

SEMI-AUTOMATIC LANDSLIDE DETECTION USING SENTINEL-2 IMAGERY: CASE STUDY IN
THE AÑASCO RIVER WATERSHED, PUERTO RICO

AN ABSTRACT

SUBMITTED ON THE 22ND OF NOVEMBER

TO THE DEPARTMENT OF EARTH AND ENVIRONMENTAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF SCIENCE AND ENGINEERING

FOR THE DEGREE


OF

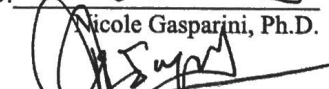
MASTER OF SCIENCE

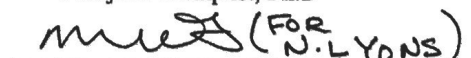
BY

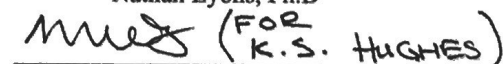

Sabrina Martinez

APPROVED:


Nicole Gasparini, Ph.D.


Torbjörn Törnqvist, Ph.D.

 (FOR N. LYONS)
Nathan Lyons, Ph.D.

 (FOR K.S. HUGHES)
K. Stephen Hughes, Ph.D.

ABSTRACT

As satellite imagery has increased in attainability as well as in spatial and temporal resolution, there is a growing effort to determine how to best use this imagery to automatically detect landslides. The field of automatic landslide detection is still in its infancy, however, and many of the proposed workflows require specialized software making them inaccessible for many geoscientists. The goal of this study is to use widely accessible software and imagery to detect landslides as well as evaluate results using robust evaluation metrics. Methods were developed in a 46 km² training site within the Añasco River Watershed in Puerto Rico and then applied across the entire 415 km² watershed. Change-detection and machine-learning classification methods were used to detect landslides triggered by Hurricane Maria in September 2017. The results illustrate that Sentinel-2 satellite imagery can be used to identify landslides, but detection methods need to be further developed and properly evaluated to increase accuracies in mapping high density landslide events. We also highlight the need for a shift toward more open and accessible means of landslide detection methods for the broader geoscience community to fully recognize the benefits of automatic landslide detection.

SEMI-AUTOMATIC LANDSLIDE DETECTION USING SENTINEL-2 IMAGERY: CASE STUDY IN
THE AÑASCO RIVER WATERSHED, PUERTO RICO

A THESIS

SUBMITTED ON THE 22ND OF NOVEMBER

TO THE DEPARTMENT OF EARTH AND ENVIRONMENTAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF SCIENCE AND ENGINEERING

FOR THE DEGREE

OF

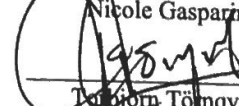
MASTER OF SCIENCE


BY

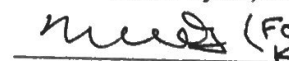

Sabrina Martinez

APPROVED:


Nicole Gasparini, Ph.D.


Torgun Törnqvist, Ph.D.

 (FOR
Nathan Lyons, Ph.D.

 (FOR
K. Stephen Hughes, Ph.D.

ACKNOWLEDGEMENTS

The work presented here would not have been possible without the support and guidance of several people and organizations. I would like to thank my advisor, Nicole Gasparini, for being an amazing source of guidance and support throughout my time at Tulane University. Thank you for instilling confidence in me by recognizing my capabilities as a scientist. Thank you for always fostering a supportive and inclusive environment as well as giving me the freedom and framework to pursue research efforts that I am deeply passionate about.

Thank you to my committee: Stephen Hughes, Nathan Lyons and Torbjörn Törnqvist for offering their time and guidance throughout the duration of the project. I would like to thank Stephen Hughes for being a generous collaborator in addition to the SLIDES-PR group at the University of Puerto Rico Mayagüez for their mapping efforts. I would like to thank Nathan Lyons for being a wonderful mentor and for being so influential in this work. You are an incredible teacher and I have learned so much from you. Thank you to Torbjörn Törnqvist for being a valuable source of support and guidance throughout my time as a graduate student.

I would like to thank the Tulane University Carol Lavin-Bernick faculty grant program as well as the Department of Earth and Environmental Sciences for providing funding for myself and this work. I would also like to thank the European Space Agency (ESA) and the United States Geological Survey (USGS) for making the satellite imagery and data used in this work free and accessible for all.

Lastly, I would like to extend my gratitude to the amazing friends that I have made in my time here at Tulane. I am immensely grateful for your generosity and support over the last two years.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	iv
LIST OF FIGURES	v
1. INTRODUCTION.....	1
2.BACKGROUND.....	2
2.1 <i>History of Object-Based Landslide Detection</i>	3
2.2 <i>Landslide Inventories and Landscape Evolution Models</i>	7
3.STUDY AREA AND HURRICANE MARIA.....	9
3.1 <i>Training Site and Añasco Watershed</i>	9
3.2 <i>Hurricane Maria</i>	10
4.METHODS	11
4.1 <i>Data</i>	12
4.2 <i>Thresholding</i>	14
4.3 <i>Segmentation</i>	15
4.4 <i>Training and Classification</i>	17
4.5 <i>Accuracy Assessment</i>	18
4.6 <i>Frequency-Area and Frequency-Volume Distributions</i>	19
5. RESULTS.....	22
5.1 <i>Spatial Analysis of Error</i>	22
5.2 <i>Landslide Area and Volume Estimations</i>	29
6. DISCUSSION.....	33
6.1 <i>Consequences of Inaccuracy: Interpretations from Erroneous Estimates</i>	34
6.2 <i>Errors in this Study and Their Causes</i>	35
6.3 <i>Scaling Problems: Study Area Size and Landslide Densities</i>	36
6.4 <i>Segmentation and Thresholding</i>	37
6.5 <i>Accuracy Assessments</i>	40
7. CONCLUSIONS	41
APPENDIX.....	44
REFERENCES.....	52

LIST OF TABLES

Table 1: Accuracies at training site and watershed	22
Table 2: Details of accuracy assessment for watershed and training site.....	25
Table 3: Landslide area statistics for true positives, false negatives.....	29
Table 4: Estimated total area and volume for manual and automatic inventories.....	30
Table 5: Frequency-area distribution information.....	31
Table A1: Compilation of accuracy metrics and reported accuracies.....	50
Table A2: Threshold points used to create power-law fit	51

LIST OF FIGURES

Figure 1: Example of Mean-Shift Segmentation.....	3
Figure 2: Location of Training Site and the Añasco river watershed.....	10
Figure 3: Workflow used in this study to detect landslides.	12
Figure 4: NDVI and change detection data for training site.	16
Figure 5: Example illustrating accuracy assessment workflow.	21
Figure 6: Example of IOU values.	24
Figure 7: Spatial distribution of false positive results in the watershed.....	26
Figure 8: Spatial distribution of false positive results in the training site.	26
Figure 9: Spatial distribution of false negative results in the watershed.	27
Figure 10: Spatial distribution of false negative results in the training site.	27
Figure 11: Landslide density maps.....	28
Figure 12: Frequency-Area distributions.....	32
Figure 13: Frequency-Volume distributions.....	32
Figure 14: Volumetric estimations associated with the Wenchuan Earthquake.....	35
Figure 15: Comparison of number of landslides mapped.....	38
Figure A1: Google Earth Engine code used to download Sentinel-2 imagery.....	44
Figure A2: Thresholding and Segmentation Code.....	45
Figure A3: Accuracy Assessment Code.....	46

1. INTRODUCTION

Mobilization of large amounts of sediment by landsliding in mountainous regions can drastically alter river drainage networks and influence the growth of mountain ranges (Korup et al., 2010). The amount of material eroded from landslides and the density of landslides can vary spatially. This variation can be influenced by a variety of factors, including the triggering mechanism as well as lithologic differences (Densmore and Hovius, 2000). The climate of a region, in both space and time, can also influence the distribution and frequency of landslides. Thus, in areas that are vulnerable to events such as tropical cyclones and/or earthquakes, it is most likely that landslides play an important and primary role in long term landscape evolution (Hewitt, 1998).

One way to better understand the role of landslides in landscape evolution is to generate and carefully study landslide inventories. Landslide inventories are catalogs that are used to display the location of landslides in the landscape as either points or polygons. Knowing the location, shape, and size of previous landslides can help to better understand what areas are more prone to failure and inform disaster mitigation efforts (Parise, 2000). Additionally, these inventories can be used to assess the damage after a large landslide-triggering event. Landslide volume, which is the total amount of material removed after a mass-wasting event, can also be estimated from an inventory using landslide volume area scaling relationships (Larsen et al., 2010; Klar et al., 2011). Landslide volumes can help answer questions regarding the impact that mass wasting has on landscape evolution and highlights the merit in developing accessible and efficient methods to create landslide inventories. This goal of this study was to develop a method to automatically detect landslides from the European Space Agency's (ESA) Sentinel-2 satellite imagery

using object-based image classification methods. To develop our methods, we tested various workflows and their effectiveness at mapping landslides that were triggered by Hurricane Maria (September 20, 2017) in Western Puerto Rico. After determining a satisfactory workflow, we analyzed the results to determine where future studies should focus their efforts. We also propose new robust standards and metrics for assessing accuracy in automatic landslide detection studies to address the discrepancy in accuracy metrics used in the literature.

2. BACKGROUND

Landslide inventories are traditionally created by either mapping landslides in the field (Cardinali et al., 2006; Santangelo et al., 2010) or by visually inspecting satellite and/or aerial images (Gao and Maroa, 2010; Coe et al., 2018). Both techniques are generally time consuming and tedious. Manually mapping landslides in this manner is also subject to a considerable element of human error and subjectivity. Recent work has aimed to map landslides automatically using various remote sensing data and classification techniques. Early classification techniques involved mapping landslides using pixel-based methods (Keyport et al., 2018). Pixel based methods, however, are not effective at capturing the shape and area of landslides (Blaschke, 2010). These methods also result in errors due to the ‘salt and pepper’ effect where individual pixels are incorrectly classified (Blaschke, 2010). Thus, object-based techniques have become increasingly popular and the standard in automatic landslide detection studies. Incorporation of ancillary information such as topography and multi-temporal data has become increasingly common as well (Martha et al., 2010; Lyons et al., 2014).

Object-based techniques aim to classify features by first turning the pixels within a target image into sets of spectrally homogenous segments or objects (Figure 1). In doing

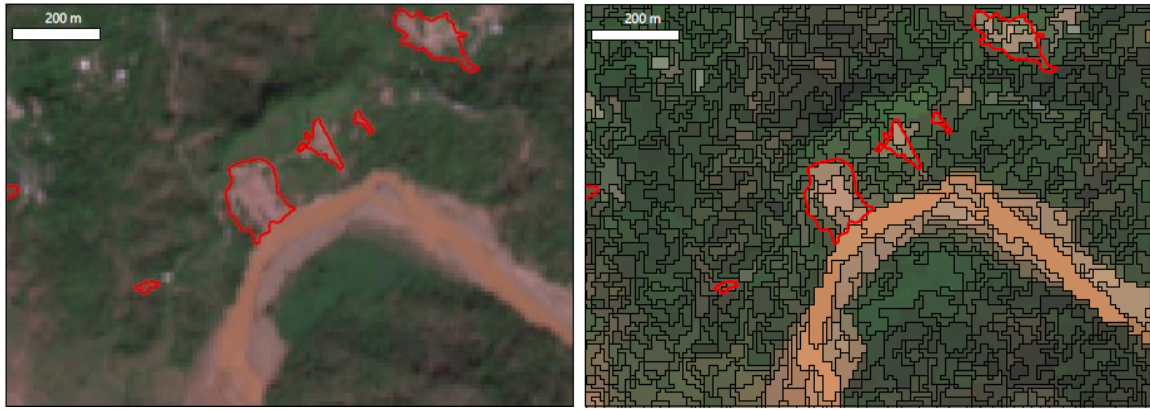


Figure 1: Example of mean-shift image segmentation performed on Sentinel-2 image. Landslides, mapped by the SLIDES-PR group (Hughes et al., 2019), are outlined in red. The left image shows landslides along a river channel and right image displays the image broken down into homogenous segments. Location of landslides is along highway 109 in the Añasco municipality in Western Puerto Rico. The exact location can be seen in figure 2, symbolized by a yellow triangle.

so objects can be classified rather than pixels in a way that mimics human interpretation (Blaschke, 2010). Object-based classification methods have thus far proven more successful than pixel-based methods in being able to automatically detect landslides (Moosavi et al., 2014). Despite this improvement, there are still uncertainties regarding the best practice(s) for identifying landslides in segmented images. While there are a wide variety of post-segmentation procedures that can be performed to identify landslides, there is no clear consensus as to which methods, if any, are versatile and applicable across a wide variety of landscapes and scenarios.

2.1 History of Object-Based Landslide Detection

Barlow et al (2003) were the first to use object-based methods to identify landslides. In their work they sought to identify landslides in the Cascade Mountains of western North America using a Digital Elevation Model (DEM) and multispectral satellite imagery

(Landsat, 30-m spatial resolution). They used a classification workflow where they first eliminate areas in the image where landslides are least likely to occur, then classify the image based on manually determined rules for each class in the image. The authors reported an accuracy of 75% in the detection of landslides over 10^4 m². While a reported success, the workflow presented was not completely automatic as the authors had to define individual rulesets for each class specifying how the classifier would identify objects. These classes and their defined rulesets were specific to the landscape classified making the method not easily transferable to other landscapes. In addition to this, their accuracy assessment was performed on only 40 randomly sampled landslides throughout the study area, so the reported accuracy may be different than it truly would be if the entire study area was assessed. The reported accuracies may also be unreliable as the images used to create the manual landslide inventory were taken five years prior to the images used for classification. In their follow up work, they used a similar workflow on a higher resolution dataset to detect specific mass wasting processes (debris slides, debris flows, and rock slides) (Barlow et al., 2006). They reported accuracies of 90% for detection of debris slides, 60% for debris flows and 80% for rock slides. While the higher resolution data was proven useful for specific mass-wasting detection, they still relied on the same methods previously mentioned and acknowledged that the workflow might not be applicable in arid regions with no vegetation.

Martha et al (2010) used object-oriented methods to classify landslides in the Himalayas. They used topographic data and multispectral imagery to identify landslides according to Varnes' classification scheme (Varnes, 1958). Their work involved first segmenting the imagery, thresholding the imagery using a manually determined cutoff

value to determine potential landslide candidates and then, using a secondary thresholding scheme, to remove false positives (areas incorrectly mapped as landslides). A drawback of their study, as mentioned in their follow-up work, is the manual thresholding technique used to determine potential landslide candidates (Martha et al., 2011). They chose landslide candidates by using the mode of Normalized Difference Vegetation Index (NDVI) values in their study area as the cutoff point that differentiates non-vegetated and vegetated areas. Thus, landslide candidates would be those areas in the images that are relatively less vegetated. This method, while effective in their study, may not be a useful threshold method for other landscapes. A data driven approach, which determines this threshold from the NDVI data for a particular landscape, is a more appropriate approach and suggested in their follow-up work (Martha et al., 2010). Another drawback of their study is that the workflow is reliant on commercial software that uses proprietary algorithms not available to those without access to Definiens eCognition (Aguilar et al., 2016). Unfortunately, the early and successful use of eCognition for the development of automatic landslide detection methods has set a precedent, and its use has proliferated in the literature.

Stumpf and Kerle (2011) explored using a machine learning classifier to identify landslides. Their workflow consisted of segmenting the target images, determining suitable classifier metrics, classifying the images and then, through error analysis, determining the amount of training samples for each class that resulted in balanced over-prediction and under-prediction errors. Their findings suggest that suitable object-metrics will vary considerably by case and suggest that, to improve automatic landslide detection, segmentation techniques need to be improved and the incorporation of pre-event imagery

would be beneficial. Since then, incorporating multi-temporal data, when pre-event imagery is available, has become common practice.

In Li et al (2016a) multi-temporal data was used to map landslides after an extreme rainstorm triggered landslides in Lantau Island, Hong Kong. In their study they aimed to create a workflow that can be universally applied to map landslides. They used change detection maps to map the landslides after a two-step thresholding and segmentation workflow. While a promising method and premise, they point out that the method chosen for final segmentation is prone to resulting in over detection since it has difficulties detecting landslides with unclear boundaries.

Qin et al (2018) were the first to use change-detection methods to test Sentinel-2 data for automatic landslide detection. With their methods they achieved reported accuracies of 95.93% and 88.03%. Their study area, which was in the Sichuan Province in China, however, was only 16.5 km². Because of their small study area, the images used to detect landslides lacked any other features, such as similarly colored river channels, urban areas or rock exposures, that may have been mistaken for landslides. These features, which present a challenge when mapping, are usually encountered in landscapes at larger scales.

In general, the consensus among studies is that object-based methods are more successful at automatically detecting landslides from satellite imagery than pixel-based methods. Most work to date has aimed to determine what slight modifiers in thresholding and segmentation may lead to more accurate results. Most of the studies, however, are undertaken in small training areas and aim to map a small number of landslides with high-resolution data. While useful in testing out new scenarios, focusing on small training sites prevents the discovery of methods that would be successful in events that trigger a large

number of landslides at once. In addition to this, the data, software and algorithms commonly used are often not readily accessible to many. While the field is continuing to test methods on case studies, the means in which they are carrying out these studies does not always make the method more accessible to geoscientists or other stakeholders who would need to create these landslide inventories.

2.2 Landslide Inventories and Landscape Evolution Models

Current landscape evolution models emphasize fluvial incision as the primary erosive mechanism in landscapes. Hillslope processes, such as landslides, are considered secondary and reactive to changes in fluvial incision. Yet, stochastic hillslope processes, such as hurricane-triggered landslides have been shown to exert a primary control on changes in the landscape (Hewitt, 1998; Dadson et al., 2004). After Typhoon Morakat in Taiwan, Yanites et al (2018) observed that wider channels occurred where channel steepness and erosion rates were higher, an unexpected result as previous models suggest a negative relationship between channel width and erosion rate (Yanites et al., 2010). They suggest that hillslopes, in this landscape, exert a primary control on the rivers. When landslide sediment is delivered to the channels it propagates through the fluvial network and aggrades the beds and banks of the river thus decreasing the channel depth and increasing the width of the channel across the former floodplain or up the valley walls (Yanites et al., 2018). This study is just one example that highlights how our current models for channel evolution are lacking important variables and drivers. Thus, gaining a better understanding of hurricane-triggered landslides is an important step in developing a better understanding of landscape evolution.

One step toward understanding the role of landslides in landscape evolution is to quantify the amount of sediment/material eroded in a large, widespread landslide event. Doing so can help derive an erosion rate representative of specific events. Hurricane Maria is an example of an unprecedented landslide triggering event that is likely to occur again as rising sea-surface temperatures (Lim et al., 2018) increase the likelihood of similar hurricanes forming. There is evidence to suggest that climate change is contributing to the growing intensity of hurricane seasons, as seen in 2017 (Lim et al., 2018). Further, the amount of rainfall due to Hurricane Maria was found to have been enhanced by climate change related factors (Ramos-Scharrón and Arima, 2019; Keelings and Hernández Ayala, 2019). Thus, understanding the erosive power of Maria can provide insight into how a landscape prone to landsliding will evolve as hurricane-driven erosion is forecasted to increase in frequency and magnitude.

Landslide volumes are an important variable to quantify as the information can be used in sediment budget studies to understand watershed dynamics and ultimately landscape evolution. An example of the insights provided by quantifying landslide volumes can be seen in the study by Larsen (2012), where he used landslide volumes in a sediment budget study to determine the influence of mass wasting on erosion rates. Landslides were the primary erosive force in each of the four watersheds studied in this investigation. One watershed, however, showed higher erosion rates. Closer evaluation of land use within the watersheds revealed that landsliding had been exacerbated by human development in the watershed with higher erosion rates. In Hovius et al (1997) downstream sediment discharge was compared to landslide volume estimates in the Western Southern Alps in New Zealand. Through this comparison, it was found that erosion in this region occurs primarily

through landsliding. These studies highlight some of the insights that can be gleaned from determining the total volume of material removed in a landscape due to landsliding.

3. STUDY AREA AND HURRICANE MARIA

Puerto Rico lies between the subducting North American and Caribbean plates (Masson and Scanlon, 1991)(Figure 2). The island is mountainous with its highest point of ~1330 m near the geographic center of the island. The basement complex of the island is composed of mostly folded and faulted volcanic and sedimentary rocks that are overlain by carbonate platform sequences to the north and south (Monroe, 1980). The northwest and north central portions of the island are marked by karst topography (Monroe, 1976). The topography of the island has an arc like shape with the southern limb of the island being steeper than the northern limb. Its location in the Caribbean region and mountainous topography leave the island prone to hazardous hurricane-triggered landslides, as made evident after Hurricane Maria.

3.1 Training Site and Añasco Watershed

Methods were developed in the Añasco river watershed (Figure 2) which is located on the west side of Puerto Rico and is approximately 415 km² in area. The lithologies within the Añasco watershed are mostly volcanoclastic rocks and basalts (Baiwec, 1998). However, there are some small exposures of granodiorite, diorite and quartz monzonite present (Baiwec, 1998). Land within the watershed is mostly undeveloped and used for rural purposes (Martinuzzi et al., 2007).

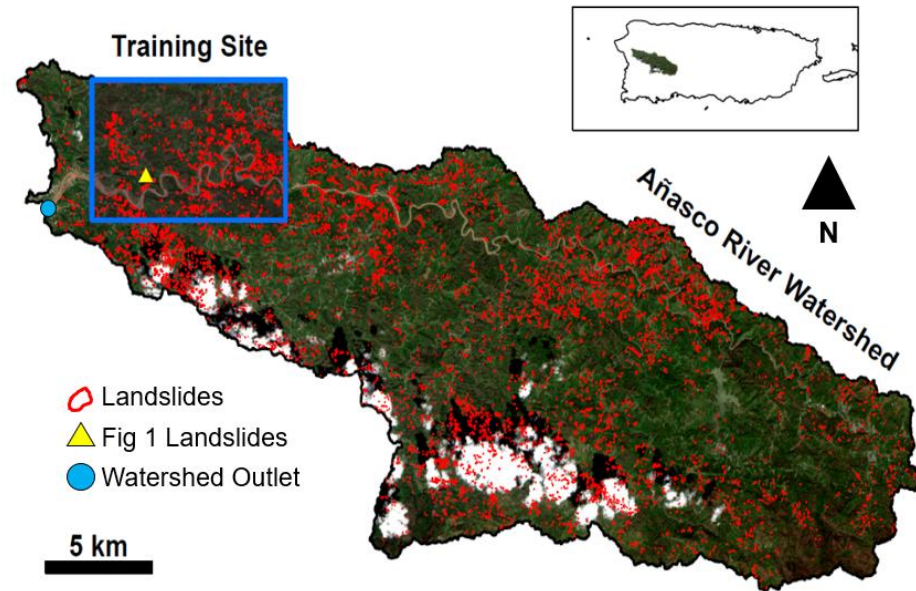


Figure 2: Location of the training site and the Añasco river watershed in Puerto Rico. Landslide polygons mapped by the SLIDES-PR group at the University of Puerto Rico, Mayagüez, shown in red. Locations of landslides in Figure 1 symbolized by a yellow triangle. Location of watershed outlet symbolized by blue circle.

3.2 Hurricane Maria

Hurricane Maria made landfall as a Category 4 hurricane in Puerto Rico on September 20th, 2017. The storm developed on September 12th off the west coast of Africa, intensified into a tropical storm on September 16th then, after being upgraded to a hurricane, followed a southeast to northwest path across the island of Puerto Rico at 155 mph while dropping up to 38 inches of rain on the island (Pasch et al., 2018). Rainfall estimates are mostly approximate, however, as winds destroyed the San Juan WSR-88D Doppler radar before the eye of the hurricane made landfall (Pasch et al., 2018). In the analysis by Ramos-Scharrón and Arima (2019) they estimated that 96% of the island

received more than 7 inches of rainfall and 10% of the island experienced more than 19 inches of rainfall.

The hurricane, which was one of the many devastating storms during the 2017 hurricane season, resulted in approximately \$90 billion in damages for the USA (Lim et al., 2018; Smith, 2018). In addition to wind, rain, and flooding, landslides triggered by the hurricane posed an additional hazard, causing damage to both structures and roads (Bessette-Kirton et al., 2019). More than 70,000 landslides were triggered on the island of Puerto Rico due to Hurricane Maria (Hughes et al., 2019). The hurricane is an example of an unprecedented landslide- triggering event that might impact the long-term landscape evolution of Puerto Rico.

4. METHODS

To semi-automatically detect landslides in the Añasco watershed using Sentinel-2 satellite imagery, a workflow was developed in ESRI ArcGIS 10.7 (Figure 3). Although ArcGIS is proprietary software, the routine incorporates common algorithms that can be applied elsewhere, such as in QGIS or Google Earth Engine. Inputs into the routine included pre- and post- event multi-spectral images, a 10-m DEM, NDVI maps, and slope maps.

A small training site was used to develop the workflow and determine a relatively successful method. Once a satisfactory workflow was determined, the method was applied to the entire watershed. The routine is as follows:

- 1) First, we remove flat areas and areas of no change in the image to be classified.
- 2) Then, we segment the image into objects.

- 3) After this, we create a small set of training data.
- 4) The training data are input into a machine learning classifier that identifies landslide candidates and other features in the image.
- 5) The landslide candidates are extracted from the classified scene, and the accuracy is assessed using carefully chosen evaluation metrics.

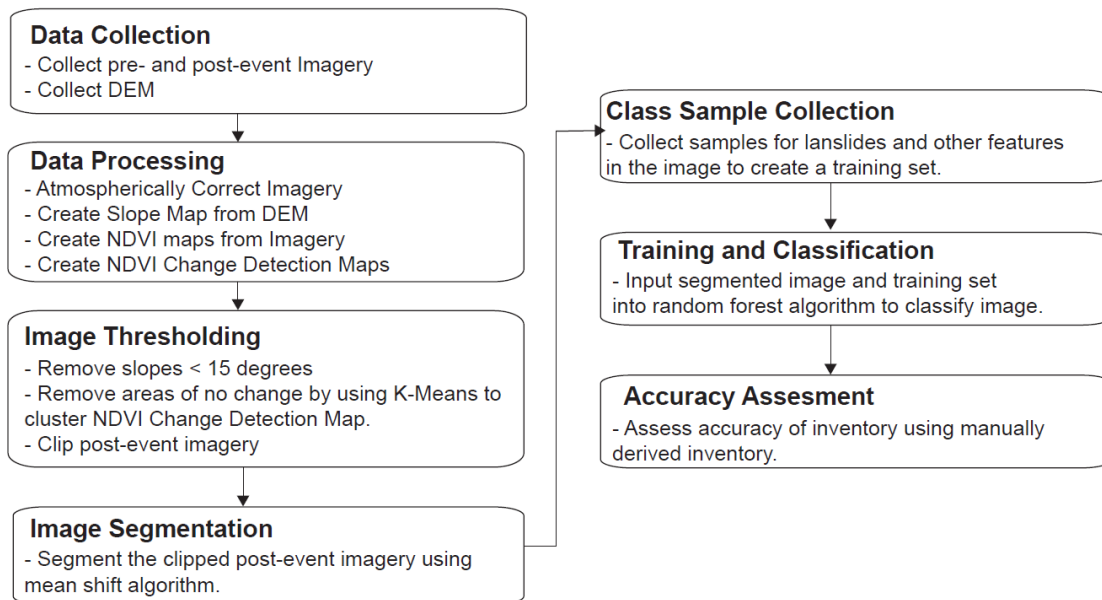


Figure 3: Workflow used in this study to detect landslides.

4.1 Data

The pre- and post-event imagery used in this study are from the European Space Agency's (ESA) Sentinel-2 multispectral instrument (MSI). The acquired images were taken on April 8, 2017 and November 11, 2017. These images were chosen on the basis of minimal cloud coverage. Ideally, images chosen would be closer to the event date but a cloud free image is higher priority for image classification. The MSI, which was launched in 2015, and is expected to be operational for 15 years, collects 13-band multispectral

images with spatial resolutions ranging from 10-60 m with a five-day revisit time (Drusch et al., 2012). The RGB and Near Infrared (NIR) bands, which we focus on in this study, are 10-m in resolution. Images used in this study were atmospherically corrected using the SEN2COR processor provided for use in the Sentinel Application Platform (SNAP) by ESA (Louis et al., 2016). Atmospherically corrected Sentinel-2 imagery was also downloaded from Google Earth Engine.

The Normalized Difference Vegetation Index (NDVI) maps used in this study were created using pre- and post-event Sentinel-2 imagery. The NDVI calculation relies on the red and NIR bands in the multispectral imagery (Deering and Haas, 1980). The NDVI values are derived for each of the pixels in the multispectral imagery according to the following formula:

$$(Eq. 1) \text{ NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}$$

Values for each pixel within the NDVI maps range from -1.0 to 1.0, where 1.0 corresponds to highly vegetated areas and -1.0 corresponds to areas of sparse or unhealthy vegetation. The resolution of the NDVI maps is maintained at 10-m.

Change detection images were calculated using the pre- and post-event NDVI maps. Pixel values within the NDVI maps are simply subtracted from one another to determine the amount of change that occurred after Hurricane Maria. Values closer to 0 in the change detection map suggest little or no change, where positive values indicate areas that changed from vegetated to non-vegetated.

The Digital Elevation Model (DEM) used in this study is a 10-m DEM from the USGS 3DEP collection (USGS, 2017). Slope maps were created from the DEM using the

slope tool in ArcMap 10.7. The slope tool fits a plane to a 3x3 neighborhood of pixels and determines the slope, in degrees, of this neighborhood using the average maximum technique (Burrough and McDonell, 1998).

4.2 Thresholding

The first step in the routine is to appropriately threshold the satellite images in order to minimize the likelihood of incorrectly identifying landslides in the final image. This simply means removing areas in the image that are least likely to have experienced landslides. To threshold the images the slope map was used to remove flat areas from the image by determining which areas consist of slopes less than 15 degrees. This value was chosen based on its effectiveness shown in other studies where it was useful in removing urban areas and lakes (Barlow et al., 2003). In removing the flat areas from the image, most urban areas and the main river channel and flood plain are also effectively removed in the study area. This is useful as segments from these areas could be mistakenly identified as landslides based on vegetation changes.

Then, K-Means clustering was used to divide the change detection image into three clusters that correspond to areas that have experienced little to no change, vegetation loss and vegetation gain. Previous studies have chosen thresholds related to NDVI manually, which makes routines less automatic and prone to human bias (Martha et al., 2010). It is difficult to cluster certain data, like NDVI and change detection maps, as their data distributions are usually unimodal, meaning there are no obvious groups within the data (Figure 4). Using K-Means to cluster our data removes the manual aspect of thresholding in relation to NDVI change detection data and gets around the issue of no groups being apparent in the data distributions. After specifying the desired cluster amount, the K-Means

algorithm begins by selecting random centroids within the data of interest (Hartigan and Wong, 1979). The Euclidian distance from each data point to these randomly chosen clusters is used to determine which cluster they fall into. After assigning clusters for the first iteration, the cluster centroids for the next iteration are then chosen as the mean of the determined clusters. This operation is repeated until the centroids become stationary and no longer change (Hartigan and Wong, 1979). After the three clusters have been chosen the cluster that corresponds to the ‘little or no change’ values is removed from the post-event image. K-means is chosen as the clustering algorithm in our work as it is easy to implement, widely used and allows for the pre-selection of clusters.

To summarize, regions that were determined to be flat and/or experienced little to no change were removed from the post-event image. Thresholding this way allows for the area in which we apply our final classification to be constrained so as to decrease the likelihood of incorrectly identifying landslides.

4.3 Segmentation

After determining which areas should not be classified (flat areas, areas of no change), and removing those areas from the post event image, we then break down the image into objects using a segmentation algorithm. For our study we implement the mean-shift segmentation algorithm using the Segment Mean Shift tool in ArcMap 10.7. The parameters for the mean-shift algorithm are spatial bandwidth (hs), spectral bandwidth (hr) and minimum segment size (M). The values of these parameters are chosen by the user. The spatial bandwidth is the window size used in the segmentation, and spectral bandwidth is the distance between segments in spectral domain. At the chosen window size, the algorithm begins at a random pixel in the image as the center. The algorithm then finds the mean of

these values in the spectral information within the chosen window and shifts the center of the window to this mean until the window is stationary, and the mean has converged. This iterative process creates an amalgamation of pixels that have similar spectral characteristics.

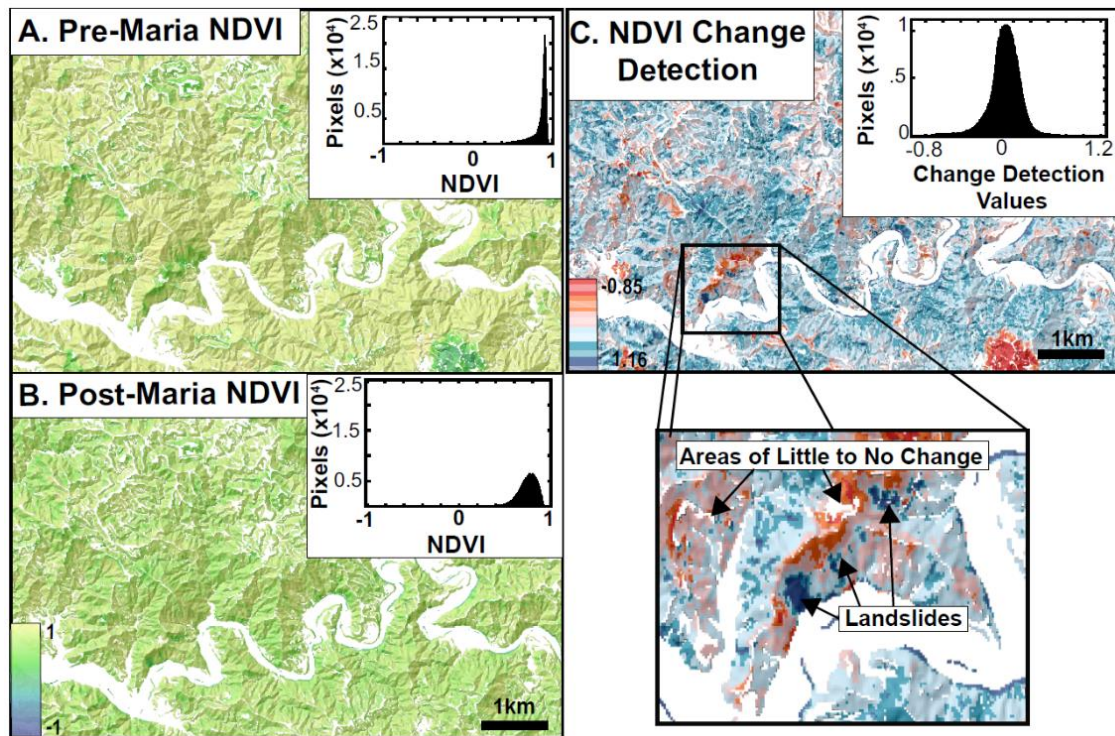


Figure 4: NDVI and change detection data for training site. (A) Pre-Maria NDVI map with distribution of NDVI values shown in top right corner. (B) Post-Maria NDVI map with distribution of NDVI values shown in top right corner. Color bar in B also applies to A. White areas in A and B have slopes < 15 degrees. (C) Change detection map with distribution of change detection values shown in top right corner.

The parameters, as previously mentioned, are chosen in our study with the goal of being able to delineate landslides as accurately as possible. To do so, we have to assign high importance to the *hs* and *hr* values and choose an appropriate minimum segment size. In ArcMap the *hr* and *hs* values are chosen on a scale from 1-20. Higher values for the *hs* parameter are chosen for spatially smoother outputs and are ideal in cases when features of interest are clustered together. Higher values for the *hr* parameter also create spectrally smoother outputs and are ideal when there are spectrally similar features that need to be

distinguished from one another. For the M parameter the value is simply the minimum amount of pixels per segment. For segmentation of our image we choose $hs = 20$, $hr = 20$, $M = 5$.

4.4 Training and Classification

After segmenting the image, a small training set needs to be created to subsequently classify the image. The training set is simply a small sample of features in the images that are classified. There are individual training sets for both the training site and the watershed. To create a training set, individual segments that correspond to certain features in the landscape are chosen. In the Añasco watershed we create a training set of $n = 30$ (30 individual features, comprised of several segments) for the following features: vegetation, landslides, clouds, shadows and tree-shed (areas where canopy was bare after Maria). After an appropriate training set is chosen, this is used to train a machine learning classifier. The classifier chosen, based on its resistance to overfitting and success in other studies, is the random forest classifier (Strumpf and Kerle, 2011; Breiman, 2001).

Our inputs for the random forest classifier within ArcMap are the segmented image, a DEM and training samples. The random forest classifier takes a set of training samples provided by the user and uses the information from the training samples to classify the image. The training samples are chosen by the user and are portions of the image that correspond to different classes that are to be mapped in the landscape. The random forest classifier works by taking the training samples and creating a subset from this training data to teach the classifier how to map the remaining portions of the image (landslides, vegetation, clouds, shadows, tree-shed). In our case, our training samples are the segments that correspond to the classes previously mentioned.

The variables associated with each of the segments used in the training sample dataset are color (RGB), compactness, rectangularity and elevation. The first three variables are associated with the imagery itself and the final variable is associated with topography. Once the subset data are created, the classifier uses this data to create multiple decision trees are based on randomly selected variables within the data set. Decision trees are just flow charts with differing queries at each level. Each decision tree decides which class each segment belongs to and once all of the trees have assigned a class to a certain segment, majority voting is used to determine what class the segment belongs to.

4.5 Accuracy Assessment

After the post event image has been classified and the landslide candidates delineated, the accuracy of the workflow is assessed by comparing the automatically mapped landslides to those mapped manually by the SLIDES-PR group (Hughes et al., 2019). Two metrics, known as recall and precision, are used to assess the accuracy of the workflow (Davis and Goadrich, 2006).

$$\text{(Eq. 2) Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{(Eq. 3) Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Where, when mapping landslides automatically, a ‘True Positive’ is a successfully mapped landslide, a ‘False Negative’ a landslide not mapped, and a ‘False Positive’ is a non-existent landslide mapped as one. These two metrics are useful in the case of landslide detection because they give us measures of how well our mapping is doing in regard to over and under detection of landslides. For recall, the measure identifies the proportion of

manually mapped landslides that were correctly identified by automatic mapping methods. For precision, the measure identifies the proportion of automatically mapped landslides that were correctly identified. We use these accuracy metrics to determine how well our method was at both recognizing the location of landslides as well as their extent, or area. Thus, we determine recall and precision for both ‘Recognition’ and ‘Extent’. Unfortunately, with the amount of landslides present in the Añasco watershed, it is difficult to manually determine these metrics. It is made especially challenging when the mapped landslides do not quite overlap with the manually mapped landslides (Figure 5). This is a common problem in computer vision studies and it is addressed by using the IOU metric (Intersection over Union) which is a simple operation to determine the degree of overlap between two shapes (Rahman & Wang, 2016).

$$(Eq. 4) IOU = \frac{\text{Intersection}}{\text{Union}}$$

$IOU > 0.20$ is designated a true positive in this study. After determining the number of landslides that are true positives, we can then determine those with IOU values < 0.20 as either being false negatives or false positives. After this, recall and precision values can be calculated for recognition and extent. Examples illustrating the accuracy assessment can be seen in Figure 5.

4.6 Frequency-Area and Frequency-Volume distributions

It is useful to analyze the area and volume characteristics of landslide inventories in terms of their frequency area-distribution (FAD) and frequency volume distribution (FVD) statistics. FAD and FVD statistics of landslide inventories allow for the comparison of large, landslide triggering events (Malamud et al., 2004). Comparisons of inventories in

this regard can lend insight into the magnitude of specific landsliding events as well as the underlying processes that may differentiate events (Tanyaş et al., 2019). Landslide area and frequency density exhibit a power-law relationship, described below.

$$(Eq. 5) p(X) = cX^{-\beta}$$

Where $p(X)$ is the frequency-density of the landslide area (X), c is a normalization constant and β is the power-law exponent (Tanyaş et al., 2018). Landslide inventories typically diverge from this power-law behavior below a threshold area. The exact cause of the divergence is still unknown and a controversial matter. The value of importance for FADs is β , the power law exponent, because it allows for comparison and characterizaion of landslide inventory size distributions. Another value of importance, is the Landslide-Event Magnitude (mLS), which corresponds to the severity of the landslide event (Tanyaş et al., 2018). The equation to determine mLS at the threshold area is shown below. In this equation c' is a constant that is determined by dividing the normalization constant previously mentioned, by the total number of landslides.

$$(Eq. 6) mLS = \log_{10} \frac{p(X)}{c'X^{-\beta}}$$

In this study, we use the methods and Matlab code provided by Clauset et al (2009) to determine the minimum threshold area (X_{min}) for which the frequency distributions exhibit power law behavior (<http://tuvalu.santafe.edu/~aaronc/powerlaws/>). X_{min} is chosen by determining which value of X_{min} results in the least amount of distance between the power-law fit and the data. Once an appropriate X_{min} value is chosen, the corresponding fit is applied to the data and the power-law exponent is determined. The informtaion is then

used to plot the FAD and determine the mLs values using the methods of Tanyaş et al (2018) and Matlab code provided by the USGS (<https://github.com/usgs/landslides-mLS>).

To determine individual landslide volumes, we use the volume-area scaling relationship from Puerto Rico as determined by Larsen and Torres-Sánchez (1998). The relationship is shown below, where A_L is landslide area and V_L is landslide volume. The FVD is determined using the same methods previously mentioned.

$$(Eq. 7) V_L = 1.826 * A_L^{0.898}$$

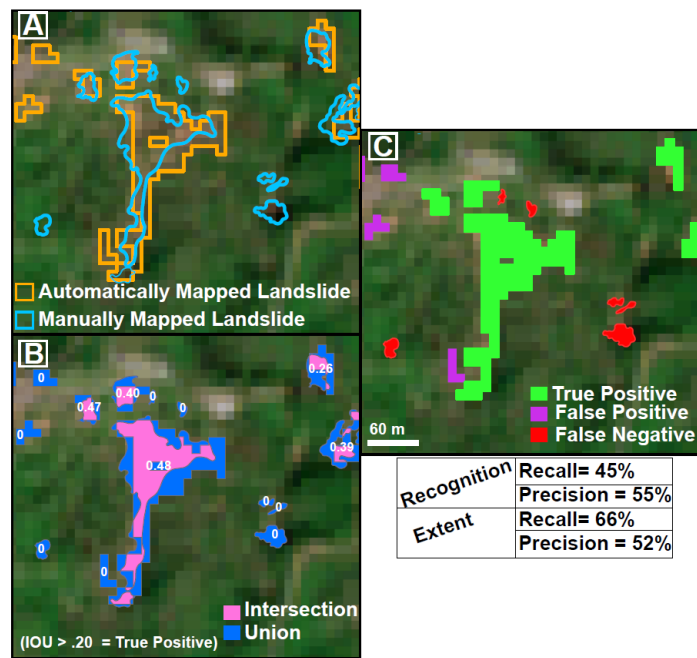


Figure 5: Example illustrating accuracy assessment metrics. (A) The landslides detected automatically are compared to those manually mapped by the SLIDES-PR group. (B) The Intersection over Union (IOU) is found for each landslide by merging the auto and manual landslides (union) into one shape, finding the intersection of the two landslides and then dividing the intersection area by union area. Those automatically mapped landslides with an IOU > 0.20 are considered true positives. All else are either false positives or false negatives. (C) After determining the count of true positives, false positives, and false negatives, recall and precision are calculated for both the recognition and extent of landslides. For extent, the individual sums of true positives, false positives and false negatives are used to determine recall and precision.

5. RESULTS

At the training site scale our method resulted in recall and precision values of 18% and 17% respectively for the recognition of landslides and 20% and 9% for detecting the extent of landslides. At the watershed scale, our method resulted in recall and precision values of 12% and 7% for the recognition of landslides and 10% and 2% for detecting the extent of landslides. There was a notable decrease in accuracies when mapping the larger scale watershed. These accuracies are summarized in Table 1. Examples of mapped landslides and their corresponding IOU values can be seen in Figure 6.

Table 1. Accuracies at training site and watershed.

	Training Site		Watershed	
	Recall	Precision	Recall	Precision
Recognition	18%	17%	12%	7%
Extent	20%	9%	10%	2%

5.1 Spatial Analysis of Error

The manually mapped inventory, within the training site, contains 865 landslide polygons while the automatic mapping resulted in an inventory with 900 landslide

polygons. Within the watershed, the manually mapped inventory contains 7,332 landslides while the automatic mapping resulted in an inventory with 12,474 sites identified as landslides. Of the landslides automatically mapped in the training site, 160 were true positives and 740 were false positives. There were 705 false negatives, or missed landslides. This information is summarized in Table 2 for both the training site and watershed.

Density maps for each of these errors, at both the training site and watershed scale, were created using a kernel density estimation to create an estimated magnitude per unit area map for each result. While not reflective of the true amount of error points at each location the maps are an easier way to visually assess their spatial distribution. The large distribution of false positives, as seen in Figures 7 and 8, seems to be caused by the classifier classifying portions of the river closest to the banks and at the edge of urban areas as landslides. The false negatives seem to be the result of the automatic methods failing to capture the smaller landslides (Table 3, Figure 9 and 10) and cloud cover masking out areas in the image (Figure 9).

In order to visually assess the difference in the overall outputs between the automatic and manual inventories, landslide density maps, using a kernel density estimation, were created at the watershed and training site scales (Figure 11). At the training site scale, the spatial distribution of automatically mapped landslide densities is roughly consistent with the spatial distribution of landslide densities seen in the manually mapped inventory. There are two localized areas of high landslide density on the east and west sides of the training site. In addition to this, the automatically mapped landslide density map excludes a high-density portion in the east because these landslides are small

and thus were undetected by automatic methods. At the watershed scale, however, there is little agreement. There is overestimation of landslides in the North West and central portions of the watershed. The southeast portions generally agree, however, and the landslide density is low in this region for both the automatically and manually mapped landslide inventories.

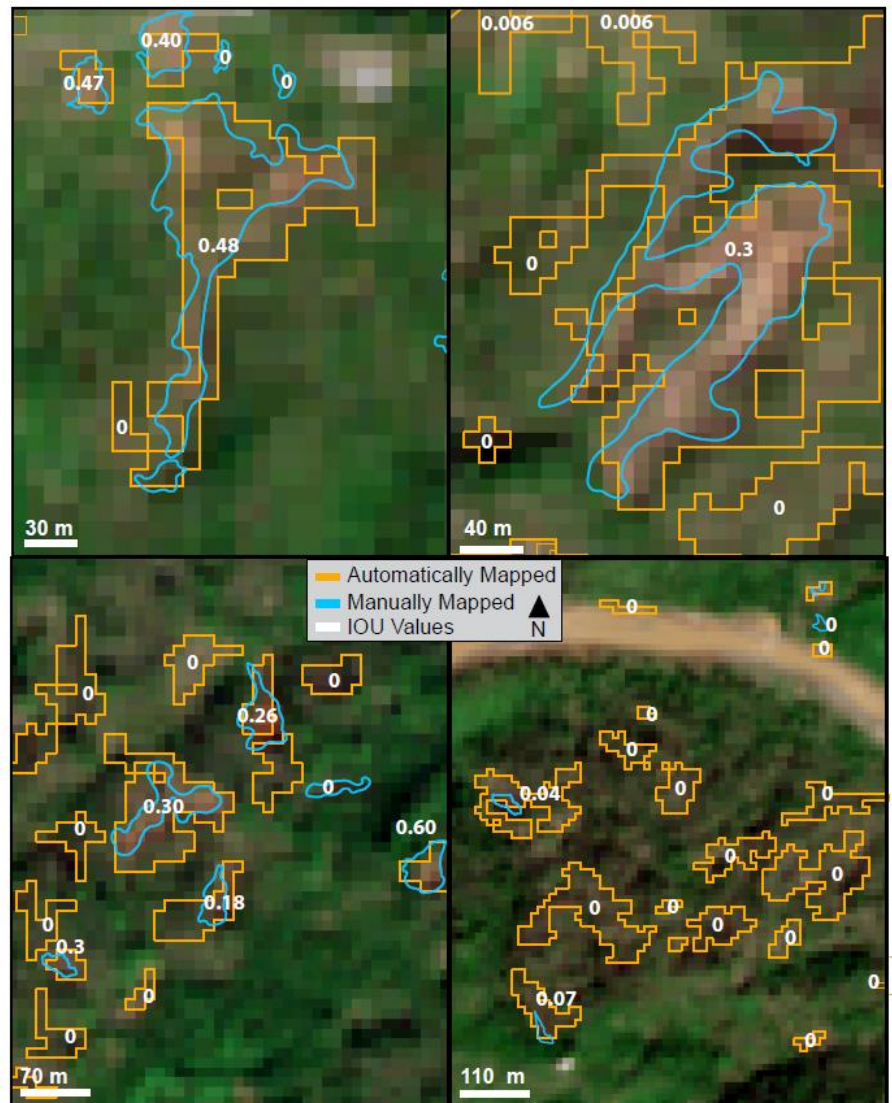


Figure 6: Example of IOU values.

Table 2. Details of accuracy assessment for watershed and training site.

	Training Site	Watershed
No. Automatic Landslides	900	12474
No. Manual Landslides	865	7332
True Positives	160	894
False Negatives	705	6438
False Positives	740	11580

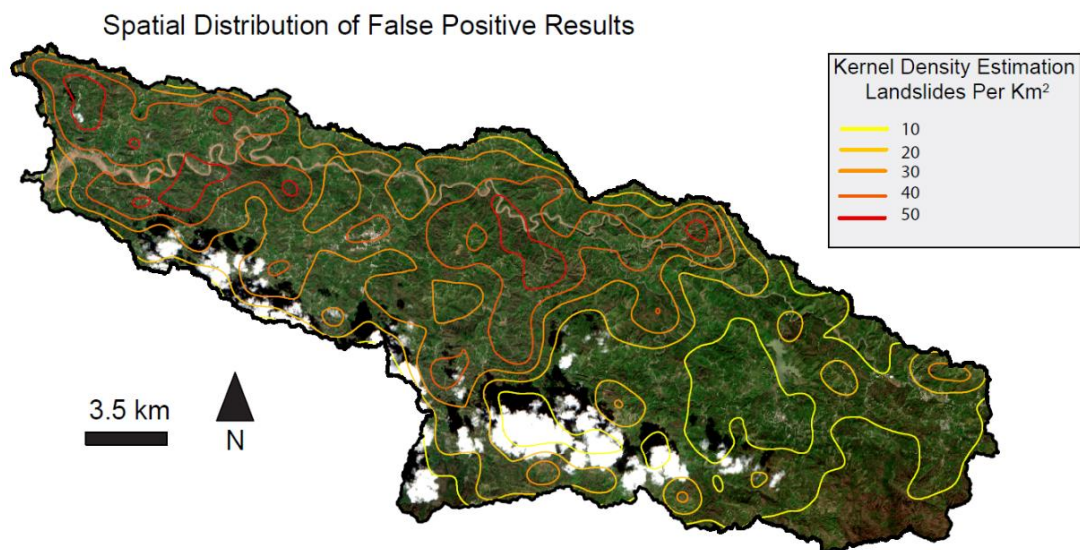


Figure 7: *Spatial distribution of false positive results in the watershed.*

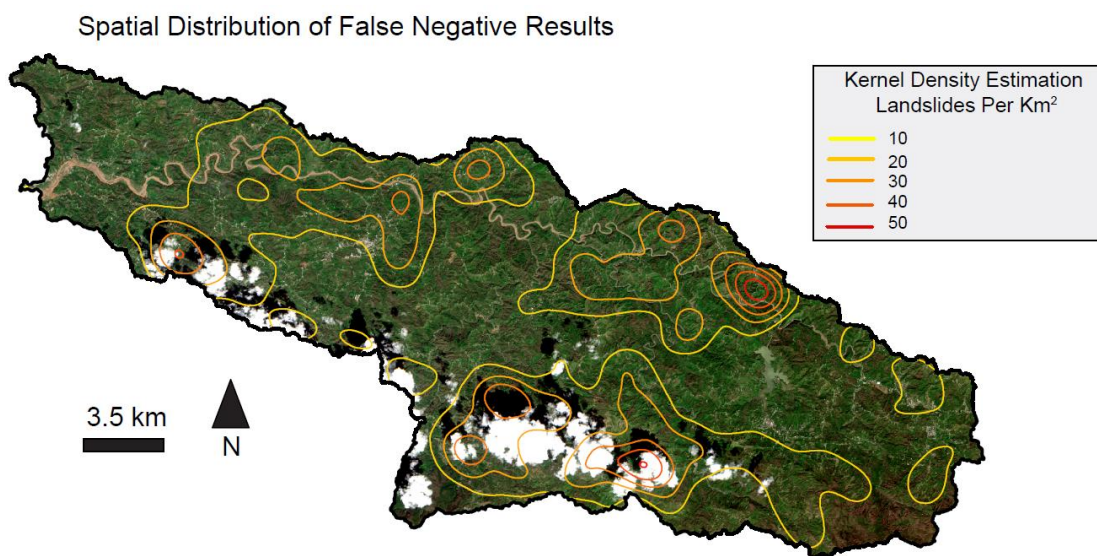


Figure 8: *Spatial distribution of false negative results in the watershed.*

Spatial Distribution of False Positive Results



Figure 9:: *Spatial distribution of false positive results in the training site.*

Spatial Distribution of False Negative Results



Figure 10: *Spatial distribution of false negative results in the training site.*

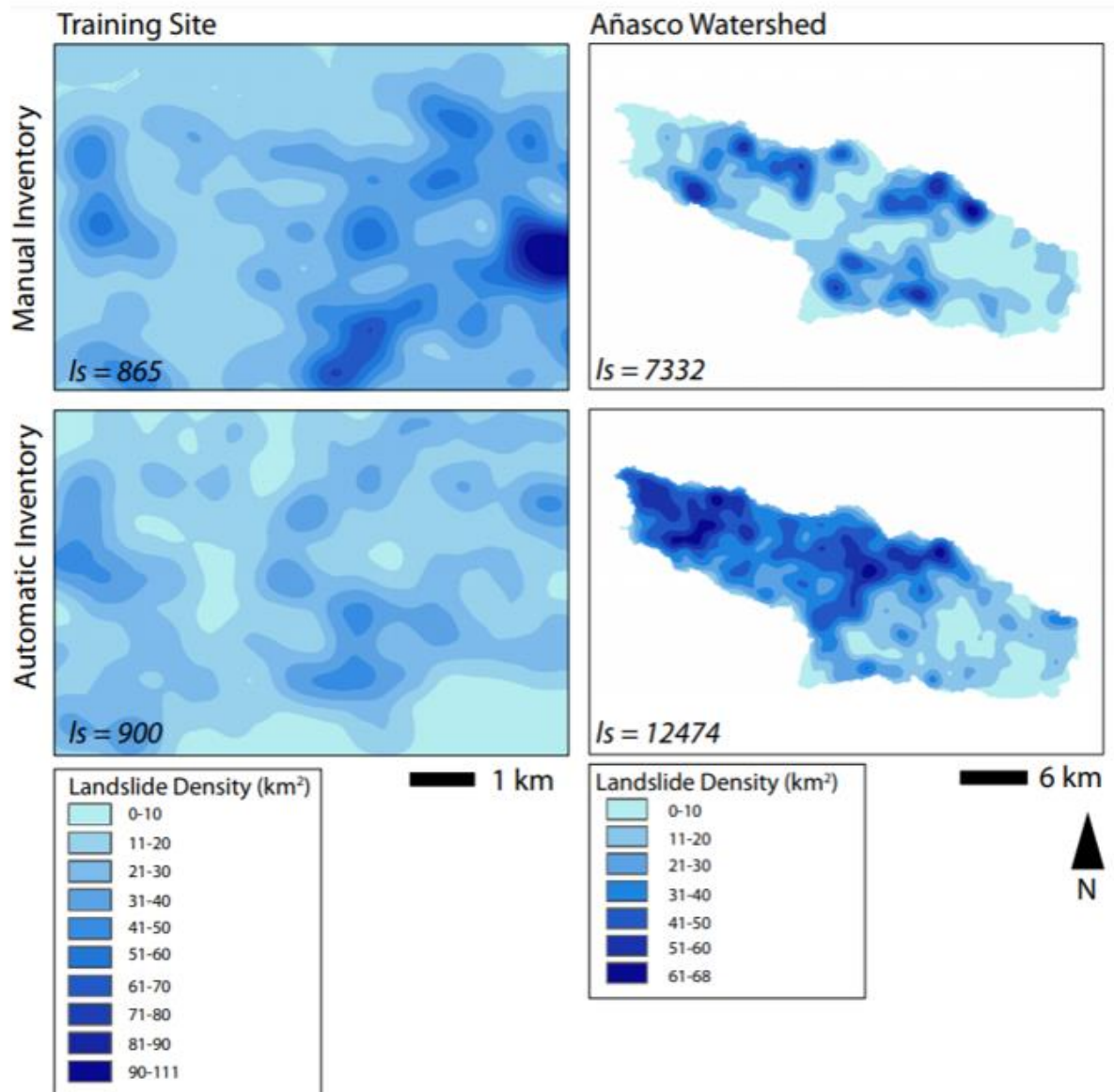


Figure 11: Landslide density maps at the training site and watershed scales for the automatically and manually mapped landslide inventories.

5.2 Landslide Area and Volume Estimations

The area and estimated volumes for both inventories are summarized in Table 4 (Larsen and Torres-Sanchez, 1998). At the watershed scale the automatic inventory overestimates areas by $1.78 \times 10^7 \text{ m}^2$, a factor of four. The overestimation in regard to volume is similar since the volume estimates are derived from the scaling relationship, at $1.39 \times 10^7 \text{ m}^3$, by a factor of more than three. The total area of automatically mapped landslides in the training site was overestimated by $8.70 \times 10^5 \text{ m}^2$, resulting in a total area that is more than double the true total area. The total volume of automatically mapped landslides overestimated by $6.70 \times 10^3 \text{ m}^3$, similarly resulting in a total volume that is more than double the true total volume.

Table 3. Landslide area statistics for true positives, false negatives.

	True Positives		False Negatives	
	Watershed	Training Site	Watershed	Training Site
$\bar{x} \text{ (m}^2\text{)}$	14436	2490	615	772
$\sigma \text{ (m}^2\text{)}$	1049	17537	475	569

Landslide inventories generally follow a power law above a threshold area. Plotting the FAD and FVD for both inventories reveals that the divergence from power-law behavior at the threshold point does occur in the manually mapped inventory but does not occur in the automatically derived inventory (Figures 12 and 13). This is most likely due to the fact that the automatic inventory, at both the training site and watershed scale, is ineffective at capturing small scale landslides due to the low resolution of the imagery used in the automatic inventory. In addition to this, our automatic method seems to be overestimating the area and volume of landslides, most likely due to the large amount of false positives as well as problems with landslide amalgamation (Marc and Hovius, 2015). Where the human interpreter can delineate individual landslides the classifier lumps them into one large landslide ultimately increasing the estimated areas and volumes for the entire inventory.

Table 4. Estimated total area and volume for manual and automatic inventories.

Training Site			Watershed	
	Manual	Automatic	Manual	Automatic
Area (m³)	7.24*10 ⁵	1.59*10 ⁶	4.23*10 ⁶	2.21*10 ⁷
Volume (m³)	6.28*10 ⁵	1.30*10 ⁶	3.86*10 ⁶	1.78*10 ⁷

Despite the large discrepancies regarding landslide area and volume, the mLS values, which corresponds to landslide event magnitude, for the automatic inventories and manual inventories seem to be in close agreement. The mLS values for the automatic and manual inventories are more similar in the training site than in the watershed, however. Summary statistics and the FAD plot can be seen in Table 5 and Figure 12 respectively. Results for the frequency volume distribution (FVD) can be seen in Figure 13. Threshold points chosen to create power-law fits are shown in Table A2.

Table 5. Frequency-area distribution information.

	Training Site		Watershed	
	β	mLS	β	mLS
Auto	-2.37	2.72	-3.40	3.45
Manual	-2.75	2.41	-3.29	2.01

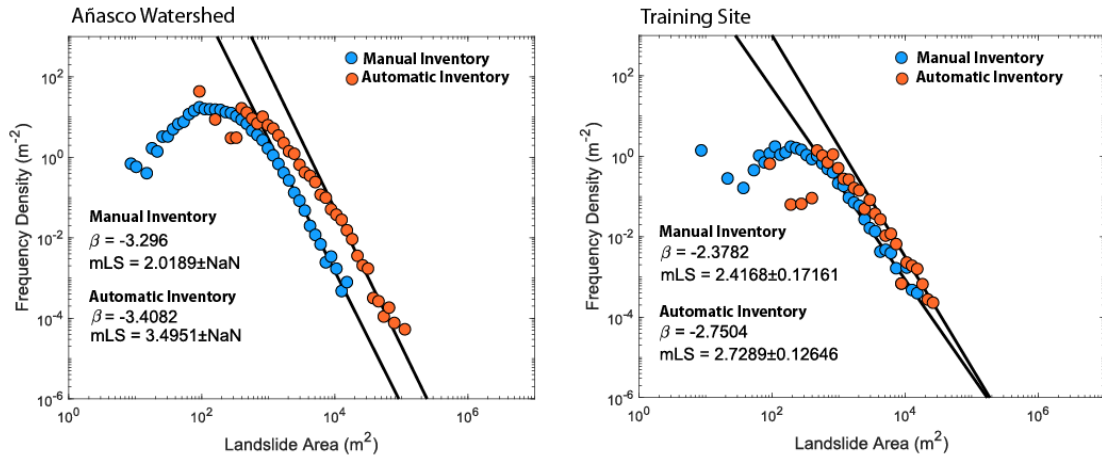


Figure 12: The frequency-area distributions at the watershed and training site scale for the automatic and manually mapped inventories.

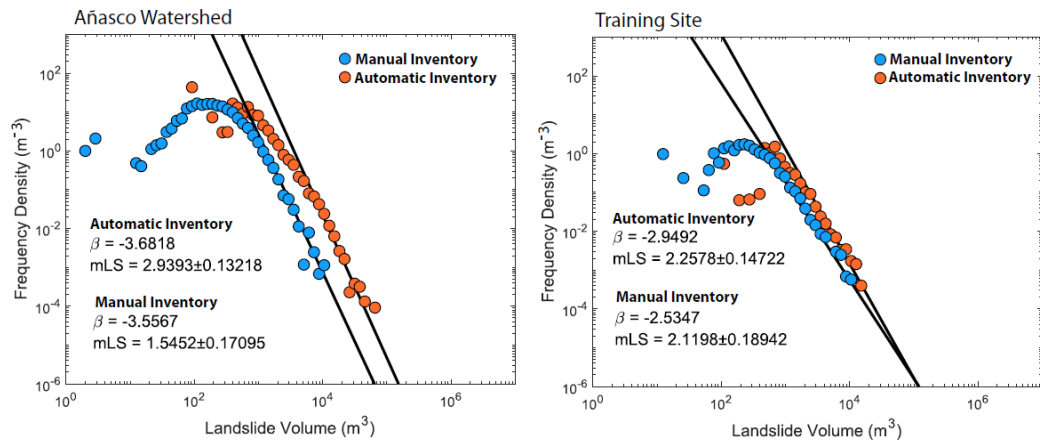


Figure 13: The frequency-volume distributions at the watershed and training site scale for the automatic and manually mapped inventories.

6. DISCUSSION

Our results show that using semi-automatic, object-based methods and Sentinel-2 data can result in large-scale landslide inventory deficiencies. The method failed to capture small landslides and overestimated a large number of landslides as well. Because we set the lower limit of our segmentation to five pixels, the limit of the smallest mapped landslides is 500 m². In addition to this, the results show that a large decrease in accuracy occurs when attempting to map landslides at larger scales revealing an inherent scaling problem that needs to be addressed in future landslide detection studies. Additionally, a limitation in this study is that the workflow can only be performed in vegetated landscapes. Despite the deficiencies, the automatic method produced landslide inventories that were similar to the manual inventory in regard to their frequency area and volume statistics. The information that can be gleaned from the FAD and FVD statistics is useful in understanding the underlying geomorphic processes for various landslide triggering events. Thus, it is promising that the automatic methods can, even when inaccurate in finer detail, produce information that may help in understanding the role that landslides play in landscape evolution.

Landslide inventories can be used to create landslide susceptibility maps and gain insight into underlying geomorphic processes as they relate to specific events and environments (Galli et al., 2006, Tanyaş et al., 2018). Information regarding the shape, size, location and amount of landslides associated with one large triggering event can help answer a myriad of scientific questions as well as aid in planning and mitigation efforts (Guzzetti et al., 2012). Because of the increasing spatiotemporal resolution of satellite imagery, efforts to develop automatic methods to quickly and accurately create landslide

inventories have increased. However, the methods proposed have not proven to be widely applicable. The methods have also been developed in study areas that are not representative of the larger scale landscapes that need to be mapped. Because of this, more robust methods need to be developed before methods can be confidently used in creating useful landslide inventories. The low accuracies achieved here highlight biases in the literature. If current methods are not further developed and applied to real scenarios, they are likely to result in erroneous interpretations and estimates. In the next sections the erroneous results in this study and their causes are explained. We also explore the implications of the estimates that would be derived from the automatic inventory. Then, we discuss future directions to help decrease the errors specific to this study as well as others.

6.1 Consequences of Inaccuracy: Interpretations from Erroneous Estimates

Parker et al (2011) used semi-automatic methods to map landslides after the Wenchuan earthquake in China. They then used their inventory to claim that landslide erosion was outpacing orogenic growth, as the estimated volume of landslides was greater than the amount of orogenic rock uplift. They estimated the total landslide volume to be 5-15 km³ and the total rock uplift at 2.6 ± 1.2 km³. In a different study, Li et al (2014) created a landslide inventory of the same event using supervised and manual classification efforts and determined the volume to be 2.8 km³, similar to the estimated volume of rock uplift (Figure 13). They attributed the overestimation of landslide volumes in Parker et al (2011) to landslide amalgamation due to automatic mapping. This example highlights the infancy of automatic detection methods and the need for further development before methods can be confidently applied, especially as our overestimation of volumes is similar to that of Parker et al (2011).

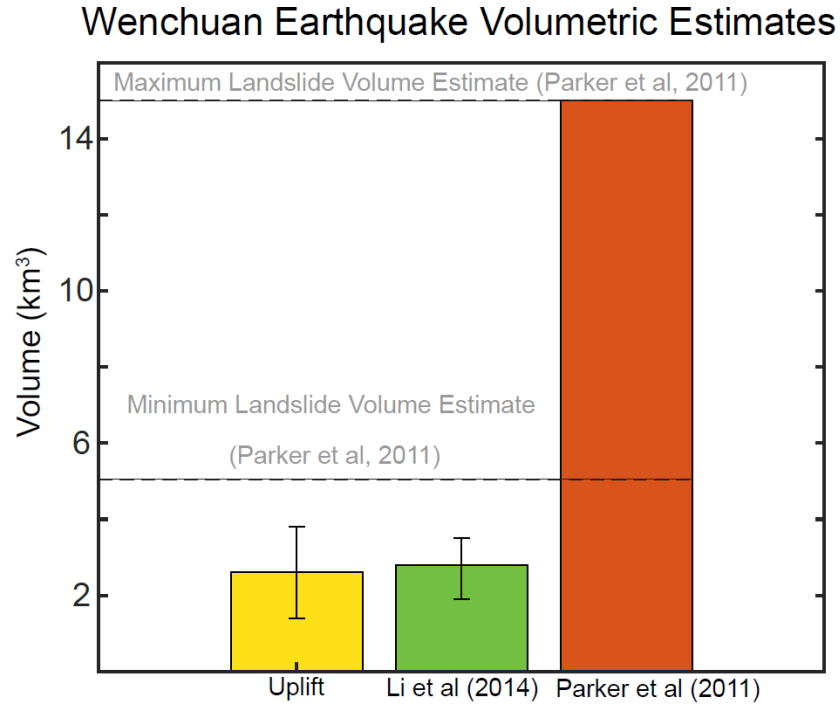


Figure 14: Volumetric estimations associated with the Wenchuan earthquake. Parker et al (2011) landslides were automatically mapped whereas Li et al (2014) landslides were mapped using manual and supervised classification methods.

6.2 Errors in this study & their causes

Similar to Parker et al (2011), large errors in our inventory and resultant volume estimates are due to overestimating the number of landslides as well as landslide areas. The overestimation can be attributed to the large amount of landslides in the study area that are difficult to distinguish from general tree-destruction and blow down that occurred after Hurricane Maria. Our errors in regard to recall are also low because we failed to capture small landslides. This may be due to the 10-m resolution of the Sentinel-2 data and lower bound on segmentation which only identifies landslides greater than 500 m².

While our methods are similar to those proposed in the literature, we attempt to map a high-density landslide event within a large study area. Thus, it is likely our large errors, despite using common object-based methods, are also due to applying these

methods to a scenario that triggered a large amount of landslides. Methods thus far in the literature have not been developed in similar scenarios.

In performing this case study, several gaps were revealed in regard to method developments in the field of automatic landslide detection. First, most studies focus on developing methods in small study areas and attempt to map a small number of landslides. Our study reveals that there seems to be an inherent scaling problem that has not been acknowledged in the context of landslide detection studies. In our study and others, a decrease in accuracy has been shown when applying methods at larger scales. Similarly, in our study, low accuracies also seem to be attributed to areas where landslide density is higher. Secondly, current methods have primarily used proprietary algorithms to develop methods, despite the existing myriad of thresholding and segmentation algorithms. Thus, a vast amount of potentially useful and widely accessible methods have gone unexplored. Lastly, there are discrepancies between accuracy metrics across studies, making it difficult to determine which methods are truly versatile and successful.

6.3 Scaling Problems: Study Area Size and Landslide Densities

The scale problem present in this study is not unique and has been observed in a myriad of contexts. The problem is officially known as the modifiable areal unit problem (MAUP). MAUP arises when choosing different scales at which to analyze data (Hay et al., 2003). Results for various analyses tend to vary based on the scale chosen to observe the phenomenon and the way the data is aggregated. Object-based methods are sometimes suggested as a solution to MAUP as objects, rather than pixels, may be less sensitive to MAUP (Hay and Castilla, 2006). However, in our study, the MAUP still seems to arise

when performing our classification at a larger scale, which is not a problem with aggregation but rather, scale chosen to apply the method. The impact of scale has been observed in similar landslide detection studies as well but is never explicitly addressed. For example, in Li et al (2016) their larger study area generally underperforms compared to the smaller subareas. To make matters worse, landslide detection studies have focused on developing methods at scales not representative of the larger landslide impacted regions. This suggests that the decrease in accuracy attributed to increasing the study area size may occur when implementing all methods proposed in the literature.

Not only have methods been developed in small areas, they have also been developed in areas that experience a small number of landslides that may not be representative of the larger scale landslide event. In Figure 15 study areas and number of landslides are compared for a variety of studies. This figure illustrates how most studies to date have developed methods in small training areas and/or in areas that experienced a small number of landslides. Our study reveals this bias and suggests that more case studies are needed that apply methods to large landslide events. Additionally, detailed studies should be undertaken determining if there is an optimal study area scale to apply methods to and determining if automatic methods could benefit by applying methods in a piecewise approach.

6.4 Segmentation and Thresholding

Segmentation and thresholding are fundamental steps in object-based image classification as they cluster the target image(s) into more easily discernible objects. Thresholding operations simply divide images into binary classes that can be used to filter out appropriate areas within the image. Segmentation operations seek to delineate

homogeneous clusters within an image. Since thresholding operations seek to segment images, the following section will use the term segmentation to refer to both operations, for the sake of simplicity.

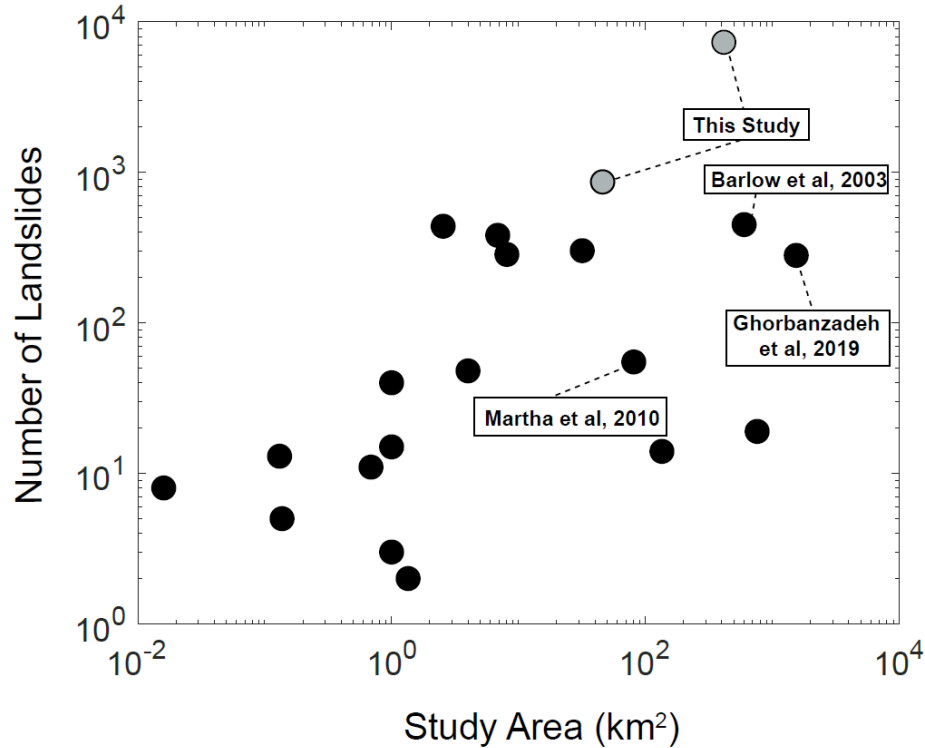


Figure 15: Comparison of number of landslides mapped within study areas across studies, shown on a log-log plot. Watershed and training site data points shown in plot.

Segmentation operations are used in a myriad of image processing studies and for a wide variety of applications. Unfortunately, within landslide detection studies, it is common practice to use software that segments images using a proprietary algorithm. This trend seems to have begun in 2000, when the eCognition software was developed for remote sensing applications. Before this time, there were not many published papers on segmentation techniques performed on remote sensing images (Dey et al., 2010). Since

then, studies have been published describing optimal metrics and workflows established using the propriety segmentation algorithm in eCognition. These methods and proposed optimal parameters, which are available only in the software, are inaccessible to those without the eCognition software (Wulder et al., 2008; Hay et al., 2003). Of the studies compiled in Table A1 in the appendix, six of 13 use the proprietary segmentation algorithm leaving a plethora of more widely accessible methods unexplored. Thus, there is potential for accessible segmentation methods to be further explored in the context of automatic landslide detection.

For example, we use the mean-shift segmentation algorithm in our study. No other landslide detection studies have used this algorithm, so there is no clear analysis on the impact this segmentation algorithm and chosen parameters may have on classification output. Future studies could focus on exploring modifications that could be made to the mean-shift segmentation algorithm in the context of landslide detection. Additionally, other segmentation algorithms, like those summarized in Dey et al (2010), could also be carefully explored to determine their impact on accuracies. The most useful pursuit would be to compare the results of the classification outputs (in the context of landslide detection) using a variety of segmentation algorithms and closely analyzing and quantifying their performance to guide further studies. Additionally, different inputs into the machine learning classifier could be explored for their effectiveness as well (such as slope, aspect or other topographic signatures) (Lyons et al., 2014). Results might suggest that one segmentation algorithm or certain topographic data may not be suited for all landscapes but instead certain landscapes. Only a detailed analysis would reveal this and allow future developers and mappers to make informed decisions on which segmentation algorithms to

implement. Evaluation metrics, which are necessary for this kind of analysis, are discussed in the following section.

6.5 Accuracy Assessments

To evaluate the effectiveness of proposed methods, standard accuracy assessments should be performed across all studies to make valid comparisons and informed decisions. Unfortunately, a careful review of accuracy assessments across automatic landslide detection studies reveals that metrics differ widely by study. Accuracy metrics used and reported accuracies across studies are shown in Table A1, located in the appendix. Below, these findings are briefly summarized, followed by a description of accuracy metrics proposed in this study.

Across 13 automatic landslide detection studies, seven unique evaluations were found to have been used. Metrics include the Users and Producers Accuracy, Errors of Commission and Omission, Completeness, Correctness and Quality, Recall and Precision as well as simple area overlap/intersection evaluations. With the large amount of variation in the metrics, the reported accuracies are nearly impossible to compare against one another. More consistency in how performance is evaluated in future studies will facilitate effective comparison and determination of which methods are successful. Below, we discuss why we advocate for more broad use of the metrics we applied in this study.

For landslide inventories, it is important that the location and the area of the landslides are mapped accurately. Thus, it is important to determine the effectiveness of methods at both recognizing landslides and mapping them in their proper locations, as well as being able to accurately capture the area of the landslide. So, any accuracy metrics

should measure the success of the inventory in regard to these two factors. For these two factors, it is also important to determine the amount that landslides are overestimated and underestimated. When mapping a large amount of landslides, it is nearly impossible to manually count the number of true positives, false positives and false negatives manually. Effectively determining these counts requires an automated method. In our study, we use the IOU metric to determine, on an individual basis, landslides with an appropriate agreement with the reference inventory, to determine true positives. Then, once this is determined, we determine recall and precision for the inventory in regard to recognition and extent. This method is proposed as it determines the effectiveness in recognizing landslides as well as their extent. It also evaluates true positives for a large inventory on an individual basis thus providing a more robust threshold for determining true positives. The routine was written as a python script that can be used in ArcMap. The code and GitHub link to the python script is available in the Appendix (Figure A3).

7. CONCLUSIONS

In regard to recognition of landslides, our method achieved recall and precision values of 18% and 17% in the training site and 12% and 7% in the watershed. In regard to the extent of landslides, our method achieved recall and precision values of 20% and 9% in the training site and 10% and 2% in the watershed. Our work reveals that currently there will be a deficiency in landslide inventories created using the Sentinel-2 data if methods are not further developed and appropriately evaluated. One caveat is that accuracies may be higher in regions where the landslide triggering mechanism does not result in tree-shed (i.e. earthquake triggered landslides). In this case there may be less features that can be

mistakenly identified as landslides. Additionally, the methods proposed here would only perform well in highly vegetated landscapes.

The Sentinel-2 imagery was successful at mapping some landslides but was not effective at capturing smaller landslides due to the spatial resolution of the data. The large amount of false positives detected were attributed to over detecting along the edges of the river channels and urban areas as well as in areas that experienced significant tree-shed after Hurricane Maria. Future work could thus focus on more accurately eliminating areas along the river banks so as to reduce false positives. In this case, care must be taken so as not to exclude those landslides that occur along the banks. Focused work on delineating landslides in noisy Sentinel-2 images could also be done to reduce the false positives detected in areas impacted by hurricanes. However, this may not be an issue in landscapes unlike Puerto Rico.

Additionally, there is an overall decrease in accuracy when applying methods at a larger scale and when attempting to map a large number of landslides. This decrease in accuracy is shown in this study and previous work (Li et al., 2016). Since previous studies have focused on developing methods in small regions this highlights the need to shift to developing methods in study areas that are more representative of the landslide events. Current methods, when applied to map landslides at a more realistic scale, may result in drastically lower accuracies than previously reported in the literature.

In previous work, there is no standard evaluation method used, making it difficult to compare the effectiveness of proposed methods across studies. We propose using the recall and precision evaluation metrics to determine the accuracy of methods in regard to identifying individual landslides as well as their areas. The metrics proposed can be used

in studies where large landslide inventories need to be assessed, as our method for determining true positives (on an individual basis) is also automatically derived using the IOU metric.

Recommendations for future work involves further exploration of segmentation and thresholding algorithms in a more methodological approach. This could better inform future users of the best practices for automatic landslide detection in differing environments. These studies should be carried out in study areas that are more representative, in regard to scale and landslide density, of actual landslide events and use open source segmentation algorithms. Evaluation metrics should also become standardized as well. This should all be done with the goal of shifting method development in a fashion that will better inform future mappers. In doing so, the community can then begin to see the benefits of being able to automatically create landslide inventories. Essentially, once geoscientists have the ability to make more informed decisions on which methods to apply in their regions of interest, the methods can become more widely used.

APPENDIX

Figure A1: Google Earth Engine code used to download Sentinel-2 imagery to create NDVI and change detection maps. <https://github.com/sabrinanicc/GEE-NDVI-Change-Detection>

```

1. //Create Desired Extent by Drawing Geometry
2. //Import Post-Event Imagery and Filter by Cloud Cover if Desired
3. var image = ee.ImageCollection('COPERNICUS/S2')
4.   //.filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 10))
5.   .filterDate('2017-11-01', '2017-11-05')
6.   .filterBounds(geometry)
7. print(image)
8. var medianpixels = image.median()
9. var medianpixelsclipped = medianpixels.clip(geometry).divide(10000)
10.   Map.addLayer(medianpixelsclipped, {bands: ['B4', 'B3', 'B2'], max:0.3}, 'Sent-2 After Maria' )
11.   Map.centerObject(medianpixelsclipped, 9);
12.
13.   //Create Post-NDVI Map
14.   var ndvi = medianpixelsclipped.normalizedDifference(['B4', 'B8']);
15.   Map.addLayer(ndvi, {min: -1, max: 1, palette: ['green', 'white']}, 'NDVI After Maria');
16.
17.   // Import Pre-Event Imagery
18.   var image3 = ee.ImageCollection('COPERNICUS/S2')
19.     //.filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 10))
20.     .filterDate('2017-04-05', '2017-04-09')
21.     .filterBounds(geometry)
22.   print(image3)
23.   var medianpixels3 = image3.median()
24.   var medianpixelsclipped3 = medianpixels3.clip(geometry).divide(10000)
25.   Map.addLayer(medianpixelsclipped3, {bands: ['B4', 'B3', 'B2'], max:0.3}, 'Sent-2 Before Maria' )
26.   Map.centerObject(medianpixelsclipped3, 9);
27.
28.   //Create Pre-NDVI Map
29.   var ndvi3 = medianpixelsclipped3.normalizedDifference(['B4', 'B8']);
30.   Map.addLayer(ndvi3, {min: -1, max: 1, palette: ['green', 'white']}, 'NDVI Before Maria');
31.
32.   // Create NDVI Difference Map

```

```

33.     var diff = ndvi3.subtract(ndvi);
34.     Map.addLayer(diff,
35.         {min: -
2, max: 2, palette: ['white', 'blue']},
36.         'NDVI Difference');
37.
38.     //Uncomment this to export the image, specifying scale and
region.Region can be imported geometry or watershed.
39.     //Export.image.toDrive({
40.         //image: diff,
41.         // description: 'Difference Map',
42.         // scale: 10,
43.         //region: geometry
44.         //});

```

Figure A2: Thresholding and Segmentation Code.

https://github.com/sabrinanicc/ls_thresholdandsegmentation

```

1. import arcpy
2. from arcpy import env
3. from arcpy.sa import *
4.
5.
6.
7. #Allow output file to overwrite any existing file of the same nam
e
8. arcpy.env.overwriteOutput = True
9. arcpy.CheckOutExtension("Spatial")
10. #Inputs
11. arcpy.env.workspace = arcpy.GetParameter(0)
12. NDVICD= arcpy.GetParameter(1)
13. DEM=arcpy.GetParameter(2)
14. Post_Event= arcpy.GetParameter(3)
15. classes = 3
16.
17. #Outputs
18. OutIso = 'Iso.tif'
19. Clip_CD='Clip_CD.tif'
20. inSQLClause1 = "Value > 15"
21. inSQLClause2 = "Value <> 2"
22. IsoEx = "IsoEx.tif"
23. Slope1 = 'SlopeMap.tif'
24. SlopeThresh = 'Slope_Thresh.tif'
25. Clip_Im = 'Post_Fin_Clip.tif'
26. Seg= 'SEGMENTED_CLIPPED_IMAGE.tif'
27. dsc = arcpy.Describe(Post_Event)

```

```

28.     #Create Slope Map
29.     outMeasurement = "DEGREE"
30.     zFactor = "1"
31.     outSlope = Slope(DEM, outMeasurement, zFactor)
32.     outSlope.save(Slope1)
33.
34.     #Remove Slopes Less than 15
35.     RemoveSlopes = ExtractByAttributes(Slope1, inSQLClause1)
36.     RemoveSlopes.save(SlopeThresh)
37.
38.     #Clip Change Detection, Removing Slopes
39.     arcpy.gp.ExtractByMask_sa(NDVICD,
40.                               SlopeThresh,
41.                               Clip_CD)
42.
43.     #Change Detection Clustering
44.     outUnsupervised = IsoClusterUnsupervisedClassification(Clip
45. _CD,classes)
46.     outUnsupervised.save(OutIso)
47.     attExtract = ExtractByAttributes(OutIso, inSQLClause2)
48.     attExtract.save(IsoEx)
49.
50.     #Clip Post Event Image
51.     outExtractByMask = ExtractByMask(Post_Event, IsoEx)
52.     outExtractByMask.save(Clip_Im)
53.
54.     #Segment Image
55.     arcpy.gp.SegmentMeanShift_sa(Clip_Im,
56.                                  Seg,
57.                                  "20",
58.                                  "20",
59.                                  "5",
60.                                  "")

```

Figure A3: Accuracy Assessment Code. https://github.com/sabrinanicc/ls_eval

```

1.  ## =====SETTING UP PARAMETERS=====
2.  == ##
3.  # Import necessary modules
4.  from __future__ import division
5.  import arcpy
6.  from arcpy.sa import *
7.  arcpy.CheckOutExtension("Spatial")
8.  #Allow output file to overwrite any existing file of the same name
9.  arcpy.env.overwriteOutput = True
10.
11. #Inputs (shapefiles of the automatically and manually derived inventories)
12. automatic = arcpy.GetParameter(0)
13. manual = arcpy.GetParameter(1)

```



```

14. arcpy.env.workspace = arcpy.GetParameter(2)
15.
16. #Outputs (shapefiles created in process to determine accuracy)
17. intersect = "Intersection.shp"
18. merge = "Merge.shp"
19. dissolve = "Dissolve.shp"
20. spatialjoin = "Spatial_Join.shp"
21.
22. #Sets spatial reference to that of the input shapefile.
23. dsc = arcpy.Describe(manual)
24. coord_sys = dsc.spatialReference
25.
26.
27. ## =====ACCURACY ASSESSMENT=====
   == ##
28.
29. #Intersect the Manual and Automatic Inventory Shapefiles
30. arcpy.Intersect_analysis(in_features=[automatic, manual],
31.                           out_feature_class=intersect,
32.                           join_attributes="ALL",
33.                           output_type="INPUT")
34.
35. #Determine Area of Intersection (Create an 'Area Field')
36. arcpy.AddField_management (intersect, "Area_Int", "DOUBLE", "", "", "", "NULLABLE", "")
37. geometryField = arcpy.Describe(intersect).shapeFieldName
38. cursor = arcpy.UpdateCursor(intersect)
39. for row in cursor:
40.     AreaValue = row.getValue(geometryField).area
41.     row.setValue("Area_Int",AreaValue)
42.     cursor.updateRow(row)
43. del row, cursor
44.
45. #Merge the Manual Inventory and Automatic Inventory Shapefiles
46. arcpy.Merge_management(inputs=[automatic,manual],
47.                          output= merge)
48.
49. #Dissolve the Merged Shapefile.
50. arcpy.Dissolve_management(in_features=merge,
51.                            out_feature_class= dissolve,
52.                            dissolve_field="",
53.                            statistics_fields="",
54.                            multi_part="SINGLE_PART",
55.                            unsplit_lines="UNSPLIT_LINES")
56.
57. #Determine the Area of the Union (Create an 'Area Field')
58. arcpy.AddField_management (dissolve, "Area_Diss", "DOUBLE", "", "", "", "NULLABLE", "")
59. geometryField = arcpy.Describe(dissolve).shapeFieldName
60. cursor = arcpy.UpdateCursor(dissolve)
61. for row in cursor:
62.     AreaValue = row.getValue(geometryField).area
63.     row.setValue("Area_Diss",AreaValue)
64.     cursor.updateRow(row)
65. del row, cursor #Clean up cursor objects
66.
67.
68.
69. #Create a spatial join of the intersection and the dissolve.
70.
71. fieldmappings = arcpy.FieldMappings()

```

```

72. fieldnamestosum = ["Area_Int"]
73. fieldmappings.addTable(dissolve)
74. fieldmappings.addTable(intersect)
75.
76. keepers = ["Area_Diss", "Area_Int"]
77. for field in fieldmappings.fields:
78.     if field.name not in keepers:
79.         fieldmappings.removeFieldMap(fieldmappings.findFieldMapIndex(field.name)
80.         )
81. for fieldName in fieldnamestosum:
82.
83.
84.     fieldIndex = fieldmappings.findFieldMapIndex(fieldName)
85.     fieldMap = fieldmappings.getFieldMap(fieldIndex)
86.     fieldMap.mergeRule = 'Sum'
87.     fieldmappings.replaceFieldMap(fieldIndex, fieldMap)
88.
89.
90. arcpy.SpatialJoin_analysis(target_features=dissolve,
91.                             join_features=intersect,
92.                             out_feature_class=spatialjoin,
93.                             join_operation="JOIN_ONE_TO_ONE",
94.                             join_type="KEEP_ALL",
95.                             field_mapping = fieldmappings,
96.                             match_option="CONTAINS",
97.                             )
98.
99. #Create a new column in the spatially joined file.
100. arcpy.AddField_management (spatialjoin, "IoU", "DOUBLE", "", "", "", "NU
    LLABLE", "")
101.
102.     #Calculate the IoU value.
103.     arcpy.CalculateField_management(in_table=spatialjoin,
104.                                     field="IoU",
105.                                     expression="[Area_Int] / [Area_Diss]",
106.                                     expression_type="VB",
107.                                     code_block="")
108.
109.     #Count how many landslides are above threshold IoU VALUE.
110.
111.     fields = "IoU"
112.     count = 0
113.     with arcpy.da.SearchCursor(spatialjoin, fields) as cursor:
114.         for row in cursor:
115.             if row[0] > .20:
116.                 count += 1
117.
118.     #Count True Positives
119.     TP = count
120.     #Count how many landslides in Manual Inventory
121.     NL_MM = int(arcpy.GetCount_management(manual).getOutput(0))
122.     #Count how many landslides are in Automatic Inventory.
123.     NL_AM = int(arcpy.GetCount_management(automatic).getOutput(0))
124.     #Determine False Negatives(TP-NL_MM).
125.     FN = NL_MM - TP
126.     #Determine False Positives.(TP-NL_AM).
127.     FP = NL_AM - TP
128.     #Determine Recall. (TP/(TP+FN)*100)
129.     Recall = (TP/(TP+FN))*100
130.     #Determine Precision. (TP/(TP+FP)*100)

```

```

131.         Precision = (TP/(TP+FP))*100
132.
133.         # Sum the area of manual inventory landslides.
134.         summed_total = 0
135.         with arcpy.da.SearchCursor(manual, "Shape_Area") as cursor:
136.             for row in cursor:
137.                 summed_total = summed_total + row[0]
138.
139.         Area_M = summed_total
140.
141.         # Sum the area of automatic inventory landslides.
142.         summed_total_A = 0
143.         with arcpy.da.SearchCursor(automatic, "Shape_Area") as cursor:
144.             for row in cursor:
145.                 summed_total_A = summed_total_A + row[0]
146.
147.         Area_A = summed_total_A
148.
149.
150.         #Sum the area of true positives by summing areas in auto mapping above th
reholds.
151.         field1 = "IoU"
152.         field2 = "Area_Int"
153.         summed_total2 = 0
154.         with arcpy.da.SearchCursor(spatialjoin, [field1, field2]) as cursor:
155.             for row in cursor:
156.                 if row[0] > .30:
157.                     summed_total2 = summed_total2 + row[1]
158.
159.
160.         Area_TP = summed_total2
161.
162.
163.         #False Negatives. (Area_FN = Area_M - Area_TP)
164.         Area_FN = Area_M - Area_TP
165.         #False Positives. (Area_FP = Area_A - Area_TP)
166.         Area_FP = Area_A - Area_TP
167.         #Recall. (Area_TP/Area_TP + Area_FN)
168.         Area_Recall = (Area_TP/(Area_TP + Area_FN))*100
169.         #Precision. (Area_TP/Area_TP + Area_FP)
170.         Area_Precision = (Area_TP/(Area_TP + Area_FP))*100
171.
172.
173.
174.         ## =====PRINTS ACCURACY VALUES TO RESULTS WINDOW =====
===== ##
175.
176.         print arcpy.AddMessage('====RECOGNITION ACCURACY====')
177.         print arcpy.AddMessage('Recall: ')
178.         print arcpy.AddMessage(Area_Recall)
179.         print arcpy.AddMessage('Precision: ')
180.         print arcpy.AddMessage(Area_Precision)
181.
182.
183.         print arcpy.AddMessage('====EXTENT ACCURACY====')
184.         print arcpy.AddMessage('Recall: ')
185.         print arcpy.AddMessage(Area_Recall)
186.         print arcpy.AddMessage('Precision: ')
187.         print arcpy.AddMessage(Area_Precision)
188.
189.

```

190.
 191. ## =====
 ===== ##

Table A1: Compilation of accuracy metrics and reported accuracies used in other studies.

Source	Metrics Used	Reported Accuracies
Barlow et al., 2003	$\text{Error of Omission} = \frac{\text{False Negative}}{\text{False Negative} + \text{True Positive}}$ $\text{Error of Comission} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Positive}}$	EO= 75% EC=77.5%
Barlow et al., 2006	$\text{Error of Omission} = \frac{\text{False Negative}}{\text{False Negative} + \text{True Positive}}$ $\text{Error of Comission} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Positive}}$	EO =90%,80%,60% EC=80%,80%,80%
Daneels et al., 2007	Percent Recovered Area, Omitted Area, Commission Error	60%, 40%, 45%
Martha et al., 2010	$\text{Recognition} = \frac{\text{True Positives}}{\text{Total Landslides}}$ $\text{Classification} = \frac{\text{Correctly Recognized and Classified by Landslide Type}}{\text{Total Landslides}}$ $\text{Error of Omission} = \frac{\text{False Negative}}{\text{False Negative} + \text{True Positive}}$ $\text{Error of Comission} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Positive}}$	Number of Landslides: [Recognition =76.4%, Classification = 69.1%, EO=23.6%, EC=56.4%] Extent of Landslides: [Recognition =69.9%, Classification = 69.5%, EO=3.7%, EC=9.2%]
Mondini et al., 2011a	Area Under Receiver Operating Characteristic Curve (A_{ROC})	Training Area:[A_{ROC} >860] Validation Area:[A_{ROC} >.800]
Mondini et al., 2011b	Percentage of automatically detected landslides that intersect manually detected landslides.	57.43%
Stumpf and Kerle, 2011	$\text{Users Accuracy} = \frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}}$ $\text{Producers Accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$ $F - \text{Score} = \frac{2 * \text{Users Accuracy} * \text{Producers Accuracy}}{\text{Users Accuracy} + \text{Producers Accuracy}}$	Object-Detection:[F-Scores = 89.7%, 80.5%, 73% ,76.5%] Areas:[F-Scores = 87.1%, 81.2%, 73.7% and 77.9%]
Lu et al., 2011	$\text{Users Accuracy} = \frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}}$ $\text{Producers Accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$	Number of Landslides:[UA=75.9%.P A=69.9%] Spatial Extent:[UA=81.8%,PA=69.5%]

Li et al., 2016a	Completeness = $\frac{P_{lm}}{P_r}$, Correctness = $\frac{P_{lm}}{P_l}$, Quality = $\frac{P_{lm}}{P_l + P_{rum}}$	Completeness = 73.6%, Correctness = 93.8% and Quality = 67.1%
Li et al., 2016b	Completeness = $\frac{P_{lm}}{P_r}$, Correctness = $\frac{P_{lm}}{P_l}$, Quality = $\frac{P_{lm}}{P_l + P_{rum}}$	Not Explicitly Reported
Qin et al., 2018	Completeness = $\frac{P_{lm}}{P_r}$, Correctness = $\frac{P_{lm}}{P_l}$, Quality = $\frac{P_{lm}}{P_l + P_{rum}}$	Completeness= 95.93% Correctness= 88.03% Quality =76.08%
Yong Lv et al., 2018	Completeness = $\frac{P_{lm}}{P_r}$, Correctness = $\frac{P_{lm}}{P_l}$, Quality = $\frac{P_{lm}}{P_l + P_{rum}}$	Completeness, Correctness and Quality Respectively for each study area:[89.57%, 90.71%, 82.04%], [90.01%,80.15%,73.61%], [95.75%,60.06%,58.50] [92.25%,89.05%,82.85%]
Ghorbanzadeh et al., 2019	$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$ $\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$ $\text{F1 - Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$ $\text{mIOU} = \text{Mean of IOU values}$	<p>F1 Measure</p> <p>=87.8%</p> <p>mIOU = 78.26%</p>

Table A2: Threshold points used to create power-law fit seen in figures 12 and 13.

Minimum Area Value			Minimum Volume Value		
Training Site	Auto	3016.9	Training Site	Auto	2465.6
	Manual	608		Manual	517.4
Watershed	Auto	11999	Watershed	Auto	8406
	Manual	1803.4		Manual	1532.7

LIST OF REFERENCES

- Aguilar, M.A., Aguilar, F.J., Guirado, E., Betlej, M., Cichon, P., Nemmaoui, A., Vallario, A., Parente, C., 2016. ASSESSMENT OF MULTIREOLUTION SEGMENTATION FOR EXTRACTING GREENHOUSES FROM WORLDVIEW-2 IMAGERY XLI, 12–19. <https://doi.org/10.5194/isprsarchives-XLI-B7-145-2016>
- Barlow, J., Franklin, S., Martin, Y., 2006. High Spatial Resolution Satellite Imagery , DEM Derivatives , and Image Segmentation for the Detection of Mass Wasting Processes 72, 687–692.
- Barlow, J., Martin, Y., Franklin, S.E., 2003. Detecting translational landslide scars using segmentation of Landsat ETM + and DEM data in the northern Cascade Mountains , British Columbia 29, 510–517.
- Bawiec, W. J. (1998). Geology, geochemistry, geophysics, mineral occurrences, and mineral resource assessment for the commonwealth of Puerto Rico.
- Bessette-Kirton, E.K., Cerovski-Darriau, C., Schulz, W.H., Coe, J.A., Kean, J.W., Godt, J.W., Thomas, M.A., Hughes, K.S., 2019. Landslides Triggered by Hurricane Maria: Assessment of an Extreme Event in Puerto Rico. GSA Today 29, 4–10. <https://doi.org/10.1130/gsatg383a.1>

- Blaschke, T., 2010. ISPRS Journal of Photogrammetry and Remote Sensing Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16.
<https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Burrough, P. A., McDonnell, R., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of geographical information systems*. Oxford university press.
- Cardinali, M., Galli, M., Guzzetti, F., Ardizzone, F., Reichenbach, P., Bartoccini, P.,
2006. Rainfall induced landslides in December 2004 in south-western Umbria,
central Italy: Types, extent, damage and risk assessment. *Nat. Hazards Earth Syst.*
Sci. 6, 237–260. <https://doi.org/10.5194/nhess-6-237-2006>
- Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical
data. *SIAM Rev.* 51, 661–703. <https://doi.org/10.1137/070710111>
- Coe, J.A., Bessette-Kirton, E.K., Geertsema, M., 2018. Increasing rock-avalanche size
and mobility in Glacier Bay National Park and Preserve, Alaska detected from 1984
to 2016 Landsat imagery. *Landslides* 15, 393–407. <https://doi.org/10.1007/s10346-017-0879-7>
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space
analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 603–619.
<https://doi.org/10.1109/34.1000236>

- Dadson, S.J., Hovius, N., Chen, H., Dade, W.B., Lin, J.C., Hsu, M.L., Lin, C.W., Horng, M.J., Chen, T.C., Milliman, J., Stark, C.P., 2004. Earthquake-triggered increase in sediment delivery from an active mountain belt. *Geology* 32, 733–736.
<https://doi.org/10.1130/G20639.1>
- Danneels, G., Pirard, E., 2007. Automatic landslide detection from remote sensing images using supervised classification methods.
<https://doi.org/10.1109/IGARSS.2007.4423479>
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240.
- Deering, D.W., Haas, R.H., 1980. USING LANDSAT DIGITAL DATA FOR ESTIMATING GREEN BIOMASS. NASA Tech. Memo. 80727.
- Densmore, A.L., Hovius, N., n.d. Topographic fingerprints of bedrock landslides 371–374.
- Dey, V., Zhang, Y., Zhong, M., Engineering, G., 2010. A Review on Image Segmentation Techniques With Remote Sensing Perspective. ISPRS TC VII Symp. – 100 Years ISPRS XXXVIII, 31–42.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES

Operational Services. *Remote Sens. Environ.* 120, 25–36.

<https://doi.org/10.1016/j.rse.2011.11.026>

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>

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019.

Evaluation of Different Machine Learning Methods and Deep-Learning

Convolutional Neural Networks for Landslide Detection.

<https://doi.org/10.3390/rs11020196>

Guzzetti, F., Malamud, B. D., Turcotte, D. L., & Reichenbach, P. (2002). Power-law correlations of landslide areas in central Italy. *Earth and Planetary Science Letters*, 195(3), 169–183. [https://doi.org/https://doi.org/10.1016/S0012-821X\(01\)00589-1](https://doi.org/https://doi.org/10.1016/S0012-821X(01)00589-1)

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. *Earth-Science Rev.* 112, 42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.

- Hay, G. J., Marceau, D. J., Dube, P., & Bouchard, A. (2001). A multiscale framework for landscape analysis: object-specific analysis and upscaling. *Landscape Ecology*, 16(6), 471–490.
- Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS J. Photogramm. Remote Sens.* 57, 327–345. [https://doi.org/10.1016/S0924-2716\(02\)00162-4](https://doi.org/10.1016/S0924-2716(02)00162-4)
- Hay, G.J., Castilla, G., 2006. OBJECT-BASED IMAGE ANALYSIS: STRENGTHS, WEAKNESSES, OPPORTUNITIES AND THREATS (SWOT) G.J. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 36, 4.
- Hewitt, K., 1998. Catastrophic landslides and their effects on the Upper Indus streams, Karakoram Himalaya, northern Pakistan. *Geomorphology* 26, 47–80. [https://doi.org/10.1016/S0169-555X\(98\)00051-8](https://doi.org/10.1016/S0169-555X(98)00051-8)
- Hovius, N., Stark, C. P., & Allen, P. A. (1997). Sediment flux from a mountain belt derived by landslide mapping. *Geology*, 25(3), 231–234.
- Hughes, K.S., Bayouth García, D., Martínez Milian, G.O., Schulz, W.H., and Baum, R.L., 2019, Map of slope-failure locations in Puerto Rico after Hurricane María: U.S. Geological Survey data release, <https://doi.org/10.5066/P9BVMD74>.

- Keellings, D., & Hernández Ayala, J. J. (2019). Extreme rainfall associated with Hurricane Maria over Puerto Rico and its connections to climate variability and change. *Geophysical Research Letters*, 46(5), 2964–2973.
- Keyport, R.N., Oommen, T., Martha, T.R., Sajinkumar, K.S., Gierke, J.S., 2018. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinf.* 64, 1–11.
<https://doi.org/10.1016/j.jag.2017.08.015>
- Klar, A., Aharonov, E., Kalderon-Asael, B., Katz, O., 2011. Analytical and observational relations between landslide volume and surface area. *J. Geophys. Res. Earth Surf.* 116, 1–10. <https://doi.org/10.1029/2009JF001604>
- Korup, O., Densmore, A.L., Schlunegger, F., 2010. The role of landslides in mountain range evolution. *Geomorphology* 120, 77–90.
<https://doi.org/10.1016/j.geomorph.2009.09.017>
- Larsen, M. C. (2012). Landslides and sediment budgets in four watersheds in eastern Puerto Rico: Chapter F in Water quality and landscape processes of four watersheds in eastern Puerto Rico. In S. F. Murphy & R. F. Stallard (Eds.), *Professional Paper*.
<https://doi.org/10.3133/pp1789F>
- Larsen, M., & Torres Sanchez, A. J. (1998). The frequency and distribution of Larsen, M.C., Torres Sa recent landslides in three montane tropical regions of Puerto Rico. *Geomorphology*, 24309e331, 309–331.

Li, G., West, A. J., Densmore, A. L., Jin, Z., Parker, R. N., & Hilton, R. G. (2014).

Seismic mountain building: Landslides associated with the 2008 Wenchuan earthquake in the context of a generalized model for earthquake volume balance.

Geochemistry, Geophysics, Geosystems, 15(4), 833–844.

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 Sens. 7, 9705–9726. <https://doi.org/10.3390/rs70809705>

Li, Z., Shi, W., Lu, P., Yan, L., Wang, Q., Miao, Z., 2016. Remote Sensing of

Environment Landslide mapping from aerial photographs using change detection-based Markov random field. *Remote Sens. Environ.* 187, 76–90.

<https://doi.org/10.1016/j.rse.2016.10.008>

Li, Z., Shi, W., Myint, S.W., Lu, P., Wang, Q., 2016. Remote Sensing of Environment

Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sens. Environ.* 175, 215–230.

<https://doi.org/10.1016/j.rse.2016.01.003>

Lim, Y.K., Schubert, S.D., Kovach, R., Molod, A.M., Pawson, S., 2018. The Roles of

Climate Change and Climate Variability in the 2017 Atlantic Hurricane Season. *Sci.*

Rep. 8, 1–10. <https://doi.org/10.1038/s41598-018-34343-5>

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U.,

Cadau, E., Gascon, F., 2016. Sentinel-2 SEN2COR: L2A processor for users. *Eur.*

Sp. Agency, (Special Publ. ESA SP SP-740, 9–13.

- Lu, P., Stumpf, A., Kerle, N., Casagli, N., 2011. Object-Oriented Change Detection for Landslide Rapid Mapping 8, 701–705. <https://doi.org/10.1109/LGRS.2010.2101045>
- Lu, Z.Y., 2018. Landslide Inventory Mapping From Bitemporal High-Resolution Remote Sensing Images Using Change Detection and Multiscale Segmentation 11, 1520–1532. <https://doi.org/10.1109/JSTARS.2018.2803784>
- Lyons, N.J., Mitsova, H., Wegmann, K.W., 2014. Improving mass-wasting inventories by incorporating debris flow topographic signatures. *Landslides* 11, 385–397. <https://doi.org/10.1007/s10346-013-0398-0>
- Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties. *Earth Surf. Process. Landforms* 29, 687–711. <https://doi.org/10.1002/esp.1064>
- Marc, O., Hovius, N., 2015. Amalgamation in landslide maps: Effects and automatic detection. *Nat. Hazards Earth Syst. Sci.* 15, 723–733. <https://doi.org/10.5194/nhess-15-723-2015>
- 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.
- Martha, T.R., Kerle, N., Westen, C.J. Van, Jetten, V., Kumar, K.V., 2011. Segment Optimization and Data-Driven Thresholding for Knowledge-Based Landslide

- Detection by Object-Based Image Analysis. *IEEE Trans. Geosci. Remote Sens.* 49, 4928–4943. <https://doi.org/10.1109/TGRS.2011.2151866>
- Martinuzzi, S., Gould, W. A., & Gonzalez, O. M. R. (2007). Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. *Landscape and Urban Planning*, 79(3-4), 288-297.
- Masson, D. G., & Scanlon, K. M. (1991). The neotectonic setting of Puerto Rico. *Geological Society of America Bulletin*, 103(1), 144–154.
- Mondini, A. C., Chang, K.-T., & Yin, H.-Y. (2011). Combining multiple change detection indices for mapping landslides triggered by typhoons. *Geomorphology*, 134(3), 440–451. [https://doi.org/https://doi.org/10.1016/j.geomorph.2011.07.021](https://doi.org/10.1016/j.geomorph.2011.07.021)
- Mondini, A.C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., Ardizzone, F., 2011. Remote Sensing of Environment Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens. Environ.* 115, 1743–1757. <https://doi.org/10.1016/j.rse.2011.03.006>
- Monroe, W. H. (1975). Geology of the middle Tertiary formations of Puerto Rico (No. 75-313). *US Geological Survey*,.
- Monroe, W. H. (1976). The karst landforms of Puerto Rico (No. 899). *US Geological Survey*.

- Moosavi, V., Talebi, A., Shirmohammadi, B., 2014. Geomorphology Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method 204, 646–656.
- Murphy, S.F., Stallard, R.F., Larsen, M.C., Gould, W.A., 2012. Physiography, geology, and land cover of four watersheds in eastern Puerto Rico. *Water Qual. Landsc. Process. four watersheds East. Puerto Rico* 1–24.
- Parise, M., 2001. Landslide mapping techniques and their use in the assessment of the landslide hazard. *Phys. Chem. Earth, Part C Solar, Terr. Planet. Sci.* 26, 697–703. [https://doi.org/10.1016/S1464-1917\(01\)00069-1](https://doi.org/10.1016/S1464-1917(01)00069-1)
- Parker, R.N., Densmore, A.L., Rosser, N.J., De Michele, M., Li, Y., Huang, R., Whadcoat, S., Petley, D.N., 2011. Mass wasting triggered by the 2008 Wenchuan earthquake is greater than orogenic growth. *Nat. Geosci.* 4, 449–452. <https://doi.org/10.1038/ngeo1154>
- Pasch, R. J., Penny, A. B., & Berg, R. (2018). National Hurricane center tropical cyclone report: Hurricane Maria. *TROPICAL CYCLONE REPORT AL152017, National Oceanic And Atmospheric Administration and the National Weather Service*, 1–48.
- Qin, Y., Lu, P., Li, Z., 2018. LANDSLIDE INVENTORY MAPPING FROM BITEMPORAL 10 m SENTINEL-2 IMAGES USING CHANGE DETECTION BASED MARKOV RANDOM FIELD XLII, 7–10.

- Rahman, M. A., & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. *International Symposium on Visual Computing*, 234–244.
- Ramos-Scharrón, C. E., & Arima, E. (2019). Hurricane Maria's Precipitation Signature in Puerto Rico: A Conceivable Presage of Rains to Come. *Scientific Reports*, 9(1), 1–7.
- Santangelo, M., Cardinali, M., Rossi, M., Mondini, A.C., Guzzetti, F., 2010. Remote landslide mapping using a laser rangefinder binocular and GPS. *Nat. Hazards Earth Syst. Sci.* 10, 2539–2546. <https://doi.org/10.5194/nhess-10-2539-2010>
- Smith, A. B. (2018). 2017 U . S . billion-dollar weather and climate disasters : a historic year in context. The Historic U . S . Billion-dollar Disasters of 2017 Hurricane Harvey : Hurricane Irma : Hurricane Maria : Looking at the bigger picture. 8816.
- Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397–399.
- Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment*, 115(10), 2564–2577.
- Tanyaş, H., Allstadt, K.E., van Westen, C.J., 2018. An updated method for estimating landslide-event magnitude. *Earth Surf. Process. Landforms* 43, 1836–1847. <https://doi.org/10.1002/esp.4359>

- Tanyaş, H., van Westen, C.J., Allstadt, K.E., Jibson, R.W., 2019. Factors controlling landslide frequency–area distributions. *Earth Surf. Process. Landforms* 44, 900–917. <https://doi.org/10.1002/esp.4543>
- Varnes, D.J., 1996. Landslide types and processes. Spec. Rep. - Natl. Res. Coun. Transp. Res. Board 247, 36–75.
- Weirich, F., Blesius, L., 2007. Comparison of satellite and air photo based landslide susceptibility maps. *Geomorphology* 87, 352–364. <https://doi.org/10.1016/j.geomorph.2006.10.003>
- Wulder, M.A., White, J.C., Hay, G.J., Castilla, G., 2015. Pixels to objects to information: Spatial context to aid in forest characterization with remote sensing, in: *Object-Based Image Analysis*. Springer, pp. 345–363. https://doi.org/10.1007/978-3-540-77058-9_19
- Yanites, B.J., Mitchell, N.A., Bregy, J.C., Carlson, G.A., Cataldo, K., Holahan, M., Johnston, G.H., Nelson, A., Valenza, J., Wanker, M., 2018. Landslides control the spatial and temporal variation of channel width in southern Taiwan: Implications for landscape evolution and cascading hazards in steep, tectonically active landscapes. *Earth Surf. Process. Landforms* 43, 1782–1797. <https://doi.org/10.1002/esp.4353>
- Yanites, B.J., Tucker, G.E., Mueller, K.J., Chen, Y.G., Wilcox, T., Huang, S.Y., Shi, K.W., 2010. Incision and channel morphology across active structures along the Peikang River, central Taiwan: Implications for the importance of channel width. *Bull. Geol. Soc. Am.* 122, 1192–1208. <https://doi.org/10.1130/B30035.1>

BIOGRAPHY

Sabrina Nicole Martinez was born in Houston, TX on June 12th, 1995. She discovered her passion for geology while pursuing her undergraduate degree at the University of Houston. There she earned a B.S. in Geology while developing an interest in remote sensing and GIS applications. This led her to participate in two internships at NASA's Lunar and Planetary Institute where she studied the geology of the Moon and Venus using satellite data. She then attended Tulane University to study how satellite imagery can be used to aid landslide hazard mapping.