than the 5m DEMs. In cases of obscuring shadow or low-resolution imagery, DEMs did not greatly aid mapping except for large landslides because of insufficient resolution. This likely resulted in underrepresentation of landslides, particularly smaller, more frequent failures, in areas of poor image quality, which would influence concentration of modeled hazard using landslide buffers and density inputs.

Other issues were found in the methods used to automatically extract landslide lengths, however, manual measurements for so many landslides would have been far too time-consuming. The flow length tool allowed rapid measurement of all landslides and increased accuracy over other automatic options such as default polygon lengths, polygon perimeters, bounding polygons, or polygon axis lengths, and avoided misapplication of “length” for landslide complexes along-slope with combined width-to-length ratios greater than one, or length-to-width less than one. Potential error exists in cases of flow path diversion within the slide mass, flow alteration due to post-slide erosion, or attribution of a maximum length to an entire connected complex comprised of multiple landslides with varying lengths. Inaccuracies were observed in calculated lengths and widths, each being lower than true dimensions due to the nature of the flow length tool, and to losing area where empty space between pixels did not completely align with polygon edges after conversion of features to raster format. However, this methodology should allow relatively high precision as each step is based on DEM computations or feature class measurements, provided consistent projection, classification, and tool parameters are used.

The maximum calculated length of landslides is 824 m, and, in one example of a landslide exceeding automated measured length, the runout distance from the most down-dip portion of the excavated face alone was 1 km long (see Figure 4). It presumably went farther, but was at that point

eroded by a river in the valley into which the deposits flowed. Lengths and widths are typically underestimated by automated geoprocessing methods. In the case of length, this occurs where the flow path of the measured flow length exits the landslide boundaries before reaching its lowest point due to flow paths resulting from deformation as the slide mass moved, or was subsequently eroded (Figure 16).

Performance of models also changed when varying included data layers or changing means of calculating weights. The Heuristic_2 model, using absolute value of the sum of weights, increased performance relative to the original Heuristic model, whereas performance of Heuristic_3 decreased using only the sum of weights. In the bivariate model, sign of the weight allows for a relative increase or decrease in susceptibility depending on a pixel in a layer, where a relative decrease indicates a stabilizing factor. Allowing negatives to influence order of bin weights in a layer for informed Heuristic models serves this purpose. However, allowing negative values in determining significance of layers relative to each other decreased performance, suggesting the magnitude of each layer's contribution is most important irrespective of sign in assigning a layer's percent contribution relative to other layers. Ultimately, improving the Heuristic method with statistical weights still underperformed the bivariate methods; being that this improvement to Heuristic models requires use of the bivariate method, it is better to use the bivariate method alone for assessment if a landslide inventory is available for calculating the statistical weights.

Plots were constructed using landslide dimensions calculated in ArcMap and may at present facilitate development of predictive length or runout scaling relationships relative to area in lieu of frequency-magnitude curves that require multi-temporal landslide data. In the future, scaling relationships combined with depth-to-volume calculations and frequency-magnitude curves derived from multi-temporal inventories (inventories that do not now exist) may further assist modeling and prediction of landslide hazard. If magnitudes of triggering events such as rainstorms and earthquakes can be similarly related, it may be possible to provide first responders with valuable real-time estimates

of risk. This may also be achieved despite the current affairs in Afghanistan preventing much by way of field work if high-resolution image catalogs can be obtained. These would likely need to be purchased, licensed, or donated for specific areas by commercial imaging companies, as those currently available for streaming in ESRI's basemaps lack temporal control. If available though, such images would allow for mapping landslides over time and for estimating recurrence intervals, and could help associate newly detected landslides with likely triggering events, such as rainfall or earthquakes. This association could help develop respective, predictive relationship curves between trigger and landslide magnitudes, or the number of landslides expected to be caused by a trigger of a given magnitude. Inventory products could be published with association to developing database projects, such as the LAND-deFeND database by Napolitano et al. (2018), or otherwise follow similar formatting.

A research thrust not implemented in this study would be to assess landslide risk to villages with respect to flow direction, flow path, or flow length projections from landslide sources rather than by using overlapping buffers for distances between landslides and villages. That is, a landslide will not flow backwards or uphill and threaten a village in the opposite direction except where inertia allows deposits to run up a slope, if a new landslide occurs underneath a structure, or the headscarp of an existing landslide reactivates and excavates further into the slope such that it compromises structure foundations previously uninfluenced. Accounting for landslide flow could be achieved by following similar methods as used to calculate landslide length in this study, but instead of masking the DEM by landslide boundaries, do so with a desired maximum runout length, and then assigning single pixels at the minimum and maximum elevations as starting spill or pour points for the hydrology tools. Smaller buffers could then be used around the resulting flow areas to approximate threat from individual slides to their immediate surroundings. The flow tools, however, function by following the lowest or easiest flow paths and would not directly reflect the rheology or inertial behavior of landslides that rigorous numerical models may represent. Hazard model adherence to approximate flow limitations may provide

a better assessment of landslide risk to villages and take less processing time than physical or numerical models that may also require material property analysis. Such on-site material analyses would be difficult or impossible to acquire for Afghan landslides in the current geopolitical environment.

6 Conclusion

Over 3400 landslides were used to model landslide susceptibility in Fayzabad District, Afghanistan. Further, extracted environmental characteristics show that landslides are primarily associated with elevations ranging between 1500 m and 2000 m, slopes of 27° to 45° and/or north to northwest aspects, or are within 60 m of a stream or river. Additionally, log-linear relationships of landslide dimensions suggest it is possible to develop predictive scaling relationships useful in estimating an area at risk for a landslide of a given magnitude. Susceptibility results show models can accurately depict areas around landslides as highly susceptible to failure, with the highest proportion identified as such being 83%. Many of the district's 720 settlements are at high risk to landslide hazards, with 58% of villages within 600 m of an existing landslide.

Model results by Kayastha et al. (2013), which utilized the bivariate statistical method, identified 53% of landslide pixels as being in high susceptibility zones, and the distribution of landslides across all susceptibility zones was nearly linear. This contrasts to results herein, where landslide identification within high susceptibility zones exceeded 80% in some trials, with an exponential distribution of landslides across susceptibility zones (Figure 11). The difference in model results between this study and those by Kayastha et al. (2013) may suggest models in this study are effectively over-trained and require application of smaller training datasets. However, the comprehensive study of peer reviewed landslide susceptibility modeling literature by Reichenbach et al. (2018) states that the bivariate method significantly increases in performance with larger datasets. That this study uses a much larger landslide inventory may explain differences in resulting model patterns. Further, statistical model results herein

widely agree with inventory analysis and observations by other researchers regarding landslides in northeastern Afghanistan (Naseri & Kang, 2016, Shroder et al., 2011; Zhang et al., 2015).

Several models were constructed, with varied distribution of landslides in respective susceptibility zones. Models can be chosen for different purposes based on these landslide-index distributions, depending on levels of acceptable risk set by policy makers in disaster avoidance, mitigation, and planning efforts. Further, a methodological and programmatic framework has been setup to facilitate both reproduction of data and results, as well as easily produce new model variations. Additionally, a programmatic framework of the methodology aids in applying the bivariate model to other areas, where location-specific statistical weights are calculated with respect to location-specific input data. Although existing models for Afghanistan (Naseri et al., 2016; Schlagel, 2015; Zhang et al., 2015) are pushing the limits of freely accessible data, more data exist in the form of commercial imagery that could be applied. Doing so would allow further development without requiring expensive repeat LiDAR flights to acquire high-resolution, multi-temporal data over large areas.

References

Booth, A. M., Roering, J. J., & Perron, J. T. (2009). Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon. *Geomorphology*, *109*(3), 132-147.

Boyd, O.S., Mueller, C.S., and Rukstales, K.S. (2007). Preliminary Earthquake Hazard Map of Afghanistan: U.S. Geological Survey Open-File Report 2007-1137, 25 p.

Buckley, A., 2010. Understanding curvature rasters. ArcGIS Blogs. ESRI Mapping Center.

Claessens, L., Heuvelink, G. B. M., Schoorl, J. M., & Veldkamp, A. (2005). DEM resolution effects on shallow landslide hazard and soil redistribution modelling. *Earth Surface Processes and Landforms*, *30*(4), 461-477.

Cooley, S. (2016). GIS 4 Geomorphology. <gis4geomorphology.com>

Effat, H. A., & Hegazy, M. N. (2014). Mapping landslide susceptibility using satellite data and spatial multicriteria evaluation: the case of Helwan District, Cairo. *Applied Geomatics*, *6*(4), 215-228.

HRTE (High Resolution Terrain Elevation Data); Project Cedar and Rampant Lion II. (2008). Department of Defense, Army Geospatial Center. DVD format.

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.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), 42-66.

Kamp, U., Owen, L. A., Growley, B. J., & Khattak, G. A. (2010). Back analysis of landslide susceptibility zonation mapping for the 2005 Kashmir earthquake: an assessment of the reliability of susceptibility zoning maps. *Natural hazards*, *54*(1), 1-25.

Kayastha, P., Dhital, M. R., & De Smedt, F. (2013). Evaluation and comparison of GIS based landslide susceptibility mapping procedures in Kulekhani watershed, Nepal. *Journal of the Geological Society of India*, *81*(2), 219-231.

McKean, J., & Roering, J. (2004). Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology*, *57*(3), 331-351.

MRl, H. M. (2003). Multi-hazard loss estimation methodology: Earthquake model. *Department of Homeland Security, FEMA, Washington, DC*.

Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C., & Jaedicke, C. (2006). Global landslide and avalanche hotspots. *Landslides*, *3*(2), 159-173.

Napolitano, E., Marchesini, I., Salvati, P., Donnini, M., Bianchi, C., & Guzzetti, F. (2018). LAND-deFeND–An innovative database structure for landslides and floods and their consequences. *Journal of environmental management*, *207*, 203-218.

Naseri, M. K., & Kang, D. (2016). Statistical Landslide Susceptibility Mapping by Using GIS-based Weight-of-Evidence (WofE) Analysis in Takhar Province of Afghanistan. *International Journal of Engineering and Technical Research (IJETR)*, 2321-0869.

OCHA (2017). Afghanistan Natural Disaster Incident Reports [2012 through 2017]. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Accessed at: <https://data.humdata.org/search?q=natural+disaster+incident&ext_search_source=main-nav>

Pardeshi, S. D., Autade, S. E., & Pardeshi, S. S. (2013). Landslide hazard assessment: recent trends and techniques. *SpringerPlus*, *2*(1), 523.

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, *40*(10), 927-930.

Pradhan, A. M. S., & Kim, Y. T. (2016). Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *Catena*, *140*, 125-139.

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

Ruleman, C. A., Crone, A. J., Machette, M. N., Haller, K. M., & Rukstales, K. S. (2007). Map and database of probable and possible Quaternary faults in Afghanistan (No. 2007-1103). Geological Survey (US).

Schlagel, N.A. (2015). Development of a Preliminary Landslide Susceptibility Model for Afghanistan using Multi-criteria Evaluation. Bachelor of Science Thesis. University of Nebraska – Omaha.

Schulz, W. H. (2004). Landslides mapped using LIDAR imagery, Seattle, Washington. *US Geological Survey Open-File Report*, *1396*(11).

Schulz, W. H. (2005). Landslide susceptibility estimated from mapping using light detection and ranging (LiDAR) imagery and historical landslide records, Seattle, Washington. *US Geological Survey Open-File Report*, *1405*, 144.

Schulz, W. H. (2007). Landslide susceptibility revealed by LIDAR imagery and historical records, Seattle, Washington. *Engineering Geology*, *89*(1-2), 67-87.

Schuster, R. L., & Krizek, R. (1978). Landslides: analysis and control: Transportation Research Board Special Report 176. National Research Council, Washington D.C.

Shroder, J.F. (1968). Landslides of Utah.

Shroder Jr, J. F. (1989). Slope failure: extent and economic significance in Afghanistan and Pakistan. *Landslides: extent and economic significance in the world*, 325-341.

Shroder, J. F. (2014). *Landslide hazards, risks, and disasters*. Academic Press.

Shroder, J. F., Schettler, M. J., & Weihs, B. J. (2011a). Loess failure in northeast Afghanistan. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(16), 1287-1293.

Shroder, J. F., Weihs, B. J., & Schettler, M. J. (2011b). Mass movement in northeast Afghanistan. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(16), 1267-1286.

Sörensen, R., Zinko, U., & Seibert, J., (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences Discussions, 10(1), 101-112.

Steinshouer, D.W. et al. (2006). "Petroleum Resource Potential GIS of Northern Afghanistan". US Department of the Interior, and US Geological Survey. Open-File Report 2006-1179. <http://pubs.usgs.gov/of/2006/1179/>.

Thompson, A. L., & Hubbard, B. E. (2014). *A comprehensive population dataset for Afghanistan constructed using GIS-based dasymetric mapping methods* (No. 2013-5238). US Geological Survey.

Turner, A. K., & Schuster, R. L. (1996). Landslides: Investigation and Mitigation: Transportation Research Board Special Report 247. National Research Council, Washington, DC.

USGS. (2018). LandSAT8. Earth Explorer. <http://earthexplorer.usgs.gov/>.

Vergano, D. (2014). Mudslide buries more than 350 in Afghan village. National Geographic. Accessed 3/16/2018. <https://news.nationalgeographic.com/news/2014/05/140502-afghanistan-landslide-deaths-united-nations/>

Wilson, R. and Keefer, DK. (1985). Predicting Areal Limit of Earthquake-Induced Landsliding, Evaluating Eathquake Hazards in the Los Angeles Region-An Earth-Science Perspective. *US Geological Survery Professional Paper 1360*, 317-345.

Zhang, J., Gurung, D. R., Liu, R., Murthy, M. S. R., & Su, F. (2015). Abe Barek landslide and landslide susceptibility assessment in Badakhshan Province, Afghanistan. *Landslides*, *12*(3), 597-609.

Equations

Equation 1

$$TWI = Ln[(FlowAccumulation + 1) * cellsize / \tan(Slope * \left(\frac{pi}{2}\right) / 90]$$

Equation 2

$$NDVI = (NIR - RED)/(NIR + RED)$$

Equation 3

$$Ac = mM * Slope + bM$$

Equation 4

$$Statistical\ weight = \ln(\frac{LS\ Area\ in\ class}{Class\ Area} * \frac{Map\ Area}{LS\ Area\ in\ map})$$

Equation 5

$$Hazard = LSI + VillageBuffer + \left(\frac{LSBuffer}{2}\right) + (\frac{LSDensity}{2})$$

Tables

Table 1 Casualties and damages for disasters in Afghanistan by type. Numbers are in percent of total for each column.

	DEAD	INJURED	AFF_IND	AFF_FAM	HOUSE_DAMAGE	HOUSE_DESTROY
Avalanche	20.97	14.26	1.64	1.74	1.13	1.22
Earthquake	9.36	36.21	20.01	18.84	23.43	19.52
Extreme Weather	2.57	1.56	3.53	3.45	0.95	0.59
Flood/Flash Flood	45.51	42.01	67.59	68.42	68.19	71.71
Heavy Rainfall	2.39	2.06	2.58	2.57	2.44	1.98
Heavy Snowfall	2.39	2.67	2.34	2.31	2.90	1.70
Landslide/Mudflow	16.81	1.23	2.32	2.68	0.96	3.28

Table 2 Model parameters. Inputs only display variables not common to all models: Elevation, Aspect, Slope, TWI, curvature, plan curvature, profile curvature, densities and buffers of streams and faults.

ModelID	ModType	Method	inputs
Heuristic_1	Heuristic		
StatMod_1	Bivariate		
StatMod_2	Bivariate		critical acceleration
StatMod_3	Bivariate		critical acceleration, ndvi15
StatMod_4	Bivariate		critical acceleration, ndvi30
Heuristic_2	i-Heuristic	Abs(Sum())	critical acceleration
Heuristic_3	i-Heuristic	Sum()	critical acceleration
FayzStatMod_1/2	Bivariate		critical acceleration
FayzMod1	Bivariate		
FayzMod2	Bivariate		

Table 3 Basic statistics of landslides in the watershed areas of the district.

	min	max	mean	median
L:W ratio	0.36	57.39	6.09	4.15
Length	5.71	824.23	133.46	91.49
Area	58.68	212359.25	6573.71	1643.34
Width	2.54	360.13	30.36	19.30
Relief	1.79	446.85	71.27	51.54

Table 4 Agreed area between models. The bottom values in Green represent perfect agreement, or cells with the same value; the top, Blue values represent an allowance of +/- 1 LSI value in model outputs.

% Agreed Area Between Models

Perfect only (bottom); +/-1 LSI class tolerance (top)

ModID	Heuristic_1	StatMod_1	StatMod_2	StatMod_3	StatMod_4	Heuristic_2	Heuristic_3
Heuristic_1	--------	89.79	88.51	89.40	89.59	95.57	95.12
StatMod_1	37.21	--------	97.54	97.26	96.96	95.66	84.24
StatMod_2	37.30	82.20	--------	99.48	98.63	98.50	87.86
StatMod_3	38.92	70.53	81.27	--------	98.76	98.07	88.74
StatMod_4	38.44	68.83	75.70	70.24	--------	97.21	87.27
Heuristic_2	50.10	51.23	57.08	56.54	53.31	--------	98.42
Heuristic_3	50.11	30.62	32.85	34.38	34.01	52.91	--------

+/- 1 LSI tolerance

Percect Agreement only

Figures

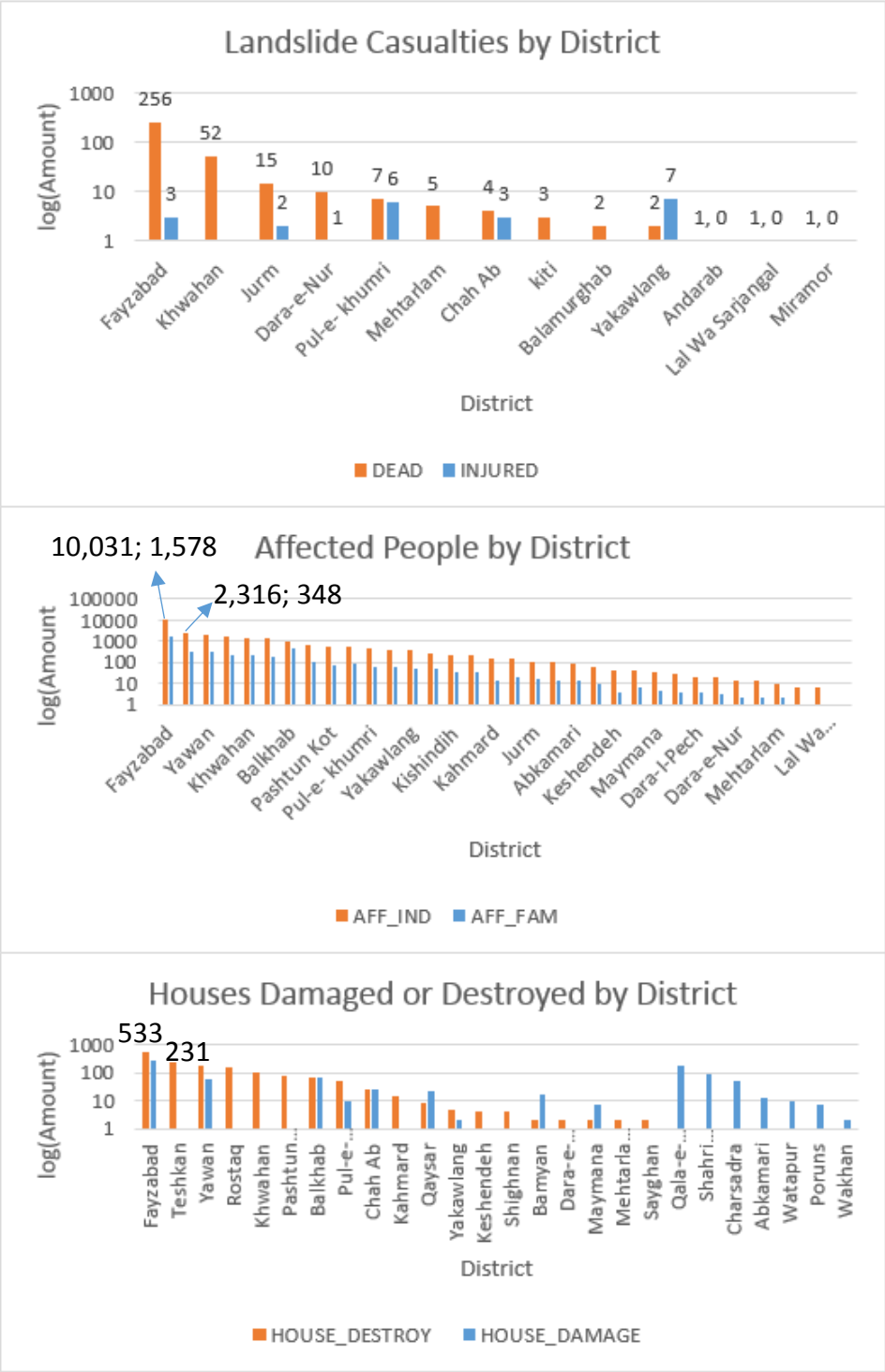

Figure 1 Landslide casualties (top), number of affected (middle), and damage to residence (bottom). At top, labels are to respective underlying columns, at middle labels are for both columns in the top two districts, at bottom labels are to destroyed homes (orange) in top two districts.

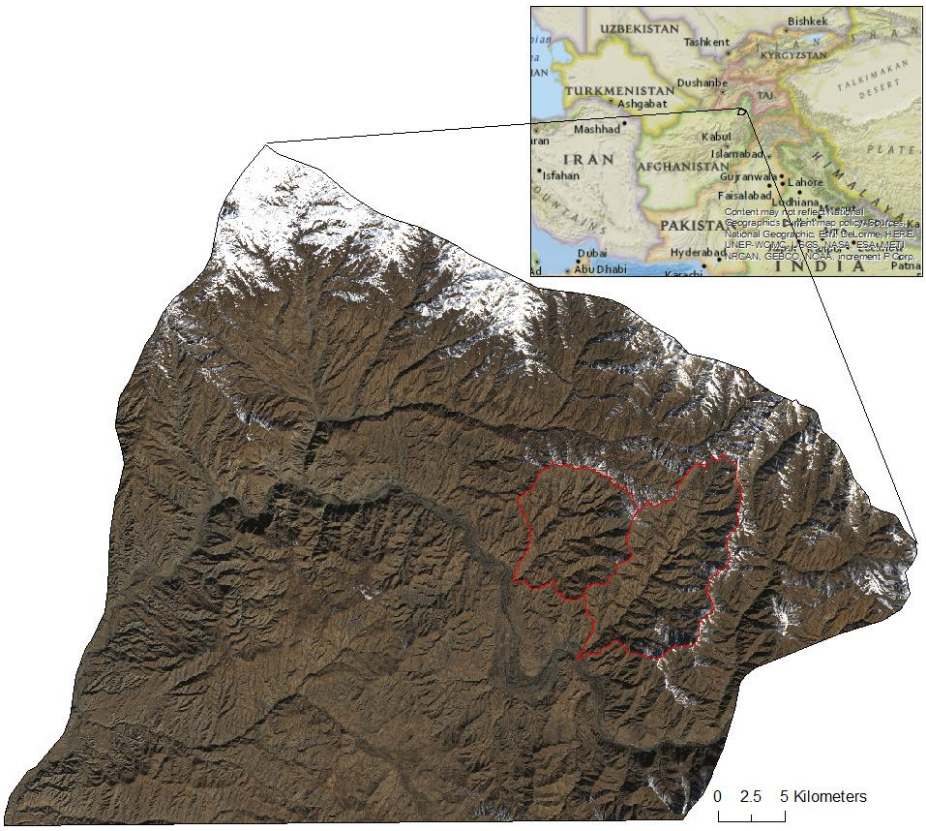

Figure 2 Fayzabad study area with regional map inset. Watershed areas used to develop model outlined in red. Note, the southern third or so of the district is cut off due to an absence of DEM data and is not modeled herein.

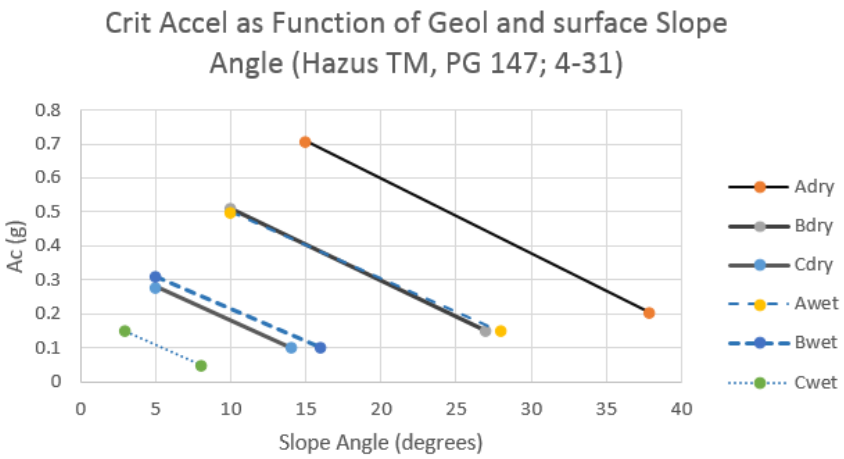

Figure 3 Critical Acceleration data used to determine material specific Ac for given ground slope angles in the study area. Material A is crystalline and strongly cemented rock; b is weakly cemented rock and sandy soil; c is shale, clayey soil, and existing landslides (MSI, 2003).

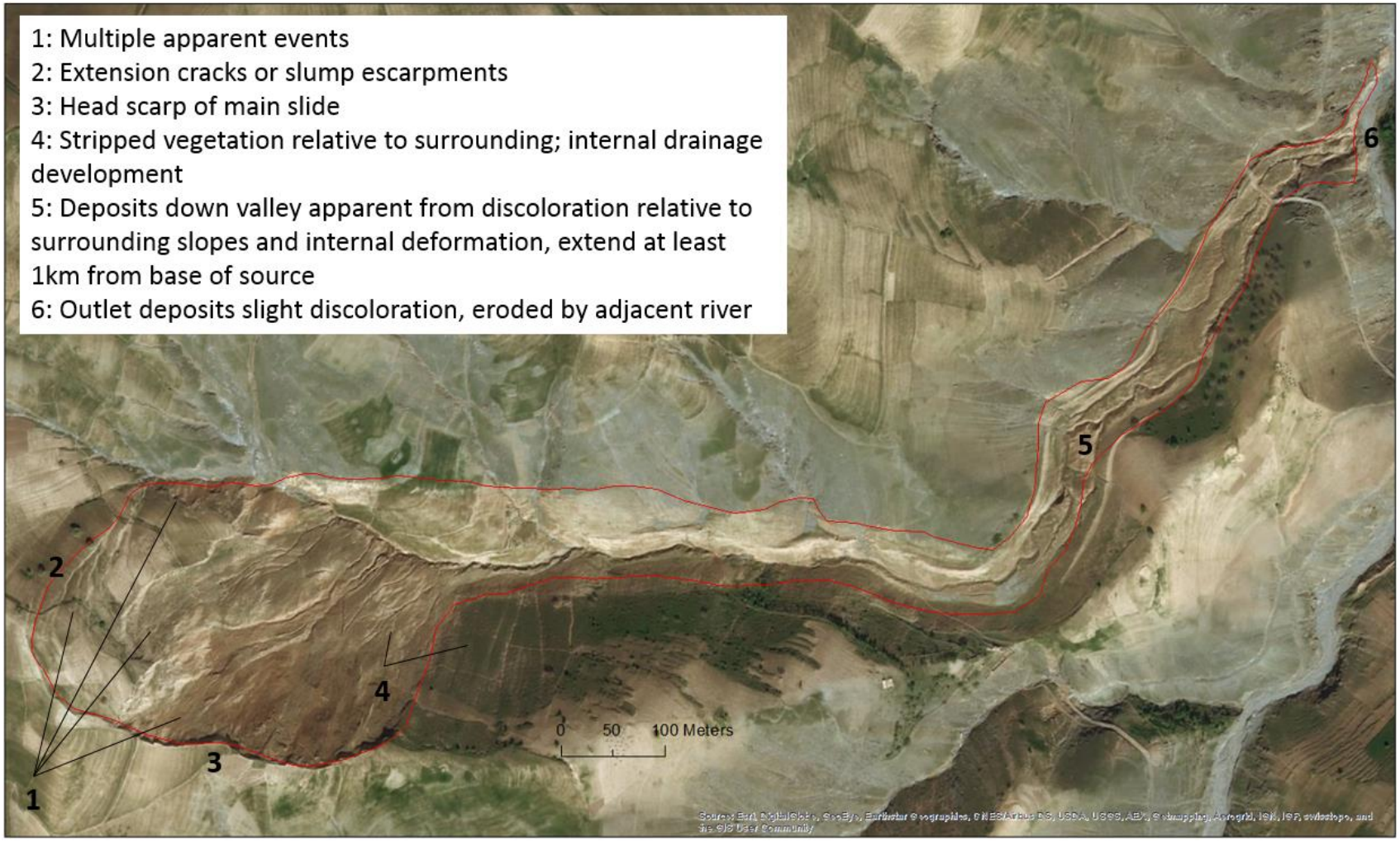

Figure 4 Criteria for landslide identification and mapping.

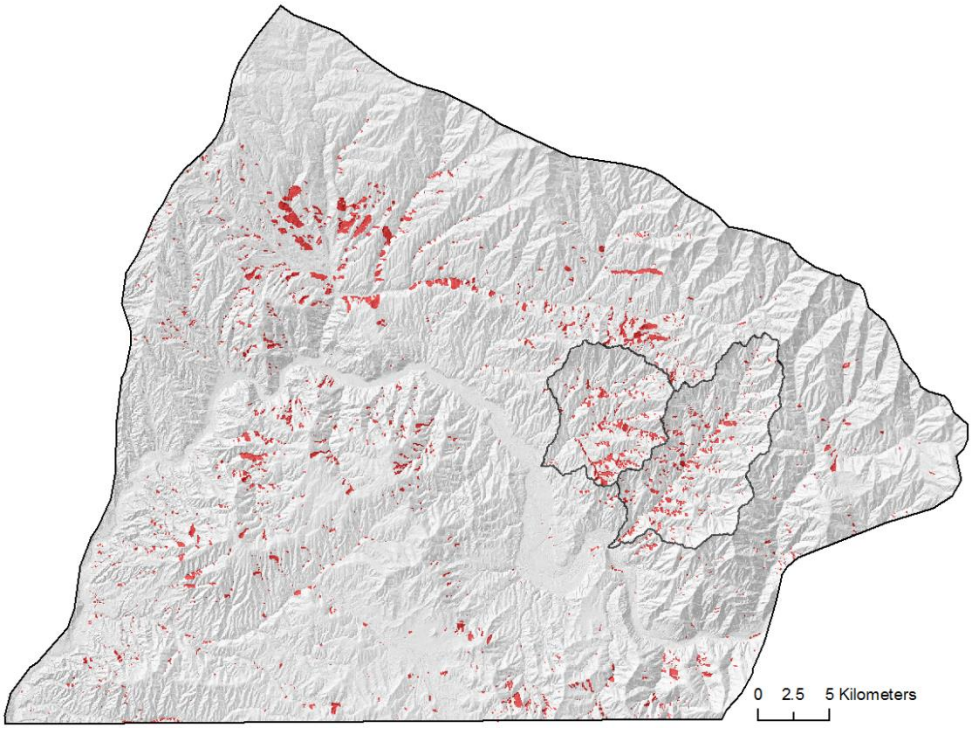

Figure 5 Landslide inventory map for Fayzabad District, Afghanistan. Watershed area where model was developed outlined in black for reference.

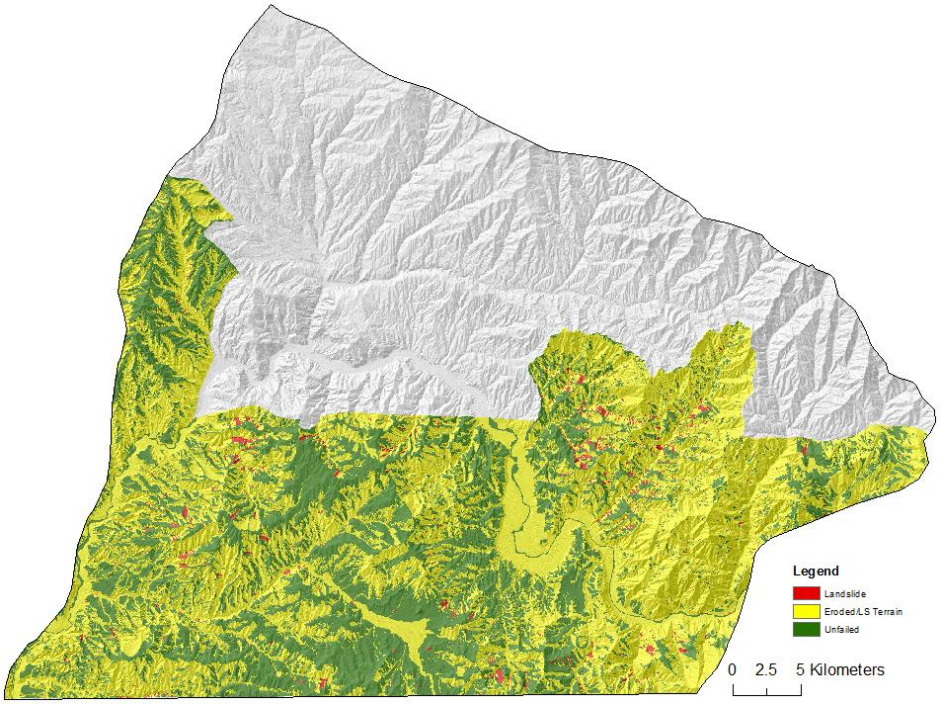

Figure 6 Surface classification map with three categories. Mapping was completed for the southern portion of the study area only. Landslides falling outside of this area have been removed from this map.

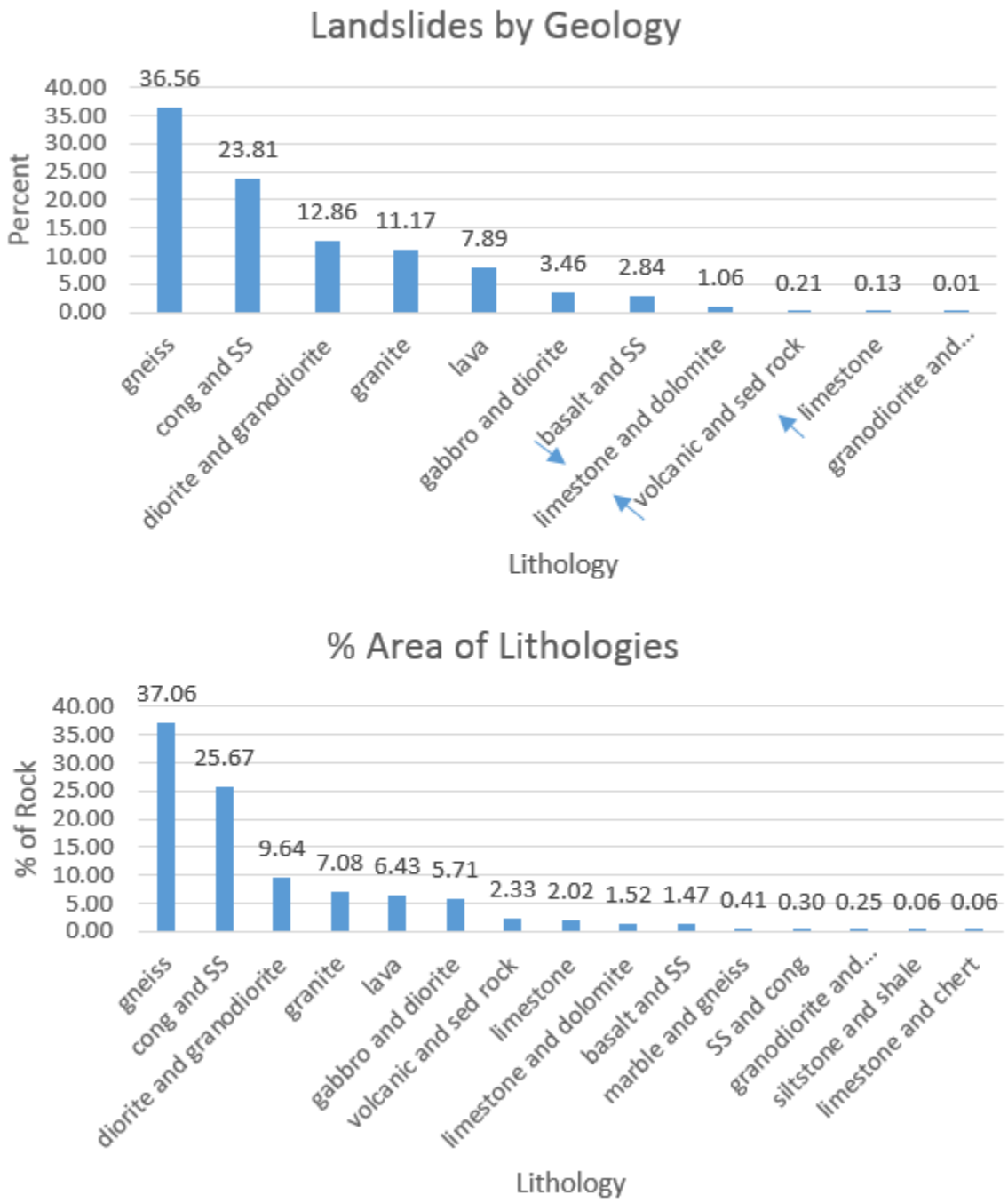

Figure 7 Distribution of landslide by lithology in terms of percent (top), and proportion of lithologies relative to overall geology (bottom). Landslide distributions roughly follow trend in dominant lithology. A few lithologies break trend and have far fewer landslides (%) than their total area, namely limestone, and volcanic/sedimentary rock. Blue arrows show changes in order of lithologies between percent of contained landslides and percent of lithology relative to regional geologic units.

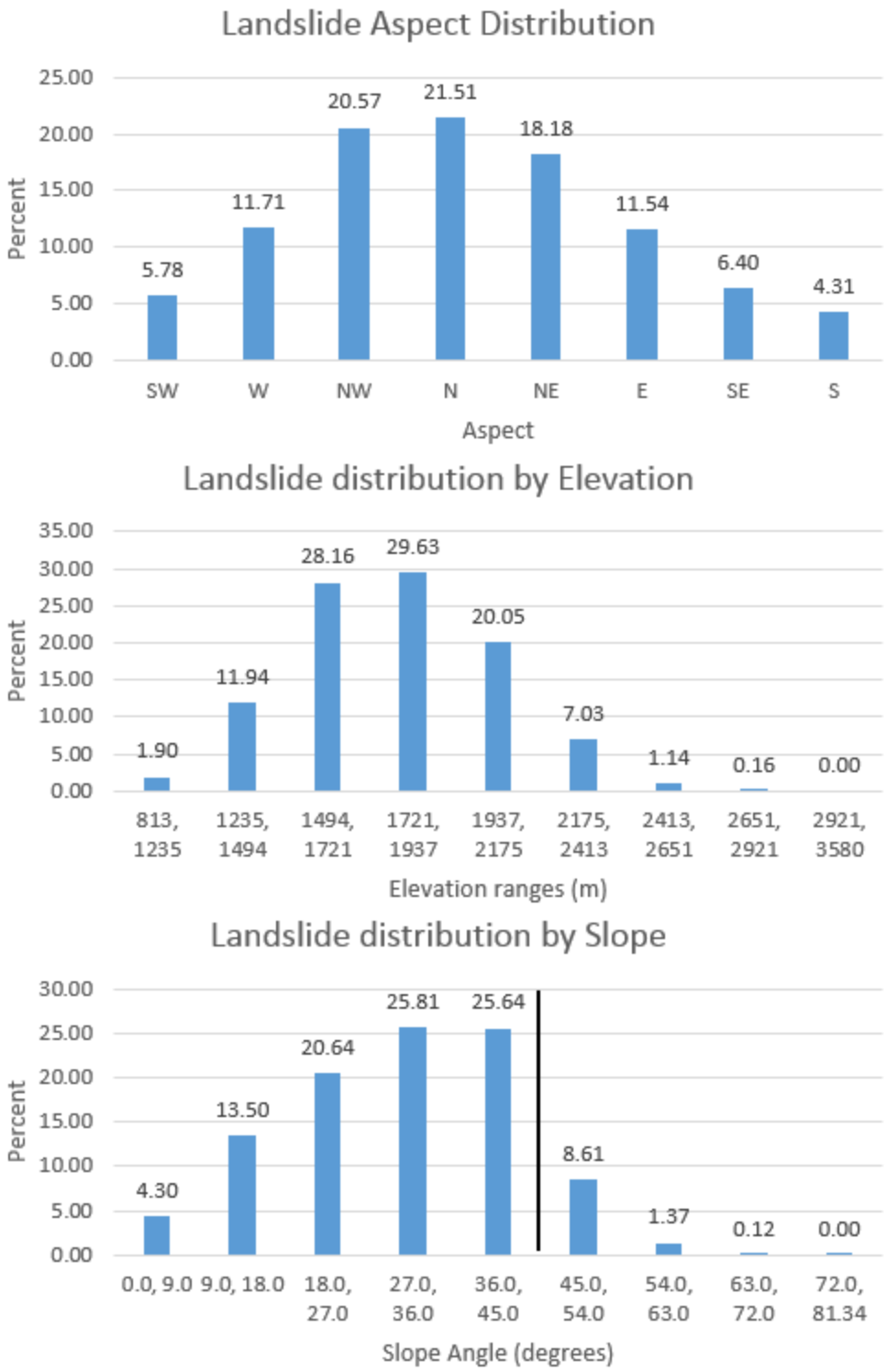

Figure 8 Distribution of landslides by key DEM derived data; aspect (top), elevation (middle), and slope (bottom). Dominant slope face directions correspond with observations by Shroder et al. (2011a), and elevation ranges are similar to those noted by Kamp et al. (2010). Sharp drop in landslides over 45 degree slopes is likely reflects common angle of repose.

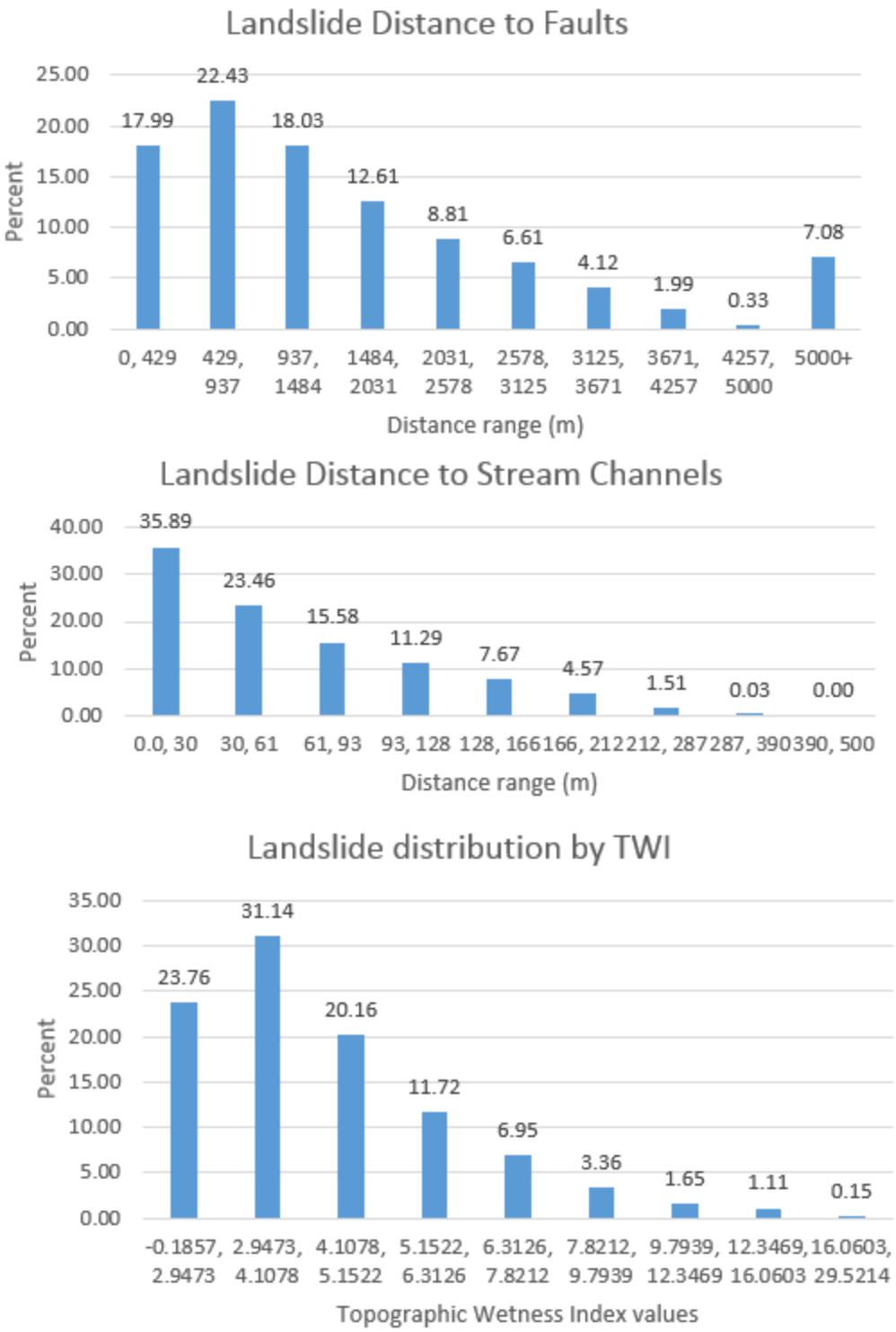

Figure 9 Landslide pixel proximity to faults (top), streams (middle), and distribution by TWI (bottom). Landslides occur primarily near faults or streams. Association with dry TWI values, however, seems to contradict this, but suggests that landslides occur near streams, but not within channels, or areas water accumulates, but rather on drainage slopes above streams. Comparison to topographic position indices (TPI) was not done quantitatively, but qualitatively observed to support this assertion.

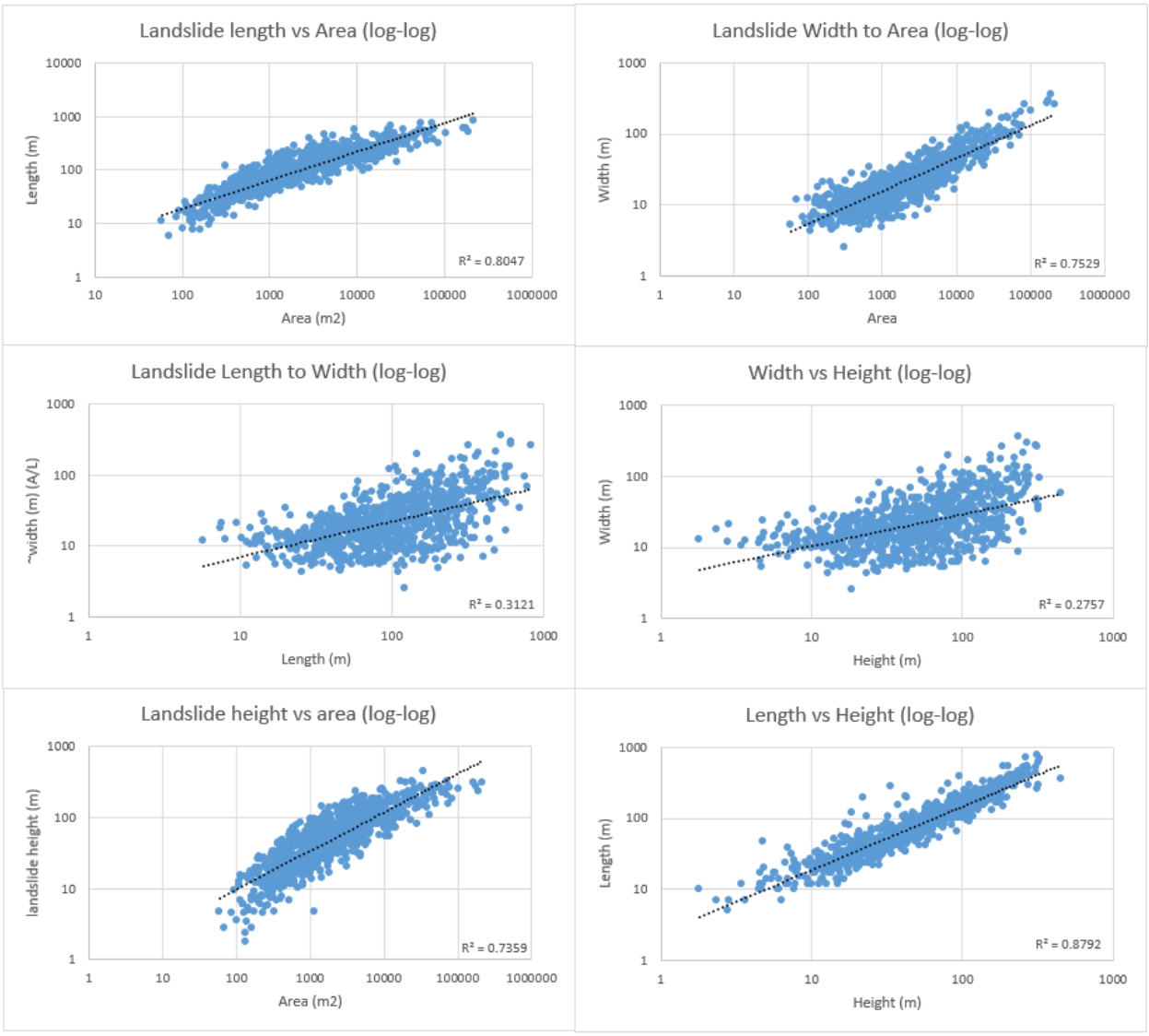

Figure 10 Landslide characteristics. Length was estimated by calculations in Flow length tools, area determined by polygon attributes, width estimated by A/L, and height by the difference of maximum and minimum elevation within landslide boundaries. Each plot is on Log-log scale, with a Power fit and R^2.

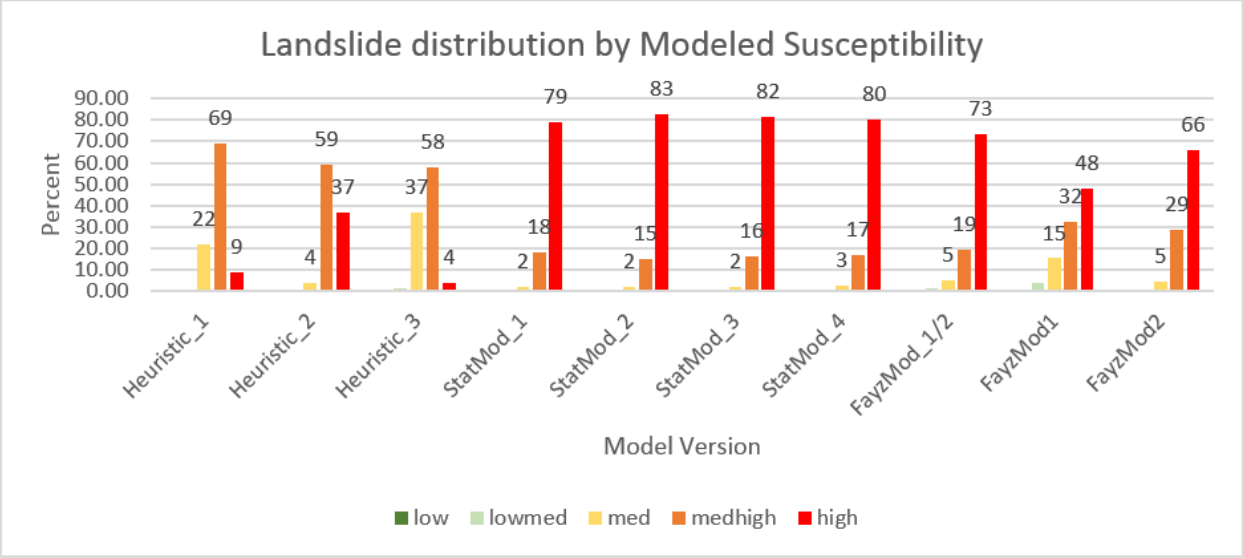

Figure 11 Distribution of landslide pixels by percent among susceptibility classes for each model.

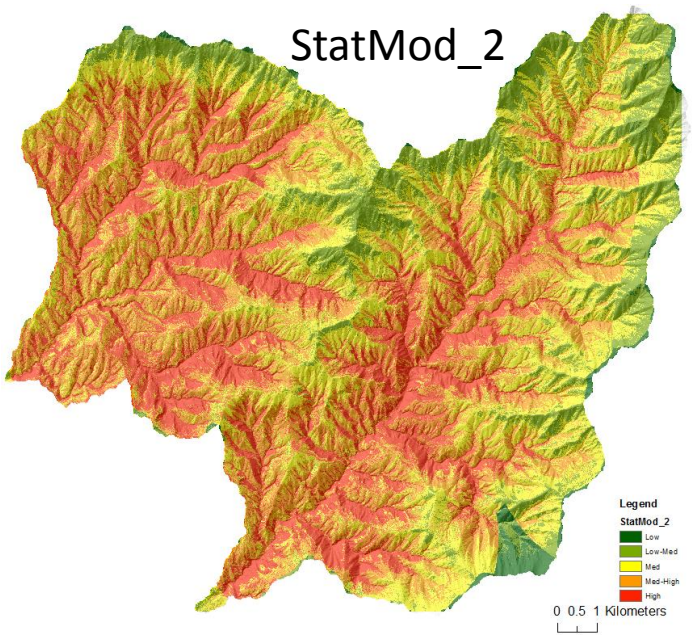

Figure 12 Watershed-scale landslide susceptibility model. StatMod_2, shown above, was the test version with greatest proportion of landslides falling in the high susceptibility category, and thus deemed to have the best performance. The parameters of this model were then chosen to scale up to cover larger portions of the district (see Figures 13 and 14).

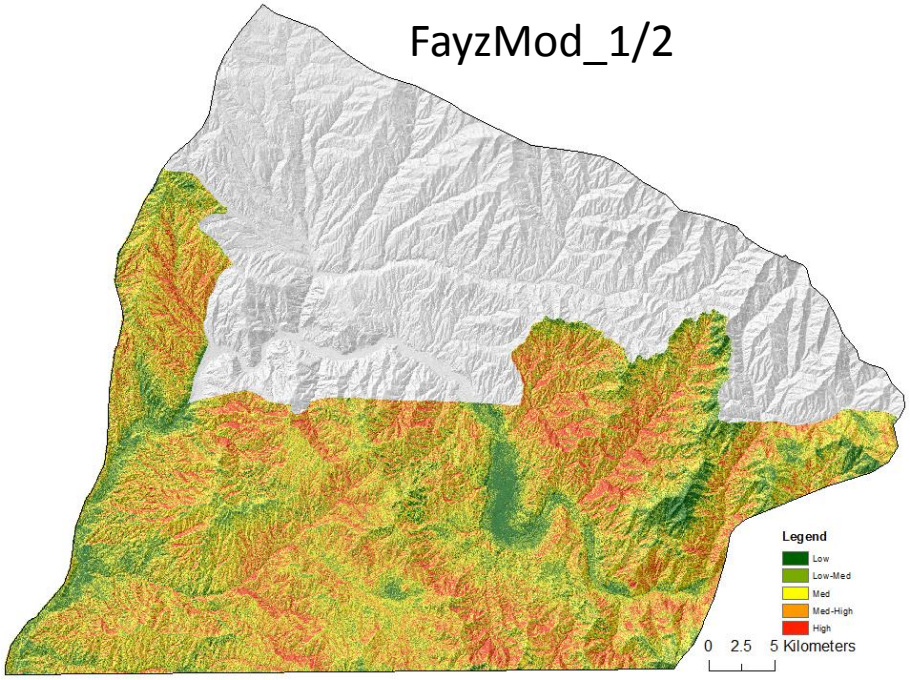

Figure 13 Landslide susceptibility model for the half of the district where surface class mapping was completed. The significance of this being the inclusion of critical acceleration data dependent on that surface class map. Of landslides in this area 73% are in areas modeled at high susceptibility, and 19% in med-high.

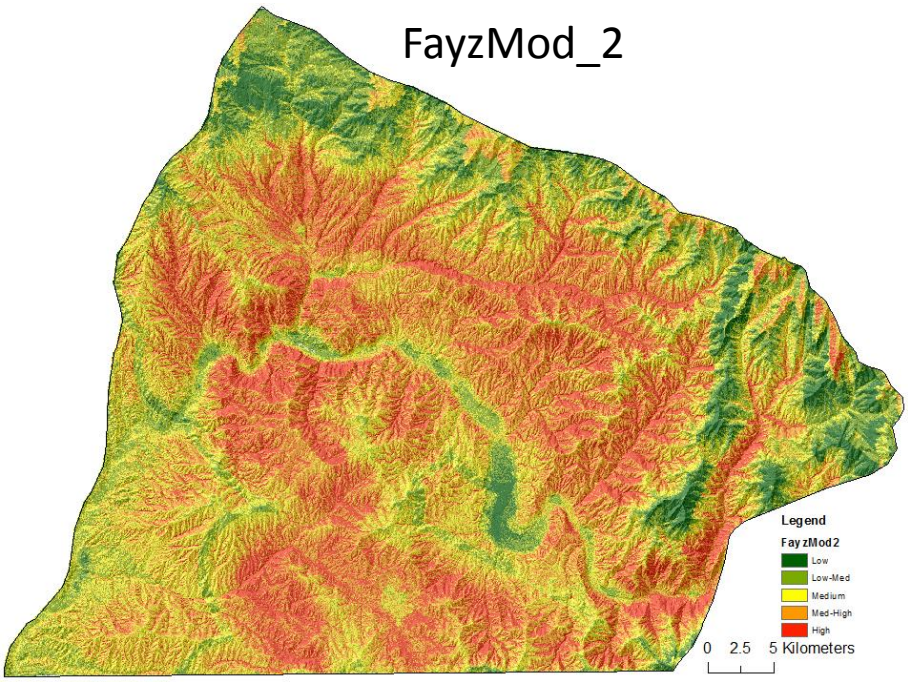

Figure 14 Final susceptibility model for Fayzabad District. The high susceptibility category contains 66% of landslide pixels, and the medium-high contains 29%. Critical acceleration data was excluded from this model at the full district scale as the requisite surface class map did not cover the northern portion (see colored extent in Figure 13).

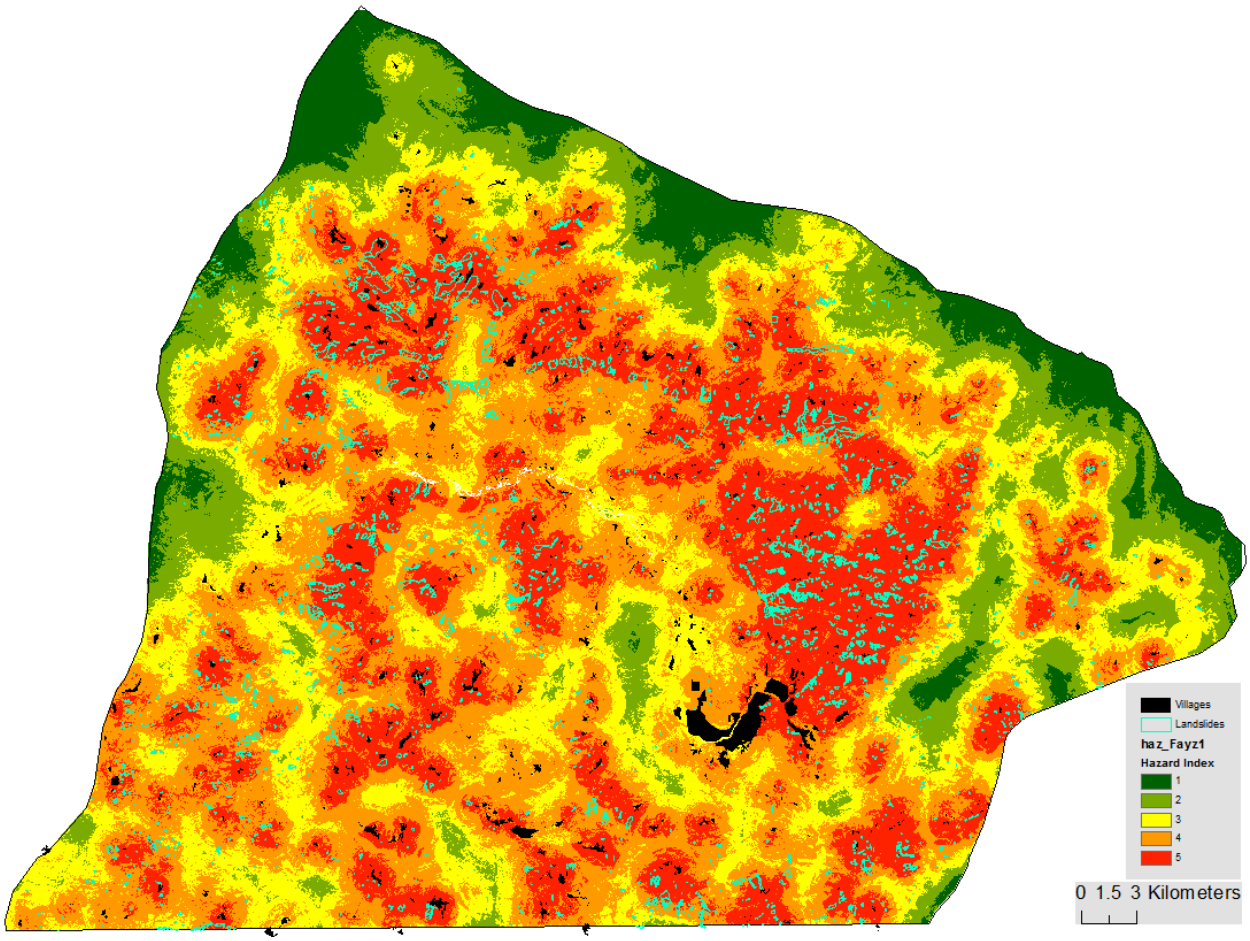

Figure 15 Hazard Index for Fayzabad District utilizing the final tested method.

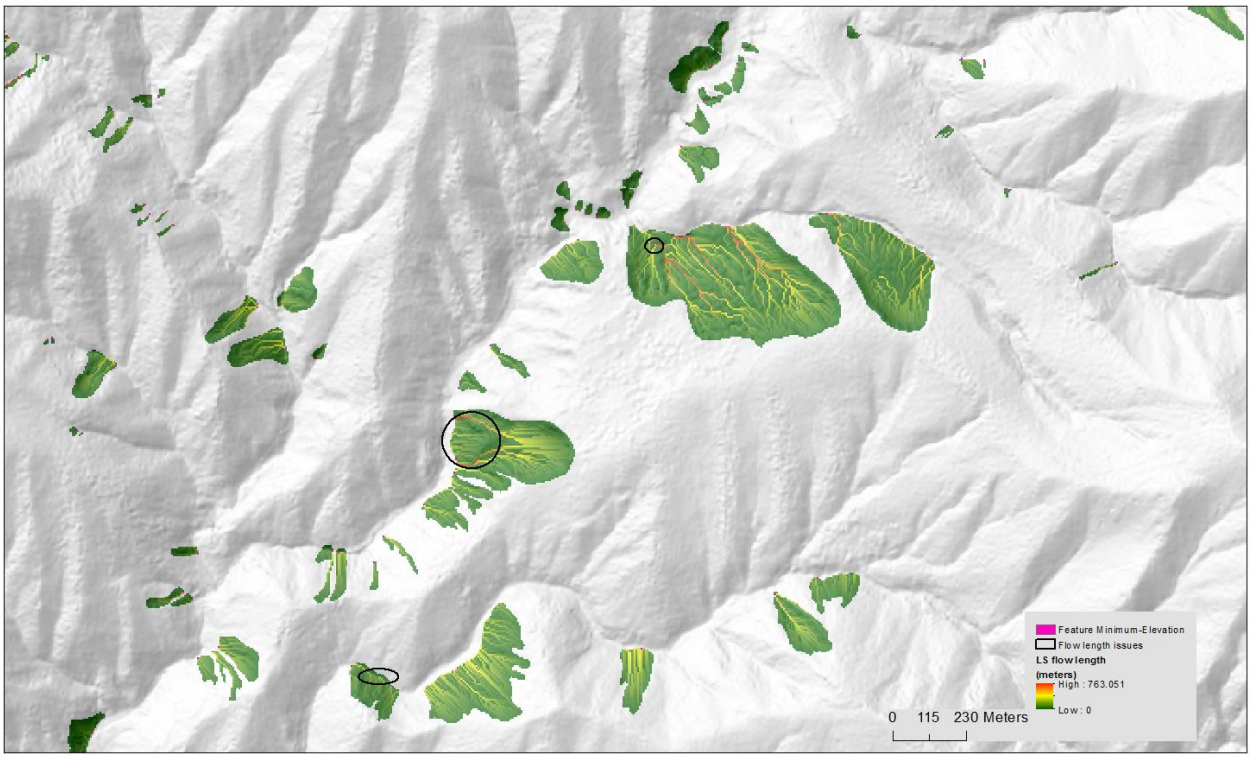

Figure 16 Length issues. Black circles show potential issues in flow length calculations, where hydrologic calculations show termination of flow within a slide mass (top circle), branches that diverge from the central axis of the landslide (middle circle), or where flow follows the bottom of valleys where deliniation extends (bottom circle). These issues are suspected to in general result in underestimation of landslide lengths.

Python Code

```
#Purpose of code to automate TabulateArea function, and calculation of associated
percent area of landslides in
#bins of an input raster (i.e. percent of mapwide landslide pixels in elevation bins 1
through 9), and
#calculation of associated statistical weights for each bin for reclassification for
use in stats wgt LSI model

import arcpy
from arcpy.sa import *
import numpy as np

arcpy.env.overwriteOutput=True
arcpy.env.workspace =
r"D:\Users\n913s392\Desktop\MasterFolder_desktop\reclass_tables.gdb"
arcpy.env.scratchWorkspace = r"D:\Users\n913s392\Desktop\MasterFolder_desktop\scratch"

arcpy.CheckOutExtension("Spatial")

################################################################################
################################
#Inputs and Variables
################################################################################
################################
#eventually make inputs get as text parameters
table_tag = "tabA_stFayz" #for naming output dynamically, tagging this to front of
source file name

in_rasters = ["rc_aspect2",
              #"rc_ac4_test",    #wip, natural breaks class with 1-9 rc vals
              "rc_curvature2",
              "rc_d2fault_2",
              "rc_d2stream",
              "rc_elevation",
              "rc_flt_lden2",
              "rc_geoFayz",     #wip, lithologies not merged, used Steinshouer's map
in area as is, rc vals in excel
              #"rc_ndvi15m",
              #"rc_ndvi30m",
              "rc_plan_2",
              "rc_profile_2",
              "rc_slope",
              "rc_strm_lden2",
              "rc_twi"]

landslide_map = "LSmap_boolean"
mask_rast = r"Mod_area_mask"
#for landslide_map, use map with SRC1 (landslides), src3 (eroded), and
src4(loess/soil)
#   landslide inventory in raster form required for running. Indexing assumes three
surfaces classes in the map;
#   landslides, erroded/stripped/landslide terrains, and unfailed/loess/soil.

out_raster_tag = "stF_" #for naming output dynamically, tagging this to front of
source file name
#out_raster = "test_RCstatwt" #has an overwrite below for full file name
"stat_rc_....."


in_fields = ["VALUE", "VALUE_0", "VALUE_1"] #VALUE_1 (index 1), VALUE_3 (index 2), and
VALUE_4 (index 3)
new_fields = ["classA_m2", "mapPerc_0", "mapPerc_1",
```

```
              "classPerc_0", "classPerc_1", "StatWgt"] # is indexable; new_fields[n]
statwt_field = "StatWgt"
LS_field = "VALUE_1"
Area_field = "classA_m2"

model_output = r"FayzMod2"

################################################################################################
##################################
"""Functions"""
################################################################################################
##################################

#split file patch and concatinate table name
def getLayerName(layer):
    desc = arcpy.Describe(layer)
    layer_name = desc.baseName; del desc
    return layer_name

#tabulate area table for layer
def myTabulateArea (in_zone_data, in_class_data, out_table):
    TabulateArea(in_zone_data, "VALUE", in_class_data, "VALUE", out_table)

#calc_ClassArea function needs converted to MODULAR follow str() convertion and
concatenation from calc_PercentSRC()
def calc_ClassArea(table,field_name):# ...(table, field, expression)
    expression = '!VALUE_0!+!VALUE_1!' #hard code, needs to be made modular
    arcpy.CalculateField_management(table, field_name, expression, "PYTHON")
    #now works

#sum area of a "Value_n" column in table - used within calc_percent()
def sum_area(table, in_field):
    np = arcpy.da.TableToNumPyArray(table, in_field)
    area = np[in_field].sum()
    del np
    return area
    #now works

#ex: calc %area of LS in each class of the variable (i.e. elevation)
def calc_PercentSrcTypesInMap(table, in_fields, f, new_fields, n):
    totalarea = sum_area(table, in_fields[f])
    expression = str("!"+in_fields[f]+"!/"+ str(totalarea) +"*100")
    arcpy.CalculateField_management(table, new_fields[n],expression, "PYTHON")
    #now works

################################################################################################
##################################
def calc_PercentSrcByVariableClass(table, in_fields, f, new_fields,n):
    expression = str("!"+ str(in_fields[f])+"!/!"+str(new_fields[0])+"!*100")
    arcpy.CalculateField_management(table, new_fields[n], expression, "PYTHON")
    # still wip, one field not calculating
################################################################################################
##################################

#calculate statwt column of table
def calc_StatWeight(table, in_fields, ls_field, class_area_field, new_fields, n):
    codeblock = "import numpy as np"
    total_area = sum_area(table, new_fields[class_area_field])
    ls_area = sum_area(table, in_fields[ls_field])
    expression = str("np.log((!"+ str(in_fields[ls_field])+"!/"+
"!"+str(new_fields[class_area_field])+"!)"+ \
                "*"+  str(total_area/ls_area) + ")")
    arcpy.CalculateField_management(table,
```

```
new_fields[n],expression,"PYTHON",codeblock)
    #now works

def remove_infinity(in_array, field):
    n = 0
    for i in in_array[field]:
        if i > 1000000:
            in_array[field][n] = 0
        elif i < -1000000:
            in_array[field][n] = 0
        n += 1
    return np

#apply reclass for statwt raster to use in model
def doRc_statwt(in_raster, varclass, statwt, out_raster):
    n=0
    for x in varclass:
        if n==0:
            out_temp = Con(in_raster == x, statwt[n], in_raster)
        else:
            out_temp = Con(in_raster == x, statwt[n], out_temp)
        n+=1
    arcpy.env.overwriteOutput = True
    out_temp.save(out_raster)

#place function for adding reclassed StatWt layers together (the statwt model)
def StatWtModel (in_stat_wt_rasters, file_output): #in_stat_wt_rasters, mask_rast,
file_output    mask_rast removed
    #model = mask_rast
    counter = 0
    for i in in_stat_wt_rasters:
        i_rast = Raster(i)
        if counter == 0:
            model = i_rast
        if counter >0:
            model += i_rast
        counter+=1
    model.save(file_output)

################################################################################
################################
#Run Code
"""Call functions to perform opperations"""
################################################################################
################################
in_stat_wt_rasters = []
for layer in in_rasters: #layer as path
    layer_name = getLayerName(layer)
    out_raster = str(out_raster_tag + layer_name)
    table = str(table_tag + layer_name[2:]) # if following convention "rc_...." where
.... is the data name (ex: elev, slope)
    in_raster = Raster(layer)
    myTabulateArea(in_raster, landslide_map, table)
    print layer_name

    n=0 #for tracking position in new_fields list during loops
    f=0 #for tracking position in source src fields VALUE_1 (index 1), VALUE_3 (index
2), and VALUE_4 (index 3)
    #Run add fields and tabulate areas
    for name in new_fields:
        arcpy.AddField_management(table, name, field_type = "DOUBLE")
        if n==0:#Calculate total area in variable class
            calc_ClassArea(table,name)
```

```
        if n>0 and n<3:#calculate percent of single Src type throughout map (i.e. all
variable classes)
            calc_PercentSrcTypesInMap(table, in_fields, f, new_fields,n)

        if n>2 and n<5: #calculate percent of each SRC within a single variable class
            calc_PercentSrcByVariableClass(table, in_fields, f, new_fields,n)

        if n==5: #calc stat weight - table, in_fields, ls_field (as index),
class_area_field (as index), new_fields, n
            ls_field = 2
            class_area_field = 0
            calc_StatWeight(table, in_fields, ls_field, class_area_field,
new_fields,n)

        n += 1
        f += 1
        if f == 3:  # reset f to index beginning of in_fields list to calculate n>4
and n<7 next.
            f = 1
        #print "n = "+str(n)+", f = "+str(f)


################################################################################
#############
    # Get stat wt values, and variable classes to lists and reclass via RasterCalc con
statements
    del table; table = str(table_tag + layer_name[2:]); input = in_fields+new_fields

    np = arcpy.da.TableToNumPyArray(table, input)
    np = remove_infinity(np,statwt_field)
    print np[statwt_field]

    varclass = []
    statwt = []

    for i in np:
        varclass.append(i[0])  # can also index as i["VALUE"] and i["StatWgt"]
respectively, would not depend on getting
        statwt.append(i[8])   # the number or order correct if using number indeces
    print varclass;print statwt

    doRc_statwt(in_raster, varclass, statwt, out_raster)
    #error in above, cannot save as FGDB file due to already existing (even with
overwrite output enabled), delete first file seems to fix, rest overwrite correctly...
    in_stat_wt_rasters.append(out_raster)  #rasterList = arcpy.ListRasters("*",
"GRID")

    del varclass, statwt, np, table, in_raster, out_raster

arcpy.env.overwriteOutput=True
#Call stat weight model function
print "Model Output: " + str(model_output) + "\nModel Inputs:
"+"\n"+str(in_stat_wt_rasters)
StatWtModel(in_stat_wt_rasters, model_output)


#Add function for tabA mod output to LS inventory to see accuracy; including new
fields and calculations

arcpy.CheckInExtension("Spatial")
print "Done."
################################################################################
```

```
################################


"""
#Snippet for running just the model again after the rest has run
import arcpy;
from arcpy.sa import *;
import numpy as np

arcpy.env.overwriteOutput = True
arcpy.env.workspace =
r"D:\Users\n913s392\Desktop\MasterFolder_desktop\reclass_tables.gdb"
arcpy.env.scratchWorkspace = r"D:\Users\n913s392\Desktop\MasterFolder_desktop\scratch"
arcpy.CheckOutExtension("Spatial")

#Snippet for running just the model again after the rest has run
def StatWtModel(in_stat_wt_rasters, file_output):  # in_stat_wt_rasters, mask_rast,
file_output    mask_rast removed
    # model = mask_rast
    counter = 0
    for i in in_stat_wt_rasters:
        i_rast = Raster(i)
        if counter == 0:
            model = i_rast
        if counter > 0:
            model += i_rast
        counter += 1
    model.save(file_output)


model_output = r"FayzMod2"
zmodelInputs = ['stF_rc_aspect2', 'stF_rc_curvature2', 'stF_rc_d2fault_2',
                'stF_rc_d2stream', 'stF_rc_elevation', 'stF_rc_flt_lden2',
                'stF_rc_geoFayz', 'stF_rc_plan_2', 'stF_rc_profile_2',
                'stF_rc_slope', 'stF_rc_strm_lden2', 'stF_rc_twi']
StatWtModel(zmodelInputs, model_output)
arcpy.CheckInExtension("Spatial")
"""
```