How Does Land Cover Classification In Google Earth Engine Compare With Traditional Methods Of Land Cover Classification? What Are The Tradeoffs?

Carlos Sebastian Reyes

University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd

Part of the Databases and Information Systems Commons, Environmental Sciences Commons, and the Geographic Information Sciences Commons

Recommended Citation
https://scholarworks.utep.edu/open_etd/3331

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.
HOW DOES LAND COVER CLASSIFICATION IN GOOGLE EARTH ENGINE COMPARE WITH TRADITIONAL METHODS OF LAND COVER CLASSIFICATION? WHAT ARE THE TRADEOFFS?

CARLOS SEBASTIAN REYES
Master’s Program in Environmental Science

APPROVED:

__________________________________________
Deana Pennington, Ph.D., Chair

__________________________________________
Hugo A Gutierrez, Ph.D.

__________________________________________
Jayajit Chakraborty, Ph.D.

__________________________________________
Stephen Crites, Ph.D.
Dean of the Graduate School
HOW DOES LAND COVER CLASSIFICATION IN GOOGLE EARTH ENGINE COMPARE WITH TRADITIONAL METHODS OF LAND COVER CLASSIFICATION? WHAT ARE THE TRADEOFFS?

by

CARLOS SEBASTIAN REYES, B.A.

THESIS

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

Department of Earth, Environmental, and Resource Sciences
THE UNIVERSITY OF TEXAS AT EL PASO
May 2021
DEDICATION

I dedicate my thesis to my family and friends that supported me.

To my parents

who worked hard to help put me through college,

as well as believing in me and encouraging me to strive for my dreams.

To my cousins

who made my life more fun and encouraged me when I was losing motivation.

To my mentors and coworkers

who encouraged me to get more involved in research and pursue acquiring new skills and

experiences through internships and traveling.

To my friends and my mentor

who encouraged me to keep pushing to finish my degree despite many setbacks.
ACKNOWLEDGEMENTS

This material is based upon work that was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2015-68007-23130. The work was also supported by a summer internship at the Texas A&M Forest Service funded through the Diana Natalicio Environmental Internship program at the University of Texas at El Paso. The author gratefully acknowledges Dr. William Hargrove, Dr. Stanley Mubako, Dr. Deana Pennington, and Omar Belhaj for their assistance in the MRG project as well as Aaron Stottlemyer and Rebekah Zehnder for their assistance in the PDA analysis.
ABSTRACT

The project focuses on comparing land cover classification of traditional methods such as ArcGIS with newer ones such as Google Earth Engine (GEE) as well as discussing any potential tradeoffs. Two studies were performed in both platforms, the first involved analyzing land cover change in the Middle Rio Grande (MRG) region of southern New Mexico, far west Texas, and northern Chihuahua, Mexico. The MRG study focused on urban and agricultural change in the region using two different classification methods. The second study focused on creating a post-hurricane damage assessment (PDA) with the goal of developing an automated method of estimating the location and quantity of debris (focusing on trees). Overall, the research shows the pros and cons of each platform as well as which is better given user’s overall standing in terms of equipment, prior experience, project size, and time constraints.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS.................................................................................................v

ABSTRACT.................................................................................................................. vi

TABLE OF CONTENTS.................................................................................................... vii

LIST OF TABLES........................................................................................................... ix

LIST OF FIGURES .......................................................................................................... x

CHAPTER 1: INTRODUCTION .........................................................................................1

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW ..............................................4
  Unsupervised and Supervised Classification ...............................................................4
  Accuracy Assessment .................................................................................................5

CHAPTER 3: METHODS ..................................................................................................7
  Study Areas ................................................................................................................7
  Middle Rio Grande LCLU Classification ..................................................................10
  ArcGIS .......................................................................................................................10
    Selecting and Preprocessing the Imagery .................................................................10
    Unsupervised Classification ....................................................................................11
    Supervised Classification ......................................................................................12
  GEE ..........................................................................................................................13
    Selecting and Preprocessing the Imagery .................................................................13
    Unsupervised Classification ....................................................................................14
    Supervised Classification ......................................................................................15
  South Texas PDA Classification ..............................................................................16
    Selecting and Preprocessing the Imagery for ArcGIS ............................................16
    Tree Detection Process in ARCGIS ......................................................................16
    Selecting and Preprocessing the Imagery for GEE ................................................17
    Tree Detection Process in GEE ............................................................................18
  Comparison Across ArcGIS and GEE Approaches ..................................................19
LIST OF TABLES

Table 1. Comparison of hardware requirements................................................................. 26
Table 2. Code to export images and tables from GEE........................................................... 33
Table 3. Algorithm and Accuracy Assessment Comparison between ArcGIS and GEE......... 38
LIST OF FIGURES

Figure 1. Map showing the study area of the MRG project......................................................... 8

............................................................................................................................................. 9

Figure 2. Map showing areas impacted by Hurricane Harvey.................................................... 9

Figure 3. The flowchart on the left is how to prepare imagery for classification in ArcGIS. The
flowchart on the right is how to pick which classification method to use in ArcGIS. ............. 11

................................................................................................................................................ 14

Figure 4. Flowchart showing how satellite imagery is prepared in GEE ............................. 14

Figure 5. Flowchart of the preparation of imagery before conducting the tree detection process in
GEE........................................................................................................................................... 17

Figure 6. Flowchart showing the Tree Detection process for the PDA in GEE...................... 19

Figure 7. GEE Unsupervised and Supervised classification maps ........................................ 23

Figure 8. Error matrix of MRG classification (Mubako et., al 2018)................................. 24

Figure 9. Map showing the original and expanded study area boundaries......................... 35
CHAPTER 1: INTRODUCTION

As time progresses new environmental problems arise around the globe. There is a need for environmental scientists to find solutions to respond to these problems in a timely manner using the voluminous data available (Argent 2004; Pennington et al. 2020). There are many different types of technologies and data that can be used to conduct analysis of human-environmental problems. Advances in computing environments such as “cloud computing” have made high-end technical resources available to scientists at minimal cost “without the need of owning any infrastructure” (Rodriguez and Buyya 2017). Environmental databases are “growing in complexity, size, and resolution,” requiring scientists to develop capabilities for using such high-end resources (Gibert et al. 2018). In addition, the geographic scale of human-environmental science research projects often encompasses large, regional study areas. When these studies are conducted using remotely sensed imagery of the earth’s surface collected from satellites, airplanes, or drones, the resulting datasets often are quite voluminous. A remotely sensed image is comprised of pixels, each of which is assigned one-to-many values measured by sensors placed on the vehicle. A wide variety of sensors are available to collect data from many different regions of the electromagnetic spectrum. The size of each pixel (spatial resolution), the number of measurements made (spectral resolution) and the spatial extent captured determine the size of the raw dataset. The raw dataset must be preprocessed then analyzed, resulting in many more images. To address the challenges that arise from these voluminous datasets there is an increasing need to use new technologies and software that can handle the size and complexity of the data. This thesis investigates new cloud computing technologies available to conduct spatial analysis of remotely sensed imagery, Google Earth Engine (GEE), and compares the benefits, challenges, and tradeoffs with approaches using a traditional desktop Geographic Information Systems (GIS), specifically, the Environmental System Research Institute’s (Esri) ArcGIS. ArcGIS is widely recognized as one of the most common platforms for GIS spatial analysis techniques as well as being recognized as one of the industry standard platforms to store and process “Big Data” (Singleton and Arribas-Bel
ArcGIS has capabilities for geographically referenced data input, editing, processing, analysis, and display (Smith and Paradis 1989), including remotely sensed imagery (Butt et al. 2015). ArcGIS provides a comprehensive set of functions for image preparation, analysis and mapping (Boori, Voženílek, and Choudhary 2015).

GEE is a cloud-based platform that provides access to high-performance computing resources for processing large geospatial datasets (Gorelick et al. 2017). The web portal can store and create vector data and process the data using programming/processing algorithms for analysis. Being cloud-based, users are able to take advantage of the large remote sensing databases that are available for free from multiple US government agencies including the USGS, NOAA, and NASA (Woodcock et al., 2008). These databases also have access to a variety of satellite data such as Landsat, MODIS, and Sentinel among other types of geophysical, demographic, and climate & weather data products that can be found in GEE’s public data catalog (https://earthengine.google.com/datasets/).

This investigation compares land use and land cover (LULC) classification techniques in both ArcGIS and GEE. LULC classification techniques are applied to categorize the type of surficial covering at every location across the study area. LULC classification methods can be used for various reasons (Rwanga and Ndambuki 2017), including use as input for environmental modeling, for instance, models dealing with climate change and policies development (Dengsheng Lu et al., 2004; Disperati and Virdis 2015). LULC classification results in each pixel being classified into a single land use or cover type. Classification occurs through application of an algorithm that analyzes the spectral data in each pixel of the image, then places that pixel in a LULC class. Many algorithms are available using a variety of different techniques, and emerging and rapidly changing methods in artificial intelligence, statistics, machine learning and data mining have led to a proliferation of available algorithms (Pennington et al., 2020). Many of the newest algorithms are available through GEE. In contrast, ArcGIS provides more traditional algorithms that have been used in many studies previously and have been adapted to the software. Hence, this investigation will compare LULC approaches in GEE and ArcGIS across a number of dimensions,
including algorithm availability and accuracy, general characteristics regarding human usage (for example, time required to learn the approach), ease of data accessibility, processing efficiency, algorithm availability, and classification outcomes. It will explore the benefits and tradeoffs of both platforms in order to determine which classification method and technology is best given the circumstances of funding, training, or time constraints as well as the project scenario.
CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

Unsupervised and Supervised Classification

There are many forms of LULC classification methods (Talukdar et al. 2020). The main ones used are broadly categorized as supervised and unsupervised classification. Supervised classification uses “ground truth” training samples to identify and classify each pixel in an image. Unsupervised classification (sometimes called “clustering”) does not rely on ground truth training samples and instead uses the data itself to identify patterns in the data.

In supervised classification, prior to classification the analyst identifies target LULC classes using prior knowledge of land use types in the study area (Mubako et al. 2018). The location of ground truth samples are selected for each class using a reference map, with the goal of comprehensively representing the land cover in the study area. Ground truth sample points are located in the field using Global Positioning System (GPS) units (Mubako et al. 2018). At a minimum, the LULC class at that pixel location is verified. The spectral signature of that pixel in the image is then used as a reference for the spectral signature of the class. In more detailed studies the actual spectral signatures of each class are measured in the field using a spectrometer. The ground truth sampling approach depends on the geographic scale, terrain, and time constraints. Each pixel in the image is then classified based on how its spectral signature corresponds to the reference signatures of the different land cover types (Rwanga and Ndambuki 2017). Different algorithms approach this comparison in different ways. In general, supervised classification algorithms attempt to minimize the error between the observed spectral signature in the pixel and the reference spectral signatures of the potential classes. Supervised classification results are typically reported as the percentage of the total study area occupied by each LULC class.

In contrast, unsupervised classification is a type of classification that doesn’t use prior knowledge of the study area because it is entirely data driven (Kantakumar and Neelamsetti 2015). Unsupervised classification finds data similarities across many different measurements and groups
data together into classes with similar measurements (Omran, Engelbrecht, and Salman 2005). The computer analyzes and groups the clusters based on their spectral similarity into their respective classes with an output of an unsupervised classification map. Unsupervised classification algorithms are iterative, with initial groups chosen at random and then progressively modified until little change in group affiliation occurs (“Unsupervised Classification” 2019). One advantage of using unsupervised classification is being able to find classification patterns that the user may not have otherwise noticed (Chang et al. 2020). However, despite the advantage of a fast classification with little to no prior knowledge of the study area, the spectral classes will not usually correspond with the known LULC classes resulting in classifications that are easily shown to be in error and in time being spent in a post-classification “cleanup” process to identify and label all spectral classes (Zhao, Yuan, and Wang 2019).

**Accuracy Assessment**

Once the LULC classification is completed, the final step is to conduct an accuracy assessment to quantitively assess the effectiveness of the method in correctly assigning the pixels to the proper land cover classes (Rwanga and Ndambuki 2017). Accuracy assessments are one of the most important steps of a classification because it validates the output classification product as well as the quality of the data itself. This is done by comparing the pixels of the classified image with ground truth data and is used in both supervised and unsupervised classification algorithms. Ground truth data are created using the same methods described in the preceding section. In the case of supervised classification, the full set of training data collected in the field is divided into two subsamples, one used for algorithm training and the other used for error testing so that the same sample is never used for both training and testing (Geiß et al., 2017). In the case of an unsupervised classification, this may be accomplished with a visual classification from a reference image or by field ground truthing. An error matrix is generated by comparing the LULC type calculated by the algorithm for a given pixel with the true LULC class identified by the ground
truth sample. For each class, the number of correct and incorrect calculated pixels are tabulated. The error matrix is a simple grid that lists the target classes and their respective number of correct and incorrect pixel classifications. Error as a percent of total number of pixels of each class is calculated. An overall classification error (kappa coefficient or similar statistic) is usually also calculated (Feng et al. 2018). Error matrices are useful for organizing and presenting accuracy assessment information in a descriptive and effective manner (Stehman 1997).

In addition, many studies report confusion matrices, which like error matrices, show the percentage of correctly classified pixels. In addition, confusion matrices show which pixels may have been misclassified as another class type, also known as the “producer accuracy” (Ahmad and Quegan 2013). Confusion matrices enable analysis of patterns of misclassification, for example, bare ground misclassified as pavement, or agricultural areas misclassified as forest. Using confusion matrices researchers can have a better idea of how accurate the overall classification is as well as common errors that may have occurred that can sometimes be corrected through incorporating additional information into the classification analysis.
CHAPTER 3: METHODS

Study Areas

Two separate classification problems were conducted in both ArcGIS and GEE. The first analyzed land cover change in the Middle Rio Grande (MRG) region of southern New Mexico, far west Texas, and northern Chihuahua, Mexico (Figure 1). The MRG study area was selected based on the extent of an ongoing research project funded by the US Department of Agriculture (see acknowledgements) that focused on urban and agricultural change in a water stressed desert region. LULC maps were needed as input to hydrological models analyzing the impact of climate change on water resources. The second study analyzed land cover change in southern Texas where Hurricane Harvey hit on August 17, 2017 and impacted a broad area including the city of Houston and its surrounding counties including Galveston, Harris, and Montgomery counties to name a few (Figure 2). Harris county, one of the affected areas, experienced “unprecedented rainfall totals,” accumulating approximately 52 inches of rainfall over the course of 3-4 days causing widespread flooding (Oldenborgh et al. 2017). In this study, work focused on post-hurricane damage assessment (PDA), especially with respect to downed trees blocking transportation routes. The study had the goal of developing an automated method for estimating the location and quantity of debris (especially trees) in order to calculate the funding and equipment needed to cleanup. The project was conducted in collaboration with the Texas A&M Forest Service and was funded by the Diana Natalicio Environmental Internship Program (see acknowledgements).
Figure 1. Map showing the study area of the MRG project.
Figure 2. Map showing areas impacted by Hurricane Harvey.
Middle Rio Grande LCLU Classification

In the MRG study, target land cover classes were selected using the National Land Cover Database (NLCD) which provides Landsat imagery nationwide on LULC change at a 30-m spatial resolution (Homer, Dewitz, and Danielson n.d.). In the NLCD data for this region, LCLU is comprised of 20 classes. In the study by Mubako et al. (2018) these were simplified into four classes: Water, Developed, Agriculture and other vegetation, and Upland mixed vegetation.

ArcGIS

SELECTING AND PREPROCESSING THE IMAGERY

Landsat data for five time periods between 1990-2015, one date for each year during the growing season from June through August, was acquired from the USGS Global Visualization Viewer (GloVis) which is an online database containing remote sensing data available to the public (“GloVis - Home” n.d.). Landsat imagery was selected if it fit the following criteria: cloud cover was less than 10 percent, having a 30 meter spatial resolution, and the date of acquisition was between the months of June and August. This resulted in 200-300 images for the analysis; 8 scenes were needed to cover the study area for all 25 years. Because GloVis has limited imagery, it was not always possible to find an image of a scene with less than 10% cloud cover. Some scenes would have one date with half of the scene covered in clouds. As a result of this, sometimes multiple images were needed from different dates in order to merge the areas with less than 10% cloud coverage. Following the extraction of the Landsat imagery, it was then imported into ArcGIS to be prepared for preliminary processing which included cropping the imagery to the study area boundaries, followed by creating mosaics which involves merging the imagery to create one continuous image, and color correction (Mubako et al. 2018). Following the preparation of the imagery, a classification method was picked based on if the user has prior knowledge of the area (Figure 3).
Figure 3. The flowchart on the left is how to prepare imagery for classification in ArcGIS. The flowchart on the right is how to pick which classification method to use in ArcGIS.

**UNSUPERVISED CLASSIFICATION**

Each Landsat image was then classified in ArcGIS using unsupervised classification approaches. An unsupervised classification was attempted using the Iterative Self-Organizing Data Analysis Technique (ISODATA) which is the only unsupervised classification available in
ArcGIS. ISODATA, also known as the “iso cluster algorithm”, computes the “minimum Euclidean distance when assigning each candidate cell to a cluster” (“How Iso Cluster Works—Help | ArcGIS for Desktop” n.d.). The algorithm runs over multiple iterations. Clusters are initially assigned random spectral centroids and each pixel is assigned to the nearest cluster centroid. In subsequent iterations, the new spectral mean of each cluster is calculated from the currently assigned pixels and each pixel is then reassigned to the nearest centroid. Iterations continue until the change between iterations is minimal (the algorithm converges on a solution) or a specified number of iterations occurs. Inputs to ISODATA include: the number of iterations of recalculating the Euclidean distance of each cell/pixel to a cluster; and the minimum class size which is the lowest number of cells/pixels that can be attributed to a cluster. Clusters that had fewer cells than the specified “minimum class size” would be eliminated and merged into the closest neighboring cluster. A unique feature of the ISODATA algorithm is that the number of final clusters is determined by the algorithm. Once the produced output map is created, each cluster is then reclassified into its corresponding LCLU class.

**Supervised Classification**

The supervised classification used was the Maximum Likelihood (ML) algorithm (Sathya and Abraham 2013) using thirty ground truth samples from each class. Following the classification, the classified maps were cleaned up using a process which involved ground checking using additional field samples and an accuracy assessment that compared ground truth classes with the automated classification results. Detailed methods and results from the supervised analysis into four classes in ArcGIS are reported in Mubako et al. (2018).
GEE

SELECTING AND PREPROCESSING THE IMAGERY

The analysis was replicated in GEE using Landsat data from the same years as the ArcGIS analysis, selected from the months of June to August, one date from each year. The Landsat data was acquired from the GEE data catalog (“Landsat Collections in Earth Engine | Earth Engine Data Catalog” n.d.). The selection criteria were similar to ArcGIS including the selected imagery dates. However, due to the Landsat data being acquired from Google’s vast database the imagery selected was able to be more precise bringing the cloud coverage from 10% to 0%. Some of the selected imagery were from the same dates as in the MRG ArcGIS analysis if it had 0% cloud data, but any imagery that was previously selected with 1-10% cloud coverage was discarded in favor of satellite imagery with 0% cloud cover. This was not possible previously due to the USGS Glovis Landsat imagery database having fewer images available to select from. However, one benefit of USGS GloVis is that you are able to specify what day you would like the imagery. In GEE you are able to set a range of dates you would like the imagery to be selected from, but you are not able to see the exact date the imagery is from unless specified using the filter and looking at the GEE console output. The Landsat imagery was imported into GEE as an image collection. Cloud cover was reduced during import by using a “rudimentary cloud scoring algorithm… which [scores] Landsat pixels by their relative cloudiness” with 0 being not cloudy to 100 being mostly cloudy (“Landsat Algorithms | Google Earth Engine” n.d.). By setting a cloud score threshold, in this case 0, any pixel scoring higher than the threshold will be masked and merged with another image from the same area that doesn’t have any clouds (Altaweel 2017). Following the import of Landsat imagery, we imported the study area boundary from ArcGIS by uploading that boundary to the cloud. The Landsat image collection was then clipped to fit inside the boundary in preparation for the classification (Figure 4).
**Figure 4.** Flowchart showing how satellite imagery is prepared in GEE.

**UNSUPERVISED CLASSIFICATION**

The next step was classification, including both supervised and unsupervised classification, since both approaches could easily be conducted on GEE. The unsupervised classification generated a cluster dataset using the k-means algorithm which is based on determining the number
of clusters as well as the max number of iterations by defining the initial centroid value which will define the performance of the classification (Syakur et al. 2018). In contrast with the ISODATA algorithm, the main process of the k-means clustering begins by the user setting the target number of clusters. The user also specifies the number of iterations the algorithm will run in order to group all the cells of a study area into a cluster. k-means allows the user to specify a neighborhood size, which is the amount of spatial overlap the clusters can have which works in tandem with grid size. Grid size limits the size of clusters to be the specified size or lower. The result was then displayed with a random visualizer that assigned a random color to each unique class. In order for the output classification to correspond with the MRG classification classes and colors, the map’s visual parameters and color palette were changed. In the GEE analysis, ground truth data was obtained by visually selecting 50-150 samples from each cluster from the classified imagery using a random stratified sample algorithm. The stratified random sample algorithm extracts a “specified number of samples for each distinct value discovered within the ‘classBand’” (“Ee.Image.StratifiedSample | Google Earth Engine” n.d.). However, the GEE k-means unsupervised classification method resulted in similar misclassifications as the ArcGIS ISODATA results and would have required substantial post-classification cleanup, therefore a supervised classification approach was taken.

**Supervised Classification**

The Landsat imagery was classified using the Statistical Machine Intelligence and Learning Engine (SMILE) and Classification and Regression Tree (CART) algorithms, simplified as the “smileCart” combination in GEE. SMILE is a recently developed classification method that highlights the importance of features while requiring a balanced distribution among the reference data to ensure a higher overall classification accuracy (Qu et al. 2021). The color palette for each of the classes was set so the output displayed the proper color for each LULC class as done in the ArcGIS classification. In GEE, the supervised classification output map defaults to encompass the
entire globe so each classified map was clipped to the study area boundary. Following the classification, GEE validated the classification results using a confusion matrix.

South Texas PDA Classification

SELECTING AND PREPROCESSING THE IMAGERY FOR ArcGIS

The PDA required satellite imagery of the affected area before and after the hurricane. Initially, 30 m Landsat data from the USGS Glovis website were tested for this analysis, however the spatial resolution was too coarse to identify individual trees. Therefore, very high resolution (1m) USDA National Agriculture Imagery Program (NAIP) imagery provided by the Texas Forest Service (TFS) was tested. This dataset was collected over the study region within two weeks prior and after the hurricane made landfall. This imagery was imported into ArcGIS and a LCLU change analysis was attempted in ArcGIS.

TREE DETECTION PROCESS IN ARCGIS

An edge-detection algorithm was developed to pick out structures (buildings, etc.) and trees as well as other types of vegetation. The next step in the analysis was to create a color filter implemented with a line detection algorithm that would differentiate damaged trees from ones that were completely torn apart/downed and needed to be salvaged. However, due to the extremely large data volume the computer was unable to handle the algorithm processing. Despite the machine specifications being well above the recommended level for ArcGIS, the processing was taking several hours before ultimately freezing and crashing. Hence, the LULC analysis was transferred to GEE.
SELECTING AND PREPROCESSING THE IMAGERY FOR GEE

In GEE, the NAIP imagery had to be imported from the GEE database as an image collection, filtering only the dates needed. Hurricane Harvey affected the region between 8/17/17-9/2/17, therefore, the filter was set for the months of August and September to incorporate before, during, and after the hurricane hit (Figure 5).

Figure 5. Flowchart of the preparation of imagery before conducting the tree detection process in GEE.
**Tree Detection Process in GEE**

Once the data was collected, a tree detection process needed to take place to differentiate urban developed land from trees. To start, an edge detection was applied using the Canny edge detection algorithm which “uses four separate filters to identify diagonal, vertical, and horizontal edges” (“Edge Detection | Google Earth Engine” n.d.). In doing so we were able to identify most buildings and other types of urban development in order to eliminate them from the image to prevent complicating further analysis. The next step was to use color filtering to further distinguish trees and other types of vegetation. While GEE does not have a color filtering algorithm, we set a visualization parameter on the imagery and changed the band combination to 5, 4, 3 in order to create a false color composite that highlighted vegetation such as trees (“Image Visualization | Google Earth Engine” n.d.). A tree cover detection algorithm was used from the Hansen Global Forest Change to detect the canopies of trees that were still upright as well as those that may have fallen (Hansen et al. 2013). We conducted this process on days before and after the hurricane hit using the updateMask algorithm in GEE to obtain a “loss image” of sorts where we were able to compare the differences in trees that were either damaged or completely uprooted as a result of the hurricane (“Compositing, Masking, and Mosaicking | Google Earth Engine” n.d.) (Figure 6).
Comparison Across ArcGIS and GEE Approaches

Both ArcGIS and GEE were used in the MRG and PDA projects using various combinations of supervised and unsupervised classification methods. Despite using similar
methods in both platforms there were different processes to setup the classification as well as varying output results. To compare the processes across platforms, the following steps were taken:

First, we compared the ease of access to the platform. This includes the usability of the platform, the time and effort needed to setup/install anything prior to use, any costs needed to acquire the software, as well as learning how to use it. This is important because the easier it is to get access, the faster the user can begin learning. Or, in the case of organizations, the easier it is to transition into a larger group of people into using the platform to work on a larger research project. Calculating the “true cost… may be difficult to quantify” (Goldberg et al. 2013), but some aspects are a bit easier to quantify such as the cost of a license for software, which may be a one-time payment or yearly subscription at a fixed price. However, other costs are more difficult to calculate such as the cost of training, maintenance for the platform, or the equipment needed to use the platform. In terms of training needed to learn the platforms, this comparison included all methods of learning such as classes, video guides, and other self-taught methods including engagement on community forums, which depends on how active each platform’s community is.

Second, we compared the workflow of the projects beginning with acquiring the input data as well as ground truth reference data, both required to perform LULC classifications and accuracy assessment. The comparison included the types of reference data available on each platform as well as the ability to acquire data from outside sources and import data into the platform. In addition, the comparison included the temporal characteristics of the data available, especially with respect to very recent data and how often the databases are updated with new and up to date imagery.

The third comparison between the two platforms included the performance as well as the flexibility of the platforms. Performance refers to the “functionality, service response time, and accuracy” (Garg, Versteeg, and Buyya 2011) of the LULC analysis performed. Flexibility refers to how easily the user can “make changes and additions to the data sources and methods used by the system” (Goldberg et al. 2013) as well as the agility of the platform itself, including how adaptable it is in regards to the portability of the platform as well as any data associated with it.
The flexibility of the platform includes if there are any alternative software or product services compatible with the ones being compared.
CHAPTER 4: RESULTS

LULC Results

GEE CLASSIFICATION MAPS

Despite using similar methodologies in GEE, the output maps varied in display and accuracy. An unsupervised classification was started in GEE in order to save time from having to import and recreate the training samples in GEE. However, examination of the results indicated there were many misclassifications. The unsupervised classification was able to correctly identify the major urban areas of the region including Las Cruces, El Paso, and Ciudad Juarez. However, the unsupervised classification misidentified the bottom right region of the study area as a large developed area as well (Figure 7). This may have been due to similarities of spectral signature that unsupervised classifications are unable to differentiate without ground truthing training samples. Due to the amount of time it would take to conduct the post-classification clean up processing, it was more efficient to conduct a supervised classification. A supervised classification was then conducted in which the urban areas were correctly identified as well as a reduced amount of misclassifications. An error matrix was conducted on the supervised classification which produced an overall accuracy of 98% and a kappa coefficient of 97.5%.
To quantify the land use changes over the years we analyzed the error matrix which showed any classification errors that may have occurred such as a pixel being misclassified. The four classes that were used (Agriculture, Mixed vegetation, Water, and Developed) resulted in over a 90% accuracy for the Agriculture and Mixed Vegetation classes. However, accuracy was somewhat lower in the other two classes, with 87% and 80% accuracy for Developed and Water classes, respectively (Figure 7). The reason for the low accuracy of water was due to the small area comprised of water and the low number of water training samples (5), resulting in one misclassified pixel reducing a large percentage of accuracy (20% reduction).
Ease of Use Comparison

As scientific problems arise, researchers need to adapt to using new software/platforms that can streamline data analysis in a timely manner which is why ease of access is an important thing to consider.

COST/HARDWARE REQUIREMENTS

In terms of accessing the software, ArcGIS requires a license and has various subscription options available depending on what content and capabilities the user(s) would like access to (Table 1). For individuals ArcGIS has a “Personal Use” as well as a “Student Use” license available for $100 per year (“ArcGIS Online Pricing for Teams & Individuals | Subscription Cost” n.d.). Both licenses provide ArcGIS Desktop which includes the ArcMap software as well as many popular and commonly utilized extensions used in creating LULC maps such as Image Analyst and Geospatial Analyst among other features for data storage and analysis.

As a software, ArcGIS Desktop needs to be installed onto a Windows computer as well as any other features or extensions that come with the license up to the user’s discretion. It can be run on a Macintosh only if the Mac is partitioned with the Windows OS installed or is set up with a Windows virtual environment. ArcGIS has some hardware requirements that may not be suitable
for lower end computers in order to run the software. This includes having a CPU with 2.2 GHz minimum although a multi-core processor is recommended, while 4 GB of memory/RAM is considered the minimum. 8 GB or higher is recommended. ArcMap only takes up 4 GB to install although more storage is recommended to save files. Lastly, the video/graphics needed at a minimum is 64 MB while 256 MB and higher are recommended (“ArcGIS Desktop 10.8.x System Requirements—ArcMap | Documentation” n.d.). Overall, ArcGIS hardware requirements are dependent on the type of processes and analysis that will be run. If conducting basic mapping the minimum requirements will suffice. However, if conducting LULC geoprocessing, more CPU and RAM will be needed such as an i5-i7 processor or the equivalent with a preference towards an i7 as well as 8-16 GB RAM (“Computer Recommendations for GIS Students and Professionals” 2018). The reason for having a higher tier processor is that it comes with more cores that will divide the workload rather than over working fewer processors, potentially causing the software to crash or the computer to overheat. Higher memory/RAM allows for better multitasking and may speed up computation processing saving time for larger datasets or study areas.

To access GEE, rather than paying for a license, a form needs to be filled out at: https://signup.earthengine.google.com/ (Table 1). Depending on when the developers review the application form, the acceptance email will be sent between a couple of hours or a few days titled “Welcome to Google Earth Engine” with instructions on how to get started (“FAQ – Google Earth Engine” n.d.). While GEE is free for research purposes, education, and nonprofit usage, it cannot be used for commercial purposes where any data products generated by GEE are sold unless the user/organization has applied for the commercial license program. To apply for a paid commercial license, users will have to contact earthengine-commercial@google.com.

In terms of hardware requirements, GEE does not have high hardware requirements due to the fact that most, if not all processing is done on the cloud and Google’s servers. Because all the processes and algorithms of GEE are run on Google’s servers there is also a significantly lower time to complete various types of analyses. GEE is a server side cloud platform and as a result does not have any specific hardware requirements. However, there are common hardware
requirements to use a server side platforms that most computers including those on the lower end should be able to run with relative ease (“Hardware and Software Requirements - Pleasant Solutions” n.d.). These include having a Dual – Core 1.6 Ghz CPU at a minimum although higher is recommended. For RAM 2 GB is needed although more RAM allows users to have multiple windows/tabs open that they can use to look at API/tutorials while writing code. Lastly, there are no video requirements as it is not applicable to GEE’s cloud platform. One thing to note is that any processes done through GEE will not take up any physical storage space when exported to the Google’s cloud storage. However, once the data is downloaded from the cloud it will need a physical storage device such as a USB or external hard drive.

<table>
<thead>
<tr>
<th>Table 1. Comparison of Hardware Requirements.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons</td>
</tr>
<tr>
<td>Cost</td>
</tr>
<tr>
<td>Platforms Available on</td>
</tr>
<tr>
<td>CPU (Minimum/Recommended)</td>
</tr>
<tr>
<td>RAM (Minimum/Recommended)</td>
</tr>
<tr>
<td>GPU (Minimum/Recommended)</td>
</tr>
</tbody>
</table>
TRAINING NEEDED INCLUDING ALL METHODS OF LEARNING

There are various ways to learn a software. These include through an in-person class, workshop, online tutorial videos, guides, or even an active community that may help teach a new user. ArcGIS has a lesson gallery on their main website which you can filter to suit your needs such as what capabilities you would like to learn, what product you are using, as well as other filters (“Lesson Gallery | Learn ArcGIS” n.d.). Filtering by ArcMap there is a “Get started with ArcMap” lesson that teaches the basics in 3 hrs. After learning the basics of the software, there is another lesson called “Introduction to Imagery and Remote Sensing” which can be completed in 18 hrs. There are three other lessons that would aid in learning about classification methods beginning with “Get started with imagery” which explores various types of Landsat imagery around the world in just 30 minutes. Following that, “Extracting information using image classification” completable in approximately 6 hrs and “Distributed analytics for imagery” that teaches about raster, image collections, and distributed processing completable in approximately 2 hrs. There are many other lessons that ArcGIS offers online. Combining all of the lessons stated will take approximately 30 hrs to learn the basics of ArcMap, the basics of imagery, remote sensing, and classification.

There are many universities that offer classes in ArcGIS with some having degrees/certificates dedicated to the craft. In one semester a new user can learn the basics of ArcGIS including simple imagery analysis and classification methods. However, in order to learn some of the more advanced techniques multiple classes may need to be taken. ArcGIS also has an active “global community of ESRI users where you can find solutions, share ideas, and collaborate to solve problems with GIS” (“Esri Community | GIS Professional Community” n.d.). Currently there are over 190 thousand posts with 37 thousand solutions available on the community, meaning any common problem new users may encounter are likely to have been solved previously in the community.

As a result of GEE being a newer platform there are fewer methods of learning the cloud computing platform as compared to ArcGIS. GEE offers many “self-paced tutorials” which serve
as an introduction to the JavaScript API that is used while assuming the new user has no background in programming (“Tutorials | Google Earth Engine” n.d.). Following that, GEE offers some video tutorials meant to be completed after the self-paced ones that will further introduce users to the platform as well as being able to perform some basic geospatial analysis. These tutorials include “Introduction to Earth Engine (condensed),” “Tables and Vectors,” “Importing and Exporting,” “Classification,” “Machine Learning,” among many others that will take approximately 19 hrs to view them all. Although there are courses at some universities that incorporate GEE into their content, these are currently not common.

There are other sources of “various tutorials and documentation… available for using [GEE] including more than “20 video tutorials with corresponding notebook examples… [as well as over] 360 Jupyter notebook examples for using [GEE]” available at the following links: https://github.com/giswqs/geemap/tree/master/examples for the video tutorials and the notebook examples at: https://github.com/giswqs/earthengine-py-notebooks (Wu 2020).

Workflow Comparison

ACQUIRING THE SATELLITE IMAGERY

The first step of either method is acquiring the ArcGIS software or access to the GEE cloud servers. Following that, satellite imagery is needed in order to conduct most studies including any types of classification. Each platform has different ways to gather satellite imagery whether it’s on the platform itself or an external source that needs to be imported.

TYPES OF REFERENCE DATA AND TEMPORAL CHARACTERISTICS

Depending on the type of study, different satellite imagery may be needed. Satellite imagery varies in their “spatial and spectral resolution, geographic and temporal coverage, cloud cover, security regulations, and price” (Boyle et al. 2014). While high resolution (usually less than 1m) tends to “outperform lower-resolution imagery” there are some complications that arise such
as more shadows coming from canopies as well as “[complications during] multi-image comparisons and processing” (Dennison, Brunelle, and Carter 2010). High resolution imagery is generally more difficult and expensive to acquire leading to lower resolution imagery being used more often for LULC with large study areas. Classification of land cover “at a high spatial resolution (1–5 m) over large areas can be challenging due to large data volumes, computational load, processing time, complexity of developing training and validation datasets, data availability, and heterogeneity in data and landscape conditions” (Maxwell et al. 2019). As a result of higher resolution not necessarily being better when conducting land cover projects, Landsat Thematic Mapper imagery is often used when “mapping global land cover with 30m resolution” (Gong et al. 2013). For more precise classification or processing involving vegetation and trees other types of satellite imagery are more effective such as the NAIP imagery used in the PDA analysis. NAIP imagery has a resolution of “0.5 to 1m spatial resolution with four spectral bands: red, green, blue, and near infrared (NIR),” the cloud cover is also said to be less than 10% but is generally much lower (Maxwell et al. 2019). However, while the NAIP 1m resolution imagery was more useful in distinguishing the trees in the PDA analysis, NAIP imagery is collected far less often than its Landsat counterpart and generally only during growing seasons. NAIP imagery is also only available to the public after 30 days of processing. While the government may be able to acquire the imagery sooner, the temporal periods of when it and other high resolution imagery are acquired are not consistent with applications that require rapid change analysis across short timespans such as after a natural disaster.

**Arcgis Imagery Acquisition**

Arcgis contains basemaps which is often used as a reference map that “overlays data from layers and [helps users] visualize geographic information” (“Basemaps—ArcGIS Pro | Documentation” n.d.). While basemaps are useful for reference when creating maps there are not many available as well as limited options on when the imagery was taken and the temporal
characteristics. ArcGIS does not have a database containing satellite imagery leading users to seek outside sources. There are many sources for satellite imagery both free and paid. The following is a list of some of the free sources of “high-quality satellite imagery”: USGS Earth Explorer (also known as USGS Glovis), Sentinel Open Access Hub, NASA Earthdata Search, NOAA Data Access Viewer, DigitalGlobe Open Data Program, Geo-Airbus Defense, NASA Worldview, NOAA Class, National Institute for Space Research (INPE), Bhuvan Indian Geo-Platform of ISRO, JAXA’s Global ALOS 3D World, VITO Vision, NOAA Digital Coast, Satellite Land Cover, UNAVCO (“15 Free Satellite Imagery Data Sources” 2016). Another source is ESRI’s online remote sensing data collection which includes various types of satellite imagery in the form of “maps, apps, and ready-to-use data layers” (“Imagery Content | Access the Largest Online Remote Sensing Data Collection” n.d.). Once the satellite imagery is acquired from an outside source it will need to be imported into ArcGIS using the “Add Data” feature.

**GEE IMAGERY ACQUISITION**

Google Earth Engine has a “public data archive [which] includes more than forty years of historical imagery and scientific datasets, updated and expanded daily” (“Earth Engine Data Catalog” n.d.). GEE’s vast database includes various types of satellite imagery including Landsat, Sentinel, MODIS, and NAIP. These vast collections of imagery are known as “ImageCollection[s]” which can be loaded into the code by copy/pasting the GEE asset ID assigned to the corresponding ImageCollection found in the data catalog (“ImageCollection Overview | Google Earth Engine” n.d.). To import imagery from an outside source the “image or other georeferenced raster datasets [will need to be uploaded] in GEOTIFF or TFRecord format” (“Importing Raster Data | Google Earth Engine” n.d.). These images can be uploaded into GEE’s asset manager that can be used in any codes using the corresponding folder or collection ID such as in the following example: “/users/name/folder-or-collection-id/new-asset”.

30
**Process Once Imagery is Acquired**

Once the imagery has been acquired, the user can then begin learning how to use both platforms beginning with the basics and later learning more advanced forms of analysis such as classification methods that were used for the MRG and PDA projects. Once the platforms have been acquired and learned, at a high level the classification workflows are similar in both ArcGIS and GEE. The projects begin with acquiring the imagery of the study areas. If the imagery is acquired from an outside source from the platform itself, it needs to be exported from the source and imported into the platform. Following that, the imagery may need to be “cleaned up” by being “preprocessed through several procedures, namely, atmospheric correction, image mosaicking, geometric correction, and image subsetting” (Liu and Yang 2015). This process needs to be done for imagery from each year (1990-2015). While users could repeat the process for each year manually it would take a lot of time. Most platforms have a process that will ‘loop’ and repeat the same processes automatically on each image. ArcGIS allows looping using ModelBuilder and Python scripting. ModelBuilder allows users to create iterations “often referred to as looping or batch processing, means to repeat a process over and over with some degree of automation” (“Iterators—ArcGIS Pro | Documentation” n.d.). GEE discouraged using loops because “the same results can be achieved using a ‘map()’ operation where you specify a function that can be independently applied to each element” (“Functional Programming Concepts | Google Earth Engine” n.d.). Despite the discouragement of looping, GEE allows the creation of an iteration for an image collection allowing users to collect imagery from multiple years without needing to rewrite the code to filter each year individually. However, the iteration function is limited on what it can perform, in particular, “it can’t modify variables outside the function; it can’t print anything; it can’t use… ‘if’ or ‘for’ statements” and as a result atmospheric correction and mosaicking need to be coded individually for each year (“Iterating over an ImageCollection | Google Earth Engine” n.d.). After all the preprocessing of the imagery is completed, then depending on the type of classification method being used, ”a sufficient number of training samples are prerequisites for a successful classification” (D. Lu and Weng 2007). Once all the training samples have been
acquired from target classes the image classification can begin. Once the classification is completed it will need to go through a “post-classification correction” in which the “misclassified pixels… [are] re-evaluated and correctly reclassified” (Manandhar, Odeh, and Ancev 2009). To assist the post classification cleanup, ground truthing or an accuracy assessment may also be performed in order to ensure that the post cleanup of any reclassified pixels improved the overall accuracy of the output LULC map.

ABILITY TO EXPORT OUTPUT DATA

Lastly, the LULC map or data may need to be exported/extracted from the platform into another software or wherever the resulting output will be used. ArcGIS offers several methods for exporting data including “using the context sensitive menu in the Catalog tree, using geoprocessing tools, using the Extract Data Wizard in ArcMap, and using the Export Data command in ArcMap” (“A Quick Tour of Exporting Data—ArcMap | Documentation” n.d.). However, one thing that is important to note no matter the data type is that any dependent data is exported as well. Exporting feature datasets, classes, and tables can be done as an XML or ZIP file. Exporting into an XML file is an uncompressed text tile while ZIP files are compressed leading to saved space and easier transfer of data, ZIP files should be used if the data exceeds 4GB.

GEE allows the exporting of “images, map tiles, tables, and video… sent to your Google Drive account, Google Cloud Storage, or a new [GEE] asset” (“Exporting Data | Google Earth Engine” n.d.). Images and tables exported from GEE need to be run through a slightly differing code depending on the export method (Table 2). Video and maps can be exported in a similar way. However, video cannot be exported to a GEE asset folder and maps can only be exported through Google Cloud Storage.
Table 2. Code to export images and tables from GEE

<table>
<thead>
<tr>
<th>Where the data will be exported to</th>
<th>Export Image Code</th>
<th>Export Table code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Drive</td>
<td>Export.image.toDrive()</td>
<td>Export.table.toDrive()</td>
</tr>
<tr>
<td>Google Cloud Storage</td>
<td>Export.image.toCloudStorage()</td>
<td>Export.table.toCloudStorage()</td>
</tr>
<tr>
<td>GEE Asset Folder</td>
<td>Export.image.toAsset()</td>
<td>Export.table.toAsset()</td>
</tr>
</tbody>
</table>

Performance and Flexibility Comparison

The last comparison between ArcGIS and GEE is performance and flexibility. While referring to flexibility we will also be mentioning other alternatives to both platforms.

Performance

In terms of performance, we will be referring to the functionality, service response time, as well as accuracy of the analysis that were performed in both platforms. Both platforms have useful features that help with geospatial analysis and creating LULC maps. Functionality-wise each platform is different. Because ArcGIS is a software its performance is dependent on the hardware of the computer it is running on. With limited RAM ArcGIS is not able to multi-task many processes before freezing up and eventually crashing the software. GEE on the other hand is a cloud platform and is not held back by hardware limitations. One of the prime examples of these limitations are when the study area for the MRG was expanded (Figure 9). ArcGIS was able to perform classifications of the original study area with little to no issues. However, once the study area was expanded, ArcGIS started having performance issues with the hardware as stated previously. The classification maps produced would vary in processing time varying from a few hours up to 4 days in ArcGIS, GEE on the other hand could classify the same study area in under 30 minutes. However, it only works with an internet connection. ArcGIS as a result be better if
working in an environment where internet is not available or as reliable. GEE is a better option if the user has a good internet connection but is limited by hardware specifications. When processing analyses ArcGIS varies in the time taken to complete ranging from a few minutes/hours if conducting a smaller analysis to even days if conducting large processes such as a LULC classifications. GEE takes less time to conduct the same process/analysis because it goes through the Google servers and takes at most 30 minutes before sending an error/fail message. After conducting the classification, both platforms are able to conduct accuracy assessments using confusion matrices.
Figure 9. Map showing the original and expanded study area boundaries.
**Flexibility**

Flexibility as referred to previously is how easily changes can be made to the data and methodology as well as the portability of the platform and data. ArcGIS has an editing tool that allows users to “edit feature data stored in shapefiles and geodatabases, as well as various tabular formats” (“What Is Editing?—Help | ArcGIS for Desktop” n.d.). However, certain features may require additional licensing to edit. For GEE, the JavaScript editor formats and highlights code as you type while underlining anything that might create problems. GEE doesn’t have an editing tool, however, editing is as simple as rewriting code on features or selecting any shapefiles creating and being able to move them around, resize, as well as deleting them.

For portability, ArcGIS can work on laptops/tablets in offline mode and can theoretically do any process/analysis as long as the hardware of the device can handle it such as creating maps. However, for more heavy-duty tasks a laptop or tablet may not have the minimum hardware specifications resulting in more frequent freezing and crashing of the software. For mapping purposes there are also other software that are more ‘mobile friendly’ such as QGIS which is one of the open source alternatives to ArcGIS. However, if working with large amounts of data a USB or larger external hard drive may need to be carried with the device as well. GEE is a cloud based platform meaning the only limitation to portability is internet service. GEE can be run on the majority of devices ranging from computers, laptops, and tablets and with any operating system. One of the benefits of the cloud is there is no need to carry any additional storage since all of the data is on Google’s cloud servers. However, users will not be able to access their code or data if they cannot connect to the internet.

**Alternatives to Platforms**

The last part of flexibility is the availability of alternative software or product services whether paid or open source as well as their compatibility (if any) with ArcGIS and GEE. There are many reasons why users might look for alternatives to prominent software in their respective field of study. One such reason may be that the licensing is too expensive. Another may be that
the software/platform was designed for a different operating system. ArcGIS for example was designed for Windows meaning that Linux and Mac are unable to run the software without either creating a “virtual environment” or installing a “Boot Camp Assistant” to run a Windows partition on Mac (“Run ArcGIS Pro on a Mac—ArcGIS Pro | Documentation” n.d.).

**ALTERNATIVES TO ArcGIS**

Some alternatives of ArcGIS include Geographic Resources Analysis Support System (GRASS), uDig, or a well-known/used alternative known as QGIS. GRASS has been developed and used by the US government since 1982 and continues to be used today by the academic community. Because GRASS was created so long ago the interface is a terminal window with most of the code written in C/C++, however, more recent modules have been written in Python. GRASS uses Python commands for data analysis and geoprocessing and can be used as an extension for many applications. uDig otherwise known as User-friendly Desktop Internet GIS is a stand-alone program which works like GIS in terms of the UI and framework. It was developed to make your own GIS app that meets a user’s needs directly. Lastly, QGIS is a good alternative open source to ArcGIS that supports raster/vector formats as well as containing various plugins that add more functionality including a Python interface. QGIS is a great mobile version of ArcGIS that allows for mapping in the field on a tablet as well as editing the features later.

**ALTERNATIVES TO GEE**

To find alternatives to GEE we will be looking at software/platforms that allow storage of geographic data in a cloud environment, analyzing data, creating maps, while being able to edit and render data quickly and efficiently. One alternative is similar to ArcMAP, ArcGISOnline, which is essentially a cloud based version of ArcMAP being able to create and render maps while visualizing each change in analysis made instantly. CartoDB is a cloud platform with a spatial database that has a simple interface to import data, create a map, and how to best visualize it.
CartoDB also contains dynamic rendering allowing visualization to be seen as data is updated. Lastly, iSpatial is a cloud platform that also offers map creation, visualization, storage, and analysis. The platform also uses Google Earth and Google Maps to solve complex problems.

**Classification Comparison**

Despite different platforms and methodology, both ArcGIS and GEE were able to produce a similar LULC map. However, there were slight disparities in regards to the final accuracy assessment. For unsupervised classification ArcGIS used the ISODATA algorithm while GEE used the k-means algorithm. In both cases, while a map was produced from the unsupervised classification an accuracy assessment was not performed because visual checks indicated there were many errors. For supervised classification ArcGIS used the ML classification methodology while GEE used a combination of the SMILE and CART algorithms. An accuracy assessment was performed on the two output supervised classification maps of both platforms. ArcGIS produced an error matrix of 93% overall accuracy and a kappa coefficient of 89%. GEE produced an error matrix of 98% and kappa coefficient of 97.5% (Table 3).

<table>
<thead>
<tr>
<th>TABLE 3. ALGORITHM AND ACCURACY ASSESSMENT COMPARISON BETWEEN ARCGIS AND GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Unsupervised Algorithm</td>
</tr>
<tr>
<td>Supervised Algorithm</td>
</tr>
<tr>
<td>Overall Accuracy</td>
</tr>
<tr>
<td>Kappa Coefficient</td>
</tr>
</tbody>
</table>
CHAPTER 5: CONCLUSION

The study used similar methods of land cover classification to compare traditional methods in ArcGIS software with newer platforms on cloud servers such as GEE. While most users currently use traditional classification methods, new platforms are being created improving on geospatial analysis whether it be ease of access, the workflow, performance, flexibility, and final accuracy results. LULC classifications were performed on both ArcGIS and GEE as well as a PDA with each platform having pros and cons. In terms of ease of use, GEE is free, easier to acquire, and has lower hardware minimum requirements to run as a result of being a cloud platform. However, ArcGIS has a more active community, online tutorials, as well as classes available in universities because it is considered a more traditional software to use for geospatial analysis and classifications, meaning it is easier to train new users to learn. For workflow both platforms had similar analysis processes with only slight variations to how data was acquired as well as the classification algorithms used. One of the key differences was exporting methods and how ArcGIS needed an external source (USB or external hard drive) to transport data between computers while GEE was easier to move data given everything was on the cloud on Google’s servers. In regard to portability, both platforms have strengths and weaknesses. ArcGIS can work offline whereas GEE cannot. GEE however is more portable with the fact that everything is on the cloud.

Overall, each platform has strengths and weaknesses. There is no better platform outright as each has advantages based on the user’s equipment, prior experience, project size, and time constraints. If the user does not have the money to buy a better computer or licensing, GEE or open source are great alternatives to ArcGIS. However, while GEE has more potential, there is a much steeper learning curve. There are more ways to learn ArcGIS including classes at universities, online tutorials, and a more active community to help with any problems. When comparing workflows, GEE has an easier way of acquiring satellite imagery with less cloud coverage as well. Both ArcGIS and GEE used similar classification methods with different
algorithms. However, after an accuracy assessment was conducted on the final LULC map, GEE was found to be more accurate overall by approximately 5%. Exporting the data was easier on ArcGIS as it could be done using a ZIP file. However, GEE had the advantage of being on the cloud and therefore easier to transport data albeit requiring an internet connection and an extra step to extract the data from the cloud storage. Users trying to make a decision on which platform is better for their purpose will need to look at their overall standing in terms of equipment, prior experience, and project size, as well as time constraints. For users working on a project with a large study area and not a high end computer, GEE would be a better choice. However, if users are working on a smaller study area, do not have a lot of prior experience coding, nor the time to learn, ArcGIS may be more suitable as there is not as steep of a learning curve as well as the tools/analysis available being easier to find and conduct rather than manually writing the code.
REFERENCES


https://learn.arcgis.com/en/gallery/#.


VITA

Carlos S. Reyes attended the University of Texas at El Paso in 2014 where he received an internship at the Center for Environmental Resource Management (CERM) that led him to become an undergraduate research assistant. During his time as a research assistant for CERM he was able to participate in a large project that resulted in his first publication as a co-author titled: “Monitoring of Land Use/Land-Cover Changes in the Arid Transboundary Middle Rio Grande Basin Using Remote Sensing.” He completed his Bachelor’s in Environmental Science with a concentration in Geology in Fall 2017. The following semester he enrolled in and completed his Geospatial Information Science and Technology (GIST) Certificate in Spring 2018. He then enrolled in graduate school in Fall 2018. The summer of 2019 he was able to get an internship at the Texas A&M Forest Service funded through the Diana Natalicio Environmental Internship program and worked on a project that later became a part of his thesis. During his time in graduate school, he became a graduate research assistant as well as a teaching assistant. He received his Master of Science in Environmental Science in Spring 2021

Contact Information: carlossreyes1@gmail.com