Portland State University
PDXScholar

Dissertations and Theses | Dissertations and Theses

3-25-2021

Evaluation of Manual and Semi-Automated Deep-Seated Landslide Inventory Processes: Willapa Hills, Washington

Tiffany E. Justice
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Geomorphology Commons, and the Remote Sensing Commons
Let us know how access to this document benefits you.

Recommended Citation
Justice, Tiffany E., "Evaluation of Manual and Semi-Automated Deep-Seated Landslide Inventory Processes: Willapa Hills, Washington" (2021). *Dissertations and Theses.* Paper 5683.
https://doi.org/10.15760/etd.7555

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Evaluation of Manual and Semi-Automated

Deep-Seated Landslide Inventory Processes:

Willapa Hills, Washington

by

Tiffany E. Justice

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Geology

Thesis Committee:
Adam Booth, Chair
Susan Shaw
John Bershaw

Portland State University
2021

© 2021 Tiffany E. Justice

Abstract

Recent advances in remote sensing data and technology have allowed for computational models to be designed that successfully extract landforms from the landscape. The goal of this work is to create one such semi-automated model to extract deep-seated landslides located in complex geomorphic terrain. This is accomplished using geographic object-based image analysis (GEOBIA) techniques, considered by leaders in the field of image analysis to have an advantage over traditional automated classification methods. GEOBIA methods can mimic human visual interpretation by including more characteristic features used to assess the relationship between image data and the ground surface such as color reflectance (spectral), texture, shadow, location, pattern, height, tone, context, size, and shape.

The standard method for identifying and mapping landslides in the Pacific Northwest is for professional geologists to manually delineate landform features using remote sensing data, referred to as remote mapping. The method is currently employed by United States Geological Survey (USGS), Washington State Department of Natural Resources (WA DNR), and Oregon Department of Geology and Mineral Industries (DOGAMI). The question remains if semi-automated models can perform as well as independent manual mappers when identifying landslides, while reducing bias due to interpretation discrepancies between mappers.

To test this hypothesis, two modeled landslide datasets are created. The first, using a model design that was not influenced by manual mapping efforts, and the second created using manually mapped landslides for visual reference. These two modeled

datasets are then compared to a manually mapped landslide inventory, created with input from four professional geomorphologists. Differences in landslide numbers, densities, geometries, and extents, that were delineated by the geologists, reflected the range of professional backgrounds. The overlapping spatial area of landslides delineated by all geologists is used as reference to verify landslide areas from the first modeled dataset. Agreement statistics (i.e., accuracy) suggest 81% of the modeled landslide area are appropriately delineated by the model in this study. The second set of modeled landslides are verified by comparing the spatial dataset to all landslides inventoried by the geologists (i.e., any terrain delineated as a landslide). Use of all landslides in the inventory eliminated the need to filter the data, thus introducing bias into the reference inventory. Agreement statistics (i.e., accuracy) suggest 78% of the modeled landslide area are appropriately delineated. Perhaps more interesting, agreement statistics for recall, emphasizing correct identification for all reference landslides, suggest 69% of the area is correctly identified as a landslide. This is compared to landslide area recall between the four manual mappers which ranges from 35-99% within the study area.

The data suggest the model is objectively using a set of morphometric characteristics to map the landslides, while the professional geomorphologists have developed interpretation style biases that lead to a large range in area mapped as a landslide. Incorrectly identifying terrain as stable could have negative impacts on public safety, suggesting more research is necessary to determine the true population of landslides that exist on the landscape. Automated models can be useful with that effort.

Acknowledgements

Many people have played key roles in making this thesis a possibility. I have benefited the most from the support of the following individuals:

This project and personal growth opportunity would not have been possible without Susan Shaw's trust, guidance, motivation, and enthusiasm. She has been an outstanding mentor and inspiration. She has pushed me to explore innovative ideas, while reminding me of the important things in life. I will be forever grateful for Susan's friendship during every step of this long journey.

Adam Booth provided an unparalleled level of technical expertise and created a pathway for me, a student above the traditional age, to succeed. He devoted the time and effort to learn a complex modeling technique, allowing me to pursue this project. He was considerate when the research path blurred and always enthusiastic to get it back on track. When I was faced with logistical and personal challenges, he responded with the utmost patience and courtesy. He has made an overwhelmingly positive difference in my life.

A large component of this research was the process of creating a landslide inventory that was used to test the model. I am extremely thankful for the team of geologists, Ted Turner, Jason Hinkle, Adam Booth, Rachel Pirot, and Susan Shaw, who made time in their busy schedules to provide many service hours devoted to mapping and technical discussions of landslide identification. This process heightened my awareness

of the intricacies of office-based landslide detection and I appreciate the valuable information that was shared.

I'd like to thank all of my managers at Weyerhaeuser (Suzanne Cuthbert, Julio Rojas, Jessica Homyack, Mike Rochelle, and Susan Shaw) for sponsoring my education and allowing me the flexibility to complete it. A lot of work was put in behind the scenes to make my education a priority. I would not be where I am without their support.

Finally, I would like to thank my daughter, Madison, and my family for supporting me over the last few years. Graduate school was a massive commitment. My family stepped in to help with everything I could no longer take care of, and probably most importantly provided laughter when I needed it most.

Table of Contents

Abstract…….…….…...i

Acknowledgements…...……..…...iii

List of Tables……..…...vii

List of Figures……...……...……viii

1. Introduction……..…..1

1.1 Motivation……..1

1.2 Definition of Deep-Seated Landslide & Brief History of Inventory Methods……..2

1.3 Morphological Evolution of Deep-Seated Landslides……...7

1.4 Semi-Automated Landslide Models……..9

2. Methods…….13

2.1 Study Area & Background……...13

2.2 Geologic History……..15

2.3 Overview of Object Creation for Image Analysis Routines……18

2.4 Preliminary Feature Extraction……25

2.5 Landslide Feature Extraction……...26

2.5.1 Data……26

2.5.2 Image Analysis Techniques & the Interpretation Process…….29

2.5.3 Model Methods……..34

2.6 Manual Inventory Methods……..38

3. Results……...44

3.1. Manual Inventory…..……..44

3.2. Modeled Inventory……..51

3.2.1. Model Strengths……53

3.2.2. Model Weakness……...58

3.2.3. Agreement Between Modeled and Mapped Landslide Inventories……..61

4. Discussion…….68

4.1 Landslide Activity Level, the Root Cause of Interpretation Disagreement…….68

4.2 Evolution of Semi-Automated Model……..72

5. Conclusion……84

Bibliography…….87

Appendix: Object-based Extraction of Stream, Road, and Ridge Features……..94

Bibliography for Appendix…….108

List of Tables

Table 1. Description of the four qualifiers (i.e., thematic attributes) for the manually mapped deep-seated landslide inventory. The dataset includes 2.3 km^2 of landslides inventoried in two survey areas (test site 1: Figure 12 & test site 2: Figure 13)…...……43

List of Figures

Figure 1. Clockwise from top left: temporal progression of landslide surface expressions from (A) recently failed, active landslide terrain to (D) old, inactive feature persisting on the landscape. The term “headscarp” used in this thesis coincides with the feature labeled “main scarp” (modified from McCalpin, 1984)……..7

Figure 2. Extent of lidar data footprint used in study. Black hatches delineate preliminary study extent. Red hatches delineate model verification extent. Blue polygons delineate manual mapping extent. Black box in inset map showing location of study in southwest Washington State……...14

Figure 3. Geologic map of project study area and surroundings (Wells & Sawlan, 2014). Black outlines indicate both preliminary and verification study extents in Figure 2……17

Figure 4. General GEOBIA workflow with arrows to show order in which image pixels are segmented into objects and then classified……. 20

Figure 5. Example of parameter selections and object results for three segmentation algorithms used in this study……..22

Figure 6. (A) Lidar-derived slope image of topography typical of the study area provided for visual reference; (B) Illustration of features excluded from landslide analysis.……..26

Figure 7. Illustration of use of image object hierarchies in object-based workflows (Ouyang et al., 2011)…….30

Figure 8. (A) 3d view of aerial photo in location of deep-seated landslide using lidar elevations for ground height reference; (B) Illustration of conceptual deep-seated landslide components (Cruden & Varnes, 1996)……...…32

Figure 9. Lidar derivatives, converted to intensity values on the scale of 0-255, shown in same extent as Figure 8. (A-C) curvature images with bright pixels representing convex features and dark for concave (TAN: curvature in the direction of contour, PRO: curvature in direction of slope, BOLSTAD: ratio of the number of concave to convex pixels within a defined window); (D) gradient image with labels for key landslide features (GRAD: topographic gradient)…….33

Figure 10. Illustration of iterative object creation within eCognition software.……35

Figure 11. Classifications for deep-seated landslides (Varnes, 1978). (A-C) records slide-type movement while (D) records flow movement……...40

Figure 12. Results from composite, manually mapped deep-seated landslide inventory; test site 1. Four point black line delineating extent of manual mapping effort…….46

Figure 13. Results from composite, manually mapped deep-seated landslide inventory; test site 2. Four point black line delineating extent of manual mapping effort. Lidar coverage ends to the west of the survey area, creating irregular boundary……..47

Figure 14. Comparison of agreement in landslide inventory maps produced by the four individual mappers A-D. Reference data in columns and verified data in rows. X's exist in cells where duplicate values are computed, or where a value of 1 is computed when the reference data is identical to the data being verified..……49

Figure 15. GEOBIA modeled landslides in the extent of model verification test site number 1……52

Figure 16. GEOBIA modeled landslides in the extent of model verification test site number 2……53

Figure 17. Multi-scale feature extraction example. Small landslide centered in small black rectangle on the left side of image. Large landslide on right side of image centered in larger black rectangle. Misclassified landslide in center of black circle. All objects classified with same routine and parameters…….55

Figure 18. Example of feature extraction in complex topography……58

Figure 19. Misclassified landslide at location of arrow…….60

Figure 20. Deep-seated landslide that was removed from classification due to large relative object boundary (purple polygon) to steep, stream features…….61

Figure 21. Spatial overlap of manually mapped landslides (composite inventory) and GEOBIA modeled landslides from the first model run in the extent of comparative site number 1……63

Figure 22. Spatial overlap of manually mapped landslides (composite inventory) and GEOBIA modeled landslides from the first model run in the extent of comparative site number 2……64

Figure 23. Landslide model verification using manually mapped landslides as the reference dataset. Landslide area for the reference dataset include test site 1 & 2. The reference landslide data was divided into two columns (Definite, and all landslides comprised of Definite+Split-decision+Non-consensus)……66

Figure 24. Area of true positive (TP) and false negative (FN) shown using mapper C landslide areas as the reference dataset and mapper B landslide areas for verification dataset. Spatial extent of data shown in test site 1…….70

Figure 25. GEOBIA modeled landslides from the second model run in the extent of comparative test site 1……74

Figure 26. GEOBIA modeled landslides from the second model run in the extent of comparative test stie 2……75

Figure 27. Spatial overlap of manually mapped landslides (composite inventory) and GEOBIA modeled landslides from the second model run in the extent of comparative test site number 1……..77

Figure 28. Spatial overlap of manually mapped landslides (composite inventory) and GEOBIA modeled landslides from the second model run in the extent of comparative test site number 2……..78

Figure 29. Statistical metrics comparing second model run to all manually, mapped landslides (reference data) on left side of figure. Comparison to first model run statistics computed for all manually, mapped landslides provided on right side of figure to note model improvements.……79

1. Introduction

1.1 Motivation

Landslide inventories are critical to understanding the spatial distribution and geometric properties of unstable hillslopes. The purpose of this paper is to introduce a new classification approach used for automating the process of creating deep-seated landslide inventories from high-resolution lidar data. A key component of this research is an in-depth exploration of inventories created by professional geologists, which are used to test model performance. This tool builds on automated landform mapping techniques presented by Shaw et al. (2017) where a suite of landforms, including deep-seated landslides, are successfully mapped using algorithms that define surface topography in terms of morphological attributes. Model advancements I make include introduction of indices created to classify landslides using a hierarchical data processing scheme. I use Trimble Geospatial' s eCognition software to design the Geographic Object-based Image Analysis (GEOBIA) routine. The empirical, image analysis approach I use mimics human perception of characteristic topographic features allowing for deep-seated landslides to be identified and mapped systematically. I then compare the model output with a compilation of manually mapped landslides created by a group of four geomorphologists with close to 100 years of combined professional landslide-mapping experience. Study design allows for quantification of agreement between professional interpretations, setting up the comparison to the model-produced inventory. The result of agreement comparisons between observers, in addition to comparisons being made

between the observers and the model, highlights inconsistencies in active vs. inactive landslide interpretation. Topography where the inconsistencies are occurring can be described using morphometric signatures suggesting that semi-automated modeling approaches may have an advantage in objectively classifying landslides.

1.2 Definition of Deep-Seated Landslide & Brief History of Inventory Methods

Comprehensive and precise landslide inventories are essential tools for various agencies including academia, emergency response, public works, land use, planning, and development (Corominas et al., 2014; Joyce et al., 2014; Razak et al., 2013; Smith & Petley, 2009). The purposes requiring landslide inventory maps vary from studying the contribution of landslides in landform evolution processes, to predicting serious public safety hazards (Galli et al., 2008). For instance, landslide inventory maps are used to evaluate the spatial distribution of landslides, thereby allowing for density and abundance information to be reported according to the temporal scale of the study. Inventory maps also allow for statistics, such as slope failure frequency-area relationships, to be quantified as an aid in examination of erosion contributions for a given area (Galli et al., 2008; Guzzetti et al., 2002). Applied geosciences capitalize on the inventory data when predicting local and regional landslide susceptibility, which in turn improves public safety and resource protections (Burns & Madin, 2009; Burns & Mickelson, 2016; Lacasse et al., 2010; Stumpf & Kerle, 2011).

While studying the role of landslides on the landscape, scientists make a distinction between shallow and deep failures because deeper landslides tend to involve different mechanisms and rates of movement, have a unique set of morphological surface

expressions, may occur on different parts of the landscape, and can have different responses to human and climatic activities (Miller, 2017). I distinguish between shallow and deep failures based on the relation of the failure plane to the rooting depth of trees, with shallow failures occurring within the upper soil layers occupied by tree roots (to an average depth of 3 m below the surface) and deep-seated occurring below that depth (WFPA, 2020). Potentially unstable hillslopes have the potential for causing large losses of life when they experience movement. For instance, 43 people devastatingly lost their lives due to a deep-seated landslide that recently occurred near Oso, Washington on March 22, 2014 (Aaron et al., 2017; Hibert et al., 2015; Iverson et al., 2015; Keaton et al., 2014; Stark et al., 2017; Wartman et al., 2016). This research will focus on automating the identification of deep, gravity driven movements of materials along a surface of shear that occurs up to hundreds of meters below ground. This sudden movement is triggered when the driving shear stress due to weight of soil, rock, and surface materials such as trees exceeds the residual shear strength of the hillslope materials (Terzaghi & Peck, 1948). Shear strength acting to hold the hillslope intact is reduced by positive pore water pressure caused by infiltrated precipitation (i.e., rainfall), which is considered to be the main trigger of historic landslides in the Pacific Northwest (LaHusen et al., 2020). Pore water pressure is also increased by longer term increases in water table elevations that can be due to anthropogenic or climatic changes (Iverson & Major, 1987).

Traditional methods for mapping deep-seated landslides used for evaluating and reducing landslide hazards or risks are completed by interpreting historical data and examining landslide features in the field (Wieczorek, 1984). Geologists first conduct a

background review of the site including recorded geologic history, existing inventory or hazard data, and soil property data. A historical investigation typically includes an office screen for previous failures through examination of aerial photographs (often stereo-pair images) and topographic map data. The geologist generally completes a field verification exercise to identify features indicative of deep-seated instability such as earthen cracks, vegetation type and associated key structural changes associated with disturbance regimes, disrupted drainages, soil structure changes, undrained depressions, pressure ridges, hummocks, deposited sediments, headscarps, and lateral margins. With the data collected, landslide type can be established based on exposed material (rock, soil, or debris) in combination with noted movement, for instance planar/translational sliding, rotational sliding, intact block sliding, or flowing patterns (Hungr et al., 2014; Varnes, 1978).

Creating an exhaustive, ground-based, deep-seated landslide inventory is challenging and time consuming due to issues of scale and logistical roadblocks. Ground-based inventories are created by geologists who rely on professional interpretations that are inherently biased (Ardizzone et al., 2002; Booth et al., 2009; Guzzetti et al., 2000; Galli et al., 2008; Hölbling et al., 2017). With the increasing availability of high-resolution surface topography including lidar data, geology practitioners and regulatory agencies such as the Oregon Department of Geology and Mineral Industries (DOGAMI) and Washington States Geological Survey (WGS), who collect, develop, use, distribute, and preserve geologic information, have shifted towards inventory mapping based on remotely sensed data for increased accuracy and efficiency. These inventories created

based on remotely sensed data, hereafter referred to as remote-sensing derived inventories, have been described by Burns & Madin (2009) as advantageous compared to field mapping alone in that they allow mappers to use multiple spatial and temporal scales to clearly identify and isolate parts of landslide complexes that may have different activity levels (active vs. inactive) and allow for much larger geographic areas to be mapped. Additionally, some features (e.g., large-scale, low-relief deposits) can be identified and interpreted more readily from remotely sensed data (Burns & Madin, 2009). However, landslide inventories produced manually using remotely sensed data rely on professional interpretation. This interpretation introduces bias. Mapping differences occur when inventories are created by multiple professionals having a large range of experience and skill sets (Hölbling et al., 2017). Inventory inconsistencies including landslide numbers and densities as well as differences in geometries and extents, continue to be problematic (Ardizzone et al., 2002).

Root causes for inconsistencies in remote-sensing based identification of deep-seated landslides using high resolution topographic data are both terrain- based, related to differences in geology and landslide morphology, and map-based due to variability in mapper experience, final map scale, and data quality and processing (Bunn et al., 2015; Galli et al., 2008; Hölbling et al., 2017; Kozeniowska, 2017). Terrain-based mapping difficulties include identifying landslide components consistently, distinguishing failures in complex terrain, and identifying historic or inactive landslides. Exacerbating the map product inconsistencies are changes in landslide characteristics that coincide with varying lithologies, topography, vegetation, and climates (Glenn et al., 2006). While efforts to

standardize data collection, processing, and inventory protocols will minimize inconsistencies in map-based product variability, terrain-based interpretation challenges remain (Burns & Mickelson, 2016; Razak et al., 2013; Sithole & Vosselman, 2013).

The effects of landslide triggering mechanisms vary depending on elements such as slope steepness, terrain morphology, soil type, underlying geology, and anthropogenic hillslope alterations (Bobrowsky & Highland, 2013). Interpretation of landslides using remote-sensing data does not allow for consideration of soil type or underlying geology. Heterogeneous landslide features such as scarps and crowns, lateral margins, transverse ridges and cracks, and deposits, as illustrated in Figure 1, are interpreted based on what is visible in the remote datasets, for instance what an observer can "see" using a bare-earth hillshade map. A clear example of the visible surface expression characteristic of a deep-seated landslide is the presence of crisp, discernible geomorphological features such as arcuate headscarps (equivalent to "main scarps") with a clear vertical step exposing bedrock (Figure 1(A)). Active landslide benches are often riddled with deposits consisting of strikingly out-of-place hummocks or angular blocks of displaced materials that can include secondary scarps where block detachment has occurred. The active landslide body is outlined by a distinct lateral margin with vertical scarps where shear is occurring. Often, debris can be found on the depositional toe at the base of the active feature, forming a visible debris fan.

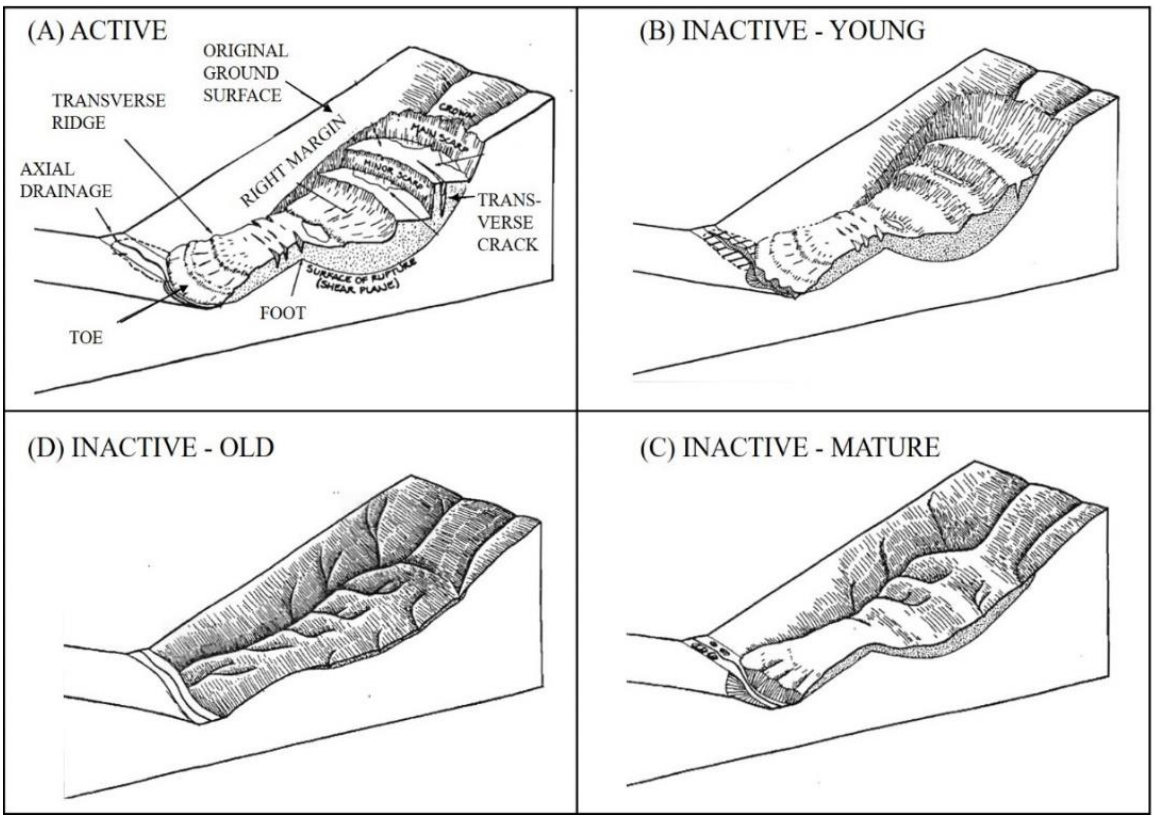


FIGURE 1. CLOCKWISE FROM TOP LEFT: TEMPORAL PROGRESSION OF LANDSLIDE SURFACE EXPRESSIONS FROM (A) RECENTLY FAILED, ACTIVE LANDSLIDE TERRAIN TO (D) OLD, INACTIVE FEATURE PERSISTING ON THE LANDSCAPE. THE TERM "HEADSCARP" USED IN THIS THESIS COINCIDES WITH THE FEATURE LABELED "MAIN SCARP" (MODIFIED FROM MCCALPIN, 1984).

1.3 Morphological Evolution of Deep-Seated Landslides

Through morphological evolution, discrete features of deep-seated landslides become more subdued and therefore harder to delineate. The weathering process smooths crisp breaks in slope near scarps and lateral margins, and also removes depositional lobes from the base of the landslide. Interpretation becomes more difficult as less evidence of failure is visible. The purpose of this study is to automate the delineation of hillslope instabilities in any activity state, so it is key to define morphological expressions of active and

inactive landslides. The following feature evolution description will serve as a target for the modeling exercise.

As a landslide ceases to move downslope as one coherent landmass, small subparallel tributary streams can occupy the margins, vegetation reappears, and internal cracks can be found in the settling deposits (Figure 1(B)). The unconsolidated toe deposits of the once active landslide are susceptible to erosion by streams in the floodplain. With additional time, the maturing inactive landslide headscarp may appear dissected as it erodes into a more gradual slope angle, losing its discernable shape (Figure 1(C)). The topography begins to smooth as soil is eroded from hummocks and fills in depressions. The stream at the base of the landslide feature is no longer eroding into toe deposits although the floodplain downstream of the toe may still retain signs of excess debris in the form of terrace blanketing or floodplain widening. As the landslide transitions into an older state of inactivity, visible signs of lateral shear along the margins disappear, the hummocky topography fades to undulating, and the disrupted drainages running through the deposits more closely resemble a stream system near equilibrium (Figure 1(D)). Streams cut down through slide debris, truncating depositional toes and redistributing debris to form river terraces. These older, more weathered features, which lose contrast to neighboring unfailed terrain as they transition into traditional hillslope and valley segments are the main cause of disagreement amongst geomorphologists who are tasked with interpreting the history of the landscape, as described more fully in section 4.1.

1.4 Semi-Automated Landslide Models

Currently, visual interpretation of remotely sensed images is the most widely used landslide inventory method. However, advancements in inventory automation are being fueled by a desire to circumvent professional mapper disagreement regarding landslide numbers, densities, geometries, and extents (Galli et al., 2008; Hölbling et al., 2017). Many attempts have been made to automate landslide inventory production. The methods capitalize on a plethora of high-resolution datasets and increasingly sophisticated computer algorithms (Passalacqua et al., 2010). Image analysis approaches designed to analyze image pixels are referred to as pixel-based methods, with the minimum analysis scale being determined by the size of the image pixels. Pixel-based methods aimed at extracting active landslides predominantly use metrics that capture terrain roughness to characterize the difference between areas that are more heterogenous than their surroundings (e.g., root mean square of elevation and slope values, eigenvalue ratios of surface-normal vectors, semivariance, co-occurrence texture statistics, windowed discrete Fourier transform (DFT), continuous wavelet transform (CWT), and wavelet lifting schemes) (McKean & Roering, 2004; Berti et al., 2013; Whelley et al., 2014). Landslide extraction methods based on surface roughness produce landslide inventories with high, in some cases greater than 80%, agreement with existing manually created inventories (Booth et al., 2009; Kasai et al., 2009; Whelley et al., 2014). The methods perform well in isolated test areas but have yet to be applied at the landscape level. Also, while simple roughness thresholding has been proven to extract landslides in terrain with homogeneous landforms, thresholding alone may not be appropriate to detect landslides in complex terrain and in features where age is influencing the roughness (Hölbling et al.,

2017; Korzeniowska, 2017). Further exploration into the relationship between roughness and landslide morphology, including exploitation of contextual information, may prove useful in expanding the geographic extent of automated inventories (Korzeniowska, 2017). Context becomes useful for differentiating inactive landslides from neighboring terrain by first classifying features with distinct roughness patterns and then expanding the classification into neighboring areas such as weathered headscarps or lateral margins with more subdued roughness.

More recently, the utilities of object-based image analysis (OBIA) or geographical object-based image analysis (GEOBIA) techniques are recognized as an enhancement to the pixel-based approaches for terrain analysis applications (Blaschke, 2010; Dragut & Eisank, 2012; Gercek, 2010; Shaw et al., 2017). One limitation of pixel-based landform classification is pixels that comprise raster-based images have geometric dimensions, unlike real landscape features, making them a poor representation of landforms observed in nature (Shaw et al., 2017). GEOBIA methods enable grouping of remote sensing data pixels into homogeneous clusters, or objects, based on similar morphometric attributes. Grouping of the image pixel values (i.e., segmentation) is the foundation of a successful computational routine. Segmentation creates objects familiar to human perception as they closely approximate real-world features (MacFaden et al., 2012). Pixel-based digital terrain models also require establishment of grid window sizes and/or characteristic scales (Shaw et al., 2017; Booth et al. 2009). GEOBIA methods do not require selection of a single analysis scale. They instead provide multi-scale image analysis techniques for segmenting and classifying data. Finally, unlike pixel-based

methods that do not retain context (neighboring feature information), shape (feature outline), texture (variance across space), pattern (configuration), or size, use of objects provides access to characteristic features that can be used to define specific classes according to geomorphometric concepts.

Image analysis by way of segmentation and classification may add an element of transparency, beneficial in a diverse landscape containing a high density of landforms with competing characteristic spatial scales (Booth et al., 2009; Perron et al., 2008; Stumph & Kerle, 2011). Object-based routines often employ repetitive trial-and-error methods to build contextual relationships describing classified and unclassified objects based on their spatial and hierarchical relationship to other objects that overlap spatially or share a common border (i.e., their neighbors) (MacFaden et al., 2012). Each step in the analysis process provides immediately viewable results, enabling instant translation of the mathematical methods used to delineated landslide footprints and classification of the components.

The work presented in this paper will build on previous GEOBIA models that have successfully automated the extraction of landslides (Blaschke et al., 2014; Bunn et al., 2019; Korzeniowska, 2017; Lahousse et al., 2011; Li et al., 2015; Martha et al., 2010; Van Den Eeckhaut et al., 2012). Those authors have successfully mapped landslides in specific geomorphic terrain types. This study uses the GEOBIA method to produce a landslide map that expands into the diverse terrain of the Pacific Northwest. Unlike the previous studies, I am designing the model to map landslides in any activity state. Taking advantage of a GEOBIA technique to exclude data from the analysis by way of classification, I define classes for landforms that could interfere with deep-seated

landslide classifications, prior to running the routine. I then design a new multi-scale roughness criterion used to classify landslides based on the density of objects with large positive and negative windowed average slope differences to their neighbors. The technique, based on statistical methods described in section 2, permits capturing of morphological features commonly used by professionals to visually identify deep-seated landslides. The result is a model that detects deep-seated landslides, as an expert would, but with greater objectivity and reproducibility, as well as more consistent agreement with both manually produced landslide inventories and classifications from alternative GEOBIA model designs.

2. Methods

2.1 Study Area & Background

The study area is situated in a geographic region known as the Willapa Hills, western Washington, United States. The area was selected because it contains an abundance of deep-seated landslides with diverse morphologies. It is located entirely on privately owned timberlands within the Coast Range of southwestern Washington (inset map in Figure 2). The temperate climate and deep, fertile soils are appropriate for growing the primary conifer tree species in the study area: 0 to 70+ year old western hemlock (*Tsuga heterophylla*) and Douglas-fir (*Pseudotzuga menziesii*) (Pringle, 1986; Turner et al., 2010). Up to 190 days of the year have measurable rainfall in the Coast Range mountains, with December and January being the wettest months (WRCC, 2020). The study area occupies the epicenter of an unusually large, but short duration, storm event which occurred in December of 2007. Within the 85 km^2 study area, 22 deep and 685 shallow landslides were triggered and digitally recorded (Turner et al., 2010). The greatest storm total rainfall amount, most of which fell in just over 48 hours, was 498.6 mm, recorded at a privately operated, Rock Creek climate station, located within the study boundary. The nearby National Weather Service rain gage at Frances (northwest corner of map in Figure 2) recorded 358.1 mm of rain, most of which fell in just over 48 hours. Maximum 24-h precipitation amounts were 70-90% of the storm total (Turner et al., 2010).

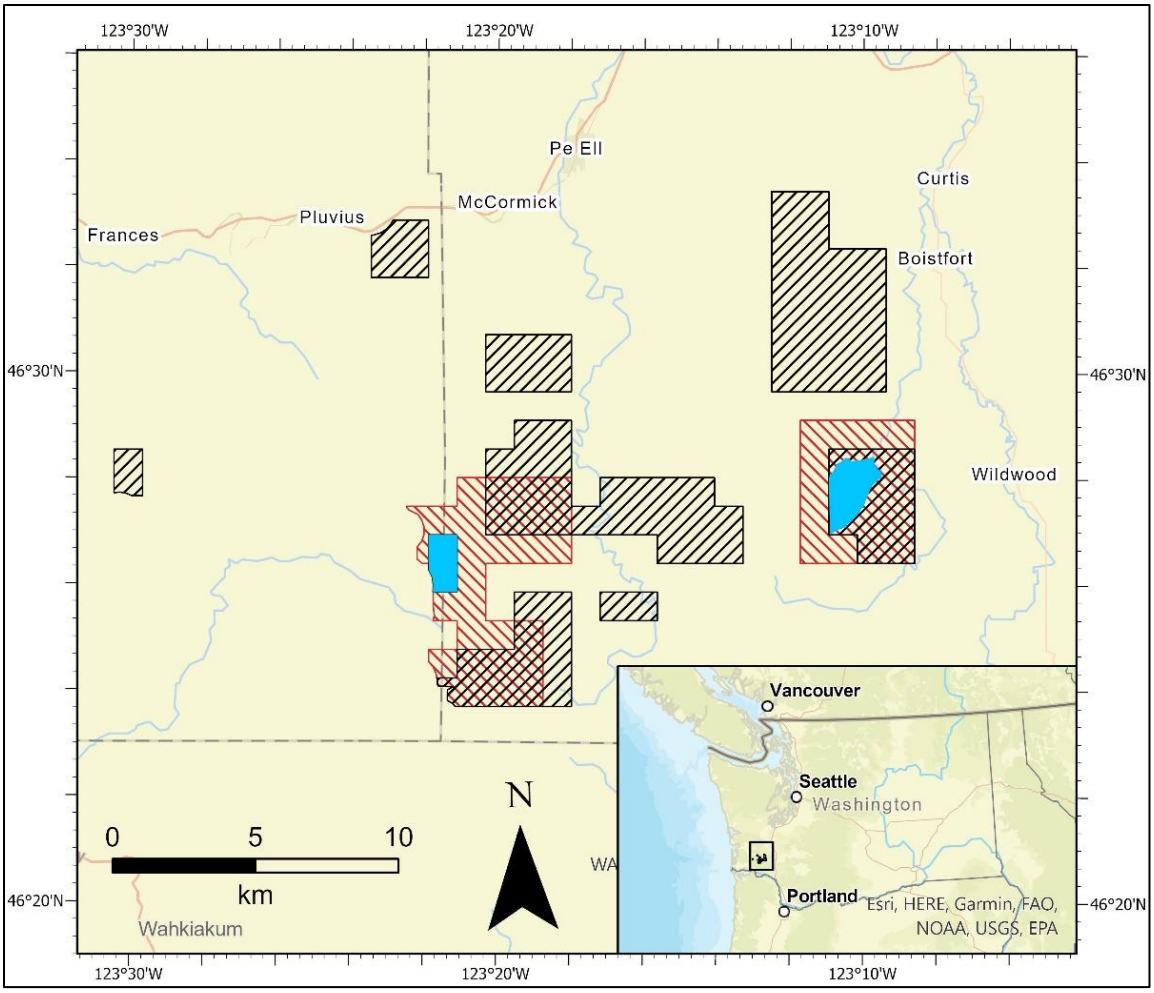


FIGURE 2. EXTENT OF LIDAR DATA FOOTPRINT USED IN STUDY. BLACK HATCHES DELINEATE PRELIMINARY STUDY EXTENT. RED HATCHES DELINEATE MODEL VERIFICATION EXTENT. BLUE POLYGONS DELINEATE MANUAL MAPPING EXTENT. BLACK BOX IN INSET MAP SHOWING LOCATION OF STUDY IN SOUTHWEST WASHINGTON STATE.

An additional consideration for study area selection was the availability of high resolution lidar data collected by Watershed Sciences, Inc in 2008, during spring, leaf-off tree conditions, through a contract with Weyerhaeuser Company. The point cloud data was processed into 2-m resolution grids by Weyerhaeuser GIS staff using a proprietary interpolation algorithm. From the gridded data, I selected all available 1x1 km lidar tiles that intersected the photo-inventoried, deep-seated-landslide initiation points, for a total

of 85 km^2, thereby ensuring ample deep-seated landslides existed within the area selected for this study (ArcGIS v.9.3). This extent is shown with black hatches in Figure 2. From the initial study area, a subset consisting of two randomly selected test areas, shown in blue in Figure 2, with moderate landslide densities, as determined by observing modeled landslide density distributions, were chosen for model verification. The two test areas represent the portion of the study area to be mapped by the four professional geologists and will be used as reference for the modeled data, testing agreement between the modeled landslide footprint and the manually mapped landslides. Lidar tiles were merged to the original study area in the vicinity of the test areas to allow for geographic connectivity, aiding the manual interpretation, and is shown with red hatches in Figure 2. This expanded model verification extent totaled 49 km^2.

2.2 Geologic History

The Willapa Hills comprises a structurally complex block of basement rock composed primarily of oceanic basalt and marine sedimentary rock that was accreted onto western North America through subduction in the mid-Early Triassic (Wells et al., 2015). The basaltic seamounts and shallow marine sediments are known as the Crescent Formation (Tcb) geologic unit, existing roughly in the center of map in Figure 3. Overlying and interbedded with the basement rocks are late and middle Eocene age tuffaceous siltstone, silty sandstone, and arkosic sandstone of the McIntosh Formation (Tmu), deposited in a near-shore marine environment. The younger rocks of the McIntosh Formation are exposed to the north and south of Tcb, with progressively

younger rocks exposed towards the edges of the map extent, implying an anticlinal structure with Tcb at the core (Figure 3) (Ebinghaus et al., 2014; Wells & Sawlan, 2014). Numerous dikes and sill-like bodies of porphyritic basalt to gabbro (Tig) intrude both the pillow basalts of the Crescent Formation and micaceous arkosic sandstones of the McIntosh formation (Ebinghaus et al., 2014). Distribution of these three geologic units within the study area is roughly quantified as 40% Tig, 27% Tcb, and 21% Tmu (Figure 3). It is qualitatively noted that deep-seated landslide densities are higher in the sedimentary units, where large earth flows are observed. It is speculated that lower densities of deep-seated landslide exist in the competent rocks of the Crescent Formation, and moderate densities occur in the Tig geologic unit.

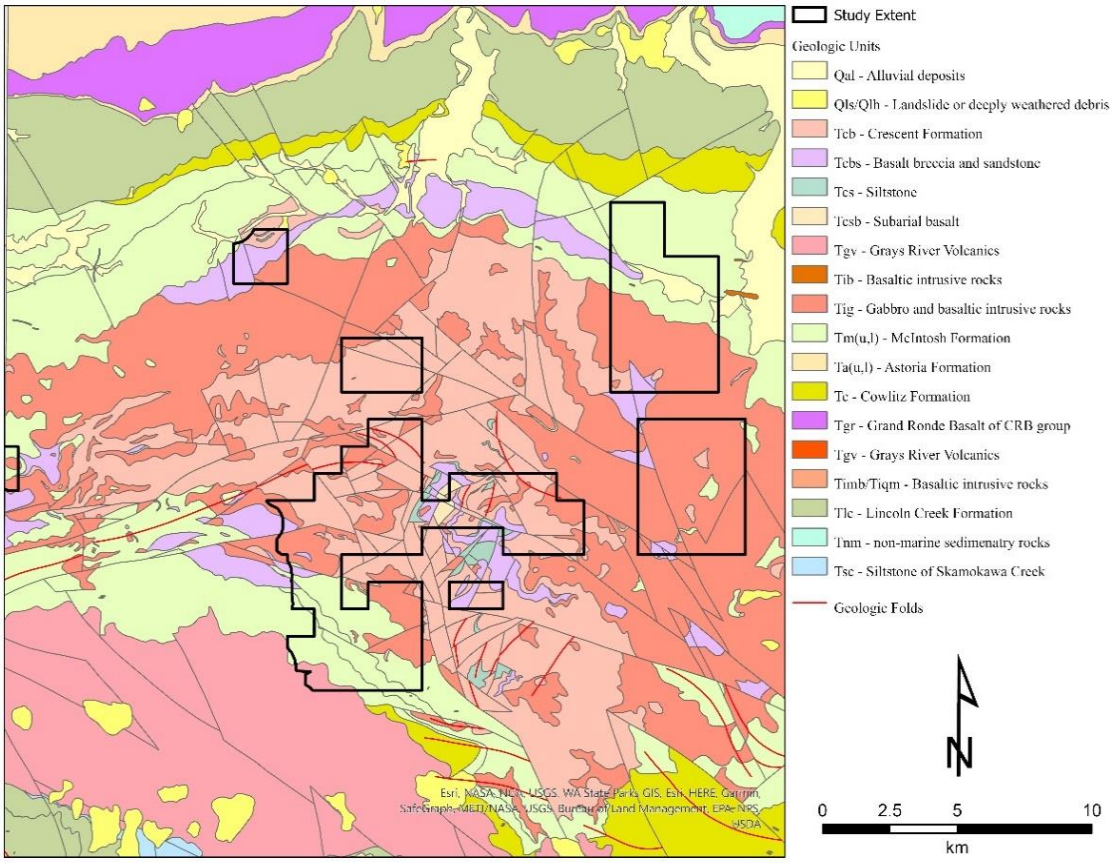


FIGURE 3. GEOLOGIC MAP OF PROJECT STUDY AREA AND SURROUNDINGS (WELLS & SAWLAN, 2014). BLACK OUTLINES INDICATE BOTH PRELIMINARY AND VERIFICATION STUDY EXTENTS IN FIGURE 2.

Elevations within the study area range from 95 m in the river valleys (Grays, Chehalis, and Willapa) to 859 m on the ridgetops. The steeper, ridge and swale topography found in the headwaters is formed by resistant volcanic rocks with the gentler slopes being comprised of sandstones and siltstones. For modeling purposes, specifically identification of headscarps, it is important to note that 32% of the topography in the study area has a topographic gradient over 60%. This steep topography will interfere with the identification of landslide components based on gradient alone. While only 3% of the landslides that initiated within this study area during the 2007 storm event were classified

as deep-seated landslide, older failures persist on the landscape that can decrease local slope stability (Turner et al., 2010). In order to determine landslide susceptibility of specific landforms, a spatially continuous map of landforms, including previously failed terrain (i.e., deep-seated landslides) as well as landforms that did not fail must be developed (Turner et al., 2010). While detailed landform mapping is beyond the scope of this study, I do intend to map deep-seated landslides across the landscape, in any activity state, so that they can be incorporated into a detailed landform map.

Anthropogenic features such as roads can have a mechanical relationship to failures, possibly decreasing the stability of the hillslope. Extensive road networks, to facilitate forest management operations, were noted during visual inspection of the data. The active network includes larger arterial roads (i.e., mainlines) as well as secondary feeder and spur roads with older, potentially abandoned roads also visible throughout the landscape. In addition to having a relationship with slope stability, roads also pose a modeling challenge. Characteristic features of disrupted roads and older, potentially abandoned roads are similar to deep-seated landslide components such as bench deposits. Care will be taken to model roads to minimize interference with landslide classification.

2.3 Overview of Object Creation for Image Analysis Routines

Before describing the algorithms, parameters, and processes used in an object-based landslide extraction model a theoretical understanding of image analysis goals including the basics of feature extraction is helpful. I've provided some insight into this process in the Appendix (Feature Extraction section). Additionally, an understanding of

how meaningful objects are created is necessary. I will discuss this concept first, followed by an overview of the specific segmentation algorithms used in this study. The goal of feature extraction routines for a computer are to use image information, often in the form of nondimensional intensity values, texture, size, location, context, patterns, or shapes, to enable automated identification of key features (Di Ruberto & Putzu, 2016). An image in this study refers to a pixel-based map derived from the lidar digital elevation model (DEM). The image pixel values alone often do not readily translate to geomorphometric features as a human perceives them (Van Den Eeckhaut, 2012). Unlike real landscape features, individual pixels have regular dimensions and, hence, are a poor representation of natural landforms (Shaw et al., 2017). Object-based modeling techniques enable the grouping of adjacent pixels with homogeneous values according to statistical characteristics. Such a group of pixels is defined as an object. The flowchart in Figure 4 describes the basic sequence for creating objects from an image to allow for classification of pixels or pixel groups.

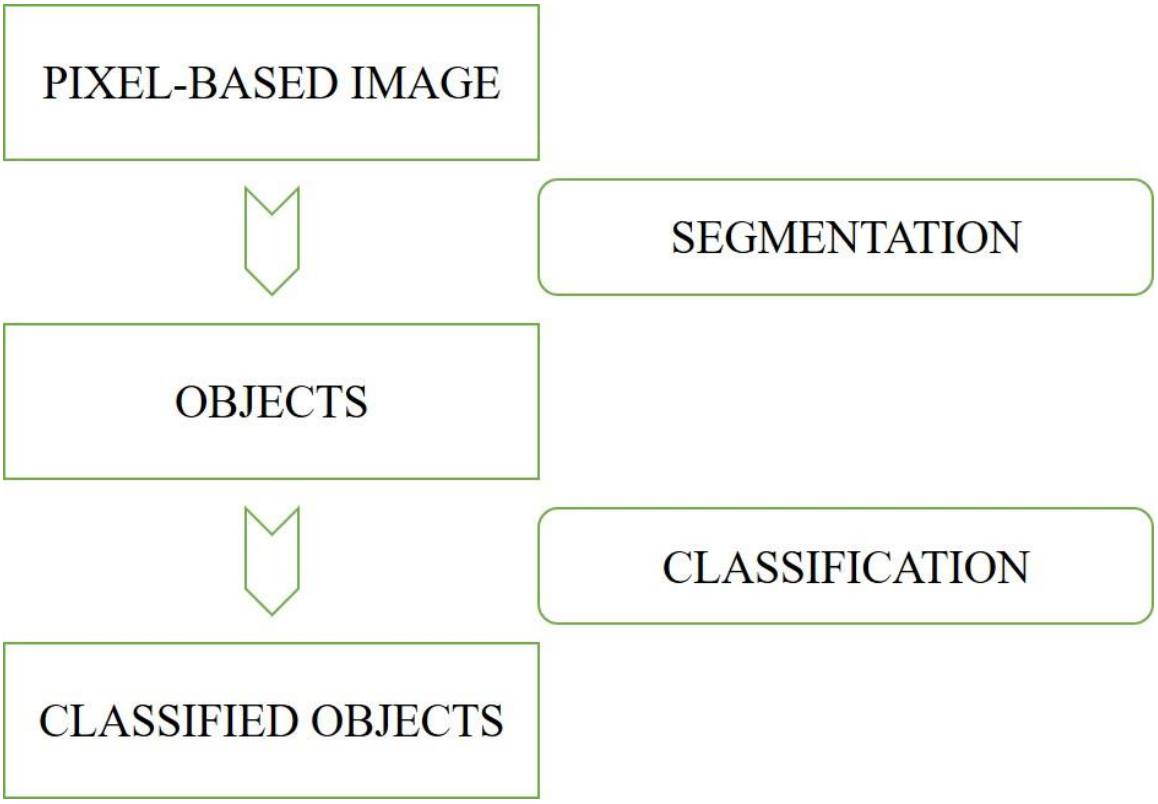


FIGURE 4. GENERAL GEOBIA WORKFLOW WITH ARROWS TO SHOW ORDER IN WHICH IMAGE PIXELS ARE SEGMENTED INTO OBJECTS AND THEN CLASSIFIED.

Creation of objects at a scale that appropriately delineates the features of interest uses object statistics, geometry, and context to successfully classify geomorphometric features. There are many methods available to create objects from images, and often image analysis workflows are designed to incorporate a variety of these methods to enable successful feature extractions. Three segmentation algorithms are used in this study: chessboard, multi-resolution, and multi-threshold. Figure 5 provides a description of the three segmentation routines used along with embedded images showing example objects resulting from the use of these algorithms. The parameters associated with each described segmentation algorithm are noted with bold type (Figure 5). Image values used to create the sample objects in Figure 5 are a slope-based difference to neighbor metric

that is defined in Appendix Eq. A2. It is a grey-level image with bright values defining pixels with a high positive difference to neighbors and dark defining high negative differences.

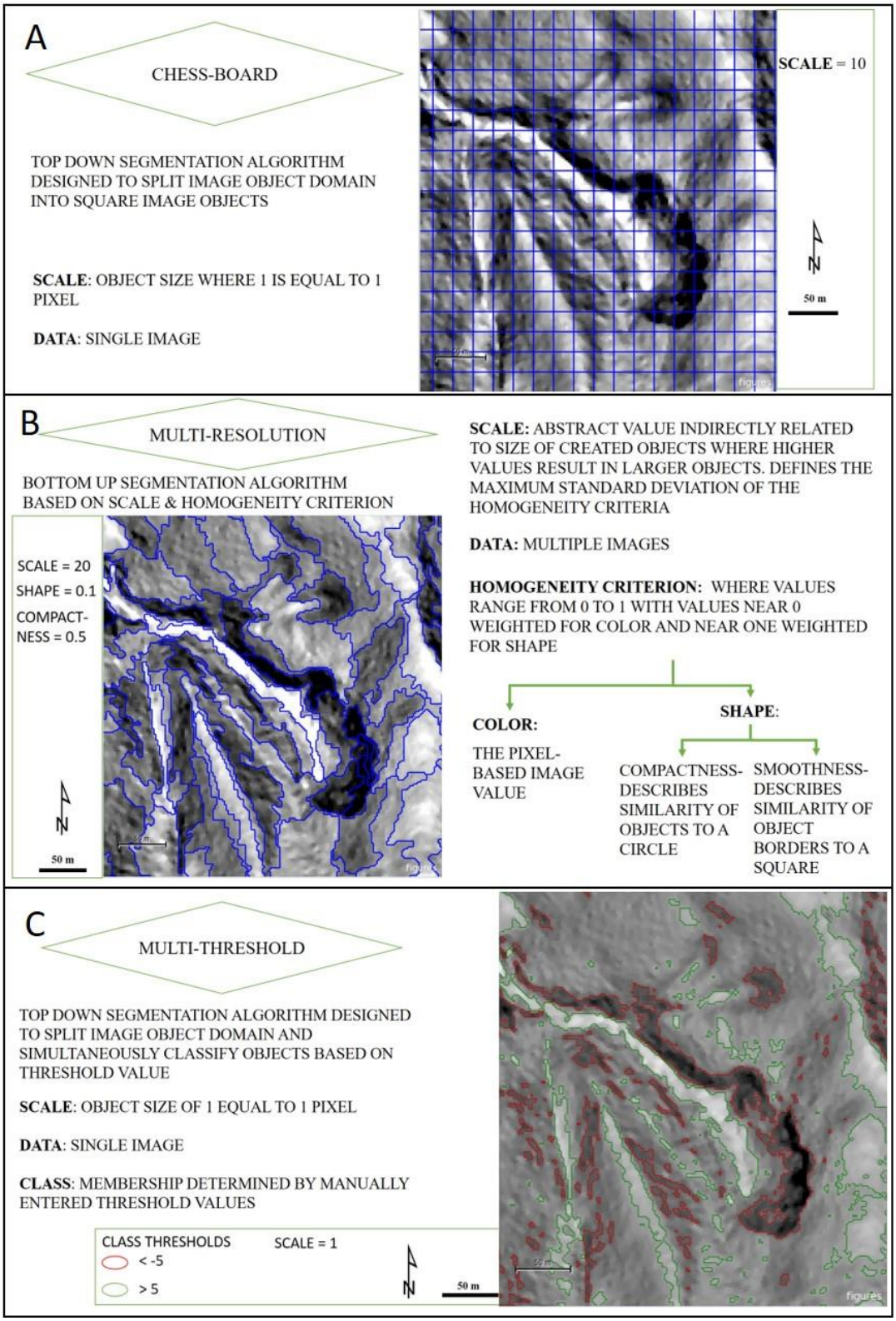


FIGURE 5. EXAMPLE OF PARAMETER SELECTIONS AND OBJECT RESULTS FOR THREE SEGMENTATION ALGORITHMS USED IN THIS STUDY.

The first and simplest segmentation, a chessboard segmentation shown in Figure 5(A), splits the image into square objects with a specified size in pixels.

The multi-resolution segmentation (MRS) algorithm shown in Figure 5(B) is a bottom-up region merging technique (as described later in this paragraph) that groups image pixels based on a manually entered homogeneity criterion and unitless scale values (Baatz & Schape, 2000). This algorithm has the capability of using multiple datasets as input. The MRS algorithm begins by considering each pixel, or combination of pixels from multiple datasets, as a separate object and subsequently merges pairs based on similarity between adjacent image objects. The homogeneity criterion that drives the merging routine is a combination of shape properties, which describe the likeness of an object to a circle or a square, and color, which for this study is simply the image pixel or intensity value (Darwish et al., 2003). The algorithm allows for flexibility in object creation by using weights that can be assigned to either color or shape. Shape can be further weighted for compactness or smoothness. The specific weight used to determine the homogeneity parameter can range from 0 to 1 with values close to 0 weighting object creation based on pixel or intensity values and values close to 1 weighting selection based on shape. For shape weights, 0 is compactness, and 1 is smoothness. For this study, the homogeneity value was set to 0.1 for all MRS algorithms, assigning 90% weight to color (i.e., lidar derivative image values). This is the default value for the algorithm, and while other values were tested, the value of 0.1 produced the most aesthetically pleasing object shapes that appeared to delineate landslides well. Equal weights were applied to shape parameters (0.5). The homogeneous merging routine described above terminates when a user-defined scale threshold (i.e., variance threshold) is reached. Scale describes a

"merging cost" that is assigned to each possible merge starting with neighboring pixels (Baatz & Schape, 2000). These costs represent the degree of fitting (*h*) in *d*-dimensional feature space, where *d* is the number of features (*f*):

$$h = \sqrt{\sum_d (f_{1d} - f_{2d})^2}. \qquad \text{Eq. 1}$$

The feature I used in segmentation is mean grey-level intensity values, but it can be defined as any characteristic feature (e.g., color, context, size, shape, shadow, texture, location, pattern, height). The pairwise merging of larger objects increases average intensity heterogeneity in the lidar derivatives, so an optimization procedure to minimize the incorporated heterogeneity at each merge is implemented in the algorithm. The degree of fitting for two adjacent image objects, denoted by the subscripts 1 and 2, is described by the change of heterogeneity (h_{diff}) such that:

$$h_{diff} = h_m - \frac{h_1 n_1 + h_2 n_2}{n_1 + n_2}, \qquad \text{Eq. 2}$$

where (*n*) is object size, and $h_m$is the average heterogeneity of image objects. The algorithm is designed to take into account different sizes of objects and different behaviors on different stages of scale since objects of interest typically appear simultaneously on different scales in an image (Baatz & Schape, 2000). By increasing the scale threshold, larger objects will be created, although their exact size and dimensions are dependent on the underlying data.

The final segmentation used in the study, multi-threshold segmentation, also splits the image into objects based on pixel values, but unlike MRS does not incorporate shape, and applies to a single image. This creates image objects and simultaneously classifies them based on user-created thresholds (Figure 5(C)). Objects can be assigned to a class or be

unclassified based on the defined threshold values. The segmentation also uses a scale parameter. At a scale equal to 1 all pixels in an image object meeting the threshold criteria are segmented and classified. As the scale increases, a larger number of pixels with a mean value meeting the threshold criteria are necessary for classification.

2.4 Preliminary Feature Extraction

To accommodate terrain variability and allow for model transferability to other geographic regions, a single, object-based workflow was designed to extract landslides across the diverse geologic setting. By employing a computational process that mimics the ability of humans to distinguish between landslide and non-landslide features, landforms interfering with landslide classification were extracted first. This excludes them from subsequent data analysis. In addition to interfering with landslide classification by adding roughness elements to the landscape, the classified features can also be used for context (i.e., proximity of objects to classified features). Use of context in this fashion provides a characteristic feature used by humans to interpret landforms, but also may be used in analysis, for example to determine proximity of a landslide feature to a road or stream. The features classified prior to landslides include roads, ridges, and streams. The Appendix describes the software packages and data used in this crucial first step of the model effort.

In addition to the landscape-scale features described above, smaller, isolated, steep features that the algorithm would confuse with landslide headscarps were also excluded. The process used contextual relationships to classify these isolated steep

slopes, defined as pockets of topography that are steeper than the surrounding topography (Difference to Neighbor (DTN) < -10 (Appendix (Eq. A2))), which occur next to previously classified road and stream features (Figure 6). The extraction of isolated steep slopes benefits the landslide classification routine in that it eliminates engineered road cut slopes and stream banks from the dataset, thereby improving landslide segmentation and subsequent classification results.

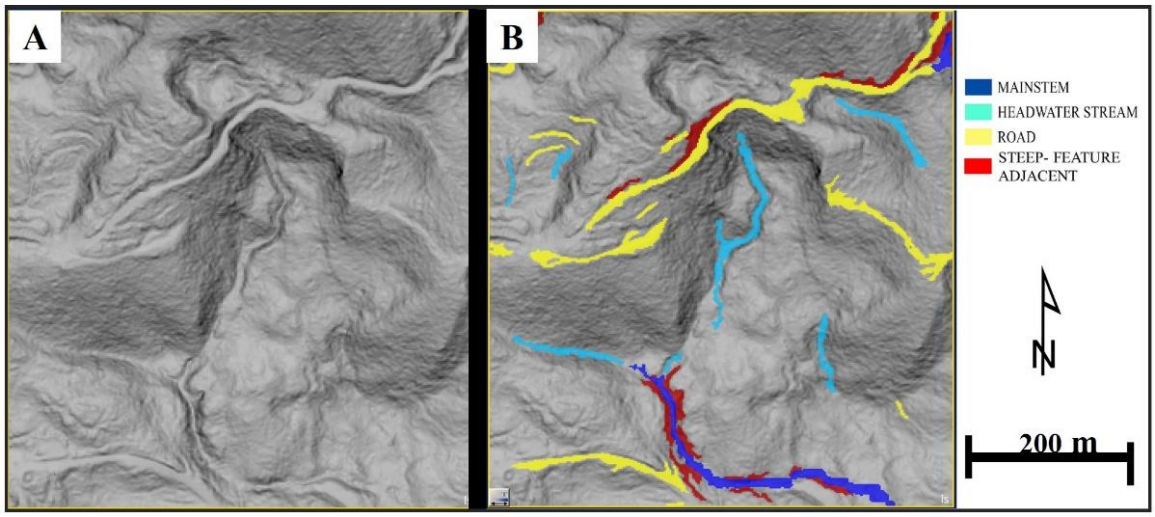


FIGURE 6. (A) LIDAR-DERIVED SLOPE IMAGE OF TOPOGRAPHY TYPICAL OF THE STUDY AREA PROVIDED FOR VISUAL REFERENCE; (B) ILLUSTRATION OF FEATURES EXCLUDED FROM LANDSLIDE ANALYSIS.

2.5 Landslide Feature Extraction

2.5.1 Data

I'll first describe the data used to extract deep-seated landslides before discussing the workflow: segmentation of landslide objects, classification of objects using a custom roughness criterion, and the cleaning process that takes advantage of the use of context in image analysis routines to refine the classification.

Three maps of curvature are used for segmentation. These metrics are derived from a lidar DEM: 1) Bolstad's variant [BOLSTAD], 2) Horizontal [TAN], and 3) Vertical [PRO] curvature. Bolstad's variant is defined as the ratio of the number of concave to convex pixels within a defined window and is computed in ArcGIS v.9.3 using the Geomorphometry and Gradient Metrics Toolbox (Evans et al., 2014). Horizontal (i.e., tangential curvature) is the curvature of a normal section tangential to a contour line at a given point of the topographic surface, and vertical (i.e., profile curvature) is the curvature of a normal section having a common tangent line with a slope line at a given point on the topographic surface (Florinsky, 2016). Horizontal and vertical curvatures are computed using a custom executable written in C++ and created by Weyerhaeuser staff. The BOLSTAD data is computed using a rectangular 15 x 15 pixel moving window (30 x 30 m at 2-m resolution), while the tangential and vertical data use a window size of 21 x 21 pixels (42 x 42 m at 2-m resolution). The large window sizes were chosen by trial and error to capture small-scale (less detailed) topographic changes. The pixel values in the three curvature datasets are converted to a scale of 0 to 255 and used as non-dimensional intensity values for computational efficiency within eCognition image analysis software, which works well with multi-band 8-bit imagery (Appendix: Eq. A1).

The second step in the landslide extraction process, classification, uses a topographic gradient [GRAD] image, calculated with the Spatial Analyst tool available in ArcGIS software packages. This tool uses a 3 x 3 cell moving window to quantify percent rise of the lidar DEM, which ranges from 0 to infinity (Burrough & McDonell, 1998).

Percent rise values are inverted prior to stretching to the 0-255 scale to arrive at the GRAD image (non-dimensional intensity values) used in eCognition. Inversion of the dataset is somewhat arbitrary, but it facilitates visualization of landform features within eCognition while the rule set is being designed. A rule set is a knowledge-based system that is created to translate understanding of an image into a process by which features can be extracted automatically. More importantly, the stretching routine does not change the distribution of the data; it simply enables all datasets used in the algorithms to have the same range. From the intensity values of the GRAD image a neighborhood similarity feature (Appendix: Eq. A2, Figure A23), mean difference to neighbor (DTN), is quantified. To computer the feature the GRAD image is segmented using the chessboard algorithm with scale value set at 1 (i.e., the native pixel resolution). The feature quantifies neighborhood similarity, or the intensity difference between an image object and its neighbors, by comparing the relationship of the center object to each neighboring object within a given window. The window size I use to calculate DTN is 15 x 15 objects (30 x 30 m at 2-m resolution). This window size is chosen by trial and error and emphasizes the small-scale topographic changes. The feature is exported as a grey-level image with values ranging from -50 to 50. Bright portions of the DTN image highlight terrain features with a large, positive difference to neighbors while dark portions of the image display large, negative differences. In this case, high positive DTN values correspond to low gradient terrain, relative to neighboring terrain.

2.5.2 Image Analysis Techniques & the Interpretation Process

In the following section I will describe the use of levels in image analysis, followed by a conceptual example that describes the correlation of human interpretation to computer interpretation of images.

In object-based image analysis routines, levels store image objects that represent information found in the image layer data. Whereas image data exist when first imported into the eCognition software package, image object levels are created to store this data as image objects and can range in size from the dimensions of a pixel up to the size of the image bounding box (i.e., the scene extent). Each level in a project may have a super-level above it, where multiple objects can be assigned to a single class, and a sub-level below it, where each object can be subdivided into multiple classes. Ouyang et al. (2011) provide a cartoon to illustrate the use of data connectivity in hierarchical processing (Figure 7(A)). In this example the project contains five levels with objects at the native image pixel size occurring on the bottom level while the top level contains one object to delineate the entire image boundary. A simple example of this processing structure occurs in vegetation mapping where a super-object on the highest processing level is classified as a forest with the sub-objects on a lower level classified as individual tree species. This hierarchical processing allows for statistics and contextual relationships describing all image object levels to be used (Figure 7(B). Context is image object networking that allows each object to know who its neighbors are, which levels and objects are above it (super-objects), and which are below it (sub-objects). In this study the technique is used to simulate perceived differences in the topography by allowing

classification of data at a sub-level for objects created at a smaller scale to be used in statistical class definitions on the super-level due to hierarchical connectivity.

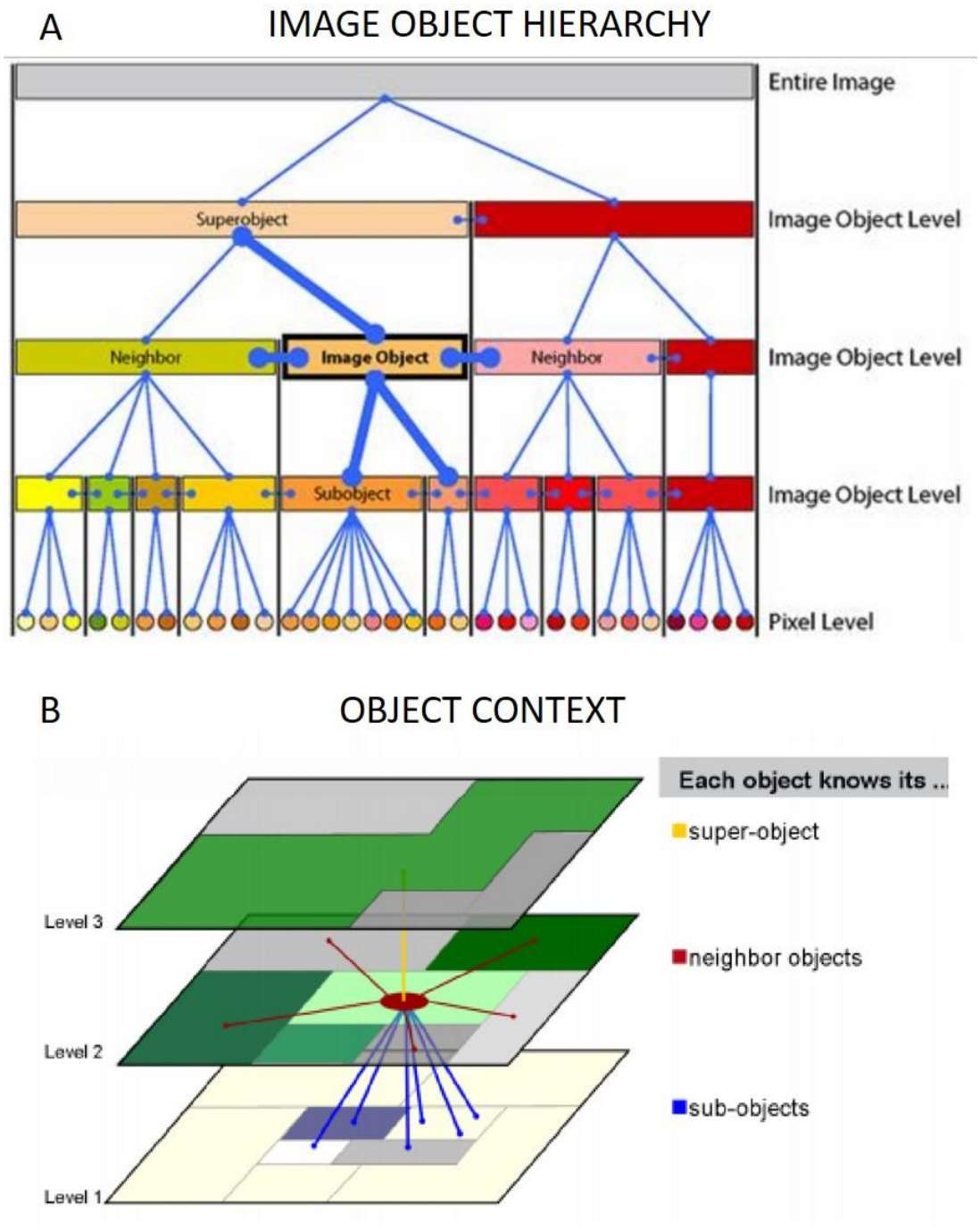


FIGURE 7. ILLUSTRATION OF USE OF IMAGE OBJECT HIERARCHIES IN OBJECT-BASED WORKFLOWS (OUYANG ET AL., 2011).

To illustrate the target for delineated object boundaries I show a conceptual model of a deep-seated landslide from the study area along with the image data used in the modeling effort. Figure 8(A) provides an aerial photo view of a landslide from within the study boundaries with components labeled in black. Upon visual inspection of the aerial photography, geologists interpret the feature as a landslide based on presence of landslide crowns, flanks, toes, and cracks. The human brain can recognize a difference in terrain based on color, texture, and contextual relationships to interpret the boundary. The semi-automated model is designed to use heterogeneity of image intensity values from the curvature data shown in Figure 9 to delineate features that a geologist "sees" in the image. A successful segmentation will draw boundaries around the landslide feature, creating objects that will later be used in classification. Figure 9(A-C) shows the curvature data used in this study at the same extent as the landslide in Figure 8. A gradient image is provided for visualization (Figure 9(D)). Curvature data that is sufficiently separated (i.e., intensity values abruptly transitioning from greys to blacks and whites) can be seen in the vicinity of the left flank of the landslide and continuing up around the crown. This area is defined by a transition from the smooth, planar topography (greys) into strikingly different curvature values (whites and blacks). A meaningful segmentation will draw a boundary on this left flank, continue up around the arcuate crown, back down the left flank and close by transecting the toe of the landslide.

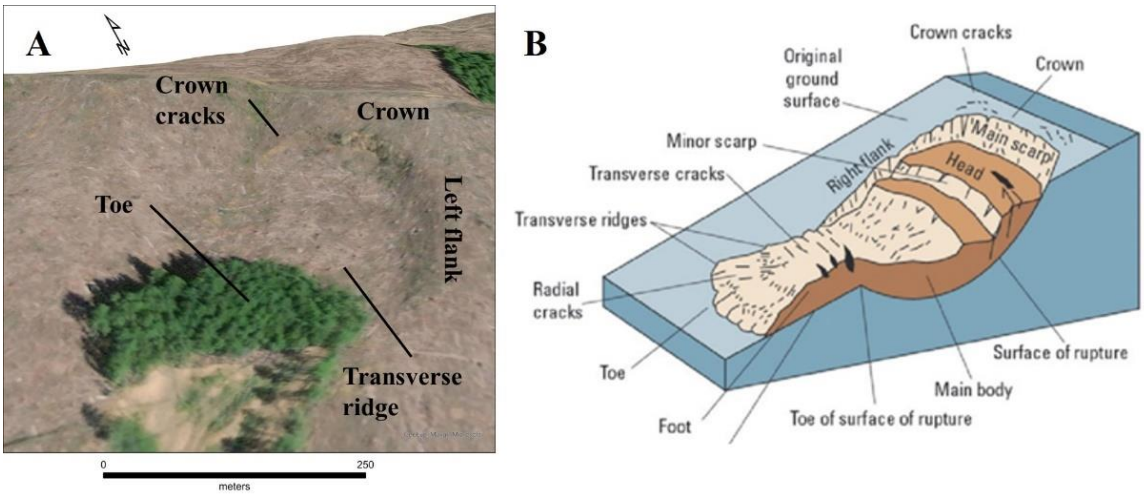


FIGURE 8. (A) 3D VIEW OF AERIAL PHOTO IN LOCATION OF DEEP-SEATED LANDSLIDE USING LIDAR ELEVATIONS FOR GROUND HEIGHT REFERENCE; (B) ILLUSTRATION OF CONCEPTUAL DEEP-SEATED LANDSLIDE COMPONENTS (CRUDEN & VARNES, 1996)

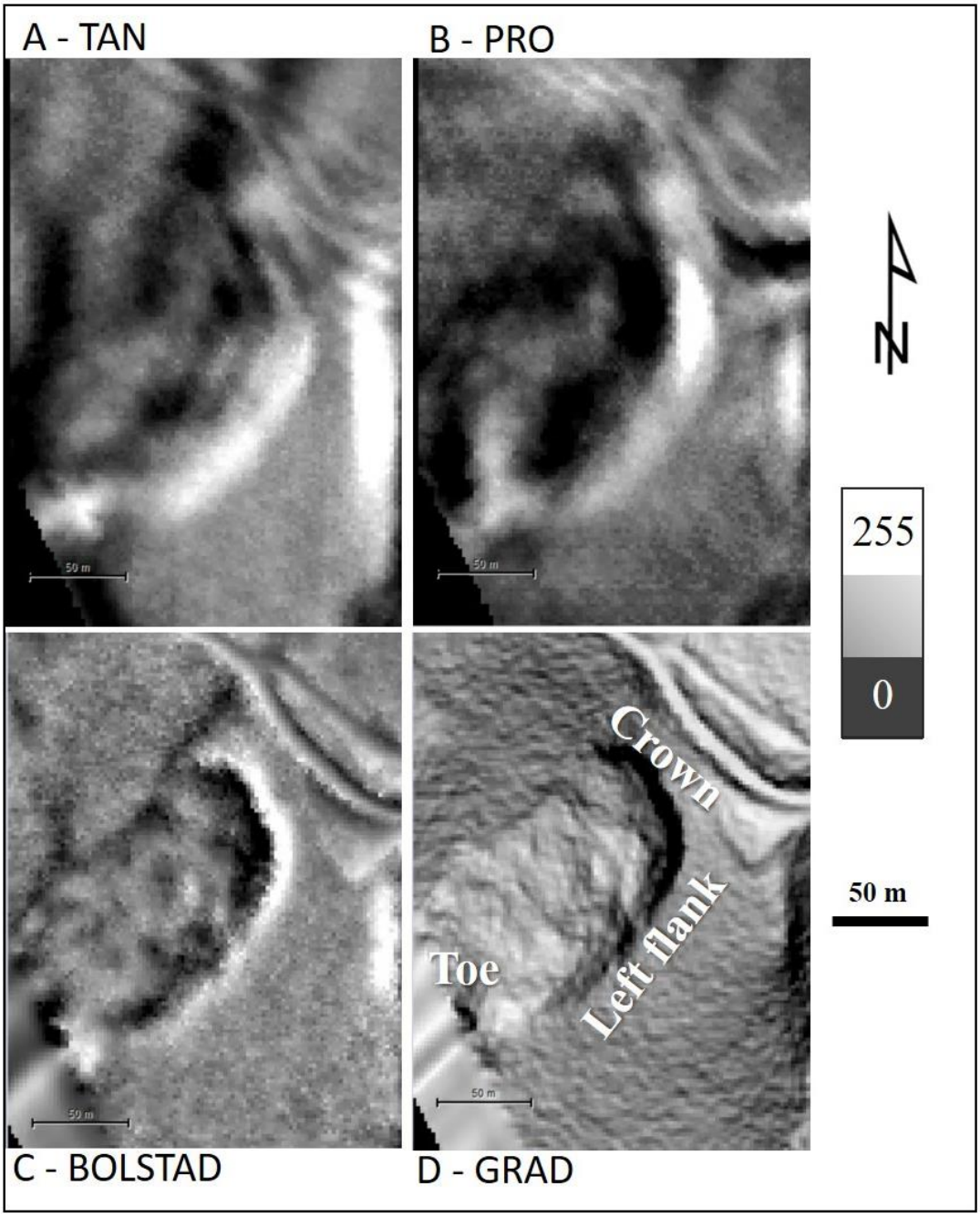


FIGURE 9. LIDAR DERIVATIVES, CONVERTED TO INTENSITY VALUES ON THE SCALE OF 0-255, SHOWN IN SAME EXTENT AS FIGURE 8. (A-C) CURVATURE IMAGES WITH BRIGHT PIXELS REPRESENTING CONVEX FEATURES AND DARK FOR CONCAVE (TAN: CURVATURE IN THE DIRECTION OF CONTOUR, PRO: CURVATURE IN DIRECTION OF SLOPE, BOLSTAD: RATIO OF THE NUMBER OF CONCAVE TO CONVEX PIXELS WITHIN A DEFINED WINDOW); (D) GRADIENT IMAGE WITH LABELS FOR KEY LANDSLIDE FEATURES (GRAD: TOPOGRAPHICAL GRADIENT).

2.5.3 Model Methods

Data is processed on two image object levels, the super-object level and the sub-object level. On the super-object level, three MRS segmentations are executed in conjunction: two object-merging routines that group curvature datasets using iteratively increasing scale factors, followed by an object-merging routine based on the DTN image. The three-step segmentation process minimizes issues of over segmentation caused by decreasing heterogeneity along the lateral margins of landslides. The first two region merging segmentations consider 3-bands of curvature data to produce the object boundaries: BOLSTAD, PRO, and TAN. The data are used collectively as the input parameter dataset, similar to using the red, green, and blue bands of spectral imagery to segment a photograph. The scale parameter is set at 20 for the initial segmentation to create the first set of objects from the pixel-based data and then is increased to 40 using the same 3-bands of data to enlarge the resulting object size (Figure 10(A) & Figure 10(B)). To define each unique landslide using a minimum number of objects, the third segmentation algorithm, a top-down region growing routine is employed to merge the objects at a scale of 30 using DTN data values based on the first derivative of the DEM (Figure 10(C)).

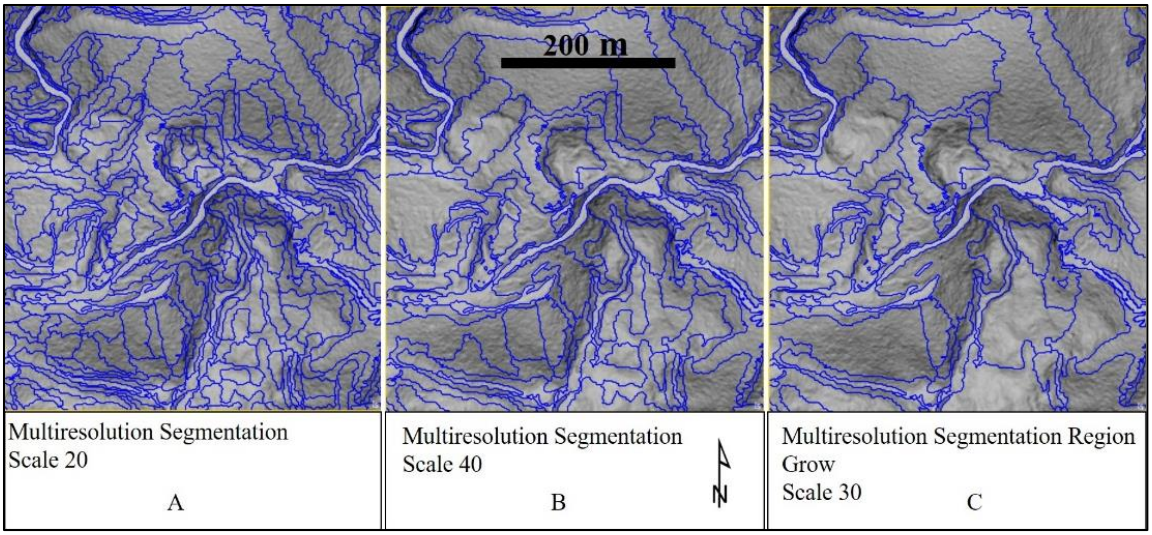


FIGURE 10. ILLUSTRATION OF ITERATIVE OBJECT CREATION WITHIN ECOGNITION SOFTWARE.

Next, the sub-object level is created to classify the objects defined by the third segmentation described above, which is based on the DTN image roughness criterion. Use of the sub-object level allows creation of new, smaller objects that do not alter the boundaries created on the super-object level. The DTN image is segmented and classified in a single step using the multi-threshold segmentation algorithm within eCognition. An algorithm scale is set at 20, and acts as a minimum mapping unit. Use of scale in the multi-threshold segmentation algorithm ensures no pixels will be classified unless 20 pixels meeting the threshold criteria exist in a cluster. This is an advantage over pixel-based thresholding in that it can help eliminate noise in the data created by isolated pixels or small clusters of pixels with large DTN values. The scale value was chosen using a trial an error approach so that classified objects replicate the approximate scale for human interpretations of roughness, or what is “seen” as rough surfaces or landslide hummocks. The classification on this sub-object level then includes two classes, any groups of DTN

image pixels with a mean intensity value under negative five, and a second class for groups of pixels with mean DTN values larger than five.

With objects delineated on the sub- and super-object levels, the next step is classification of deep-seated landslides. Class definitions are built using features available within the eCognition software package or features that are manually defined. Specifically, I used three criteria: a novel roughness criterion, a geometry criterion, and a slope threshold.

To assist with classification, I created a roughness criterion after noting that roughness can be “seen” in the DTN image. This image is chosen for use in the roughness criterion because its derivation (Appendix Eq. A2) renders it a detrended image, enabling transferability to deep-seated landslide terrain in all states of activity. Obviously, if no hummocks exist due to smoothing of the surface, alternative features of the deep-seated landslide would need to be defined and classified with additional routines, for instance, breaks in slope where weathered headscarps transition into benches. This criterion is defined using 1) a threshold for the ratio of the sum of the area of the classified sub-objects to the area of the super-object and 2) the variance of the DTN intensity values within each image object on the super-object level.

Geometry is also used in this study to refine class definitions. The eCognition Density feature describes the distribution in space of the pixels of an image object. It is defined as the square root of the number of pixels forming an object divided by its radius - approximated with an ellipse that is based on eigenvalues of the covariance matrix (1 plus the square root of variance in X plus variance in Y, where (X,Y) are the coordinates

of the center of an ellipse) (Appendix: Eq. A4). As the long axis of the bounding ellipse gets larger while the number of pixels remains low, the density value decreases. Higher values represent objects with more dense shapes, closer to that of a square.

The third feature used to refine the deep-seated landslide class is a mean slope threshold. A distinction is made between hillslopes over 60% slope and under 60% slope, as morphological properties related to landslide processes are thought to differ as the hillslope steepens. Note that this is not the gradient intensity image described earlier. The slope threshold was determined based on Rule Identified Landform (RIL) guidance in the Washington Forest Practices Act, which requires geotechnical evaluation of and specific management prescriptions for landforms that are potentially unstable (Shaw et al., 2017). Definitions for a suite of landforms present in forested terrain include slope thresholds, the lowest of which is 65% for active deep-seated landslide toes (WDNR, 2016). I chose to use a conservative value of 60%.

The iterative process of refining the deep-seated landslide class begins by defining all super-objects with a mean slope under 60% as a landslide if they contain a sub-object DTN classification area ratio of greater than 20% and DTN intensity variance (standard deviation squared) larger than 20 (i.e., roughness criterion = DTNarea >20% and DTNvar >20), along with a dense geometry (Density >1.2). The use of geometry excludes long and narrow objects such as streams or knife-edge ridges. All values used in this classification process were chosen by trial and error based on my interpretation of where landslides existed on the landscape. This initial landslide class is grown into adjacent unclassified objects with a relative border >15%, mean slope under 60%,

roughness criterion = DTNarea >15% and DTNvar >15, and Density >1.2. A second growing routine classifies dense objects (Density >1.2) that share a relative border >20% and contain >400 pixels of previously classified sub-objects with DTN values >5. Finally, all objects previously classified as landslide are unclassified if the relative border to a stream or steep, stream adjacent class is >50%.

Next, objects with a mean slope >60% are classified as landslides if roughness criterion = DTNarea >40% and DTNvar >40, and Density >1.2. These objects are grown into adjacent unclassified objects with a relative border of greater than 40%, and a roughness criterion = DTNarea >20% and DTNvar >20. A second growing routine classifies dense objects (Density >1.2) that share a relative border >20% AND contain >400 pixels of previously classified sub-objects with DTN values >5.

Finally, all classified landslides are merged, and a cleaning procedure is completed to incorporate any unclassified objects that exist as an island within the spatial distribution of classified deep-seated landslides.

2.6 Manual Inventory Methods

The landslide inventory necessary for model comparison is compiled manually by four geologists by attempting to create a consensus landslide inventory map for two test sites labeled with blue polygons in Figure 2. The purpose of seeking consensus among multiple geologists is to reduce inherent professional bias existing in maps produced by independent observers. It is recognized that the number of observers selected for this study may not have been sufficient for a validation study, but it does provide data for a

statistical comparison of mapping differences produced by diverse backgrounds. Four professional slope stability geologists, with extensive backgrounds in field and remote sensing-based landslide mapping techniques, were enlisted to complete the deep-seated landslide mapping exercise. The mappers had a range of experience, some with decades of field-based mapping expertise, and some formally trained in the identification of hillslope morphologies. During round one, the geologists were given access to the same lidar DEM used in the semi-automated landslide model. Each geologist then created derivatives from the lidar DEM at their own discretion, which included slope maps, contours, and hillshades. It was the geologist's choice to access additional information such as existing landslide inventories or imagery datasets including those provided by the National Agriculture Imagery Program (NAIP) or Google Earth Imagery. Each geologist then created an independent, digital inventory of deep-seated landslides, guided by the Varnes (1978) landslide classification system (Figure 11). For the purpose of this exercise the surface expression of the landslide needed to be visible beyond the 1:3000 map scale with an aerial footprint larger than 500 m^2. Thematic attributes collected were an identification number and landform. The identification number is simply a unique

number assigned to each deep-seated landslide, and the landforms collected were classified as landslide body, primary scarp, and secondary scarp.

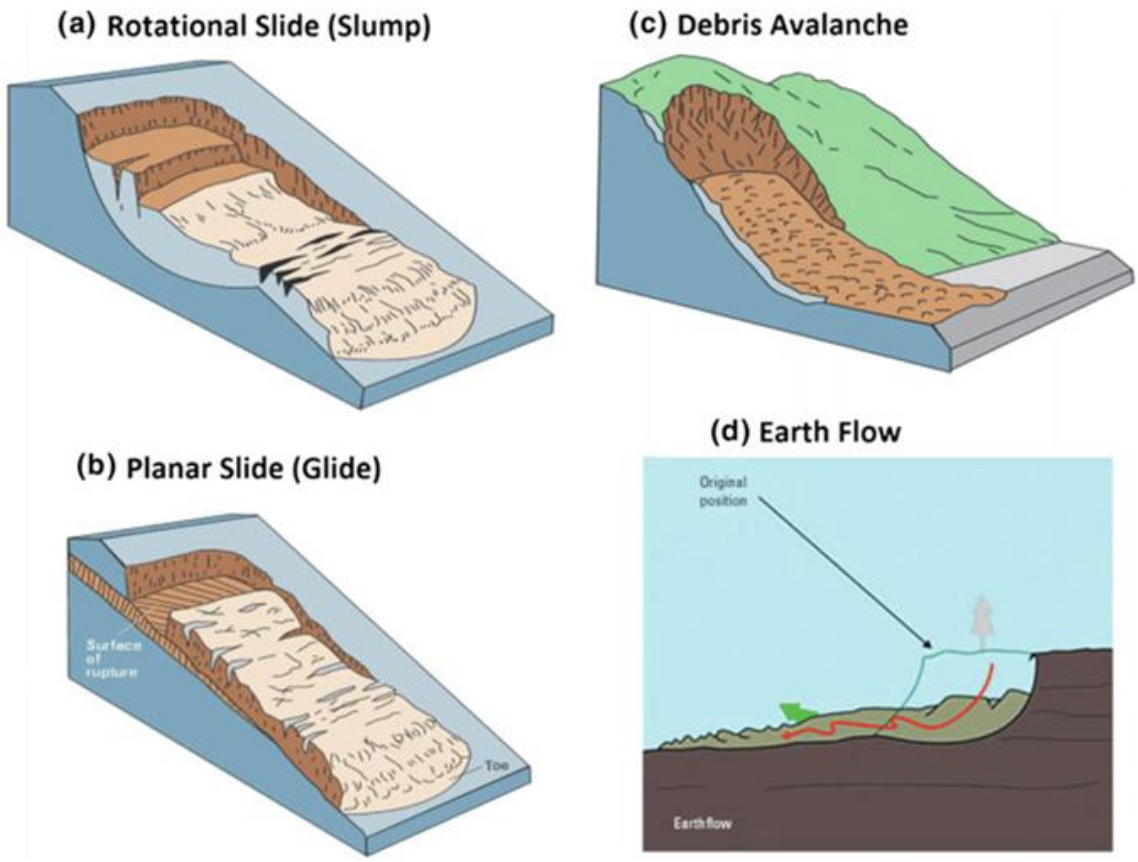


FIGURE 11. CLASSIFICATIONS FOR DEEP-SEATED LANDSLIDES (VARNES, 1978). (A-C) RECORD SLIDE-TYPE MOVEMENT WHILE (D) RECORDS FLOW MOVEMENT.

I then conducted an intermediate GIS exercise to "clean" and create a union of the independently created inventories. First, all landforms were merged, creating one polygon for each unique landslide, thereby eliminating the scarp distinctions. An original intent of the study was to evaluate landslide model performance against landslide scarps and bodies, but that was not achieved due to time restrictions. It was hypothesized that agreement between observers regarding the terrain features that constituted a landslide

scarp would be lower than what constituted landslide bodies. Next, a union of the four independently mapped inventories was performed, resulting in polygons where spatial intersections of inventoried landslide components were identified. The polygons retained thematic attributes including identification of the initial polygon observer (i.e., the dataset allowed for quantification of agreement between professional observers for each polygon). All polygons that were less than 0.001 km^2 in size were dissolved into adjacent boundaries, under the assumption that these were likely minor border discrepancies.

A qualitative review of the agreement between the independent mappers revealed that large differences existed in landslide numbers, densities, geometries, and extents. To enable simple quantification of the agreement between the modeled landslide inventory and a reference inventory, a “true” landslide population was desired. Ideally the reference inventory would consist of a map designating the entire verification extent as either landslide or not a landslide. This was attempted, during round two, by creating a composite inventory map recording the interpretation of each mapper for each landslide polygon to determine what population of landslides should exist. The goal was founded on a premise that an in-depth geomorphological discussion involving all four mappers, regarding the factors that led to the interpretations would resolve differences in interpretive styles and produce a consensus decision for each feature in the combined dataset.

The composite inventory records the evolved interpretation or modified landslide designation of the four mappers for each polygon in the inventory. This designation is reached after a group-negotiated process that results in a modified landslide designation

(i.e., attribute). The attribute can be the same designation as the independent interpretation or could have changed after considering geomorphological evidence presented by the other mappers. In the absence of a complete consensus reached for all features in the inventory, the effort resulted in derivation of four qualifiers, one of which was assigned to each feature in the landslide inventory: Definite, Questionable, Split-decision, and Non-consensus (Table 1). The two endmembers for mapper agreement are Definite, which describe spatial polygons where four observers agreed that a landslide existed, and Questionable, where all mappers requested the landslide polygon be removed from the inventory. The Split-decision category contains polygons where the evolved interpretation (i.e., interpretation of the observer after group discussions) of half of the observers recorded the feature as a landslide, while the other half did not. These features were often large, inactive-old landslides. The final category, Non-consensus, recorded polygons where either split or consensus decision was not reached for reasons that included differences of opinion regarding geomorphic processes, or map-based problems resolving features, including data processing techniques and/or data resolution. This Non-consensus category also included features where all four observers agreed that there was not enough information to make a confident decision regarding the initial mapping of the polygon as a landslide, and field verification would be necessary to make a proper determination.

LANDSLIDE QUALIFIER	DESCRIPTION	SPATIAL EXTENT
DEFINITE	All mappers' original or evolved interpretation is deep-seated landslide	1.1 km^2; 22% of survey area
QUESTIONABLE	All mappers' original or evolved interpretation is not a deep-seated landslide	0.1 km^2; 2% of survey area
SPLIT DECISION	50-75% of the mappers' original or evolved interpretation is deep-seated landslide	0.9 km^2; 18% of survey area
NON-CONSENSUS	Either 25% of mappers' interpretation is deep-seated landslide or all mappers agreed that more information, such as a field visit, was necessary to make a distinction	0.3 km^2; 5% of survey area

TABLE 1. DESCRIPTION OF THE FOUR QUALIFIERS (I.E., THEMATIC ATTRIBUTES) FOR THE MANUALLY MAPPED DEEP-SEATED LANDSLIDE INVENTORY. THE DATASET INCLUDES 2.3 KM^2 OF LANDSLIDES INVENTORIED IN TWO SURVEY AREAS (TEST SITE 1: FIGURE 12 & TEST SITE 2: FIGURE 13).

The automated inventory was qualitatively compared to each of the four attributed Landslide Qualifiers of the composite inventory listed in Table 1. Agreement between the modeled landslides (data to be verified) and the manually mapped landslides (reference data) was quantified and reported for two categories of landslides: those that were inventoried as Definite and any landslide inventoried by the manual mappers (i.e., Definite, Split-decision, or Non-consensus).

3. Results

3.1. Manual Inventory

The spatial dataset housing the composite manual inventory results allowed analysis of agreement between mappers. This design was necessary due to the lack of complete agreement between the four geologists regarding which features on the landscape should be included in the final inventory. The final dataset is grouped into the four qualifiers listed in Table 1 recording evolved interpretation for each mapped polygon.

Evaluation of the compiled inventory dataset reveals that a sub-population of landslides defined by distinct surface morphology related to recent activity (active or inactive-young, Figure 1) were consistently identified and thematically attributed by all mappers. This active landslide sub-population covered 22% of the survey area (Landslide Qualifier: Definite; Figure 12 & Figure 13). A sub-population of more subdued features that are thought to be inactive-old or inactive-mature (Figure 1) were also consistently identified and attributed by half of the mapping team (Landslide Qualifier: Split decision). These features that have undergone substantial weathering existed on 18% of the model verification study area, just 0.2 km^2 less than the Definite sub-population footprint. A Questionable attribute was assigned to a limited number of mapped polygons (0.1 km^2) comprising 2% of the study area. The designation was applied primarily to features occurring on the margins of mapped landslides, such as an extension upslope, laterally, or downslope of a mapped feature, and was applied infrequently to anthropogenic features. The final attribute (Landslide Qualifier: Non-consensus) is

applied where at least one geologist originally mapped a landslide and could not be convinced through further conversation and data exploration that the interpretation should change, such that the landslide was not removed from the final inventory. Non-consensus attributes were often assigned to features that could be interpreted as severely weathered or eroded, and they cover 5% of the survey area. It should be noted that one possible reason for not making a Definite or Questionable call could have been limitations in using remote sensing data (i.e., a field visit was deemed necessary to decide if the interpreted feature had previously failed). Ninety-six percent of the original mapped landslide areas were included in the final compilation inventory, suggesting discussions with fellow geomorphologists had limited effectiveness to persuade mappers to change an interpretation that the terrain had failed.

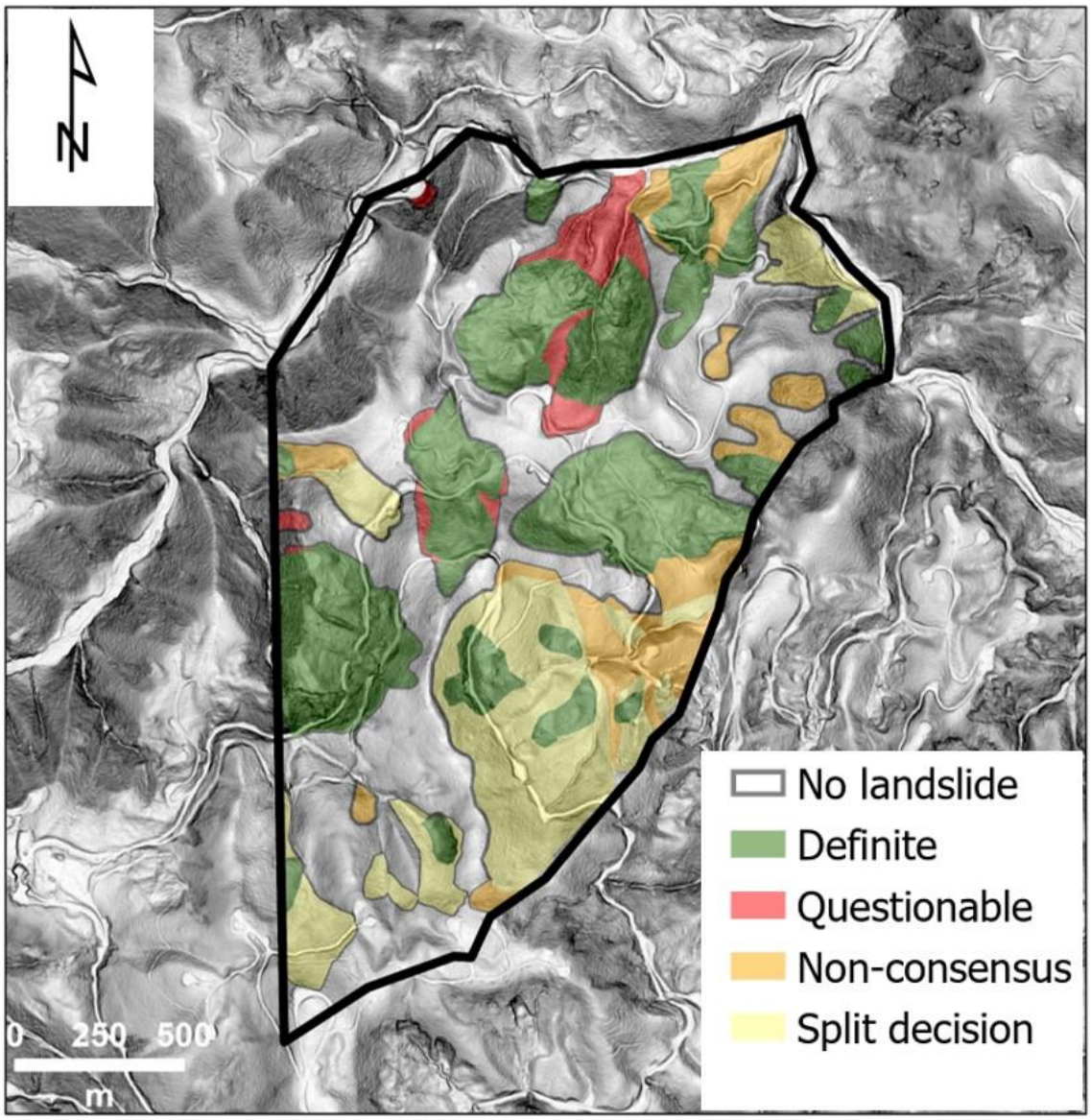


FIGURE 12. RESULTS FROM COMPOSITE, MANUALLY MAPPED, DEEP-SEATED LANDSLIDE INVENTORY; TEST SITE 1. FOUR POINT BLACK LINE DELINEATING EXTENT OF MANUAL MAPPING EFFORT.

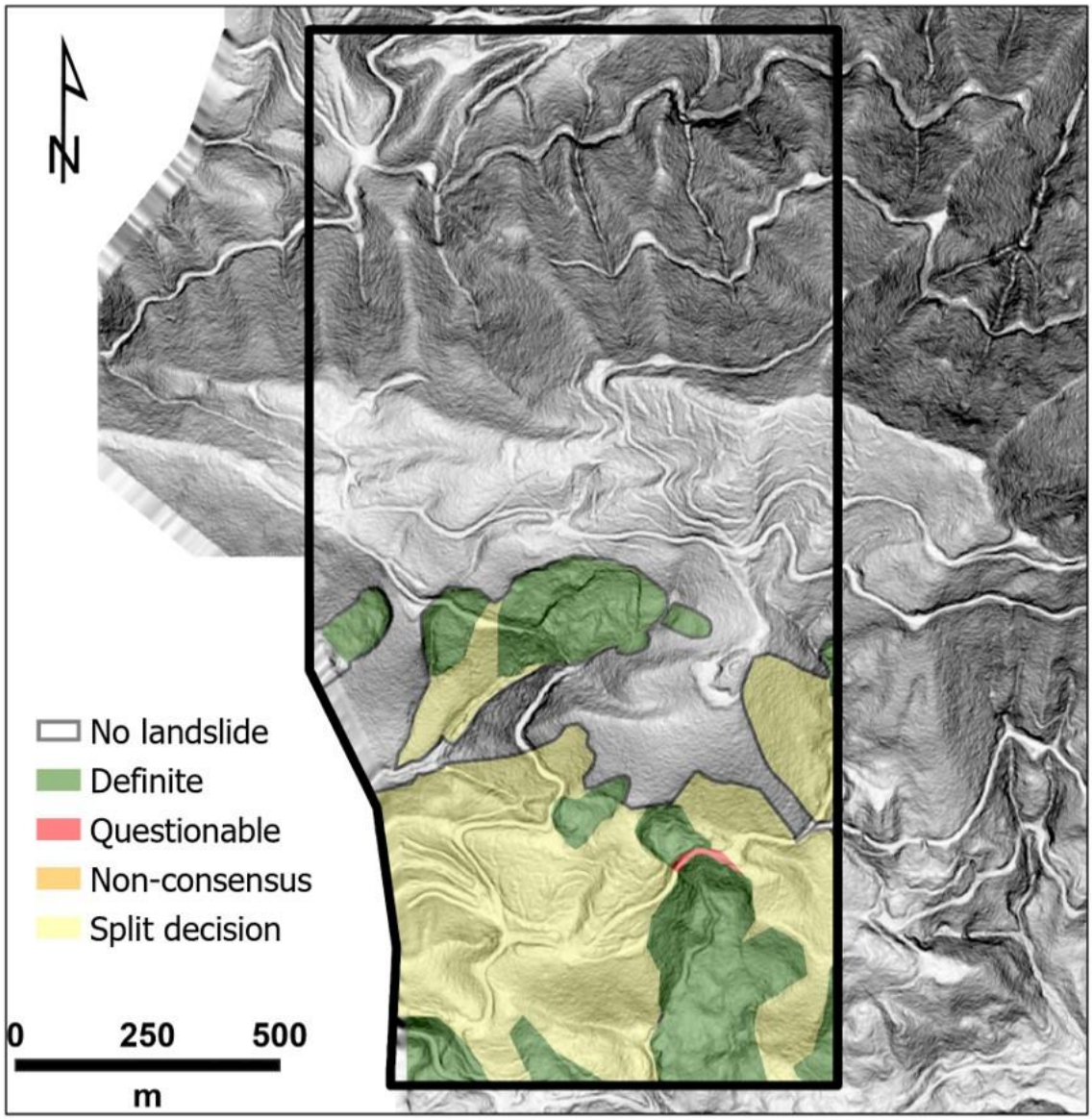


FIGURE 13. RESULTS FROM COMPOSITE, MANUALLY MAPPED, DEEP-SEATED LANDSLIDE INVENTORY; TEST SITE 2. FOUR POINT BLACK LINE DELINEATING EXTENT OF MANUAL MAPPING EFFORT. LIDAR COVERAGE ENDS TO THE WEST OF THE SURVEY AREA, CREATING IRREGULAR BOUNDARY.

Standard accuracy metrics are calculated, following previous automated landslide mapping work, to examine variability, or level of agreement, in the composite map products (Congalton & Green, 2009; Korzeniowska, 2017). Accuracy, precision, and

recall metrics are commonly used for comparing classification results in large spatial datasets. To compute the metrics, confusion matrices are populated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. True means the model, or data to be verified, correctly classified a reference landslide area (TP) or non-landslide area (TN). False means the model incorrectly classified an area recorded in the reference dataset as a landslide (FN) or non-landslide area (FP). Accuracy provides a good starting point to quantify the area of correctly predicted landslides in relation to the total area. Accuracy is defined as the sum of TP and TN divided by the basin area, and as such depends on the distribution of landslides in a given area. Precision, defined as TP divided by the sum of TP and FP, and recall, defined as TP divided by the sum of TP and FN, need to be considered to better understand the performance of the model, specifically for use in landslide hazard and risk assessments where a FN can be more disastrous than a FP. Precision, defined as the portion of modeled landslide area that intersects the reference landslide area, includes both correctly and incorrectly modeled landslide terrain (TP and FP). Precision answers the question: “Of the landslide area delineated by the model, what area was correctly identified?”. Recall, the portion of reference landslide area that was identified by the model, takes into account only reference landslide terrain (TP and FN). In the case of landslide classification, it is particularly important to consider recall considering it is advantageous to classify an area as a landslide when it may not be, as opposed to the converse if public safety or resources are at stake.

Formulas using the ratios of TP, TN, FP, and FN values are provided under each metric displayed in Figure 14 (Congalton & Green, 2009). Confusion matrices are shown

with columns representing the reference data, which is assumed to be the true population of landslides for each comparison, and rows representing data from the map being verified. Unlike the formulas for recall and precision, the formula for accuracy produces the same value if an observer's data is considered as reference or if it is considered as data to be verified. The duplicate values are show with X's in Figure 14. Computations use polygon area within landslide footprints due to availability of polygons in both verification and reference datasets. Alternatively, a per-pixel approach could have been employed. In the absence of a comprehensive landslide inventory mapped by an unbiased expert, we first evaluate how closely each mapper aligned with the other mappers in their interpretations (i.e., the agreement between mappers).

	ACCURACY				AREA
	(TP+TN)/BASIN AREA				MAPPED
	A	B	C	D	(KM²)
A	X	0.77	0.87	0.78	1.8
B	X	X	0.71	0.93	0.79
C	X	X	X	0.73	2.7
D	X	X	X	X	0.87
	RECALL				
	TP/(TP+FN)				
	A	B	C	D	
A	X	0.93	0.77	0.93	1.8
B	0.40	X	0.35	0.75	0.79
C	0.92	0.97	X	0.99	2.7
D	0.45	0.83	0.39	X	0.87
	PRECISION				
	(TP)/(TP+FP)				
	A	B	C	D	
A	X	0.40	0.92	0.45	1.8
B	0.93	X	0.97	0.83	0.79
C	0.77	0.35	X	0.39	2.7
D	0.93	0.75	0.99	X	0.87

FIGURE 14. COMPARISON OF AGREEMENT IN LANDSLIDE INVENTORY MAPS PRODUCED BY THE FOUR INDIVIDUAL MAPPERS A-D. REFERENCE DATA IN COLUMNS AND VERIFIED DATA IN ROWS. X'S EXIST IN CELLS WHERE DUPLICATE VALUES ARE COMPUTED, OR WHERE A VALUE OF 1 IS COMPUTED WHEN THE REFERENCE DATA IS IDENTICAL TO THE DATA BEING VERIFIED.

Accuracy ranged from 71-93%, documenting that each pair of mappers identified a large majority of the terrain in the same way. Recall values ranged from 35-97% with lower values for mappers who missed landslides delineated by the reference observer. Similarly, precision ranged from 35-99%, where low-end values were recorded for mappers who inventoried areas not mapped by the reference observer. A mapper with low precision compared to another mapper therefore tended to have high recall compared to that same mapper. The range in recall and accuracy metrics illustrates that a style difference, in landslide interpretation or mapping methods, exists within the mapping team. Two of the observers consistently delineated features interpreted to be older, more weathered landslides, which lose contrast to neighboring unfailed terrain. While the other 2 observers consistently delineated features interpreted to be more recent failures, identifiable by crisp, discernible geomorphological features.

The statistical metrics used to calculate agreement between the mappers delineated landslide area are sensitive to the set of mappers being compared. Accuracy values for landslide areas mapped by observers B and D are 93%. When considering landslide areas mapped by observers B and C accuracy drops over 20%. Accuracy statistics include the area of TN. In this example mapper C delineated a larger area of landslides, reducing the area of TNs in the statistical calculation, leading to reduced accuracy statistics compared to mappers recording relatively less landslide area for the same test site. Precision and recall statistics more dramatically record this sensitivity to the set of mappers being compared. Recall statistical values for mapper C was over 90% when compared to all other mappers as reference (i.e., the mapper was delineating a large

percentage of landslides recorded by the other observers). This mapper inventoried the largest amount of area in the study (2.7 km^2). However, mapper C's precision ranged from 35-77%. The lower precision values reflect relatively larger areas of FPs (i.e., non-landslides) being recorded. The opposite is true for mapper B, who mapped the least amount of study area as landslides (0.79 km^2). Mapper B's recall ranged from 35-75%, while the precision was over 83%.

The choice of which statistical metric (accuracy, precision, or recall) to base management decisions on may vary with the goal of the landslide inventory. For example, when the goal is to map all previously failed terrain (i.e., deep-seated landslides) on the landscape, high recall values are desirable, indicating that mappers are consistently classifying all deep-seated landslides that exist on the landscape. If the goal is to test the similarities in map products, precision should be considered.

3.2. Modeled Inventory

Prior to reviewing the quantitative results to verify the modeled landslide area, a qualitative assessment of the extracted geomorphometric features is summarized, providing an overview of how the model is segmenting and classifying deep-seated landslides. The model first segments the image based primarily on curvature values, and then classifies the curvature segments by assessing the amount and variability of a roughness criterion within each segment. This method allows for systematic delineation along breaks in topographic curvature. The breaks of interest, while trying to model deep-seated landslides, are the boundaries that occur where the hillslope is transitioning from

concave to planar or convex surface curvature expression. Within these concave areas, surface roughness is evaluated to determine class membership. The spatial results, presented in Figure 15 & Figure 16, suggest the object boundaries generally do a good job of following curvature breaks, at a scale that is large enough to view rough surface features within each object.

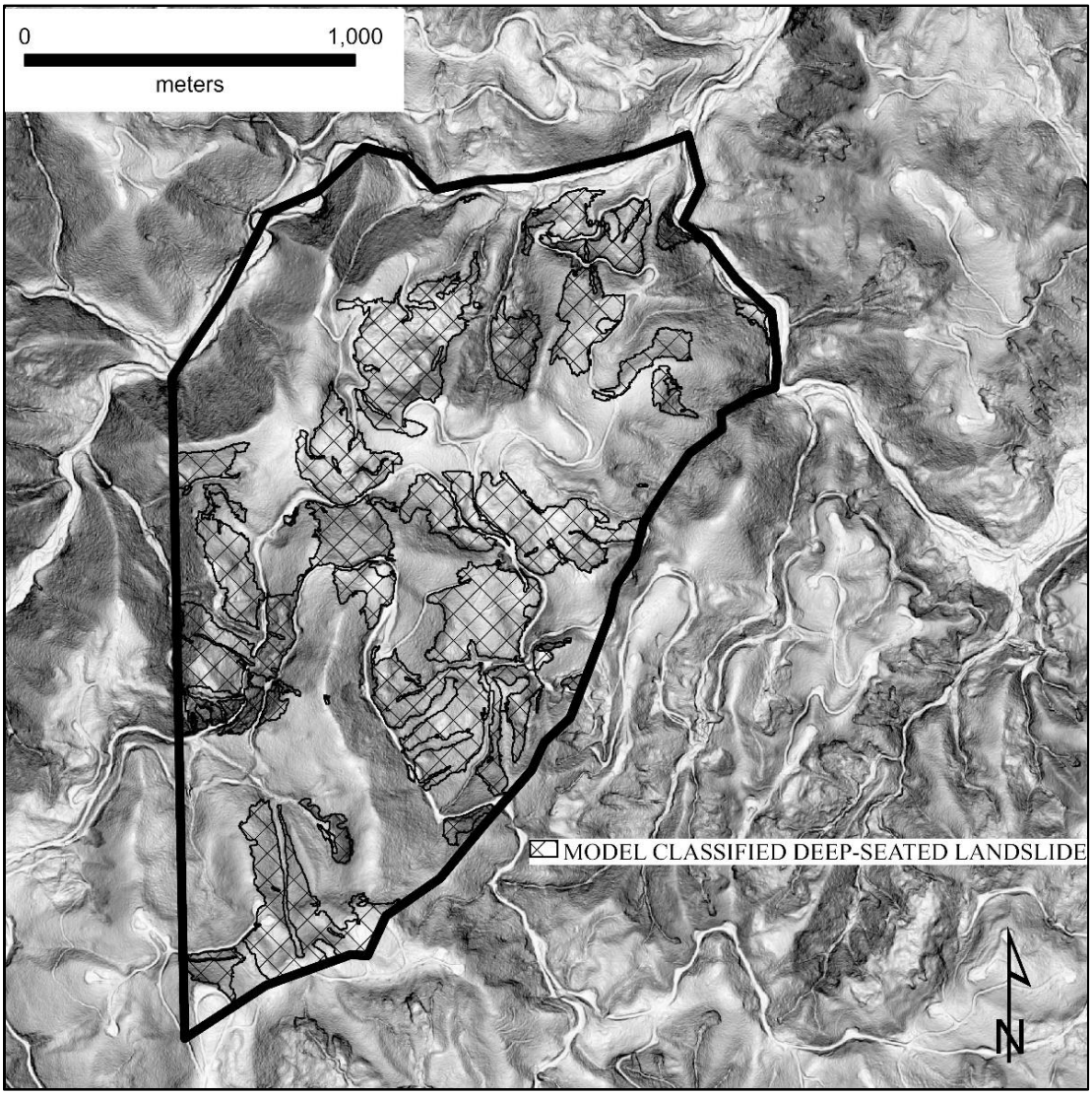


FIGURE 15. GEOBIA MODELED LANDSLIDES IN THE EXTENT OF MODEL VERIFICATION TEST SITE NUMBER 1.

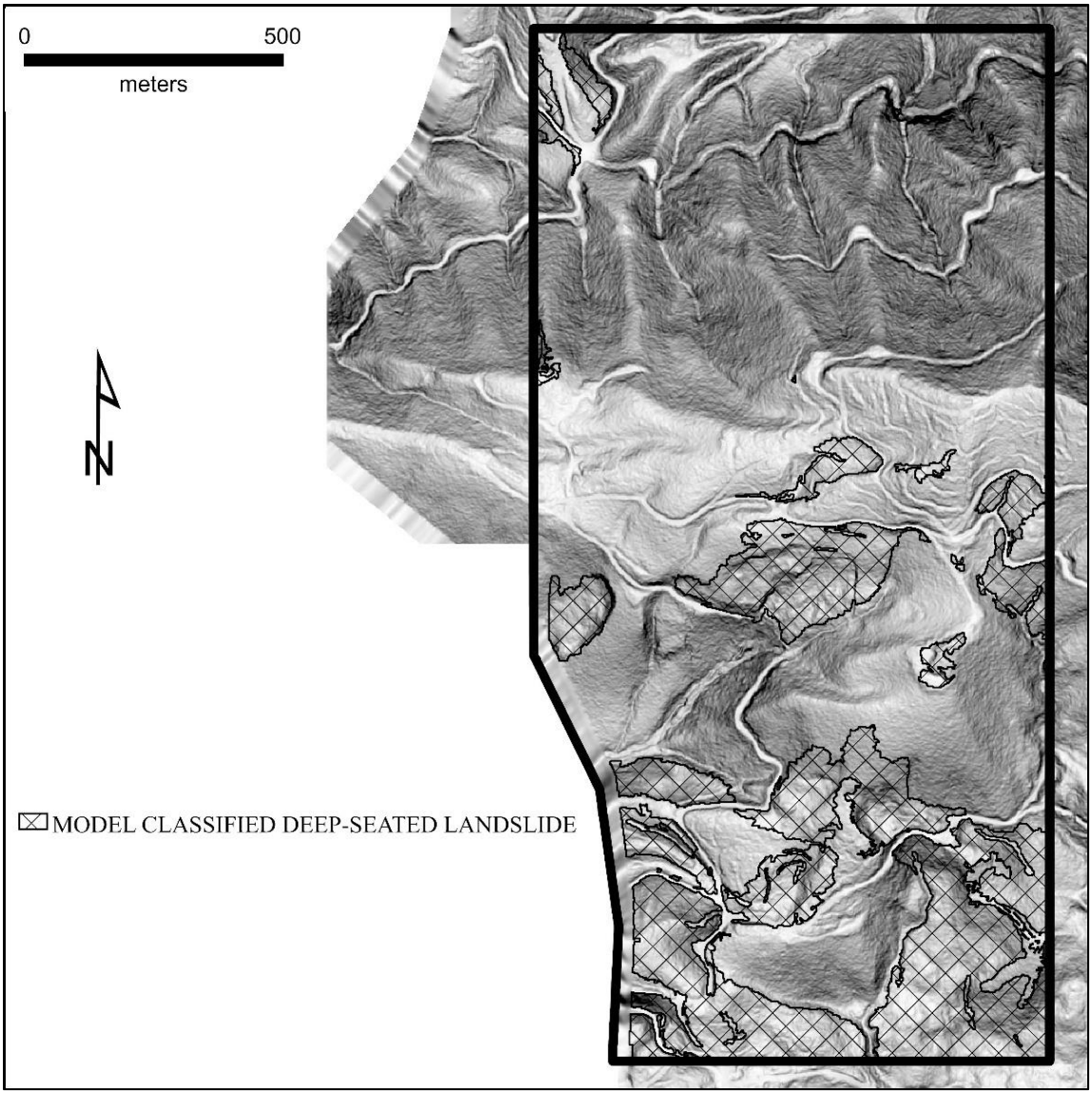


FIGURE 16. GEOBIA MODELED LANDSLIDES IN THE EXTENT OF MODEL VERIFICATION TEST SITE NUMBER 2.

3.2.1. Model Strengths

The model detects changes in curvature and slope at a range of scales enabling large and small landslides to be classified with the same routine. To illustrate, I present an example of two deep-seated landslides with drastically different areas, 0.1 km^2 vs. 560

m^2. The largest of the two classified landslides is in the south to southeast corner of the map centered in the largest of two black rectangles (Figure 17). This landslide was delineated because the model identified distinct curvature expressions as compared to neighboring image pixels, and classified because the definition for roughness captured the obvious hummocks and breaks in slopes formed by internal scarps. A much smaller road-adjacent slump, visible in the southwest corner of the map (smaller of two black rectangles in Figure 17), is also segmented and classified. The model was able to detect the dramatic change in curvature between the steep scarp and the bench to delineate the object. Landslide classification definition also captured the irregular topography on the landslide bench. The same model data, algorithms, and parameters are used to classify both landslides, highlighting the ability of this object-based approach to successfully segment and classify features at different scales.

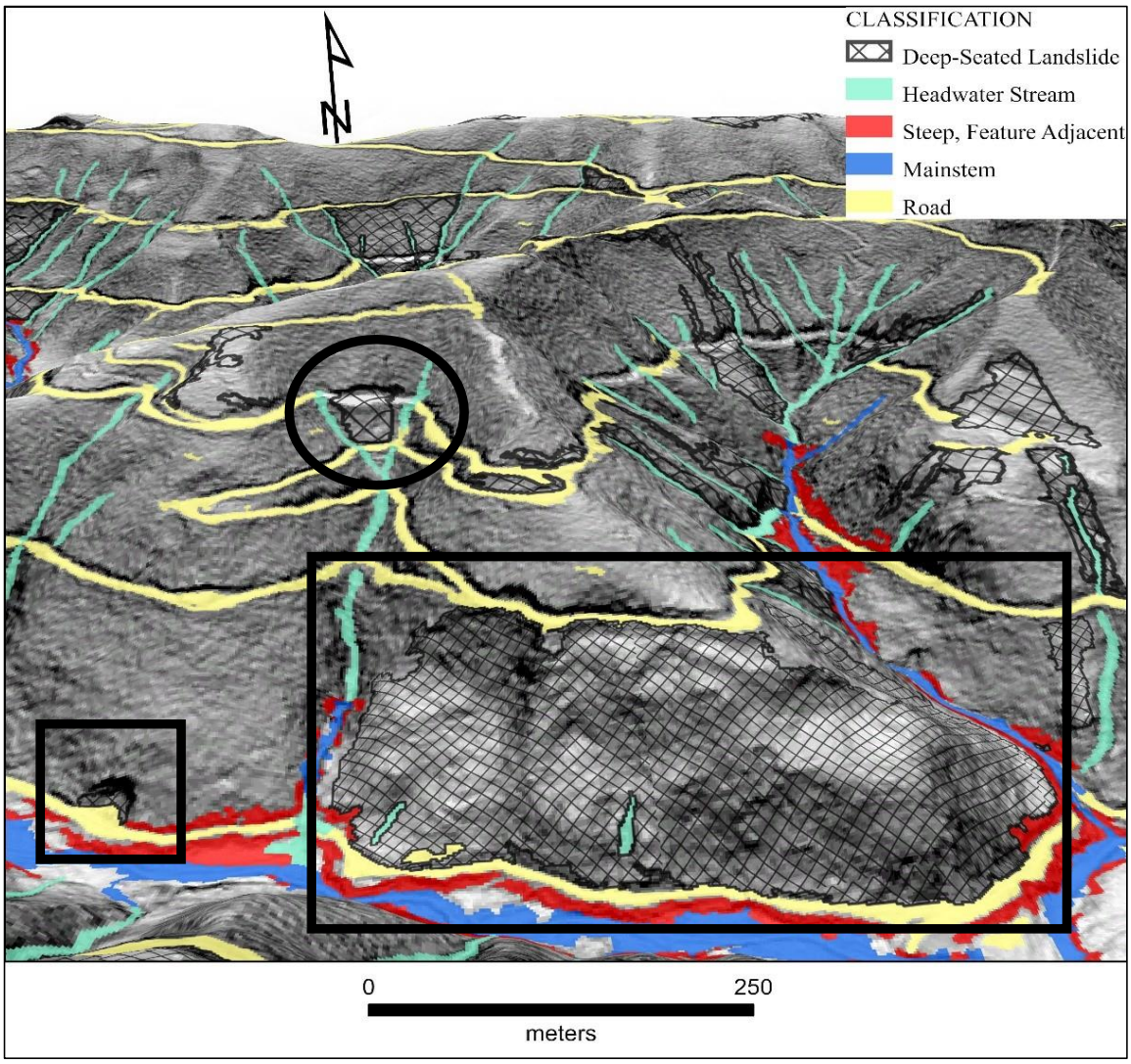


FIGURE 17. MULTI-SCALE FEATURE EXTRACTION EXAMPLE. SMALL LANDSLIDE CENTERED IN SMALL BLACK RECTANGLE ON THE LEFT SIDE OF IMAGE. LARGE LANDSLIDE ON RIGHT SIDE OF IMAGE CENTERED IN LARGER BLACK RECTANGLE. MISCLASSIFIED LANDSLIDE IN CENTER OF BLACK CIRCLE. ALL OBJECTS CLASSIFIED WITH SAME ROUTINE AND PARAMETERS.

The model relies on accurate classification of anthropomorphic features, such as roads, which are well characterized in this example. A road exists at the base of both landslides classified in Figure 17 (black hatches within black rectangles). If left unclassified the road would produce a roughness signal as the hillslope transitions to

smooth, low gradient topography, similar to the signal produced in the slump feature as the steep scarp transitions to a bench. It is likely that this roughness signal would be captured by the landslide classification, causing a misclassification of the object. It appears this was the case for at least one other object that was delineated near the head of the stream occurring between the two landslides being discussed (black circle in Figure 17). A segment of the road was not classified, leaving data values to be considered in the landslide analysis. In the absence of a road classification in the area highlighted within the black circle, the roughness definition for the landslide class was met, resulting in a portion of the hillslope being classified as a landslide that is most likely a road and road-adjacent topography. Forest roads have a similar gradient to that of landslide bench elements, and much like low gradient, bench deposits, they are distinct from their neighbors (i.e., steep slopes). An attributed example of the similarities between roads and benches, can be viewed in the DTN image, shown in Figure A32 of the Appendix. Proper classification of the roads using the geometric property of density (as described in Eq. A4 of the Appendix) is advantageous to using image values alone to minimize misclassification of landslides.

In addition to using the same data, algorithms, and parameters to successfully segment and classify features at different scales, the model is used to segment and classify different types of landslides that occur in diverse geologic terrain. The morphometric surface expression of an earthflow complex existing in the soft sediments of the McIntosh Formation is shown in Figure 18. In this terrain type there is an absence

of distinct headscarps, which would otherwise be indicated by abrupt changes in slope where a vertical step in the hillslope occurs, immediately upslope of low gradient benches within landslide deposits. Instead, the most notable earthflow features are depositional lobes occurring in a range of sizes. There are also large swaths of irregular topography, with numerous mounds and benches. Unlike the landslides occurring in steep topography, which are distinctly different from the long, smooth, neighboring hillslopes (shown in Figure 17), the earthflow complex occurs in more gentle terrain with gradients that are not drastically different from the neighboring slopes. The semi-automated model is designed to extract features that do not have excessively steep slopes by calculating slope differences within a 15X15 pixel window, effectively detrending the slope data. This allows for extraction of features having a more subdue difference to their neighbors over a large area (Figure 18).

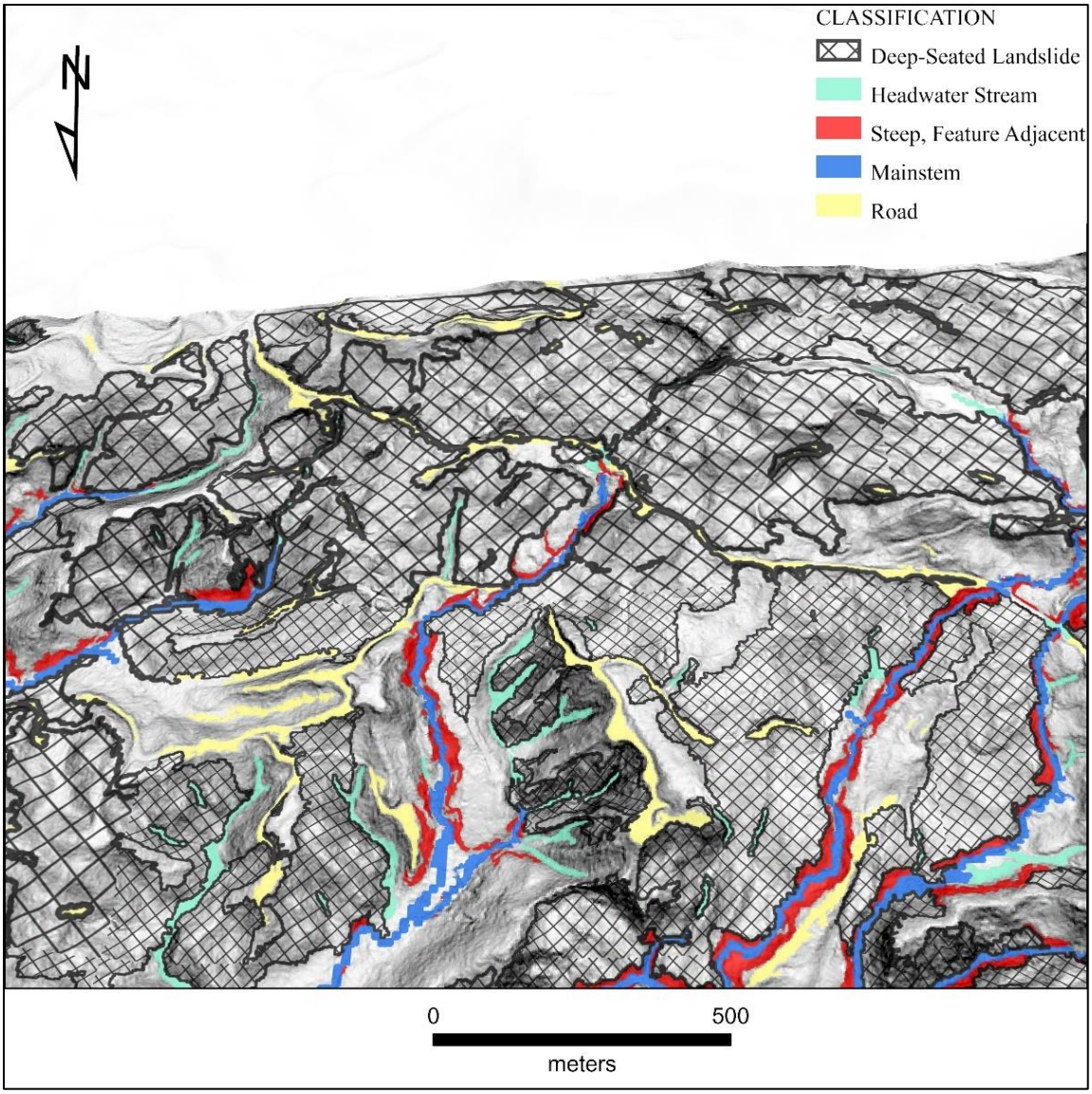


FIGURE 18. EXAMPLE OF FEATURE EXTRACTION IN COMPLEX TOPOGRAPHY.

3.2.2. Model Weakness

In this section I show an area where, in at least two instances, the use of contextual information (i.e., neighboring object relationships) led to misidentification of landslides. However, as a general observation, I have found that the use of contextual

information enhances model performance. Neighboring object classification is specifically useful in steep terrain with high contrast between hillslopes and valleys. In more subdue terrain the use of contextual information becomes more difficult as hillslopes transition gently into valley features. Irregular valley features often satisfy landslide classification definitions, designed to capture low gradient, rough surfaces.

Figure 19 illustrates one instance where the location of the road extracted in the preliminary classifications hindered the ability to use relative border relationship to stream features as a way to eliminate fluvial landforms from the landslide classification routine. Figure 19 shows a landform, identified with the black arrow, that was included in the inventory, and would have otherwise been eliminated based on context (i.e., an object with mean gradient under 60% with a large relative border to steep, stream adjacent features).

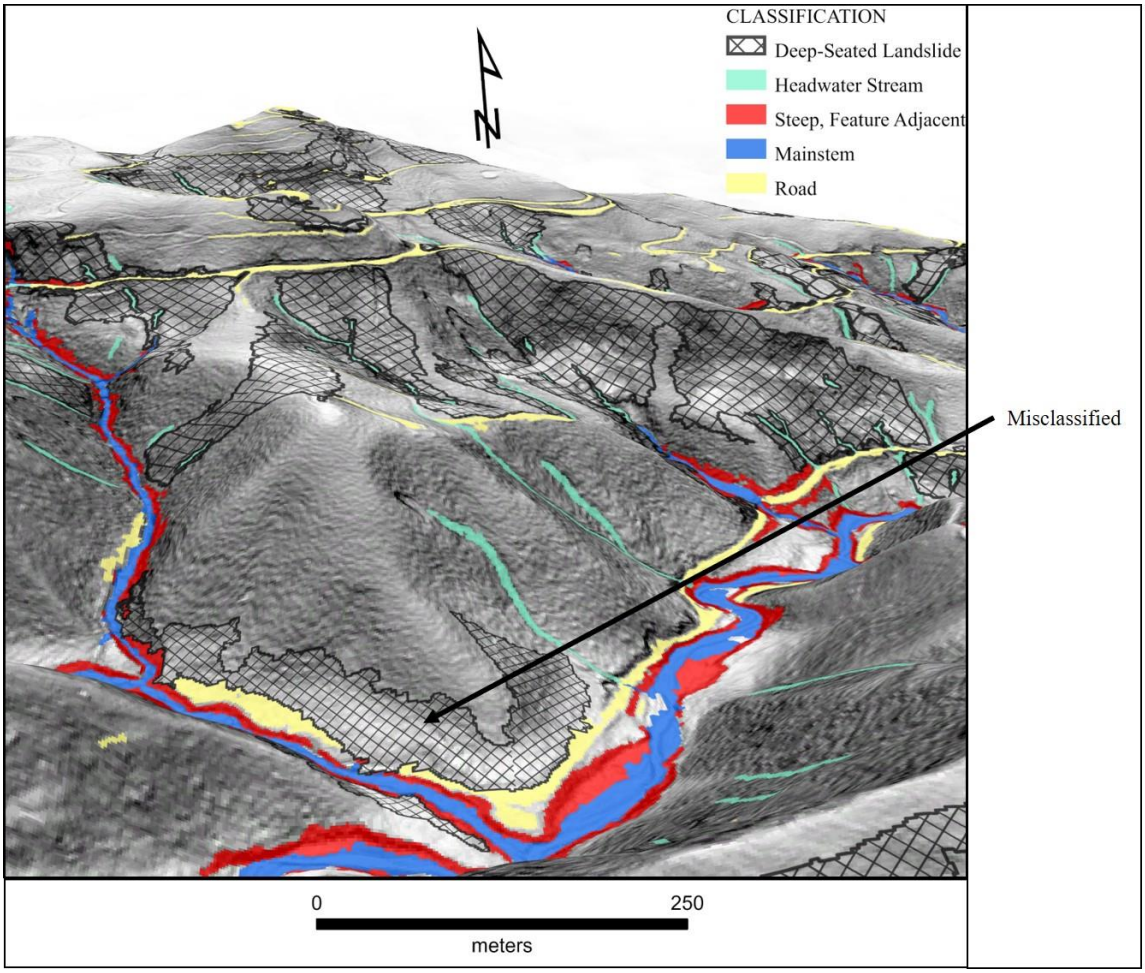


FIGURE 19. MISCLASSIFIED LANDSLIDE AT LOCATION OF ARROW.

In the second instance an object was eliminated from the landslide classification due to the use of context. This object (purple polygon) is shown at the location of the arrow in Figure 20. In this scenario the object was classified as a landslide based on the roughness criterion, but later unclassified because the large, relative object border to steep, stream adjacent features exceeded the threshold of 50%. Exacerbating the odds for misclassification is that the landslide object is bisected by a road, isolating the toe component, effectively lowering the mean object gradient, a criterion necessary to unclassify the object based on the relative border to the steep, stream adjacent feature.

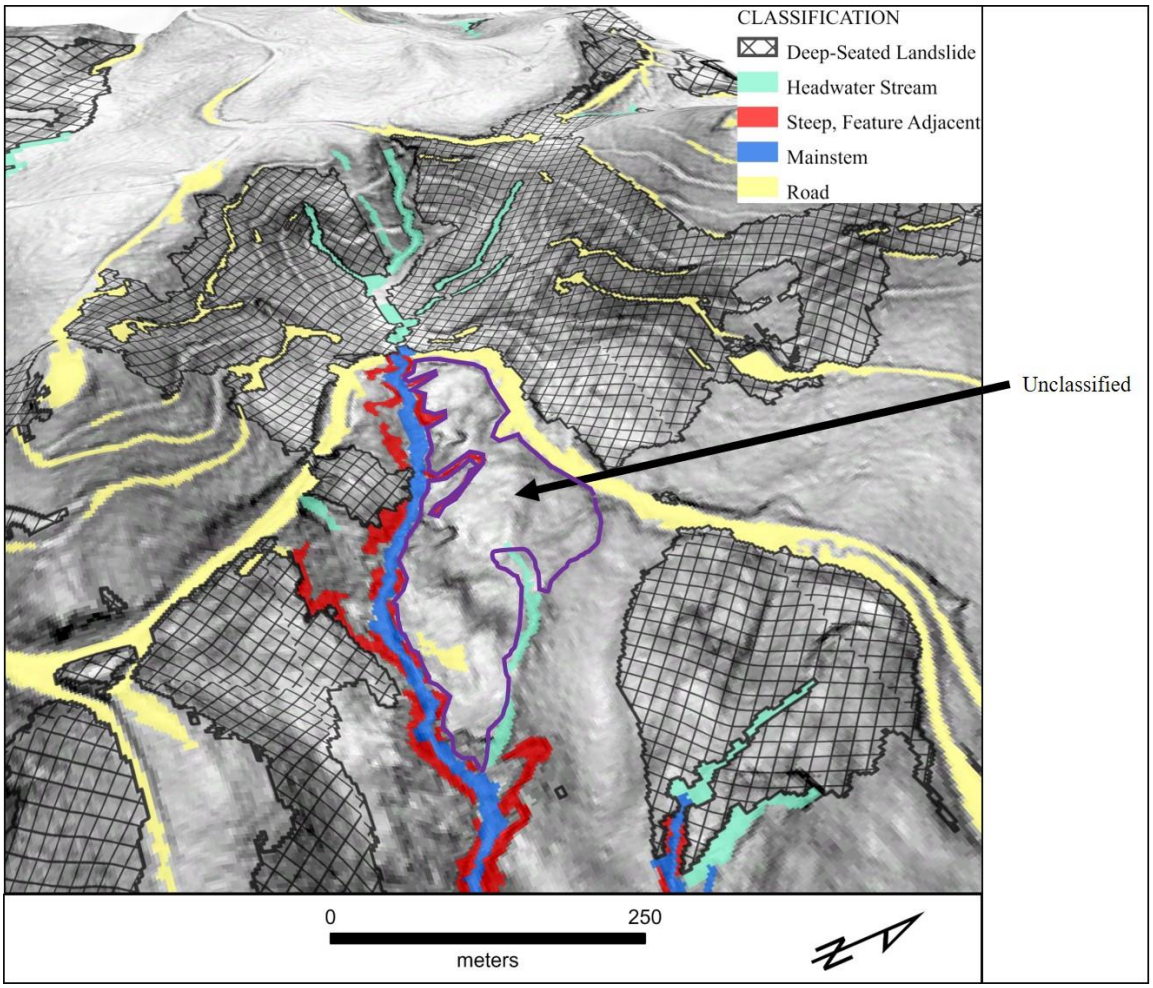


FIGURE 20. DEEP-SEATED LANDSLIDE THAT WAS REMOVED FROM CLASSIFICATION DUE TO LARGE RELATIVE OBJECT BOUNDARY (PURPLE POLYGON) TO STEEP, STREAM ADJACENT FEATURES.

3.2.3. Agreement Between Modeled & Mapped Landslide Inventories

Quantitative comparison between the modeled landslides and the manually mapped, composite inventory was completed by computing accuracy, precision, and recall of mapper-delineated polygons (i.e., reference data) and modeled deep-seated landslide polygons (i.e., data to be verified) (Figure 21 & Figure 22). To reiterate a point made in the methods section, the composite inventory includes the evolved interpretation

of the mappers (i.e., landslide interpretation after group discussions occurred). The two datasets (reference, and model data) are polygon shapefiles with overlapping geographic extents. Each polygon in the datasets contain an attribute assigning a classification that is specific to that dataset. Examples of attributes assigned to the modeled landslide data include modeled landslide, modeled road, modeled stream, modeled steep-feature adjacent, modeled ridge, or unclassified. Examples of attributes assigned to the compiled inventory include Definite, Non-consensus, Split-decision, Questionable, or no landslide. A union of the two spatial datasets was performed, resulting in polygons formed for each unique combination of attributes from the two inventories. Polygon areas for the thematically attributed data (i.e., landslide classifications from each dataset) were tabulated in Microsoft Excel for comparison and summarized in the agreement statistics. The same statistical metrics that were computed to examine mapper variability were used for this quantitative comparison (see section 3.1). The reference dataset (i.e., composite inventory) is summarized in the statistical matrices using one tabular dataset that includes only landslides thematically attributed as Definite, and one tabular dataset that includes all (i.e., any) polygon area inventoried as a landslide. This second reference dataset for all polygon area includes inventoried landslides attributed as either Definite, Split-decision, or Non-consensus. The distinction in the reference dataset was selected to compare model agreement with landslide polygons delineated by *all* manual mappers and landslide polygons delineated by *any* manual mapper. Establishment of these two sets of data allows for examination of areas where all geomorphologists agreed a landslide existed, as well areas where confidence may have been lower, but at least one professional believed enough evidence existed to constitute a landslide designation. This design ensured that I

did not have to make a decision regarding which inventory was correct, and instead all data collected was used in the analysis.

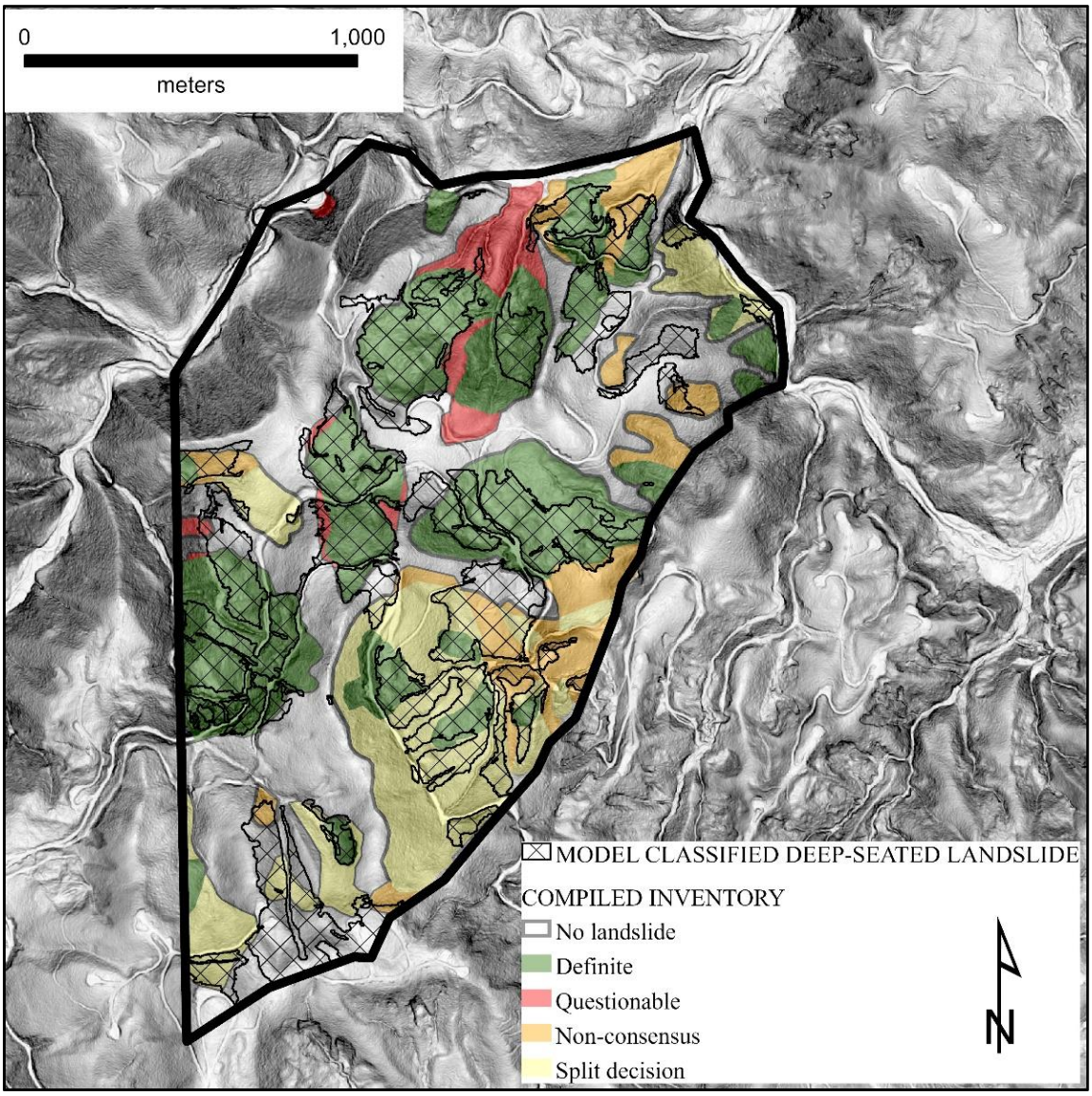


FIGURE 21. SPATIAL OVERLAP OF MANUALLY MAPPED LANDSLIDES (COMPOSITE INVENTORY) AND GEOBIA MODELED LANDSLIDES FROM THE FIRST MODEL RUN IN THE EXTENT OF COMPARATIVE SITE NUMBER 1.

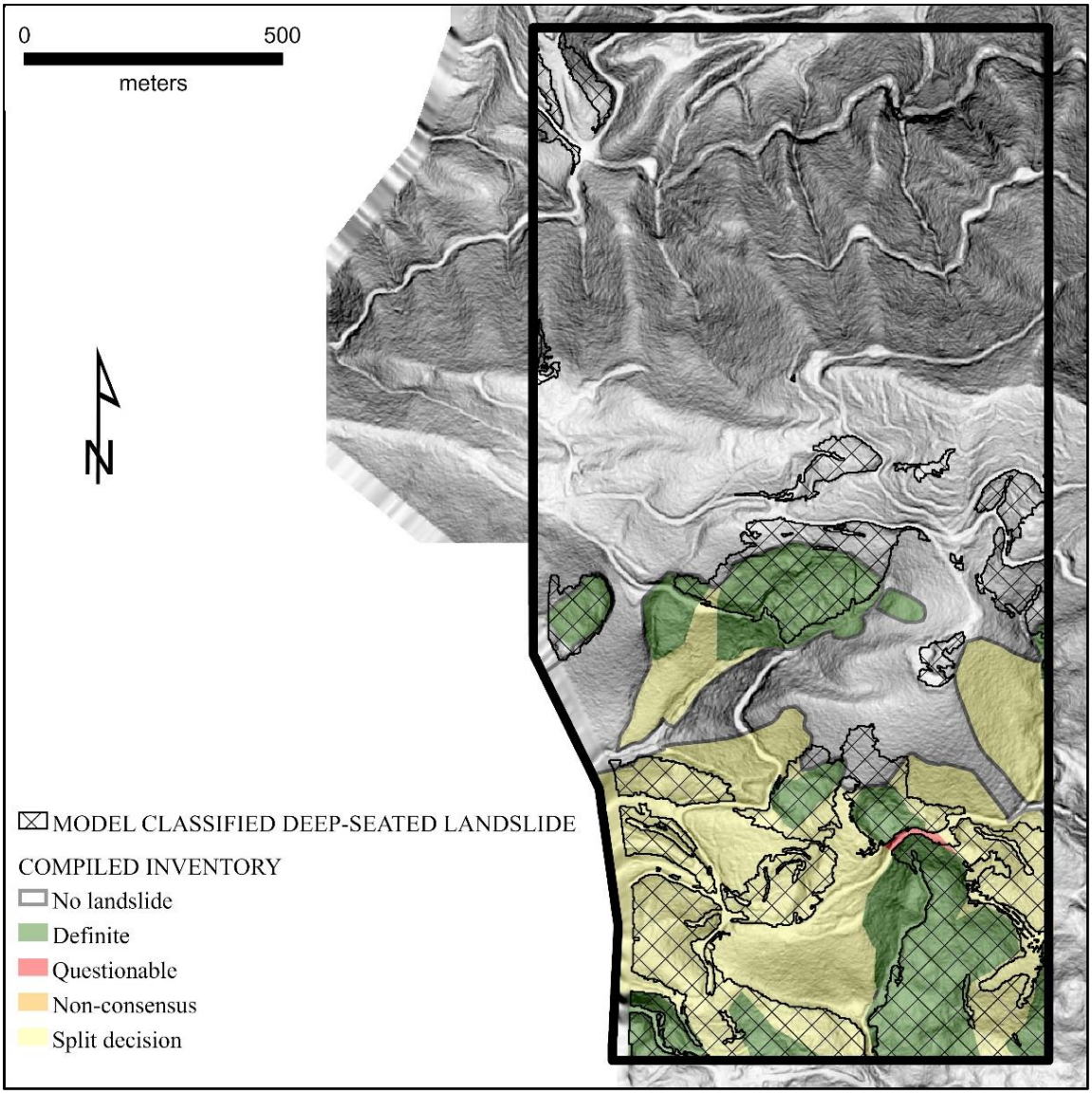


FIGURE 22. SPATIAL OVERLAP OF MANUALLY MAPPED LANDSLIDES (COMPOSITE INVENTORY) AND GEOBIA MODELED LANDSLIDES FROM THE FIRST MODEL RUN IN THE EXTENT OF COMPARATIVE SITE NUMBER 2.

A direct spatial comparison between modeled and inventoried (reference) landslides was not possible because of the ancillary feature mapping (roads, streams, ridges, and adjacent steep slopes) that occurred in the model prior to mapping landslides. Observers were not asked to interpret any non-landslide features, and results show that manually inventoried landslide polygons included features such as roads and streams

within the delineated landslide. For analysis purposes, any model-classified, non-landslide features overlapping the same geographic area as a manually inventoried landslide are tabulated as true positives. Similarly, all unclassified polygons from the composite inventory are compared only to modeled deep-seated landslide polygons. It is noted that the quantification of agreement statistics without considering all features on the landscape may skew the data in the favor of the model, since any features modeled in the same geographic extent as manually mapped landslides are considered landslides for the analysis.

Manually mapped inventory data from comparative sites 1 and 2, where identical mapping protocols were followed, is combined for use as reference data to report modeled landslide verification statistics. This study was originally designed to use one test site for model design and one test site for validation. Instead, model design remains in the verification stages, so data was used from both sites as reference. It is believed that the additional landslides inventoried in test site 2 provide valuable reference data for terrain with a higher density of landslides (earthflows in soft sediments), occurring in lower topographic gradients. This different terrain type will test the ability of the model to identify landslides in diverse settings. The model was developed to map landslides presented in results section 3.2.3, prior to completion of the manual inventory, ensuring the reference data was not used to train the model in any fashion.

Modeled landslide agreement, in terms of accuracy, to the reference data was high (77% or 81% of the comparative area, compared to Definite or all landslides, respectively), documenting that the model identified a large majority of the terrain in the

same way as the group of geologists (Figure 23). Recall, the ability of the model to map landslide areas identified in the reference datasets was lower, 76% for Definite landslides and 58% for all landslides. The model did not map all landslides well (i.e., older, inactive landslides with more subtle surface expressions). Model precision of 61% for Definite landslides, and 69% for all landslides in the comparative area suggests the model mapped additional landscape features as landslides that were not mapped by the geologists.

MODELED DATA - FIRST RUN - 1.4 (KM²)		REFERENCE DATA - COMPOSITE INVENTORY	
		1.1 KM²	2.3 KM²
		ACCURACY	
		(TP+TN)/BASIN AREA	
		DEFINITE	ALL LANDSLIDES
	GEOBIA	0.81	0.77
		RECALL	
		TP/(TP+FN)	
		DEFINITE	ALL LANDSLIDES
	GEOBIA	0.76	0.58
		PRECISION	
		(TP)/(TP+FP)	
		DEFINITE	ALL LANDSLIDES
	GEOBIA	0.69	0.61

FIGURE 23. LANDSLIDE MODEL VERIFICATION USING MANUALLY MAPPED LANDSLIDES AS THE REFERENCE DATASET. LANDSLIDE AREA FOR THE REFERENCE DATASET INCLUDE TEST SITE 1 & 2. THE REFERENCE LANDSLIDE DATA WAS DIVIDED INTO TWO COLUMNS (DEFINITE, AND ALL LANDSLIDES COMPRISED OF DEFINITE+SPLIT-DECISION+NON-CONSENSUS).

The model performed reasonably well, as recorded by the accuracy statistics, against both the Definite landslide dataset, and the dataset including all inventoried landslides reflecting different interpretive styles. Model accuracy was within the range of

individual mapper accuracies (71-93%). Recall statistics suggest the modeled landslide areas are correctly identifying landslides in the reference datasets more often than when comparing agreement (recall) between mappers with different interpretive styles. Recall is as low as 35% between manual mappers when the dataset being verified records a much smaller area of landslides (mapper B), compared to a reference dataset with a larger area of landslides (mapper C). Recall statistics also suggest the modeled landslide areas are correctly identifying landslides in the Definite reference dataset at a similar level of agreement (76% recall) as that recorded for mappers with similar interpretive styles. When comparing difference in statistical results for two mappers (A and C) who delineated the most amount of landslide area, including terrain with subtle surface expressions, use of landslide data delineated by mapper C for reference to verify landslide areas mapped by mapper A, recall is 77%. Similarly, for the two mappers who delineated the least amount of landslide area (B and D), use of landslide data delineated by mapper D for reference to verify landslide areas mapped by mapper B, recall is 75%. Precision statistics record lower values for mappers who delineate more landslide area incorrectly, compared to the reference dataset. Like recall, choice of reference dataset plays a large role in precision values. Precision statistics suggest the model is not identifying false positives (i.e., non-landslide features) as often as mappers with different interpretive styles. For example, when mapper D landslide data is used as reference to verify mapper C landslide area, precision is low (39%). This is compared to model precision using the Definite landslide dataset as reference, recording a precision of 69%.

4. Discussion

4.1 Landslide Activity Level, the Root Cause of Interpretation Disagreement

Visible geomorphic features were used by geologists to develop their interpretations of landslide presence and/or absence and to delineate and classify deep-seated landslides, that were then mapped and added to deep-seated landslide inventories. While landforms can suggest that mass movement has occurred sometime in the past, morphological evidence viewed in the remote sensing data is not always conclusive, for a variety of reasons that can include no discernable lateral margins or deposits, and severely weathered hummocks and headscarps. In the absence of fresh scars from active landslide activity, novel geomorphic features, not necessarily the ones encountered more frequently while mapping landforms associated with deep-seated landslides, are used by at least two of the mappers in this study. Recognition of inconsistent landslide feature expressions led Burns & Madin (2009) to recommend assigning a confidence level to each mapped landslide. In the case of limited morphological evidence, a mapper must decide, with some degree of confidence, if the combination of morphological features visible in the remote sensing data is enough to suggest mass movement has occurred. For example, in this study, while trying to gather evidence for interpretation one mapper raised the question of why two parallel drainages were initiating from the same spatial location but flowing in opposite directions. The symmetry of the drainage system was a key piece of logic used by the mapper to determine that mass movement had occurred, although the evidence was not overwhelming and, as such, the confidence in the

interpretation was not high. These logic-based interpretation approaches, made with relatively lower confidence than interpretation of landslides displaying a suite of classic features and/or evidence of active movement, are the source of low levels of agreement between human observers.

Landslides in all states of activity are arguably important features to identify because previously failed terrain poses a potentially higher hazard than hillslopes that have not failed (Turner et al., 2010). Low agreement between observers in the study appears to have occurred when two mappers mapped substantially more inactive, old landslides than the other two mappers. An example of the spatial distribution of the geomorphometric features mapped by the observer recording the largest amount of landslide area (mapper C), compared to the observer recording the least amount of landslide area (mapper B) is shown in Figure 24. In this example, mapper C is used as reference to computer recall statistics. Red polygons, labeled as false negatives (FN), highlight landslide terrain mapped by mapper C that was not mapped by mapper B. The result of the area comparison for mapper C and B is the lowest recall between observers, at 35%. With increasingly scarce topographic signatures indicating failure has occurred, half of the mappers were less likely to conclude that features, such as those highlighted in red, were a landslide. This remained the case even after extensive group discussions occurred.

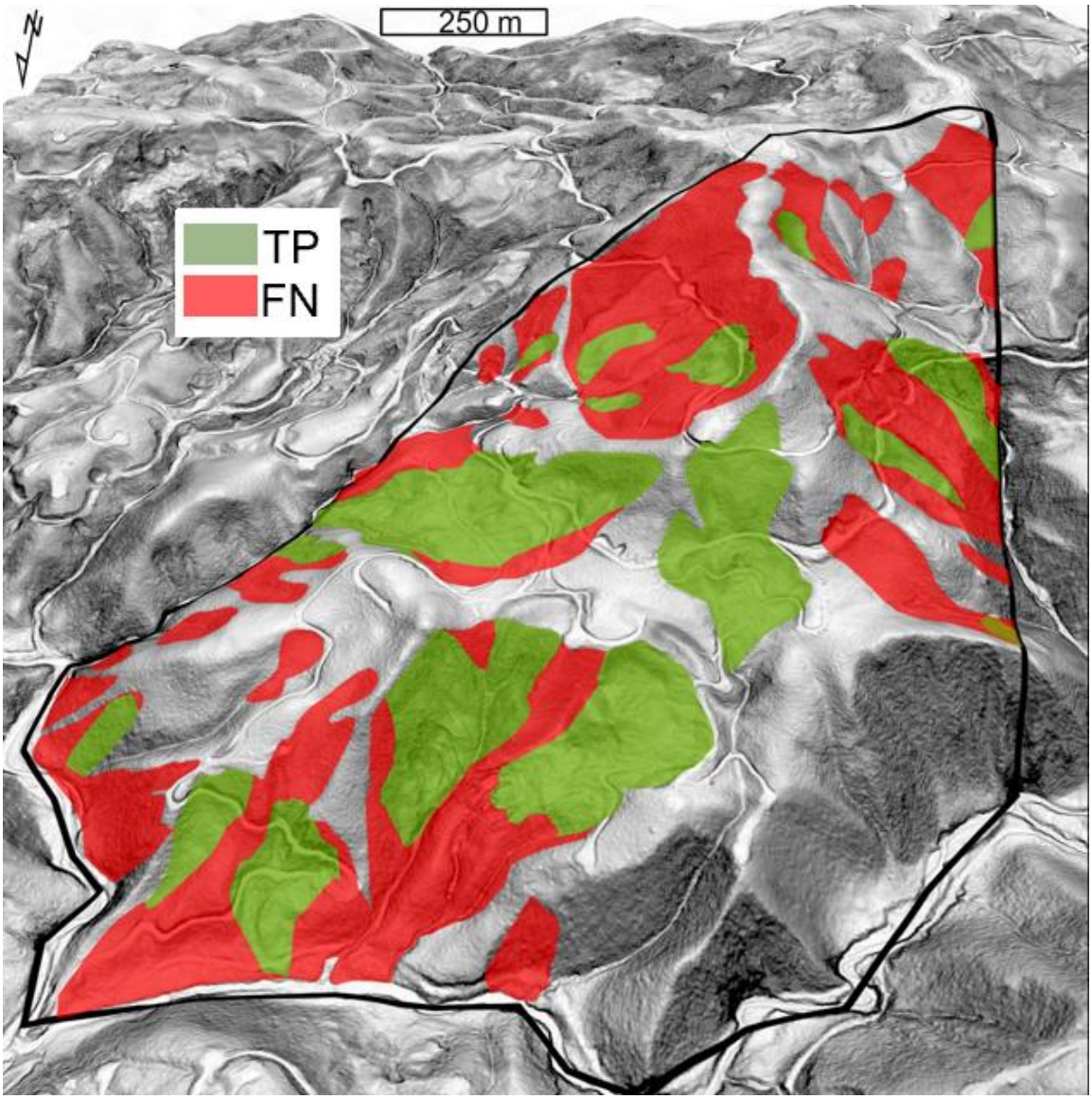


FIGURE 24. AREA OF TRUE POSITIVE (TP) AND FALSE NEGATIVE (FN) SHOWN USING MAPPER C LANDSLIDE AREAS AS THE REFERENCE DATASET AND MAPPER B LANDSLIDE AREAS FOR VERIFICATION DATASET. SPATIAL EXTENT OF DATA SHOWN IN TEST SITE 1.

Parallels may exist between confidence levels in manually mapped landslides and those mapped with automated models. Inactive, old landslides with few remaining signs of instability are very difficult to target in empirical model designs. As the morphometric expressions become scarce, designing a model using context, or object associations can be beneficial. An example of interpretation based on association, or a lack of key

evidence, occurred in at least one instance in round two of the manual mapping exercise. The human observers noted a visible scarp feature and a bench. Group discussion led to the conclusion that the toe was assumed to be located downslope, and was mapped and attributed as such, even though clear evidence of a depositional lobe was not visible in the images. Object-based model design allows for similar logic, or context, to be applied by classifying objects based on association to other classified objects. This logical thought process can be implemented in the model by adding weight to classified landslide objects that are extracted using a roughness criterion defining very rough surfaces. Landslide components that are adjacent to, or presumed to be associated with, the classified landslide objects are included in the classification after meeting a defined set of rules, much like the missing deposits were included in the manual observations. These new rules do not require an increasingly larger roughness criterion for larger objects, to meet the class definition for landslide, as would have been the case if the object were not adjacent to a classified landslide. Instead, a lower roughness criterion can be used to include components of the landslide that are weathered and subdued. The confidence in modeling these features with missing pieces of evidence may be lower, much like manual mapping, but the technique retains the ability to mimic human perception.

Like landslides, non-landslide modeled features also require interpretation. I used these non-landslide, modeled features to assist with landslide mapping. The features were also considered in the quantification of model and manual map agreement. A separate study to examine the agreement between the ancillary landform features mapped by the

model and manual mappers would be necessary prior to making changes to either the modeled or the manually mapped landslide datasets. A dataset that includes ancillary landform feature mapping would be beneficial for future analysis. This dataset could be used to determine if a sub-population of terrain features exits that are consistently misclassified as landslides by the model, or if the model is consistently misclassifying a specific terrain feature, disallowing it to be appropriately classified as a landslide.

4.2 Evolution of Semi-Automated Model

Evolution of the mapper's thought, or interpretive process, while manually mapped landslide inventories were being developed occurred, not only during group discussions, but also during the mapping process (i.e., via separate conversations with mappers). After conversations with peers, mappers changed their interpretation strategies as they gained a better understanding of landscape characteristics. A similar change in the modeled landslide mapping strategy occurred for the author during the initial model verification component of the study (section 3.2.3). After participating in the geomorphic interpretation process, and after viewing the composite, manually mapped inventory results, a second model routine was designed to account for inactive geomorphic features that were not considered in the initial model design. Model validation is not the goal of this second exercise, instead model improvements is the goal. Improvements included using neighborhood relationships to grow the existing landslide classification into areas where the deep-seated landslide surface expression became more subtle. A similar

classification strategy as that described in the methods section (first run of model) was employed during development of the revised model. For the second model run, the initial landslide population was expanded by removing the slope threshold in the first step. All dense (Density >1.2) objects on the super-object level were classified as landslides if: Roughness criterion = DTNarea >40% and DTNvar >40, and contain >400 pixels of previously classified sub-objects with DTN values >5, OR Roughness criterion = DTNarea >30% and DTNvar >30. Two growing processes were run for adjacent unclassified objects with >30% boundary and a mean object slope under 60%: 1) Roughness criterion = DTNarea >25% and DTNvar >40, and 2) variance in profile curvature (standard deviation of profile curvature, squared) >30. Objects with mean slope >60% were added to the landslide classification if Roughness criterion = DTNarea >20% and DTNvar >20. Finally, all classified objects were merged, and those with a mean slope >75% OR mean slope <60% and relative border to stream, or stream adjacent classes >40% were removed. Spatial results from the second model run are shown in Figure 24 and Figure 25.

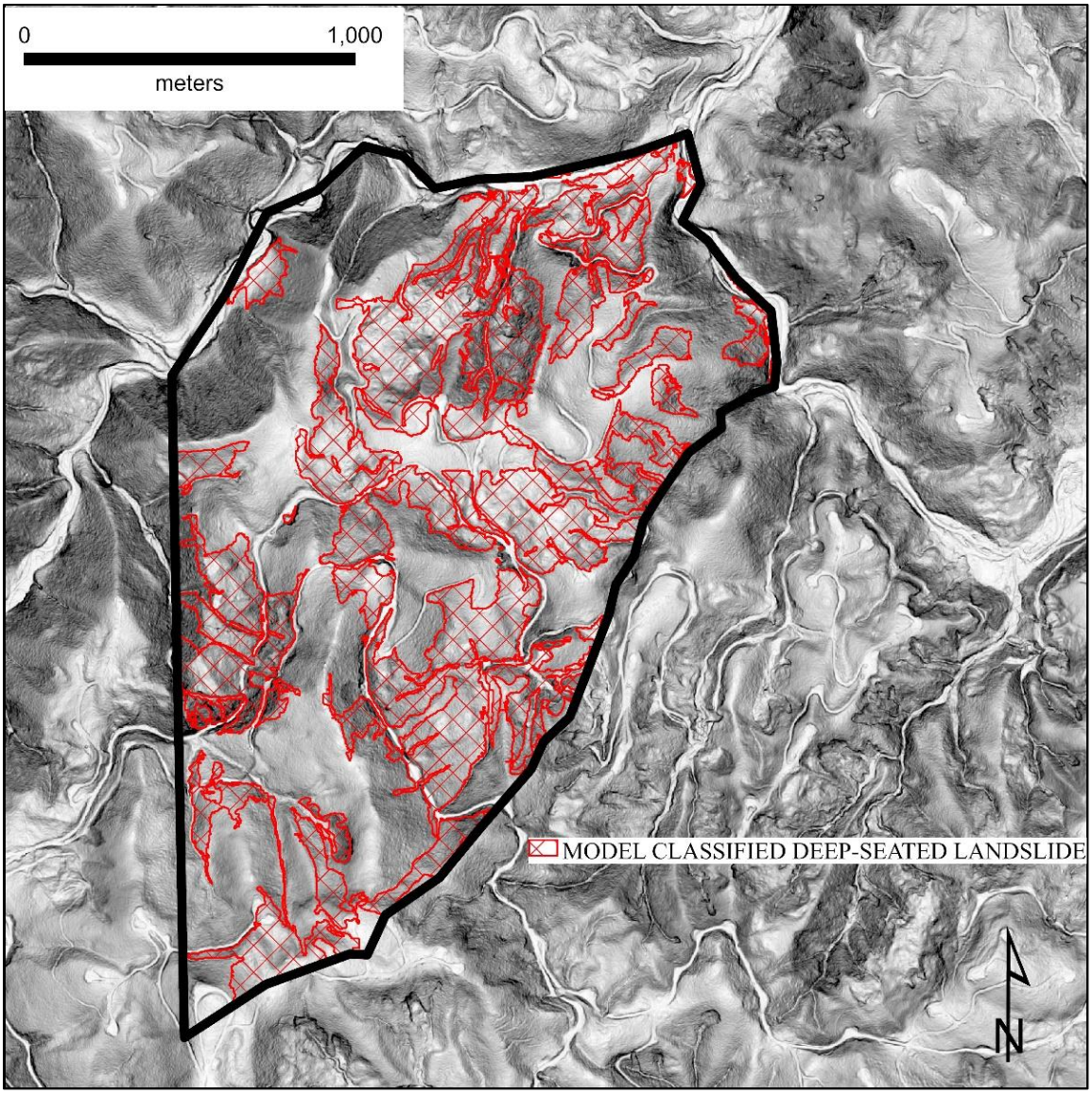


FIGURE 25. GEOBIA MODELED LANDSLIDES FROM THE SECOND MODEL RUN IN THE EXTENT OF COMPARATIVE TEST SITE 1.

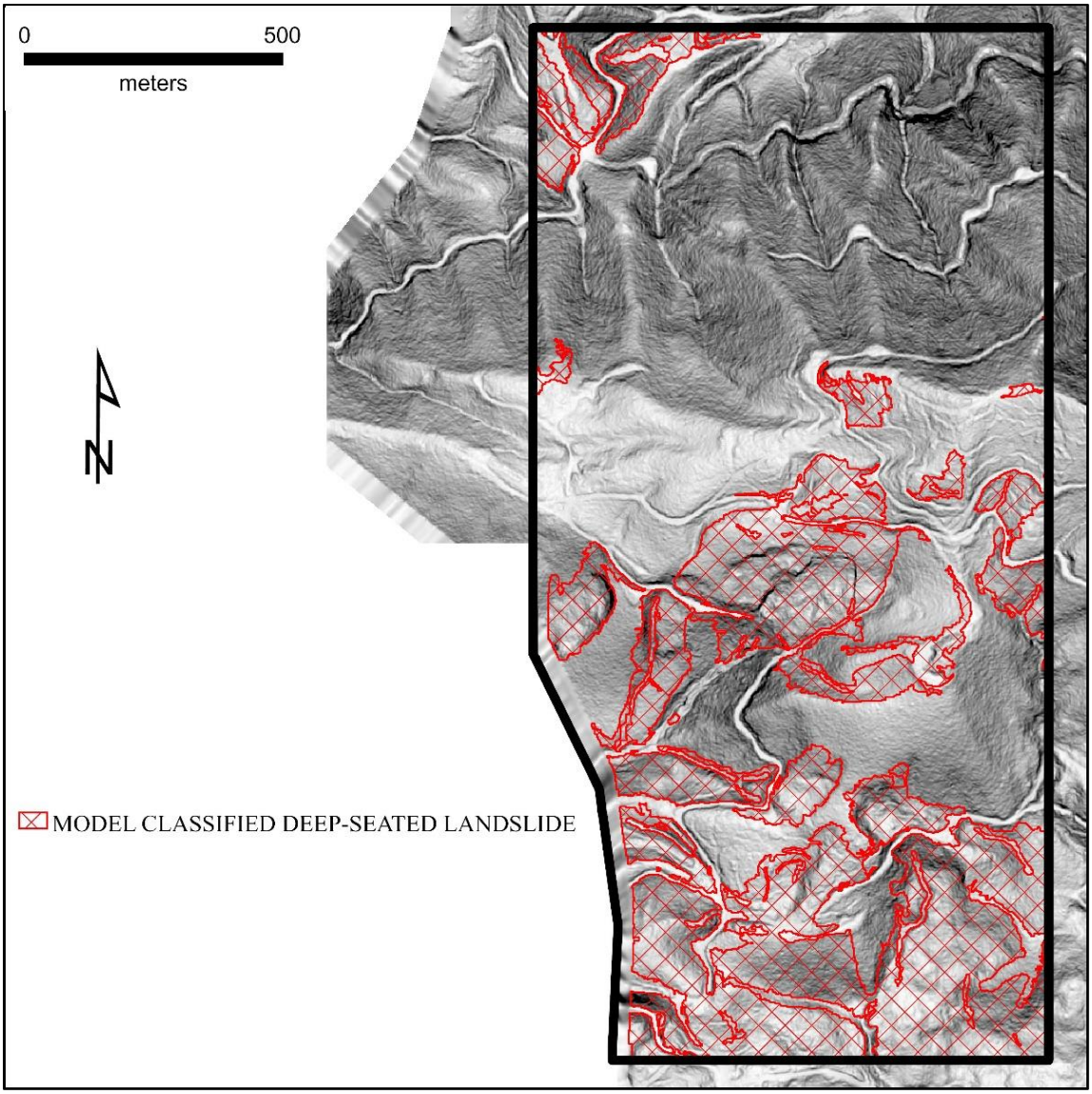


FIGURE 26. GEOBIA MODELED LANDSLIDES FROM THE SECOND MODEL RUN IN THE EXTENT OF COMPARATIVE TEST SITE 2.

A second comparative analysis is completed to test model improvements. Model design changes were made during the course of the study, after I gained a better understanding of landscape characteristics by participating in the composite mapping exercise with the professional geologists. This second analysis is not evaluating an independently developed model being tested with independent data (i.e., validation), nor

is it evaluating a model that has been trained with landslide inventory data. Instead, it is intended to verify if modification of criteria used in the first model run can result in a landslide classification that encompasses larger areas of the inactive landslide terrain mapped more consistently by two of the four observers.

Results from the second model run are draped over the manually mapped inventory for a qualitative, spatial comparison (Figure 26 & Figure 27). The modified class definitions allow the model to extract more of the mapped landslide areas as compared to the first run. The model is also extracting more terrain that was not mapped as a landslide by the geologists such as ridges, terraces and swales.

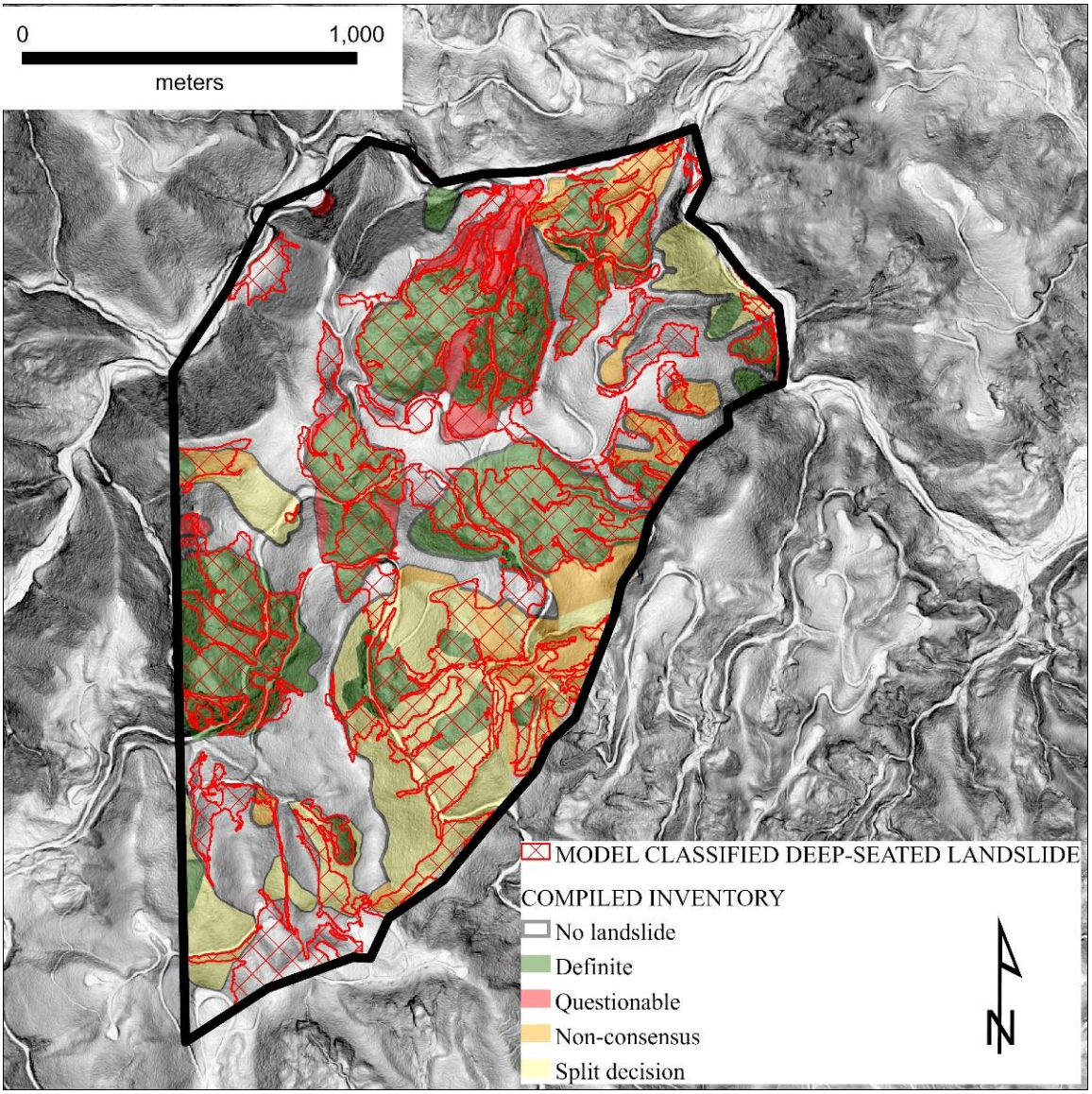


FIGURE 27. SPATIAL OVERLAP OF MANUALLY MAPPED LANDSLIDES (COMPOSITE INVENTORY) AND GEOBIA MODELED LANDSLIDES FROM THE SECOND MODEL RUN IN THE EXTENT OF COMPARATIVE TEST SITE NUMBER 1.

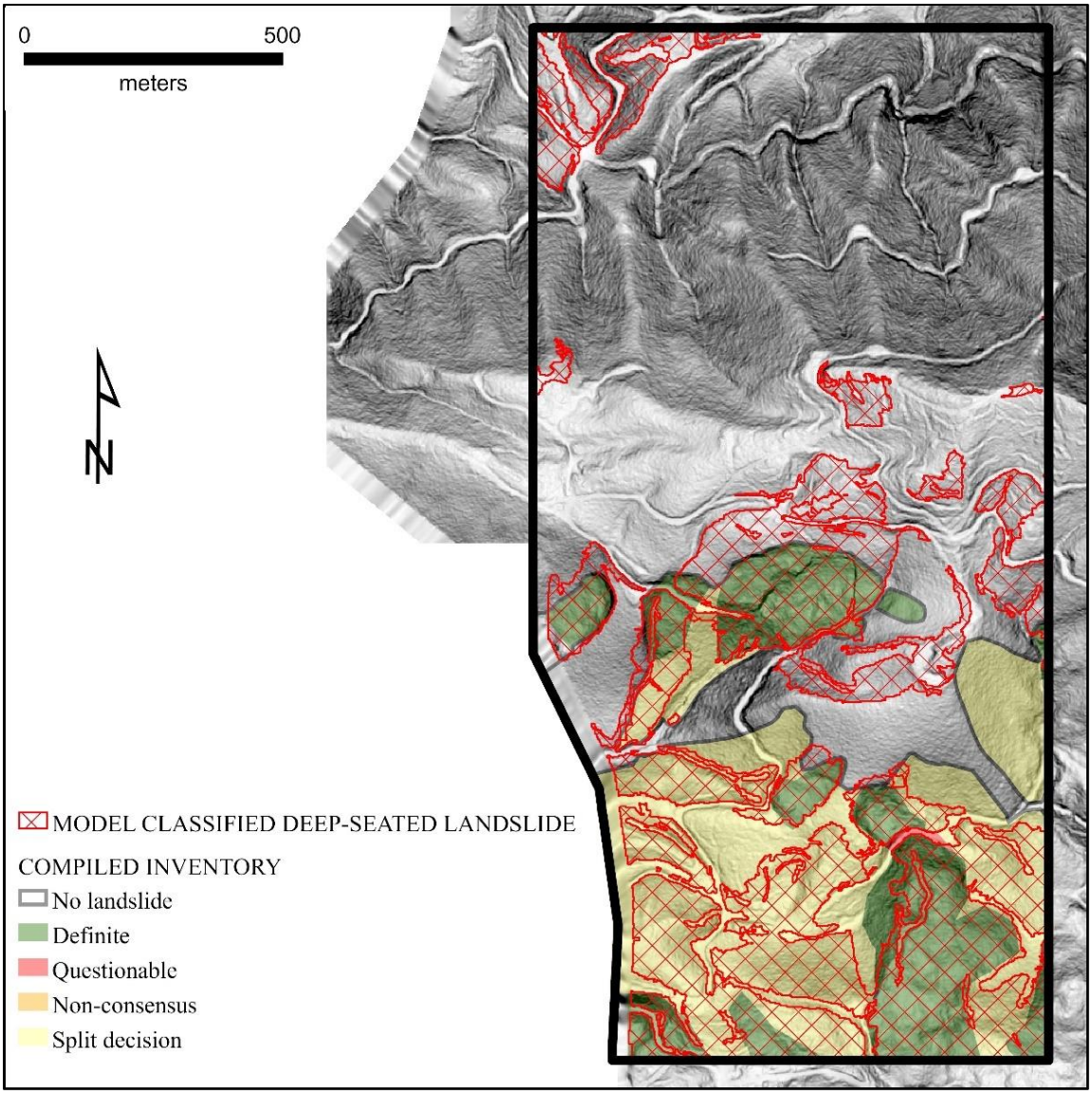


FIGURE 28. SPATIAL OVERLAP OF MANUALLY MAPPED-LANDSLIDES (COMPOSITE INVENTORY) AND GEOBIA MODELED LANDSLIDES FROM THE SECOND MODEL RUN IN THE EXTENT OF COMPARATIVE TEST SITE NUMBER 2.

The same statistics used to examine agreement between modeled landslide data and the manually mapped landslide data in section 3.2.3 show an interesting pattern between this second modeled dataset and the modeled dataset examined in the results section. I will focus this discussion on any landslide mapped (i.e., the ALL LANDSLIDES column in Figure 29) because one goal of the second model design is the

mapping of inactive landslides that were not consistently mapped by all observers. Overlapping identification of landslide areas occurred in 22% of the test area (inventoried landslide attributed as Definite). Model improvements in the second model were aimed at mapping the additional 23% of the test area where the mappers disagreed on the area delineated as landslide (inventoried landslide attributed as Split-decision and Non-consensus). Model accuracy for the second model run is almost static, improving by only 1%, while recall improved by 11%, and precision by 13%.

REFERENCE DATA - COMPOSITE INVENTORY - 2.3 (KM²)				
MODELED DATA - SECOND RUN - 1.8 (KM²)			MODELED DATA - FIRST RUN - 1.4 (KM²)	
	ACCURACY			
	(TP+TN)/BASIN AREA			
		ALL LANDSLIDES		ALL LANDSLIDES
	GEOBIA	0.78		0.77
	RECALL			
	TP/(TP+FN)			
		ALL LANDSLIDES		ALL LANDSLIDES
	GEOBIA	0.69		0.58
	PRECISION			
	(TP)/(TP+FP)			
		ALL LANDSLIDES		ALL LANDSLIDES
	GEOBIA	0.74		0.61

FIGURE 29. STATISTICAL METRICS COMPARING SECOND MODEL RUN TO ALL MANUALLY, MAPPED LANDSLIDES (REFERENCE DATA) ON LEFT SIDE OF FIGURE. COMPARISON TO FIRST MODEL RUN STATISTICS COMPUTED FOR ALL MANUALLY, MAPPED LANDSLIDES PROVIDED ON RIGHT SIDE OF FIGURE TO NOTE MODEL IMPROVEMENTS.

This second attempt at modeling, designed to mimic professional interpretations, added almost 0.40 km^2 more area classified as landslide. When compared to the agreement statistics reported for the first model run, this second modeled dataset did not

produce a difference in agreement statistics as large as those recorded between different manual observers with different interpretive styles. In the following sections I will frequently compare the statistical mapping agreement between mappers C and A, who systematically inventoried more landslide area described by subdued surface expressions than mappers B and D. Accuracy between mappers C and A is 87%. The accuracy of the second modeled landslide output area compared to any landslide mapped in the inventory is 11% lower than the accuracy recorded between mappers C and A.

In the absence of a reference dataset consisting of the true population of landslides that is used to independently validate the model, the statistics should be viewed as a tool to examine the data. The difference in landslide interpretation has been described by comparing and contrasting selection of reference datasets. This difference remains for the two mappers who identified the largest area of subtle landslide features. If mapper C is used for the reference dataset, recall between mapper C and mapper A is 77%. This is the same value for precision if considering mapper A as reference. If mapper A is selected for the reference dataset, recall increases to 92% (also the same value for precision if mapper C is selected for reference). This 15% increase in recall can be explained by mapper C delineating an additional 0.9 km^2 as deep-seated landslide. Recall statistics for the second modeled landslide area, compared to any landslide mapped, is 69%, 8% lower than the recall between mapper C and A, using mapper C as reference (77%). The change in criteria, to mimic the manual mapping of inactive, old landslides, suggests more inactive landslides that are difficult for professionals to delineate confidently, are classified using this second model design.

Discrepancy in the area mapped as a landslide was as high as 1.95 km^2 between professionals, with differences in recall as high as 64%. Modifying parameters in the object-based routine, designed using all landslides in the reference data as a visual aid, resulted in statistical comparisons of recall that increased by 11% and precision that increased by 13% for the mapped area. This positive association between modifying model parameters and producing higher agreement between all landslides in the dataset, was hypothesized to occur in the manually, mapped landslide inventory. It was thought that group discussions amongst the geologists inventorying landslides would lead to sharing of geomorphological evidence enabling improved landslide area agreement between mappers. When considering the agreement statistics for the manual-mapped landslide data, and the agreement statistics to highlight improvements made in the two model runs, it is suggested that automating the identification of landslides based on surface morphology is a more objective classification technique than enlisting professionals with varying backgrounds to manually map landslides.

A fine line exists between creating a model that is too complex to enable transferability between terrains with diverse landform characteristics and too simple to achieve accurate results. This work has demonstrated that incorporation of contextual and geometric information, use of hierarchies within the GEOBIA platform, and working at multiple spatial scales simultaneously enables successful landslide extraction in diverse terrain, without compromising accuracy. This modeling style facilitates rapid learning for the model developer regarding the influence each metric has on extracting landslides, by

allowing for immediate visualization of the georeferenced object boundaries on the landscape and the spatial distribution of the classes produced with each new definition.

While great progress has been made in this study towards automating the mapping of deep-seated landslides, establishment of the complete suite of metrics necessary to extract landslides in all states of activity across diverse terrain has not been accomplished, as shown by the recall statistics. It is the author's opinion that two possible paths towards higher recall in automated mapping exist: 1) use the model features, that show good results in this study, along with all landslides mapped in the composite dataset in a machine learning routine (i.e., Convolutional Neural Network (CNN)), or 2) take a closer look into the terrain features that are not being mapped by the automated model and refine contextual relationships, object delineation, or seek an additional metric that can assist with class definitions in more subdued landslide terrain. The use of machine learning routines can reduce bias in model development (i.e., eliminate the use of pre-defined thresholds for class definitions), while an in depth look at terrain features can help teach us more about the empirical relationships that describe features on the landscape that are indicative of old, inactive hillslope instabilities.

Improvements were made in this study to automate the mapping of deep-seated landslides in the diverse topography of the PNW. The test sites chosen for this research were in areas with moderate landslide density, as determined by observing modeled landslide density distributions. To discuss model transparency and the ability to replicate this study in different geologic terrain types, a similar study would need to be conducted in areas with significantly higher densities of landslides or significantly lower. It is

thought that studies in terrain with higher or lower landslide densities would provide a better opportunity to verify or even validate model performance. Terrain with high landslide densities would test the model's ability to detect hillslopes dominated by deep-seated landslide processes, as recorded by recall statistics. Model precision in terrain with lower landslide density would verify to what degree the model is recording false positive landslides.

5. Conclusion

Variability in the surface expression of deep-seated landslides occurs throughout the study area, and the Pacific Northwest (Glenn et al., 2016). Highly variable terrain characteristics are causing differences in professional agreement regarding the presence and extent of deep-seated landslides (Hölbling et al., 2017). Rapid improvements in remote sensing data quantity and quality are fueling development of automated models to map landslides that could reduce variability in interpretive styles and inventory products, inherent in manual landslide mapping (Martha et al., 2011). This study shows that new methods, such as GEOBIA, have the ability to incorporate professional knowledge, while remaining empirically based, to improve objectivity during the landslide inventory process.

The model presented in this paper was able to systematically identify deep-seated landslides in multiple watersheds within the Willapa Hills area using a straightforward GEOBIA routine with a procedure to identify surface roughness. Landslide detection performance from the semi-automated model, when referenced to landslide areas mapped by all of the four professional geologists, achieved 81% accuracy, which is on par with the average manually mapped inventory accuracy for each pair of mappers. Similar to the recall statistic, choice of reference dataset plays a large role in precision values. Precision statistics suggest the model is not identifying false positives (i.e., non-landslide features) as often as mappers with different interpretive styles, where precision was as low as 39%. Model precision for mapping landslide areas from the Definite landslide reference dataset is 69%.

A major component of this study was the exploration of deep-seated landslide map agreement between professional geologists. It is clear that the choice of mapper used as reference had a large impact on the statistical comparison of map products. While it is noted that mappers who were primarily recording active or inactive, young landslides produced high agreement, suggesting higher confidence in mapping these features, using only these features could bias a landslide inventory. An incomplete inventory becomes especially problematic if the stability of previously failed terrain is being studied for public safety concerns. With this in mind, any landslide mapped by a geologist was considered when examining improvements made in the model. Landslide detection performance from the semi-automated model, when referenced to landslide areas mapped by any of the geologists, achieved 69% recall (down 7% from recall of landslide area mapped by all geologists (i.e., Definite class)). The model was able to identify more reference landslide areas correctly when compared to all landslides in the reference dataset than when comparing two mappers with a different mapping style, where recall was as low as 35%. When comparing two landslide inventories produced by mappers with similar mapping styles, landslide area recall for the model is similar to landslide area recall between observers (75% and 77%).

Automated models can be used to develop morphometric indices describing object statistics, geometry, and context allowing for mapping of deep-seated landslides that mimic human interpretation of landslides. Rule sets (i.e., model algorithms) can be designed to map landslides with fresh geomorphic evidence, indicating recent failure. They can also be modified to map landslides that have weathered substantially over long

periods of time. The tools, including hierarchical data processing, use of contextual object relationships, exploitation of geometric properties, and multi-scale processing offer an alternative to subjective, manual mapping that is hard to replicate between observers. The lack of clear geomorphic evidence for these inactive, old landslides are the primary cause of manual mapped landslide inventory inconsistencies. Use of automated models provides an opportunity to classify deep-seated landslides using empirical data to consistently identify terrain features that are indicative of failure. This is an improvement over manual mapping in that it can reduce inherent bias introduced by professional judgement.

Bibliography

Aaron, J., Hungr, O., Stark, T. D., & Baghdady, A. K. (2017). Oso, Washington, landslide of March 22, 2014: Dynamic analysis. *Journal of Geotechnical and Geoenvironmental Engineering*, *143*(9), 1–10. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001748

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., & Reichenbach, P. (2002). Impact of mapping errors on the reliability of landslide hazard maps. *Natural Hazards and Earth System Sciences*, *2*, 3–14. https://doi.org/10.5194/nhess-2-3-2002

Baatz, M., & Schape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Information*, *XII*, 12–23.

Berti, M., Alessandro, C., Em, R., & Daehne, A. (2013). Comparative analysis of surface roughness algorithms for the identification of active landslides. *Geomorphology*, *182*, 1–18. https://doi.org/10.1016/j.geomorph.2012.10.022

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke, T., Feizizadeh, B., & Hölbling, D. (2014). Object-based image analysis and digital terrain analysis for locating landslides in the Urmia Lake basin, Iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(12), 4806–4817. https://doi.org/10.1109/JSTARS.2014.2350036

Bobrowsky P., Highland L. (2013). The Landslide Handbook-a Guide to Understanding Landslides: A Landmark Publication for Landslide Education and Preparedness. In: Sassa K., Rouhban B., Briceño S., McSaveney M., He B. (eds) Landslides: Global Risk Preparedness. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22087-6_5

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*(3–4), 132–147. https://doi.org/10.1016/j.geomorph.2009.02.027

Bunn, M. D., Leshchinsky, B. A., Olsen, M. J., & Booth, A. (2019). A simplified, object-based framework for efficient landslide inventorying using LIDAR digital elevation model derivatives. *Remote Sensing*, *11*(3). https://doi.org/10.3390/rs11030303

Burns, W. J., & Madin, I. P. (2009). Protocol for Inventory Mapping of Landslide Deposits from Light Detection and Ranging (lidar) Imagery, (Special Paper 42), 1–36. Retrieved from https://www.oregongeology.org/pubs/sp/p-SP-42.htm

Burns, W. J., & Mickelson, K. A. (2016). Protocol for deep landslide susceptibility mapping, (Special Paper 48), 1–63. Retrieved from https://www.oregongeology.org/pubs/sp/SP-48.pdf

Burrough, P. A., and McDonell, R. A. (1998). *Principles of Geographical Information Systems.* Oxford University Press, New York. 190 pp.

Congalton, R. G., & Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data; Principles and Practices*. CRC Press.

Corominas, J., Westen, C. Van, Frattini, P., Cascini, L., Fotopoulou, S., Catani, F., … Herva, J. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, *73*(2), 209–263. https://doi.org/10.1007/s10064-013-0538-8

Cruden, D. M., & Varnes, D. J. (1996). Landslide Types and Processes. In T. A.K. & S. R.L. (Eds.), *Landslide Investigation and Mitigation.* (Transporta, pp. 36–75). Washington, DC.

Darwish, A., Leukert, K., & Reinhardt, W. (2003). Image Segmentation for the Purpose of Object-Based Classification. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *3*, 2039–2041. https://doi.org/10.1109/igarss.2003.1294332

Di Ruberto, C., & Putzu, L. (2016). Emerging Trends in Applications and Infrastructures for Computational Biology, Biorinformatics, and Systems Bilogy. In *Emerging Trends in Computer Science and Applied Computing* (pp. 37–47).

Drăguţ, L., & Eisank, C. (2012). Automated object-based classification of topography from SRTM data. *Geomorphology*, *141–142*, 21–33. https://doi.org/10.1016/j.geomorph.2011.12.001

Ebinghaus, A., Hartley, A. J., Jolley, D. W., Hole, M., & Millett, J. (2014). Lava – Sediment Interaction and Drainage-System Development in a Large Igneous Province: Columbia River Flood Basalt Province, Washington State, U.S.A. *Journal of Sedimentary Research*, *84*, 1041–1063. https://doi.org/http://dx.doi.org/10.2110/jsr.2014.85

Evans JS, Oakleaf J, Cushman SA, Theobald D (2014). An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial. Accessed: 2020 January.

Fairbridge, R. W. (Ed.). (1968). *The Encyclopedia of Geomorphology* (Encyclopedia of Earth Sciences Series, Volume III). Reinhold, New York: Dowden, Hutchinson & Ross Inc.

Florinsky, I. V. (2016). *Digital Terrain Analysis in Soil Science and Geology – Second Edition*. Elsevier, Amsterdam. 475 pp.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, *94*, 268–289. https://doi.org/10.1016/j.geomorph.2006.09.023

Gercek, D. (2010). *Object-Based Classification of Landforms Based on Their Local Geometry and Geomorphometric Context*. Middle East Technical University.

Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D., & Dorsch, S. J. (2006). Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology*, *73*(1–2), 131–148. https://doi.org/10.1016/j.geomorph.2005.07.006

Hibert, C., Stark, C. P., & Ekström, G. (2015). Dynamics of the Oso-Steelhead landslide from broadband seismic analysis. *Natural Hazards and Earth System Sciences*, *15*(6), 1265–1273. https://doi.org/10.5194/nhess-15-1265-2015

Hölbling, D., Eisank, C., Albrecht, F., Vecchiotti, F., Friedl, B., Weinke, E., & Kociu, A. (2017). Comparing Manual and Semi-Automated Landslide Mapping Based on Optical Satellite Images from Different Sensors. *Geosciences*, *7*(37). https://doi.org/10.3390/geosciences7020037

Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, *11*(2), 167–194. https://doi.org/10.1007/s10346-013-0436-y

Iverson, R. M., & Major, J. J. (1987). Rainfall, ground-water flow, and seasonal movement at Minor Creek landslide, northwestern California: physical interpretation of empirical relations. *Geological Society of America Bulletin*, *99*(4), 579–594. https://doi.org/10.1130/0016-7606(1987)99<579:RGFASM>2.0.CO;2

Iverson, R. M., George, D. L., Allstadt, K., Reid, M. E., Collins, B. D., Vallance, J. W., Schilling, S. P., Godt, J. W., Cannon, C. M., Magirl, C. S., Baum, R. L., Coe, J. A., Schulz, W. H., & Bower, J. B. (2015). Landslide mobility and hazards: Implications of the 2014 Oso disaster. *Earth and Planetary Science Letters*, *412*, 197–208. https://doi.org/10.1016/j.epsl.2014.12.020

Joyce, K. E., Samsonov, S. V., Levick, S. R., Engelbrecht, J., & Belliss, S. (2014). Mapping and monitoring geological hazards using optical, LiDAR, and synthetic aperture RADAR image data. *Natural Hazards*, *73*(2), 137–163. https://doi.org/10.1007/s11069-014-1122-7

Kasai, M., Ikeda, M., Asahina, T., & Fujisawa, K. (2009). LiDAR-derived DEM evaluation of deep-seated landslides in a steep and rocky region of Japan. *Geomorphology*, *113*(1–2), 57–69. https://doi.org/10.1016/j.geomorph.2009.06.004

Korzeniowska, K. (2017). *Object-based image analysis for detecting landforms diagnostic of natural hazards*. University of Potsdam.

Lacasse, S., Nadim, F., & Kalsnes, B. (2010). Living with landslide risk. *Geotechnical Engineering*, *41*(4).

Lahousse, T., Chang, K. T., & Lin, Y. H. (2011). Landslide mapping with multi-scale object-based image analysis-a case study in the Baichi watershed, Taiwan. *Natural Hazards and Earth System Science*, *11*(10), 2715–2726. https://doi.org/10.5194/nhess-11-2715-2011

LaHusen, S. R., Duvall, A. R., Booth, A. M., Grant, A., Mishkin, B. A., Montgomery, D. R., Struble, W., Roering, J. J., & Wartman, J. (2020). Rainfall triggers more deep-seated landslides than Cascadia earthquakes in the Oregon Coast Range, USA. *Science Advances*, *6*(38), eaba6790. https://doi.org/10.1126/sciadv.aba6790

Li, X., Cheng, X., Chen, W., Chen, G., & Liu, S. (2015). Identification of forested landslides using lidar data, object-based image analysis, and machine learning algorithms. *Remote Sensing*, *7*(8), 9705–9726. https://doi.org/10.3390/rs70809705

MacFaden, S. W., O'Neil-Dunne, J. P. M., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*, *2012*(063567), 1–23. https://doi.org/10.1117/1.JRS.6.063567

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., & Kumar, K. V. (2010). Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*, *116*(1–2), 24–36. https://doi.org/10.1016/j.geomorph.2009.10.004

Martha, T. R., Kerle, N., Van Westen, C. J., Jetten, V., & Kumar, K. V. (2011). Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(12 PART 1), 4928–4943. https://doi.org/10.1109/TGRS.2011.2151866

McCalpin, J. (1984). Preliminary age classification of landslides for inventory mapping. *Proceedings 21st Annual Engineering Geology and Soils Engineering Symposium*, *4*(5), 5–6.

McKean, J., & Roering, J. (2004). Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology*, *57*(3–4), 331–351. https://doi.org/10.1016/S0169-555X(03)00164-8

Miller, D. (2017). *Literature Synthesis of the Effects of Forest Practices on Non-Glacial Deep-Seated Landslides and Groundwater Recharge: Prepared for the Upslope Processes Scientific Advisory Group Cooperative Monitoring, Evaluation, and Research Committe*. https://www.dnr.wa.gov/publications/bc_tfw_litsyngdsl_20170202.pdf

Ouyang, Z. T., Zhang, M. Q., Xie, X., Shen, Q., Guo, H. Q., & Zhao, B. (2011). A comparison of pixel-based and object-oriented approaches to VHR imagery for mapping saltmarsh plants. *Ecological Informatics*, *6*(2), 136–146. https://doi.org/10.1016/j.ecoinf.2011.01.002

Parker, R. N., Densmore, A. L., Rosser, N. J., de Michele, M., Li, Y., Huang, R., … Petley, D. N. (2011). Mass wasting triggered by the 2008 Wenchuan earthquake is greater than orogenic growth. *Nature Geoscience*, *4*(7), 449–452.

Passalacqua, P., Tarolli, P., & Foufoula-Georgiou, E. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. *Water Resources Research*, *46*(11), 1–17. https://doi.org/10.1029/2009WR008812

Perron, J. T., Kirchner, J. W., & Dietrich, W. E. (2008). Spectral signatures of characteristic spatial scales and nonfractal structure in landscapes. *Journal of Geophysical Research: Earth Surface*, *113*(4), 1–14. https://doi.org/10.1029/2007JF000866

Pringle, R. F. (1986). *Soil Survey of Grays Harbor County Area, Pacific County, and Wahkiakum County, Washington*.

Razak, K. A., Santangelo, M., Van Westen, C. J., Straatsma, M. W., & de Jong, S. M. (2013). Generating an optimal DTM from airborne laser scanning data for landslide mapping in a tropical forest environment. *Geomorphology*, *190*, 112–125. https://doi.org/10.1016/j.geomorph.2013.02.021

Shaw, S. C., Justice, T. E., McCarthy, R., Hinkle, J. C., Turner, T. R., Fransen, B. R., … Thornton, J. (2017). Automated, Object-Based Image Analysis (GEOBIA) Model for Landform Detection and Mapping, with Applications to Pacific Northwest USA Landslide Assessments. In *AEG; 3rd North American Symposium on Landslides* (pp. 9–20).

Sithole, G., & Vosselman, G. (2004). Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *59*(1–2), 85–101. https://doi.org/10.1016/j.isprsjprs.2004.05.004

Smith, K., & Petley, D. N. (2009). *Environmental Hazards: Assessing Risk and Reducing Disasters*. (B. J. Boruff, Ed.) (5th Editio). Retrieved from https://doi-org.proxy.lib.pdx.edu/10.1111/j.1745-5871.2009.00611.x

Sultana, N. (2020). Analysis of landslide-induced fatalities and injuries in Bangladesh: 2000-2018. *Cogent Social Sciences*, *6*(1737402).

Stark, T. D., Baghdady, A. K., Hungr, O., & Aaron, J. (2017). Case study: Oso, Washington, landslide of march 22, 2014-material properties and failure mechanism. *Journal of Geotechnical and Geoenvironmental Engineering*, *143*(5), 1–13. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001615

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment*, *115*, 2564–2577.

Turner, T. R., Duke, S. D., Fransen, B. R., Reiter, M. L., Kroll, A. J., Ward, J. W., … Bilby, R. E. (2010). Landslide densities associated with rainfall, stand age, and topography on forested landscapes, southwestern Washington, USA. *Forest Ecology and Management*, *259*(12), 2233–2247. https://doi.org/10.1016/j.foreco.2010.01.051

Van Den Eeckhaut, M., Kerle, N., Poesen, J., & Hervás, J. (2012). Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology*, *173–174*, 30–42. https://doi.org/10.1016/j.geomorph.2012.05.024

Varnes, D. J. (1978). Slope Movement Types and Processes. *Transportation Research Board Special Report*, *176*, 11–33.

Wartman, J., Montgomery, D. R., Anderson, S. A., Keaton, J. R., Benoît, J., dela Chapelle, J., & Gilbert, R. (2016). The 22 March 2014 Oso landslide, Washington, USA. *Geomorphology*, *253*(March 2014), 275–288. https://doi.org/10.1016/j.geomorph.2015.10.022

Washington Forest Practices Board Manual, Section 16*:* Guidelines for Evaluating Potentially Unstable Slopes and Landforms (WDNR). (2016). http://file.dnr.wa.gov/publications/bc_fpb_manual_section16.pdf

Washington Forest Protection Association (WFPA). (2020). Steep Slopes. Retrieved from: http://www.wfpa.org/forests-fish-law/steep-slopes/#:~:.

Wells, R. E., & Sawlan, M. G. (2014). *Preliminary geologic map of the eastern Willapa Hills, Cowlitz, Lewis, and Wahkiakum Counties, Washington*. *Open-File Report*. Reston, VA. https://doi.org/10.3133/ofr20141063

Wells, R., Bukry, D., Friedman, R., Pyle, D., Duncan, R., Haeussler, P., & Wooden, J. (2014). Geologic history of Siletzia, a large igneous province in the Oregon and Washington Coast Range: Correlation to the geomagnetic polarity time scale and implications for a long-lived Yellowstone hotspot. *Geosphere*, *10*(4), 692–719. https://doi.org/10.1130/GES01018.1

Western Regional Climate Center (WRCC). (2020). *Climate of Washington.* Retrieved October 1, 2020, from https://wrcc.dri.edu/Climate/narrative_wa.php

Whelley, P. L., Glaze, L. S., Calder, E. S., & Harding, D. J. (2014). LiDAR-derived surface roughness texture mapping: Application to mount St. Helens pumice plain deposit analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 426–438. https://doi.org/10.1109/TGRS.2013.2241443

Wieczorek, G. F. (1984). Preparing a Detailed Landslide-Inventory Map for Hazard Evaluation and Reduction. *Environmental and Engineering Geoscience*, *XXI*(3), 337–342.

Appendix: Object-based Extraction of Stream, Road, and Ridge Features

For this study deep-seated landslide extraction is completed using an object-based image analysis (GEOBIA) classification approach that groups data pixels into objects according to their spatial arrangement. The technique relies on quantification of surface variability within each of these objects to distinguish landslides from terrain that has not failed. Many features exist on the landscape that can interfere with this variability signal, such as valley bottom streams and associated low-gradient terraces, tightly confined headwater streams, roads bisecting otherwise uniform slopes, and ridge features. Confounding features are classified to exclude them from the dataset prior to executing the landslide classification process. The purpose of this document is to describe the feature extraction process for non-landslides. The image data used for feature extraction is presented first, followed by a description of the object-based routine and feature utilization.

Data Description

Data for this study is created using four software packages, ESRI's ArcGIS v.10.7.1 & ArcGIS Pro v.2.5.0, Trimble's eCognition v.9.5, and R Foundation for Statistical Computing (R). The following packages in R are accessed: TIFF, RTIFF, RASTER, RGDAL, SP, and MASS. All metrics are calculated using 1X1 km lidar dem tiles obtained from Weyerhaeuser Company lidar analysts with interpolation completed using a proprietary procedure. The geographic coordinate system for all data is Universal Transvers Mercator (UTM) North American Datum of 1983(NAD83), zone 10 north. Native raster cell size is 2-m resolution. Lidar tile processing includes mosaicking dems that intersect photo-inventoried deep-seated landslides resulting from a recent storm event occuring in the Willapa Hills of western Washington (Turner et al, 2010). This data selection process ensures deep-seated landslides are present in the study area.

The data extent consists of two overlapping areas: the preliminary study extent attributed with black hatches in Figure A30 consisting of approximately 85 km^2, and a second, smaller extent of 50 km^2 attributed with red hatches. This second extent is a subset of the original study used during model verification. The subset is clipped and then expanded to allow geographic connectivity. Subsetting the data reduced the lidar dem file size from 480 MBs to 170 MBs, and as such improved processing speed.

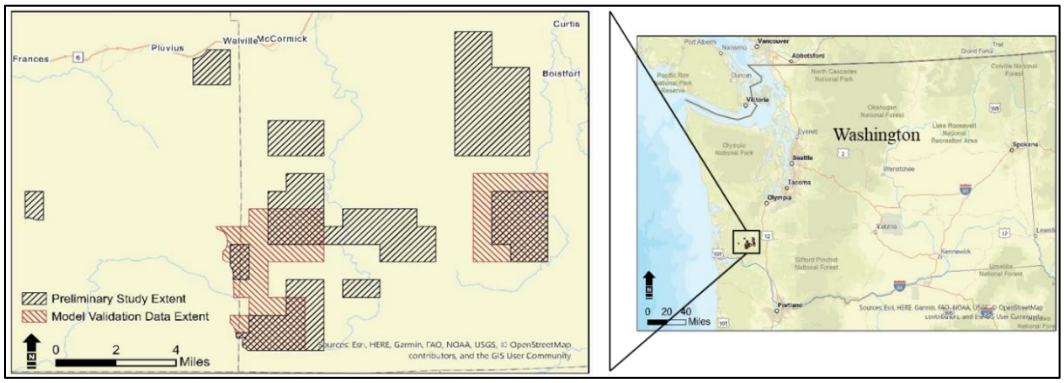


FIGURE A30. EXTENT OF LIDAR DATA FOOTPRINT USED IN STUDY.

GEOBIA is rooted in the utilization of spectral properties of pixels. Lidar derivatives are created in ArcGIS and tailored for spectral analysis by scaling 32-bit floating point data to a range of 0-255. This is accomplished using the Zonal Statistics tool to calculate dataset minimum and maximum values used in the following Raster Calculator function:

$$\left\{\frac{((VALUE - VALUE\ MIN) * 255)}{(VALUE\ MAX - VALUE\ MIN}\right\} \qquad \text{Eq. A1}$$

Images scaled in this fashion include surface curvature and slope gradient, which are used similarly to a hillshade as outlined by Burns & Madin, 2009. Datasets with non-scalable values used in classification routines are left in native format – these include slope gradient, reported as percent change, and flow accumulation data. The remaining images used in the analysis are created within eCognition software: dem's of difference, a mean difference to neighbor metric, and results from an unsupervised clustering routine, all described in detail below. Please note that dataset names referenced in the remainder of this document are bracketed.

- Slope gradient [GRAD], in percent, is calculated by accessing a tool within the Spatial Analyst toolbox included in the ArcGIS Pro software package. This method uses a 3 x 3 moving window to quantify the rate of change in the x and y direction, or slope inclination, for each cell (Burrough and McDonell, 1998).
- Flow accumulation [FAC] is derived within the ArcGIS Pro software package using the D-infinity method, calculating flow in the direction of the steepest slope on a triangular facet (Tarboton, 1997). The dataset value records the sum of contributing grid cells that naturally flow into each cell.
- Landform/surface curvature [BOLSTAD] is quantified within the ArcGIS v.10.1 software package by accessing a Geomorphometric & Gradient Metrics tool published by Jeff Evans, where curvature Bolstad's variant, is defined as the ratio of concavity to convexity of a pixel within a defined window of analysis (Evans et al., 2014). For this study a window size of 15 x 15 was used to best represent headwater streams.
- Dem residuals, or dems of difference [DOD] have been widely used for denoising dems (Hillier & Smith, 2009; Delai et al., 2013). The process decomposes the topographic surface based on predefined scales allowing for extraction of features at the scale of interest. To produce these datasets for a range of features, relief at five different scales is calculated within eCognition software by finding the median height of each cell using 5 x 5, 9 x 9, 15 x 15, 31 x 31, 55 x 55 sliding windows (kernel). The median filtered dem is then subtracted from the original dem. With scale dependent variations smoothed the resulting image separates the data, and makes extractable, remaining features as shown in Figure A31.

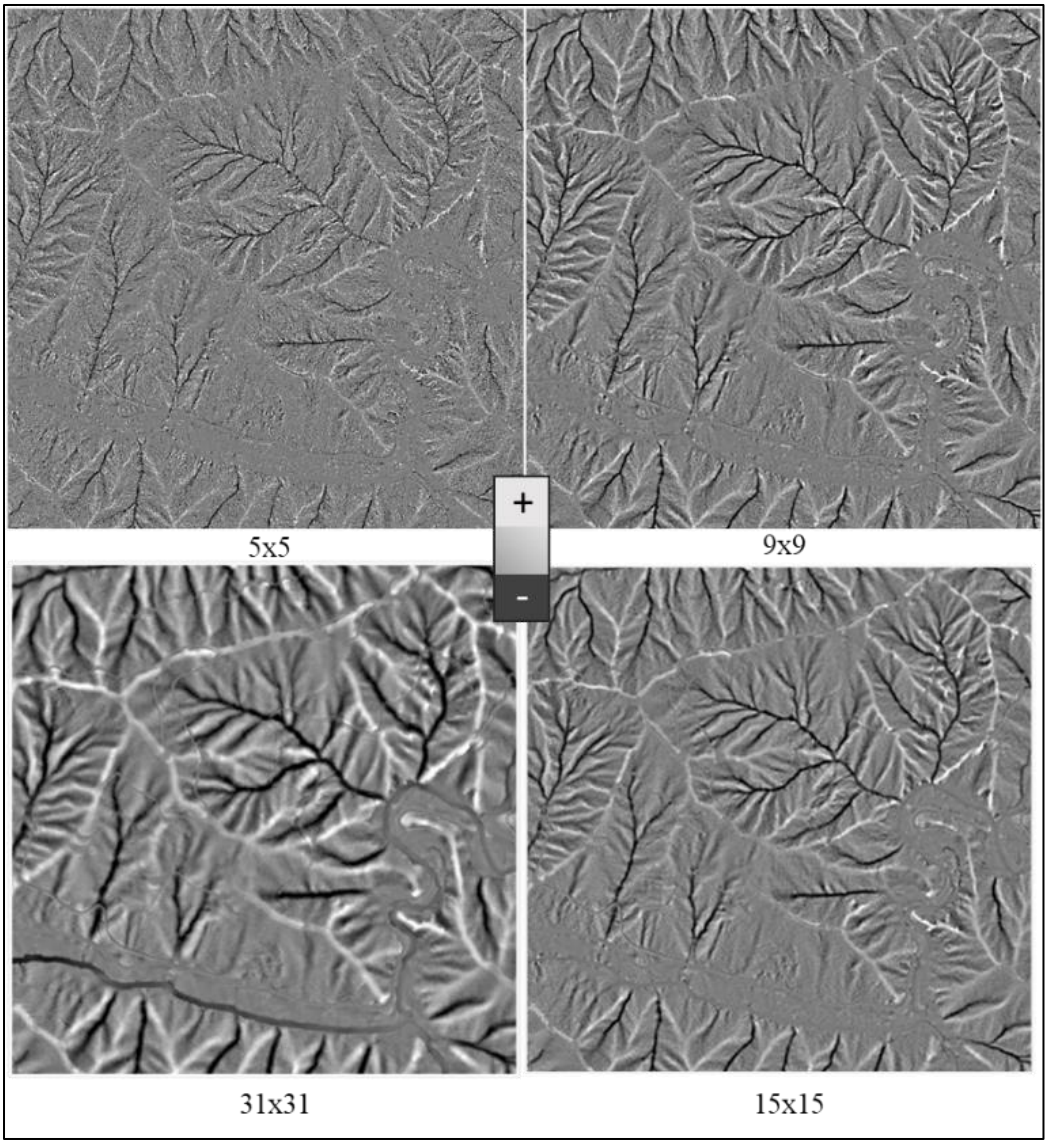


FIGURE A31. DEM OF DIFFERENCE DATA ILLUSTRATION – INCREASING KERNEL SIZE CLOCKWISE, STARTING AT TOP LEFT.

- The Iterative Self-Organizing Data Analysis Technique (ISODATA), or unsupervised classification, used for multi-spectral pattern recognition is run within the eCogntion software package. ISODATA is an iterative technique developed to sort a set of multi-dimensional data into clusters not rigidly constrained by a number of a priori assumptions. The process works to reduce between-cluster variation, as recorded by average and standard deviation, and by iteratively lumping or splitting the clusters (Ball & Hall, 1965). The 5 DOD layers described above are used to create the ISODATA image. The algorithm was allowed to iterate 20 times with the initial number of clusters set at 8, a maximum number of clusters set at 10, and a minimum size of 50 pixels.

- Neighborhood similarity is calculated using a feature available within the eCognition software; ‘mean difference to neighbors’ [DTN]. This algorithm quantifies DTN, $\bar{\Delta}_{k\,(v)}$, as the intensity difference between an image object and its neighbors, as illustrated in Figure A32, by comparing the relationship of the center object to each neighboring objects in a given window size such that (*https://docs.ecognition.com/v9.5.0/#eCognition_documentation/Reference Book/27 Object Features - Layer values/To Neighbors.htm?Highlight=mean difference to*) :

$$\bar{\Delta}_{k\,(v)} = \frac{1}{\omega}\sum_{u\epsilon N_v(d)} w_u(\bar{c}_k\,(v) - \bar{c}_k(u)), \qquad \text{Eq. A2}$$

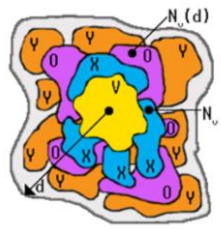


Figure A32. Sample Image Object from Trimble Reference Guide.

where,

- u, v are image objects
- $b(u, v)$ is the length of the common border between u and v
- $\bar{c}_k$ is the mean intensity of image layer k
- $c_k^{min}$is the darkest intensity value of image layer k
- $c_k^{max}$is the brightest intensity value of image layer k
- $\#P_u$is the total number of pixels contained in P_u
- d is the distance between neighbors
- w is the image weight with $w = \sum_{u\epsilon N_v(d)} w_u$, where
  - $w_u$is the weight of image object u, with
    - $w_u = \begin{cases} b(u,v), & d = 0 \\ \#P_u, & d > 0 \end{cases}$
- $N_v$is the direct neighbor to image object v, with
  - $N_v = \{u \,\epsilon\, V_i : \exists\,(x,y)\ \epsilon\ P_v\ \exists\,(x',y')\ \epsilon\ P_u : (x',y')\ \exists N_4(x,y)\}$
  - $N_v\,(d)$ is a neighbor to v at distance d with
    - $N_v(d) = \{u \,\epsilon\, V_i : d\,(v,u)\}$.

DTN is quantified for inverted gradient rasters that have been scaled from 0 to 255 and median filtered using a kernel size of 3. The gradient raster is segmented using a regular grid pattern in the native pixel resolution resulting in 2m^2 objects. The difference to surrounding pixels is calculated using a kernel size of 7, 15, and 25. The resulting DTN feature is then converted to a data layer/image. Figure A33 displays a DTN image produced using a kernel size of 15. Bright portions of the

image highlight terrain features with a large, positive difference to neighbors while dark portions of the image display large, negative differences. In this use case a high positive value will encompass low gradient terrain since the input gradient raster was inverted. As can be seen; ridges, roads, benches, and stream features can be resolved using the DTN technique with a kernel size of 15.

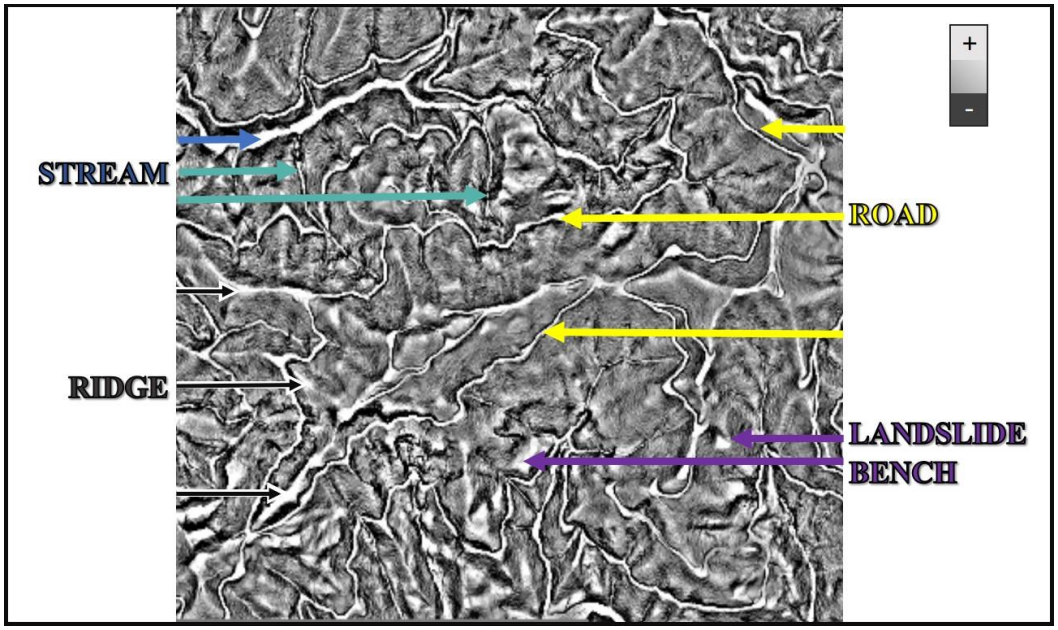


FIGURE A33. IMAGE ILLUSTRATION USING DTN TECHNIQUE. ARROWS INDICATED LANDSCAPE FEATURES WITH HIGH DTN VALUES (WHITE AREAS).

Feature Extraction

GEOBIA has been used in numerous applications to classify features on the Earth's surface including benthic habitat, vegetation, impervious surfaces, glaciers, landslides, and landforms (D'Oleire-Oltmanns et al., 2013; Fu et al., 2019; Korzeniowska, 2017; MacFaden et al., 2012; Parrish et al., 2015; Van Den Eeckhaut et al., 2012). The authors have also reported success with this technique while striving to delineate landforms with greater objectivity, reproducibility, accuracy, and precision as compared to manual classification methods (Shaw et al., 2017). A critical need exists to minimize interference from components of the landscape that are not part of a deep-seated landslide. Our spatially explicit workflow design permits? removing classes, and thus data, from the analysis in a seamless process that allows for higher order connections to be drawn, which lead to more robust classification results. While it is noted that the application of statistically derived thresholds to create morphological indicators has proven feasible, a novel approach centering on spatial connectivity is attempted (Sofia et al., 2014). Throughout this thesis a two-pronged approach is described for extracting data using a traditional threshold value classification approach in tandem with image analysis methods to add meaningful context to the extracted classes. The classification routine is built to work in feature succession, where the spatial relationship of previously classified data

can be used, making the order of these classifications important. For brevity sake, a comprehensive list of algorithms is not included in this appendix. All image-derived thresholds are included, but standard image analysis processing techniques are not fully documented, including geometric selections. Final classification results for streams, roads, and ridges can be viewed in Figure A34. In the interest of time, accuracy metrics were not able to be quantified. A detailed description of the feature extraction process, including process snapshots with classification results displayed using the inverted slope raster as a background, follows.

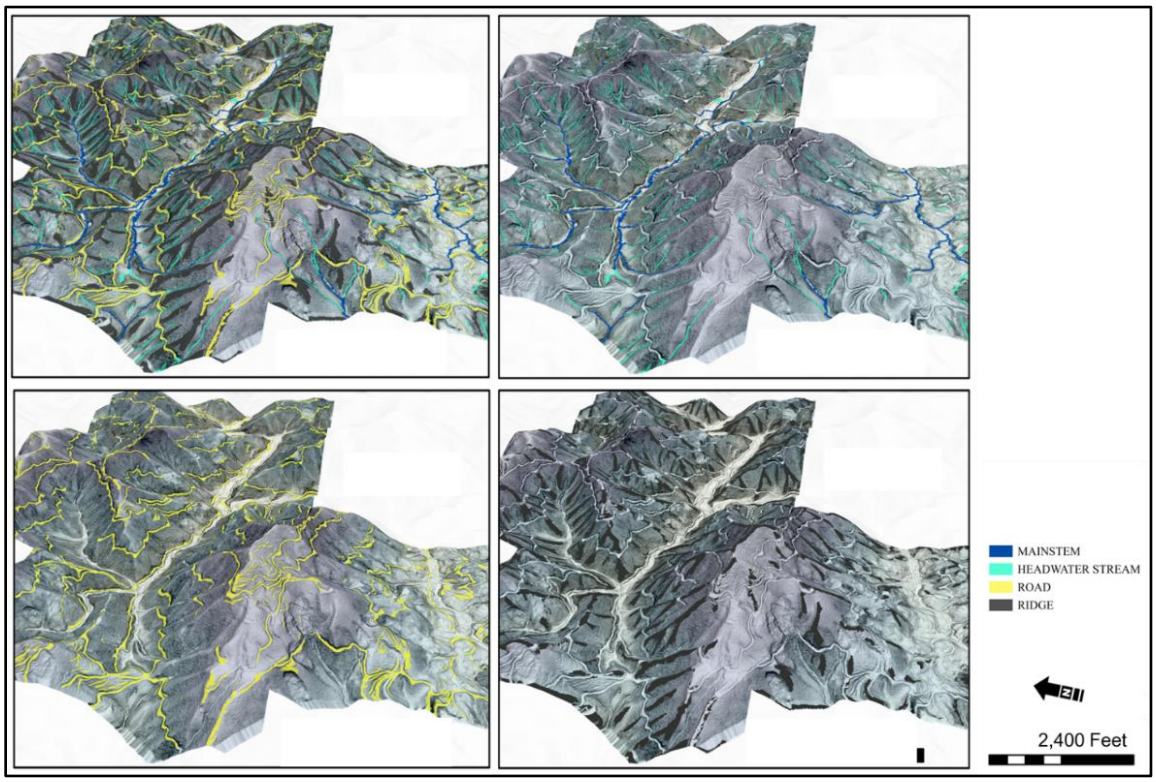


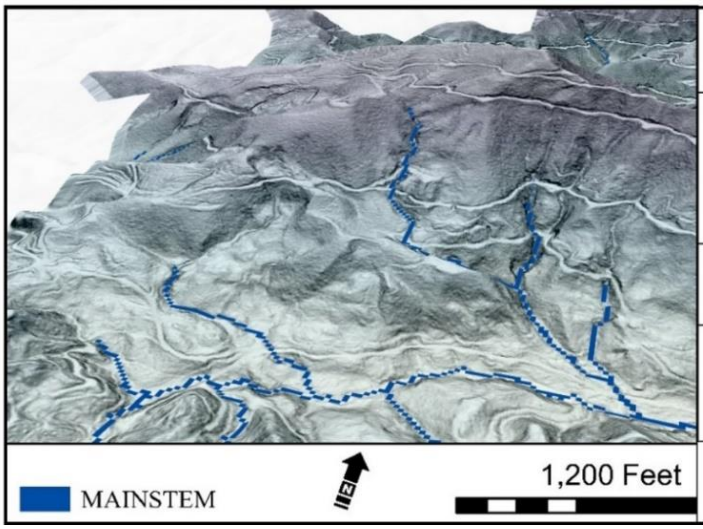
FIGURE A34. CLASSIFICATION RESULTS FOR NON-LANDSLIDE FEATURES. TOP LEFT; ALL FEATURES. TOP RIGHT; STREAMS. BOTTOM LEFT; ROADS. BOTTOM RIGHT; RIDGES.

Stream

This section is devoted to describing the evolution of our modeling approach for streams in a sequential fashion. The classification process for stream features involves four steps that capture the essence of the object-based workflow The approach is designed in recognition that streams exist on the landscape as relatively low gradient features within varying degrees of hillslope confinement where high accumulated flow is recorded. As such, grid-based, flow accumulation data is used as the foundation of our extraction strategy. Preliminary stream classifications are grown into meaningful objects by weaving in additional information from datasets including DOD, GRAD, and DTN while incorporating geometric properties such as object size. It is noted that FAC data is

sensitive to the footprint of the data and would benefit from inclusion of entire drainage basins as compared with the geometric-tile pattern to which our data are clipped.

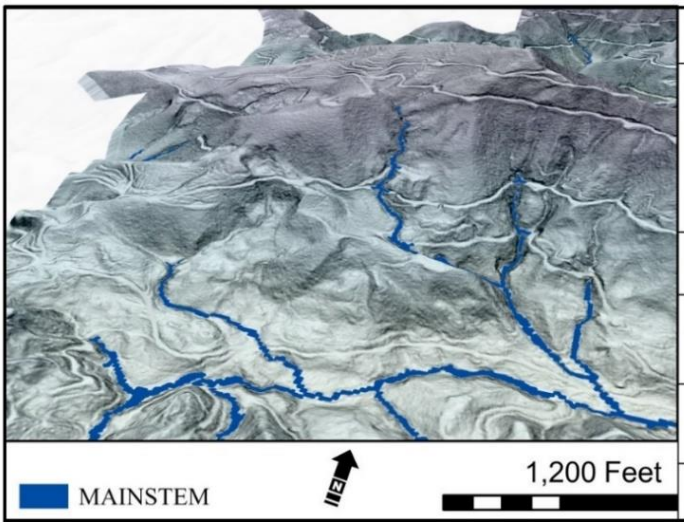
Stream classification begins by running two multi-threshold segmentation routines that identify areas with concentrated flow by splitting flow accumulation (FAC) raster data into objects and classifying those objects using arbitrary, large threshold values for both 2-meter and 10-meter resolution lidar dems. As can be seen with the blue class in Figure A35, this technique serves as a good baseline for identifying pixels where flow is accumulating, but it neither encapsulates the visible bankfull channel feature, nor does it result in a continuous stream pathway -this becomes more evident when using higher resolution lidar data. Also by using only FAC thresholds the complete morphological profile of continuous stream features are not captured, and they appear to stop in arbitrary locations (Figure A35).



ALGORITHM	DATA	SCALE/ VALUE
MULTI-THRESHOLD SEGMENTATION	FAC (2-m)	1 PIXEL (2 m²)
CLASSIFICATION	FAC (2-m)	10,000
MULTI-THRESHOLD SEGMENTATION	FAC (10-m)	1 PIXEL (2 m²)
CLASSIFICATION	FAC (10-m)	1,000

FIGURE A35. CLASSIFICATION RESULTS FOR THE FIRST STEP IN STREAM EXTRACTION PROCESS.

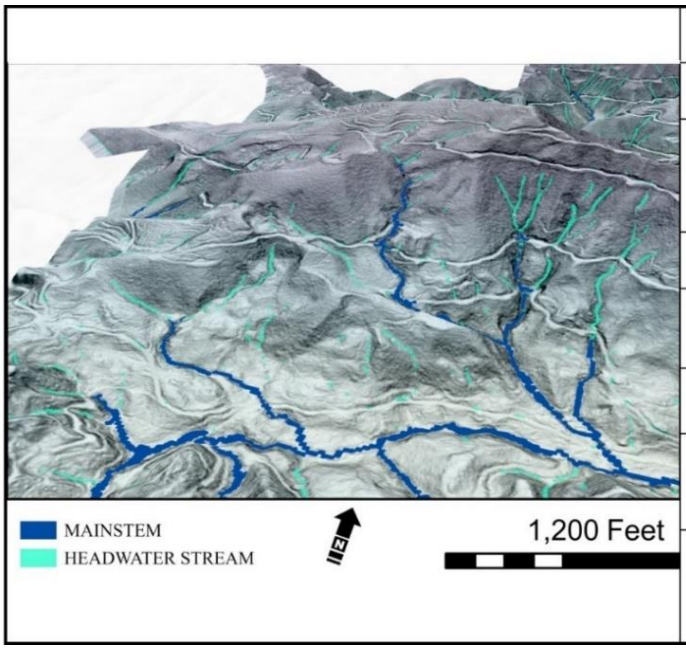
To create meaningful stream segments the initial/preliminary stream classification is used as a basis or seed. The seeds are then grown into adjacent, unclassified pixels using slope gradient (GRAD), neighborhood similarity (DTN), residual topography (DOD), and flow accumulation criteria. The growing process assigns the stream classification to all objects that share a border to the initial stream classification if they meet the criteria outlined in Figure A36. This growing procedure helps fill gaps in the classification where pixels sharing a similar spatial footprint and gradient signature to the previously classified high flow accumulation data are included. As a cleaning procedure, after all classified objects are merged, any small islands of unclassified data existing within the stream objects are included in the stream class. Finally, an object resizing process is employed to smooth pixelated classified objects, making the data more aesthetically pleasing. This process removes any single pixels that protrude from the class and does the inverse where addition of one pixel is necessary to create a smooth border.



ALGORITHM	DATA	SCALE/ VALUE
CLASSIFICATION --0R--	DTN (w7)	> 10
	DTN (w15)	
	DTN (w25)	
CLASSIFICATION	DTN (w7)	> 4
CLASSIFICATION	GRAD	< 30%
CLASSIFICATION --AND--	DOD (w9)	< -1
	GRAD	< 60%
CLASSIFICATION	FAC (10-m)	> 750

FIGURE A36. CLASSIFICATION RESULTS FOR SECOND STEP IN STREAM EXTRACTION PROCESS.

With larger order flow paths classified, a second feature extraction routine is deigned for headwater systems. This routine is similar in nature and begins by recognizing criteria indicative of headwater systems e.g., steeper topography existing within tightly confined areas with less opportunity for flow to accumulate as it's nearer the ridgetop. Seed pixels are initially classified using flow accumulation data with the added criteria of existing in a highly confined area as described by DOD data. Seed data are grown into neighboring pixels in the same fashion as the initial stream classification. The process iteratively grows the seed class into steeper and less confined areas, with attention given to spurious flow accumulation data (Figure A37). The classified data are then cleaned and smoothed in a similar fashion as described above by including small islands of data and adding or removing protruding classified border objects -again using confinement data to constrain the process. It is noted that headwater streams may exist in steeper portions of the landscape, for instance as bedrock cascades along with their?? tributary juctions, but for the purposes of identifying deep-seated landslides, extracting these features was deemed unnecessary.



ALGORITHM	DATA	SCALE/ VALUE
CLASSIFICATION --AND--	FAC (w1)	400
CLASSIFICATION --OR--	DOD (w5)	< -0.5
	DOD (w9)	< -1
CLASSIFICATION	GRAD	< 10%
CLASSIFICATION --AND--	DOD (w5)	< 0
	GRAD	< 20%
CLASSIFICATION --AND--	DOD (w5)	< 0
	FAC (2-m)	> 1000
CLASSIFICATION --AND--	GRAD	< 30%
	FAC (2-m)	> 1000
CLASSIFICATION --AND--	DOD (w9)	< -1
	GRAD	< 60%

FIGURE A37. CLASSIFICATION RESULTS FOR THIRD STEP IN STREAM CLASSIFICATION - HEADWATER STREAM EXTRACTION (AQUA POLYGONS). BLACK ARROW ON TABLE INDICATES AN ALGORITHM RAN IN CONJUNCTION.

The final step in the stream extraction process establishes continuity of the streams and eliminates spurious segments that, as a classified object, do not meet the threshold requirements. Image analysis techniques such as geometry (object size) and neighborhood relationships are relied on to define continuous stream segments. Border effects are cleaned up by incorporating single pixels into the feature that share three sides with classified streams. Classified objects that are separated by one pixel are merged. Classified stream objects that do not overly exist (i.e., do not meet threshold requirements when considering the mean object values) in a tightly confined area are removed. The data is further cleaned by eliminating small, isolated objects that do not meet confinement criteria. Finally, the classified objects are segmented based on gradient values and then re-classified using a mean value of flow accumulation to distinguish larger drainages from the headwater systems. The stream classification shown in Figure A38 result from a conservative approach aimed at minimizing over-classification of features, such as gullies and discontinuous streams, that may exist on deep-seated landslides.

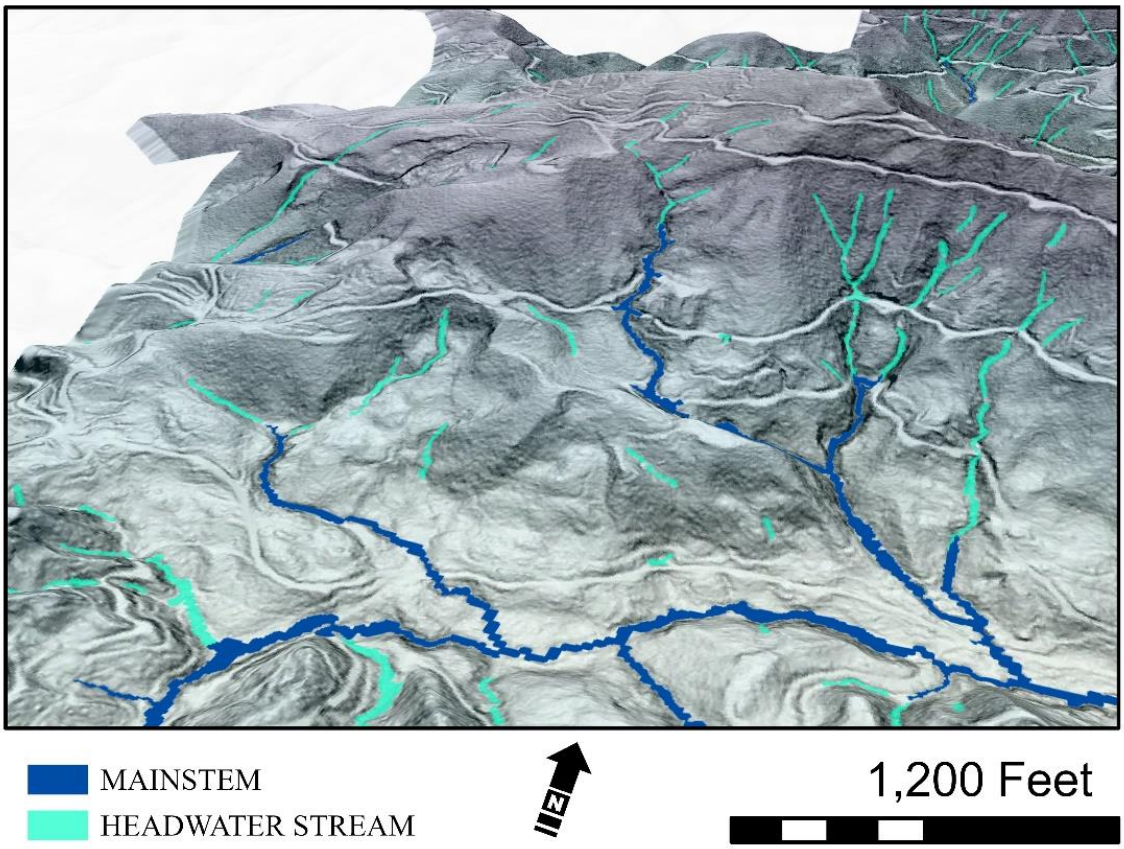


FIGURE A38. FINAL STEP IN CLASSIFICATION OF STREAM FEATURES - HEADWATER AND MAINSTEM FEATURES. THE RESULTS ARE IDENTICAL TO THOSE SHOWN IN FIGURE A34 WITH A SLIGHTLY ADJUSTED VANTAGE POINT.

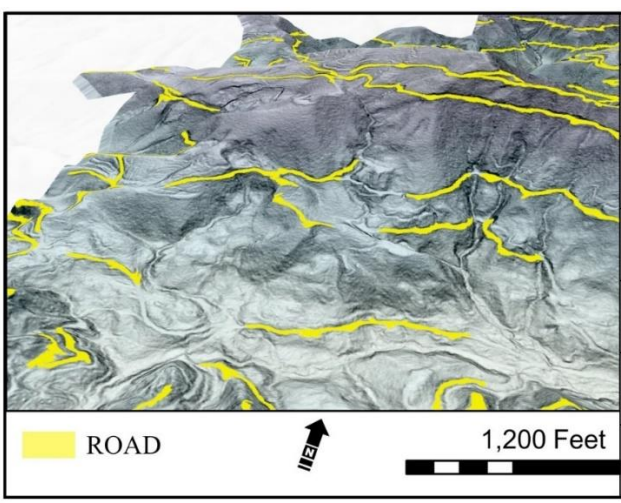
Road

Since hillslope texture is altered in the presence of features such as roads it interferes with the variability signature used to classify deep-seated landslides. A road extraction process is designed similar to stream extraction, in that it relies on neighborhood relationships and geometric properties of segmented objects. However, in the absence of network information, such as direction of steepest decent, neighborhood relationships and geometry are exploited to successfully classify the objects. Within the study area segments of main logging roads have a remarkably consistent geometry that contrast with the adjacent hillslopes. These narrow, long roads are often truncated by landings with the same uniformly low gradient surface but in a teardrop shaped pattern. In addition to main logging roads, anthropogenic features presumed to be old cat roads and/or abandoned roads are visible on the hillslopes. These often-steeper road segments can lack continuity where hillslope processes have started to recapture the surface. On lower gradient topography landscape features were observed that appear to have been traversed by machines traveling a short distance adjacent to the roads. These areas do not have crisp delineations bounding the road surface but do display a hummocky texture, presumed to

be anthropogenic in nature. The feature extraction process for roads is designed to find the equilibrium between road-related signatures and deep-seated landslide features. The multi-faceted approach is described as two steps. The first step in the process relies on data values to classify objects, while the second relies almost exclusively on geometry.

To enhance the success of the geometric classification process a barrier class, consisting of steep slopes, is used because often these portions of the hillside help distinguish road features. This class will act as a boundary to assist with the effectiveness of the road extraction process, where cuts into the hillside that are necessary to construct the roads result in distinct, isolated steep terrain features adjacent to flat, narrow road segments. A seed class is created using large negative values from the DTN image in a multi-threshold segmentation algorithm, and expanded by including neighboring steep slopes over 70%. It is further cleaned by encompassing any small islands (less than 30 pixels) of unclassified data and then removing any classified objects that were less than 10 pixels in size.

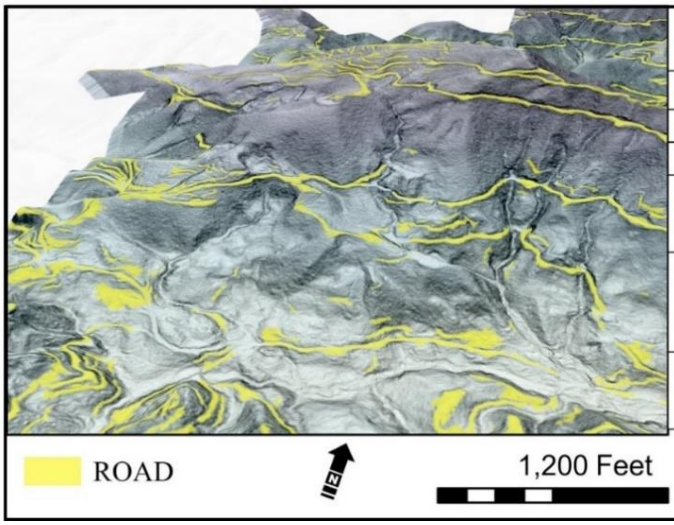
To extract main logging roads, a classification that can be thought of as the inverse to the steep slope classification is used, where large, positive DTN values are classified as seeds. These classified objects are then expanded into neighboring objects with slightly lower threshold values. The expansion is controlled by the neighboring candidates relative area to the seed class. Geometric properties of the objects are used to remove any classified objects that are not shaped like a road using a density feature described in Table A1 as pertaining to the structure of a box. Interfering knife edge ridge features are also eliminated from the road class by excluding any convex up objects using BOLSTAD curvature values. Finally, any object longer than 100 pixels is classified as a road while the seed class is left intact to allow for additional classification using alternative metrics (Figure A39). As a final step in the classification, any objects that were extracted as roads, but have high FAC values (>1500) and are concave as defined using BOLSTAD curvature values less than the 25th percentile, are transferred into the headwater class. This initial road classification required 14 algorithms to classify 2.2 km^2 of roads within the 50 km^2 study area. The remaining 1.4 km^2 of roads classified required 56 algorithms.



ALGORITHM	DATA/ FEATURE	SCALE/ VALUE
MULTI-THRESHOLD SEGMENTATION	DTN (w15)	1 PIXEL (2 m^2)
CLASSIFICATION	DTN (w15)	> 7
CLASSIFICATION	DTN (w15)	> 5
CLASSIFICATION	DENSITY	<= 1.5
CLASSIFICATION	BOLSTAD (w15)	< 90th PERCENTILE
CLASSIFICATION	LENGTH OF MAINLINE (SKELETON)	> 100

FIGURE A39. FIRST STEP IN ROAD EXTRACTION PROCESS BASED ON IMAGE VALUES.

Due to the number of algorithms involved in this process and the iterative nature of the routine a complete list of thresholds is not provided. Instead the techniques employed are described along with the features used. The process to extract the remaining road features uses the same seed class described above with the addition of a second seed class with a DTN (w7) threshold value of greater than 3. The same set of methods outlined in Figure A39 and described above are used again with final road classification being based on a suite of geometric properties listed in Figure A40.



ALGORITHM	FEATURE	RELATIVE VALUE
CLASSIFICATION	ASYMMETRY	HIGH
CLASSIFICATION	DENSITY	LOW
CLASSIFICATION	COMPACTNESS	≠ 1
CLASSIFICATION	LENGTH OF MAINLINE (SKELETON)	HIGH
CLASSIFICATION	LENGTH OF LONGEST EDGE (POLYGON)	HIGH
CLASSIFICATION	NUMBER OF EDGES (POLYGON)	LOW

FIGURE A40. SECOND STEP IN ROAD CLASSIFICATION STRATEGY BASED ON GEOMETRIC PROPERTIES.

ASYMMETRY	How well the image object fits into a bounding ellipse. The metric quantifies variance in the x- and y- directions of the object. High values indicate asymmetric, or non-uniform objects. $$\frac{\sqrt{\left(\frac{1}{4}\right)\left(Var(X)+Var(Y)\right)^2+\left(Var(X)(Y)\right)-Var(X)Var(Y)}}{Var(X)+Var(Y)},$$ where Var(X) is the variance of X and Var(Y) is the variance of Y.	Eq. A3

DENSITY	Describing the distribution in space of the pixels of an image object. It is calculated by the number of pixels divided by the approximated radius, based on the covariance matrix. The densest objects are squares with lower values representing filament shapes. $$\frac{\sqrt{\#P_v},}{1+\sqrt{Var(X)+Var(Y)}},$$ Eq. A4 where $\sqrt{\#P_v}$ is the diameter of a square object with $\#P_v$ pixels and $\sqrt{Var(X)+Var(Y)}$ is the diameter of the ellipse enclosing the object .
COMPACT-NESS	Distribution of length and width as compared to area. $$\frac{L_v W_v}{\# P_v},$$ Eq. A5 where L_v is the length of object v and W_v is the width of object v.
LENGTH OF MAIN LINE	Each object can be decomposed into a main line and branches. These segments are determined by nodes, or midpoints resulting from Delaunay triangulation. Length is the sum of distances between these nodes.
LENGTH OF LONGEST EDGE	Polygons result from the vectorization of pixels that form an image object cluseter. The vectors smooth raster-based corners. The length of longest edge feature returns the largest distance between vector nodes.
NUMBER OF EDGES	The number of lines created by nodes.

TABLE A1. GEOMETRIC FEATURES AVAILABLE WITH ECOGNTION SOFWARE PACKAGE.

Ridges

Like streams, ridges can interfere with landslide classification due to an abrupt topographic change from convex to concave creating a spike in variability. Unlike concave stream valleys in this study area, convex ridges occur at a much larger range of curvature scales, from sharp to very broad. As such a method similar to the stream extraction process while using inverted DEM's was not explored. Instead an ISODATA unsupervised classification algorithm to extract multiple scales of convex, ridge features. Within eCognition a hierarchical process is created that allows for classifications that overlap the spatial extent of the previous classifications (i.e., a level with no classifications). This allows us to segment and classify continuous ridge features without considering previously classified data. Results from the ISODATA clustering algorithm

assigned one of six possible values to all pixels in an image. The image was visually investigated and the value corresponding with ridges is used for classification. Object boundaries are then cleaned using a pixel-based object resizing algorithm to eliminate protruding pixels. A minimum mapping unit for ridge class is set to 500 pixels. This hierarchically aligned data is then reincorporated into only the remaining, or unclassified data, from the previous steps. Ridge classes are extracted at a large scale so it's possible that data generated with bigger window sizes can record convex values where steep-breaks in slope are occurring. Potential slope breaks by preserved by classifying them prior to importing the final ridge features (Figure A41). These data will be useful in the segmentation process for deep-seated landslides. Steep, breaks in slope are calculated by segmenting the DTN metric produced at a window size of 25 and classified using any value less than negative 7. The original classification is then grown into any pixels with a DTN25 value of less than negative 4 and cleaned by removing small islands and small classified objects.

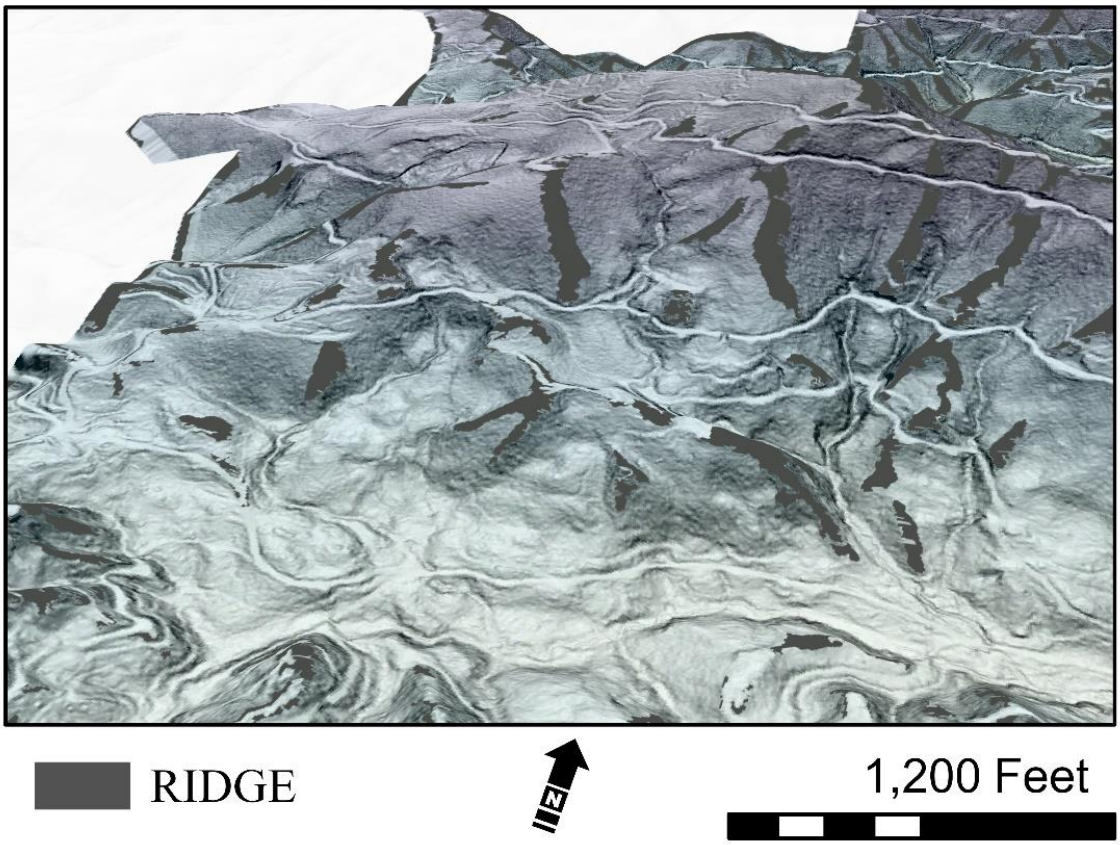


FIGURE A41. CLASSIFICATION RESULTS FOR RIDGE FEATURES.

Bibliography for Appendix

Baatz, M., & Schape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Information*, *XII*, 12–23.

Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*. Stanford Research Institute, Office of Naval Research and Information Sciences, Menlo Park, CA.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

D'Oleire-Oltmanns, S., Eisank, C., Drăgut, L., & Blaschke, T. (2013). An object-based workflow to extract landforms at multiple scales from two distinct data types. *IEEE Geoscience and Remote Sensing Letters*, *10*(4), 947–951. https://doi.org/10.1109/LGRS.2013.2254465

BranchBurrough, P. A., & McDonell, R. A. (1998). Principles of Geographical Information Systems. *Oxford University Press*, New York, 190 pp. https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-slope-works.htm

Burns, W. J., & Madin, I. P. (2009). Protocol for Inventory Mapping of Landslide Deposits from Light Detection and Ranging (lidar) Imagery, 1–36.

Evans JS, Oakleaf J, Cushman SA, Theobald D (2014). An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling. version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial. Accessed: 2019.

Fu, Y., Liu, K., Shen, Z., Deng, J., Gan, M., Liu, X., … Wang, K. (2019). Mapping Impervious Surfaces in Town – Rural Transition Belts Using China ' s GF-2 Imagery and Object-Based Deep CNNs. *Remote Sensing*, *11*(280), 17. https://doi.org/10.3390/rs11030280

Florinsky, I. V. (2016). *Digital terrain analysis in soil science and geology*. (A. Shapiro & T. Frank, Eds.) (Second). Amsterdam: Elsevier.

Hiller, J. K., & Smith, M. (2008). Residual relief separation: digital elevation model enhancement for geomorphological mapping. *Earth Surface Processes and Landforms*, *33*, 2266–2276. https://doi.org/10.1002/esp.1659

Ironside, K. E., Mattson, D. J., Arundel, T., Theimer, T., Holton, B., Peters, M., … Hansen, J. (2018). Geomorphometry in Landscape Ecology: Issues of Scale,

Physiography, and Application. *Environment and Ecology Research*, *6*(5), 397–412. https://doi.org/10.13189/eer.2018.060501

Korzeniowska, K. (2017). *Object-based image analysis for detecting landforms diagnostic of natural hazards*. University of Potsdam.

MacFaden, S. W., O'Neil-Dunne, J. P. M., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*, *2012*(063567), 1–23. https://doi.org/10.1117/1.JRS.6.063567

Parrish, C. E., Dijkstra, J. A., O'Neil-Dunne, J. P., McKenna, L., & Pe'eri, S. (2015). Coastal Education & Research Foundation, Inc. *Journal of Coastal Research*, *SI*(76), 200–208.

Shaw, S. C., Justice, T. E., McCarthy, R., Hinkle, J. C., Turner, T. R., Fransen, B. R., … Thornton, J. (2017). Automated, Object-Based Image Analysis (GEOBIA) Model for Landform Detection and Mapping, with Applications to Pacific Northwest USA Landslide Assessments. In *AEG; 3rd North American Symposium on Landslides* (pp. 9–20).

Sofia, G., Fontana, G. D., & Tarolli, P. (2014). High Resolution Topography and Anthropogenic Feature Extraction. *Hydrological Processes*, *28*, 2046–2061. https://doi.org/10.1002/hyp.9727

Team, R. C. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, *33*(2), 309–319.

Turner, T. R., Duke, S. D., Fransen, B. R., Reiter, M. L., Kroll, A. J., Ward, J. W., … Bilby, R. E. (2010). Landslide densities associated with rainfall, stand age, and topography on forested landscapes, southwestern Washington, USA. *Forest Ecology and Management*, *259*(12), 2233–2247. https://doi.org/10.1016/j.foreco.2010.01.051

Van Den Eeckhaut, M., Kerle, N., Poesen, J., & Hervás, J. (2012). Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology*, *173–174*, 30–42. https://doi.org/10.1016/j.geomorph.2012.05.024

Witharana, C., & Civco, D. L. (2014). Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy

measure Euclidean distance 2 (ED2). *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 108–121. https://doi.org/10.1016/j.isprsjprs.2013.11.006