




2021

## LANDSLIDE SITE ASSESSMENT AND CHARACTERIZATION USING REMOTE SENSING TECHNIQUES

Batmyagmar Dashbold

University of Kentucky, myagmar85@yahoo.com

Author ORCID Identifier:

 <https://orcid.org/0000-0002-1197-1668>

Digital Object Identifier: <https://doi.org/10.13023/etd.2021.369>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

### Recommended Citation

Dashbold, Batmyagmar, "LANDSLIDE SITE ASSESSMENT AND CHARACTERIZATION USING REMOTE SENSING TECHNIQUES" (2021). *Theses and Dissertations--Civil Engineering*. 113.  
[https://uknowledge.uky.edu/ce\\_etds/113](https://uknowledge.uky.edu/ce_etds/113)

This Master's Thesis is brought to you for free and open access by the Civil Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Civil Engineering by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Batmyagmar Dashbold, Student

Dr. Lindsey Sebastian Bryson, Major Professor

Dr. Lindsey Sebastian Bryson, Director of Graduate Studies

LANDSLIDE SITE ASSESSMENT AND CHARACTERIZATION USING  
REMOTE SENSING TECHNIQUES

---

THESIS

---

A thesis submitted in partial fulfillment of the  
requirements for the degree of Master of Science in  
Civil Engineering in the College of Engineering at the  
University of Kentucky

By

Batmyagmar Dashbold

Lexington, Kentucky

Director: Dr. L. Sebastian Bryson, Professor of Civil Engineering

Lexington, Kentucky

2021

Copyright © Batmyagmar Dashbold 2021

<https://orcid.org/0000-0002-1197-1668>

## ABSTRACT OF THESIS

### LANDSLIDE SITE ASSESSMENT AND CHARACTERIZATION USING REMOTE SENSING TECHNIQUES

Landslides are common and dangerous natural hazards that occur worldwide, often causing severe direct impacts on human lives, public and private properties. It is imperative to identify the landslide susceptible areas to avoid or mitigate the possible damage. Landslide prediction can be presented in a slope failure in spatial and/ or temporal terms. If it is presented in spatial term, it is considered a landslide susceptibility map (LSM) defined as the probability of spatial occurrence of slope failures. If it is presented in a combination of spatial and temporal distribution of the landslide susceptibility, it is commonly referred to as landslide hazard map (LHM). This document presents generation and comparison of LHM, and LSM using a remote sensing data. In addition, this paper shows the workflow of using multi-temporal UAV images to detect land movement and estimate soil moisture.

**KEYWORDS:** Landslide hazard, landslide susceptibility, UAV, COSI-Corr, Logistic Regression model, Remote Sensing.

---

Batmyagmar Dashbold

---

July 13, 2021

---

LANDSLIDE SITE ASSESSMENT AND CHARACTERIZATION USING REMOTE  
SENSING TECHNIQUES

By

Batmyagmar Dashbold

Dr. Lindsey Sebastian Bryson  

---

Director of Thesis and

Dr. Lindsey Sebastian Bryson  

---

Director of Graduate Studies

July 14, 2021  

---

## Dedication

I would like to dedicate this work to my father, Dashbold Yondon, my mother, Lkhamsuren Tsoodol, and my late grandmother Dolgorjav Sambuu who showed me what love and kindness is.

## ACKNOWLEDGEMENTS

I would like to thank my family for believing in me, and making me feel anything is possible. Their unconditional love and support keeps me moving forward. I would like to extend my gratitude to my adviser and mentor Dr. Sebastian Bryson for guiding me throughout the process, pushing me to be a better engineer and a better person, making me expand my knowledge and tap into my talent, and helping me see what I didn't see in myself. I would like to thank the faculty and staff in the Civil Engineering Department at the University of Kentucky for supporting me throughout the process. Last but not least, I would like to thank Dr. Michael Kalinski for teaching us advanced geotechnical courses and preparing us for the real world challenges.

## TABLE OF CONTENTS

Acknowledgements.....	iii
List of Tables .....	vi
List of Figures .....	vii
1 Introduction.....	1
1.1 Problem Synopsis.....	1
1.2 Research Objectives .....	2
1.3 Relevance of Research.....	2
1.4 Contents of Thesis.....	3
2 Landslide Hazard and Susceptibility Maps Derived from Satellite and Remote Sensing Data Using Limit Equilibrium Analysis and Machine Learning Model.....	5
2.1 Introduction.....	5
2.2 Study area.....	7
2.2.1 Area Geology .....	7
2.2.2 Site Locations.....	7
2.3 Landslide susceptibility map – machine learning analysis .....	9
2.3.1 Variables and Statistics Compilation used in the Machine Learning Analysis	10
2.3.2 Logistic Regression Model for the Creation of the LSM .....	15
2.3.3 LSM Model Performance and Validation.....	18
2.4 Landslide Hazard Map – Factor of Safety Equation.....	22
2.4.1 Factor Of Safety Equation – Creation Of Dynamic Landslide Hazard Map	22
2.4.2 Estimation of Matric Suction.....	26
2.4.3 Performance and Validation of Landslide Hazard Map.....	28
2.5 Discussion .....	39



2.6	Conclusion.....	40
3	Using Multi-Temporal UAV Images to Detect Land Movement and to Estimate Soil Moisture Data Using Machine Learning Model. ....	42
3.1	Introduction.....	42
3.2	Study area.....	44
3.3	UAV deployment and operation .....	45
3.3.1	Flight path.....	46
3.4	Image processing:.....	47
3.5	Research Methods .....	49
3.5.1	Landslide movement detection .....	49
3.5.2	Results of land displacement analysis.....	51
3.5.3	Extraction of soil parameters from optical images using logistic regression machine learning model.....	54
3.6	Discussion .....	64
3.7	Conclusion.....	66
4	Conclusions.....	68
Appendix A	Python Codes For Landslide Susceptibility Map Using Logistic Regression Model .....	70
Appendix B	Python Codes To Estimate Soil Parameters From Multi-Temporal UAV Images. ....	86
References	.....	106
Vita	.....	113

## LIST OF TABLES

Table 2-1. Features used in the logistic regression model. ....	11
Table 2-2. Statistics for each of the features in a buffer area for logistic regression model. ....	14
Table 2-3. Features that yielded highest score of accuracy after univariate selection method in the logistic regression model. ....	18
Table 2-4. Landslide occurrences modeled in LSM in Campbell County.....	21
Table 2-5. Soil properties used in the general factor of safety equation as presented in Eq. 5 at the study sites. ....	26
Table 2-6. Extraction of hydrologic parameter using the Rosetta online tool. ....	27
Table 2-7. Factor of safety at the study sites at different times. ....	29
Table 2-8. Confusion matrix of LHM (4/4/2018) vs LSM in part of northern Kentucky site area (Figure 2-17).....	38
Table 2-9. Confusion matrix of LHM (3/20/2018) vs LSM in part of northern Kentucky site area (Figure 2-18). ....	39
Table 3-1. Technical specification of the UAV .....	46
Table 3-2. The mean displacement in E (+)/W (-), N (+)/ S (-), and Euclidean distance for each pair of the images.....	54
Table 3-3. Soil parameters extracted from UAV images using machine learning model. ....	57
Table 3-4. Statistics for each band of the images for linear regression model. ....	58
Table 3-5. Features used to model soil parameters in linear regression model. ....	60
Table 3-6. Coefficients to estimate soil parameters from Linear Regression model using Eq. 10. ....	62
Table 3-7. Target variables that yielded more than 0.8 R <sup>2</sup> score in the linear regression model provided input from the variables from UAV images. ....	63

## LIST OF FIGURES

Figure 2-1. Location of the study sites: (a) landslide sites in northern Kentucky used for training the machine learning model for the LSM creation, (b) two sites in northern Kentucky (Site 6294 and Site 6501) used in validation of LHM, (c) validation site in Johnson County used in creating the LHM (Site 6396), (d) validation site in Pike County in eastern Kentucky (Site 6458) for creation of the LHM. Images were taken from DEM and Google Earth imagery using ArcGIS scene. ....	8
Figure 2-2. Soil property features from NRCS WSS: (a) one third bar water content, (b) available water capacity, (c) clay percentage, (d) saturated hydraulic conductivity, (e) liquid limit, (f) plasticity index, (g) sand percentage, (h) silt percentage, and (i) land. ...	12
Figure 2-3. Geomorphic features derived from 1.5 m LiDAR DEM: (a) aspect, (b) curvature, (c) plan curvature, (d) elevation, (e) roughness, and (f) slope. All features were used in Logistic Regression model to create LSM in northern Kentucky. ....	13
Figure 2-4. Landslide buffers (44.7m radius) around centroids of mapped landslides and a non-landslide, Campbell County, Kentucky. ....	15
Figure 2-5. Optimum number of features using univariate feature selection method in logistic regression model. ....	18
Figure 2-6. Results of Logistic Regression function. Model performance is indicated by the receiver operating characteristic curve, the area under the curve (ROC AUC) for the test data. ....	19
Figure 2-7. LSM for Campbell County, Kentucky using logistic regression model. ....	20
Figure 2-8. Zoomed view of LSM for Campbell County with landslide extents. Note, the cell size of the figure is 91 m. ....	22
Figure 2-9. Soils data in northern Kentucky used for the creation of the LHM: (a) soil friction angles estimated for the study area in vicinity of Site 6294 and 6501 using the Wood (1990) equation; (b) plasticity index data taken from NRCS WSS. ....	24

Figure 2-10. Resampled slope angle from 1.5 m to 15.2 m grid cells (taken from 1.5 m LiDAR DEM) for the study area in vicinity of Site 6294 and 6501 in northern Kentucky. .... 25

Figure 2-11. Soil data obtained from the NRCS WSS for the study area in vicinity of Site 6294 and Site 6501 in northern Kentucky: (a) sand percentage (b) silt percentage (c) clay percentage. The data were consequently used to estimate van Genuchten (1980) model parameters in Rosetta online tool..... 27

Figure 2-12. Temporal and Spatial SMAP data for the northern Kentucky area: (a) Spatial grid of volumetric water content from SMAP in Campbell County, Kentucky (b) volumetric water content at Site 6294 over six-week period during which the failure occurred..... 28

Figure 2-13. The LHM for Site 6396: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety at Site 6396 between January 3, 2017 and February 12, 2017 during which the failure happened. .... 31

Figure 2-14. The LHM for Site 6294: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between January 14, 2016 and February 23, 2016 during which the failure happened. .... 32

Figure 2-15. The LHM for Site 6501: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between March 9, 2018 and April 18, 2018 during which the failure happened. .... 35

Figure 2-16. The LHM for Site 6458: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between January 20, 2018 and March 1, 2018 during which the failure happened. .... 36

Figure 2-17. Comparison of (a) the lowest value of LHM in 5 years at Site 6501 on 4/4/2018 and (b) the LSM. .... 37

Figure 2-18. Comparison of (a) LHM at Site 6501 on 3/20/2018 and (b) the LSM..... 38

Figure 3-1. Aerial view of the study site. Landslide occurred and remediated in 1995, Garrard County, Kentucky. .... 45

Figure 3-2. UAV and its flight path used in the study area (a) Phantom 4 advanced UAV equipped with a digital camera to collect images (b) study site from oblique angle from the look point shown in (c) and the landslide boundary extent in red line marker (c) aerial view of flight path for the UAV set in flight planning mobile app. .... 47

Figure 3-3. Matching point on the ground from different overlapping images in SIFT photogrammetry technique processed in Pix4D mapper. .... 48

Figure 3-4. Coregistration of images were performed using area based matching technique in ENVI (a) tie points in the base image (8/11/2020) to which (b) tie points in secondary image (9/11/2020) were co-registered. .... 49

Figure 3-5. Horizontal displacement map of the study site for the pair images from (a) 8/11/20 and 9/11/20, (b) 9/11/20 and 10/9/20, (c) 10/9/20 and 11/10/20 using COSI-Corr technique. (d) Vertical displacement (in LOS direction of SAR satellite) of the study site between 8/6/20 and 9/11/20 using DinSAR technique, Garrard County, Kentucky. .... 53

Figure 3-6. Accumulated mean displacement in E (+)/W (-) and N (+)/S (-) direction within the metrics area for the time span of 8/11/20 and 11/10/2020..... 54

Figure 3-7. Spectral bands of digital images (9/11/2020) and their histograms of digital numbers (0-255) for surface reflectance. (a) Red (Band 1) (b) green (band 2) (c) blue (band 3) (d) histogram of band 1 (e) histogram of band 2 (f)..... 55

Figure 3-8. Mean and standard deviation for the multi-temporal images (a) mean band values and (b) standard deviation values for each image over three month period at the study site. .... 58

Figure 3-9. Predicted target values using linear regression model, actual target values from Satellite data (GLDAS, SMAP), and  $R^2$  score. (a) soil moisture percentile at root zone (b) Soil moisture at root zone (c) Vegetation Greenness fraction (d) Soil Moisture Content at 100-200 cm at study site, Garrard County, Kentucky. .... 64

---

# 1 Introduction

## 1.1 Problem Synopsis

Landslides are common and dangerous natural hazards that occur worldwide, often causing severe direct impacts on human lives, public and private properties, and lifelines (Klose et al. 2014). It is imperative to identify the landslide susceptible areas to avoid or mitigate the possible damage. Landslides occur when the shear stress along a failure plane within the geologic materials of a slope exceeds the shear strength of the material. Landslide susceptibility can be defined as the probability of spatial occurrence of slope failures, given a set of geo-environmental conditions (Guzzetti et al. 2005). Common methods for landslide susceptibility modelling include logistic regression, neural network analysis, data-overlay, index-based and weight of evidence analyses, with an increasing preference towards machine learning methods in the recent years (Reichenbach et al. 2018). Landslide susceptibility map (LSM) created by aforementioned methods stays static across the area because these methods do not account for the variation of soil characteristics such as soil moisture. The antecedent soil moisture and evapotranspiration govern the actual soil moisture regime and thus govern the factor of safety of a slope. Landslide hazard map (LHM) presents when and where the landslide is likely to occur, combining spatial and temporal distribution of the landslide susceptibility (Guzzetti et al. 2005). The aim of this paper is to produce local-scale landslide hazard map (LHM) using a limit equilibrium approach (Lu and Godt 2008) that utilized publicly available satellite and remote sensing data, and to compare it to LSM that is produced using a machine learning model. A LHM will provide insights into the temporal and spatial evolution of landslides compared to a static LSM, which only models the probability and location of the landslides. Additionally, this paper presents the workflow of using a small unmanned aerial vehicle (UAV) with optical digital camera to detect a land movement, and to extract soil parameters. Land movement and soil moisture of a slope are good indicators of the slope stability, and used as a forecasting of the slope failures and landslides.

## 1.2 Research Objectives

The following objectives were established for this study:

- Extract geomorphic variables derived from an aerial LiDAR-based 1.5 m digital elevation model (DEM), soil property variables from Web Soil Survey (WSS), and land cover data in the study area in northern Kentucky.
- Use machine learning approach to model the landslide susceptibility using the variables extracted.
- Create LSM using the trained model and validate the result using actual landslides in northern Kentucky area.
- Estimate soil strength data (internal friction angle) using plasticity index from WSS.
- Extract soil hydrologic parameters using sand, clay, and silt percentage.
- Create LHM using a limit equilibrium factor of safety equation (Lu and Godt 2008), and validate the result at study sites.
- Compare LSM to LHM.
- Acquire multi-temporal images of a landslide area using UAV.
- Detect a horizontal land movement using an image correlation technique (COSI-Corr).
- Detect a vertical land movement using Synthetic Aperture Radar (SAR) images.
- Estimate soil parameters such as soil moisture using machine learning technique using the multi-temporal images.

## 1.3 Relevance of Research

Landslide susceptibility map is often created using logistic regression, neural network analysis, data-overlay, index-based and weight of evidence analyses, with an increasing preference towards machine learning methods in the recent years (Reichenbach, P et al., 2018). However, landslide hazard map gives close-to-real time assessment of the slope stability because it takes account of the variation in soil moisture, which plays important roles in limit equilibrium factor of safety equation. With the advent of the remote sensing technologies, there is a lot of open source spatial and temporal data such as 1.5 m DEM, soil database from WSS, and land cover information. This document demonstrates the use

of remote sensing data to create LSM and LHM. The latter gives us evolution of the slope stability condition compared to the static LSM. The use of landslide hazard map helps identify the risk of landslide at any given time. In addition, this paper shows the use of UAV to detect land movement which is the most common indicator of a future landslide occurrence, and also it shows the extraction of soil moisture using the multi-temporal images collected by the UAV. The soil moisture extracted from the UAV images can be utilized in limit equilibrium factor of safety equation to assess the slope stability.

#### **1.4 Contents of Thesis**

Chapter 2-3 consist of papers that are to be submitted for publication and the contents is verbatim.

Chapter 2 presents landslide hazard and susceptibility maps derived from satellite and remote sensing data using limit equilibrium analysis and machine learning model. This paper shows that using publicly available data, we can create a multi-temporal landslide hazard map that can produce a close-to-real time landslide susceptibility map. The landslide hazard map derived from the limit equilibrium analysis tells us the evolution of the landslide development temporally and spatially, whereas the landslide susceptibility map derived statistically indicates the locations of possible landslides in the long term.

*Dashbold, B., Bryson, L.S., and Crawford, M.M. (2021). Landslide hazard and susceptibility maps derived from satellite and remote sensing data using limit equilibrium analysis and machine learning model. Landslide [to be submitted]*

Chapter 3 establishes a method to use a small unmanned aerial vehicle (UAV) to detect land movement utilizing photogrammetric techniques, and to estimate soil moisture data using machine learning model. The result of land movement detection analysis showed a mean movement of 9.23 cm, 3.03 cm, and 0.23 cm in the study area for three pairs of images that were taken one month apart, respectively. Using the UAV multi-temporal images, linear regression model estimated the soil parameters including soil moisture at root zone (0-100 cm) ( $R^2=0.823$ ), soil moisture at 100-200 cm ( $R^2=0.906$ ) at 100-200 cm, and vegetation greenness fraction ( $R^2=0.988$ ). The land movement and soil moisture data



can be used to gain significant information about the active landslide, and the stability of a slope.

*Dashbold, B. and Bryson, L.S. (2021). Using multi-temporal UAV images to detect land movement and to estimate soil moisture data using machine learning model. Engineering Geology [to be submitted]*

---

## **2 Landslide Hazard and Susceptibility Maps Derived from Satellite and Remote Sensing Data Using Limit Equilibrium Analysis and Machine Learning Model**

### **2.1 Introduction**

Landslides are common and dangerous natural hazards that occur worldwide, often causing severe direct impacts on human lives, public and private properties, and critical infrastructure (Klose et al. 2014). Direct costs of landslides in Kentucky are conservatively estimated to be between \$10-\$20 million annually (Crawford 2014; Crawford and Bryson 2017). Therefore, it is imperative to identify the landslide susceptible areas to avoid or mitigate the possible damage. Landslides occur when the shear stress along a failure plane within the geologic materials of a slope exceeds the shear strength of the material. Landslides is commonly triggered by intense short period rainfall or prolonged rainfall, earthquakes, or human activities. Landslide susceptibility mapping (LSM), which describes the spatial distribution for the probability of landslide occurrence in each area according to the geographical environment, is considered a common countermeasure for mitigating the effects of landslides (Huang and Zhao 2018; Merghadi et al. 2020). More simply put, susceptibility can be defined as the probability of spatial occurrence of slope failures, given a set of geologic and climatological conditions (Guzzetti et al. 2005).

Several researchers (e.g., Guzzetti et al. 2005; Alvioli et al. 2016; Reichenbach 2018) have suggested a variety of approaches and methods for landslide susceptibility in different geological and climatic settings. With an increasing preference towards machine learning methods in the recent years, the more common methods for landslide susceptibility modelling include logistic regression, neural network analysis, data-overlay, index-based and weight of evidence analyses (Reichenbach et al. 2018). While LSM provides the spatial distribution of landslide probability, landslide hazard map (LHM) gives the probability that a landslide of a given magnitude will occur in each period and in a given area (Segoni et al. 2018). LHM presents when and where the landslide is likely to occur, combining spatial and temporal distribution of the landslide susceptibility (Guzzetti et al. 2005).

Rainfall thresholds are the most widely used triggering factor in landslide hazard assessment and early warning tools (Crozier 1997; Segoni et al. 2018). However, in many cases, early warnings based solely on rainfall is not adequate because antecedent soil

moisture conditions play a crucial role in the initiation of landslides (Zhuo et al. 2019). Some researchers (Zhuo et al. 2019; Wicki et al. 2020; Guzzetti et al. 2020) have investigated the soil moisture threshold for triggering landslides directly, accounting for the temporal evolution of soil moisture before and after slope movement. These studies mainly used machine learning models to investigate the correlation between the soil moisture threshold and the onset of landslides. However, these types of approaches do not consider the statics and fundamental mechanics associated with landslides. Approaches that utilize fundamental mechanics, such as limit equilibrium, are required to better understand landslide occurrences. However, the use of a limit equilibrium equation to create and verify a LHM at known locations and known dates is very minimal.

The aim of this paper is to produce local-scale landslide hazard map (LHM) using a limit equilibrium approach (Lu and Godt 2008) that utilized publicly available satellite and remote sensing data, and to compare it to LSM that is produced using a machine learning model. A LHM will provide insights into the temporal and spatial evolution of landslides compared to a static LSM, which only models the probability and location of the landslides.

This paper presents a two-part approach, where satellite and local site geologic data were combined to create a LHM and an LSM. A machine learning model was used to create the LSM using geomorphic variables derived from an aerial LiDAR-based 1.5 m digital elevation model (DEM), soil property data from Natural Resources Conservation Service (NRCS) Web Soil Survey (WSS), and land cover data in northern Kentucky area where the Kentucky Geology Survey (KGS) has an extensive landslide inventory. The produced LSM was verified against the known landslide occurrences yielding an accuracy rate of 84.1 percent. The LSM is based on relatively static parameters at failure conditions (i.e., the onset of a landslide). Therefore, the LSM represents the probability of the failure at the worst case of scenario.

A limit equilibrium factor of safety equation for a hillslope with transient infiltration conditions (Lu and Godt 2008) was used to produce the LHM with 15.2 m spatial resolution using soil moisture data from the National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) mission, slope angles derived from 1.5 m LiDAR DEM, and soil physical properties from WSS. The LHM was verified temporally and

spatially against the landslides. The LHM on the date with the lowest factor of safety values was compared to the LSM produced by a machine learning approach to assess how the factor of safety values compared to the probability of the landslide susceptibility map, which does not change temporally.

## **2.2 Study area**

### **2.2.1 Area Geology**

Several landslide sites in northern Kentucky were used to comprise the study area for this study. The north-central area of Kentucky is characterized by weathered limestone bedrock of the Ordovician strata that has been pushed toward the crest of the Cincinnati Arch and is often exposed at the ground surface (McGrain 1983). This area is defined as Outer Bluegrass physiographic region of Kentucky. The defining geologic elements of the Outer Bluegrass region consist of late Ordovician and Silurian-age limestones, dolomites, and shales (McDowell 1986), deep valleys formed by erodible shale that exists above non-erodible rock types, rich soils for agricultural uses, and gentle rolling hills formed by slopes ranging from 20 percent to 30 percent. Topographic relief averages approximately 150 m, ranging from steep slopes (i.e., slope angles between 20 to 35 degrees) along the Ohio River, gently sloping uplands (i.e., slope angles < 6 degrees) and broad dissected valleys. Shaly bedrock in the region weathers easily and produces thin to thick, clayey colluvial soils. Landslides typically occur within the colluvium or along the colluvial-bedrock contact (Crawford and Andrews 2012). Geotechnical reports in northern Kentucky show that the average depth to the bedrock in the hillslope soil is about 4.6 m. Therefore, it is used in the slip surface depth in the factor of safety equation.

### **2.2.2 Site Locations**

Landslide data used for this study were obtained from the Kentucky Geological Survey (KGS) Landslide Inventory (Kentucky Geological Survey, 2019). Landslide sites in Campbell County and Kenton County, northern Kentucky were used as training points for the machine learning model to create the LSM. Other sites in Campbell County not used as training points were as validation sites. Two sites in northern Kentucky (identified in

the KGS Landslide Inventory as Site 6501 and Site 6294) and two sites in eastern Kentucky (identified in the KGS Landslide Inventory as Site 6396 and Site 6458) were used to validate the results of LHM, which was created using the limit equilibrium analysis.

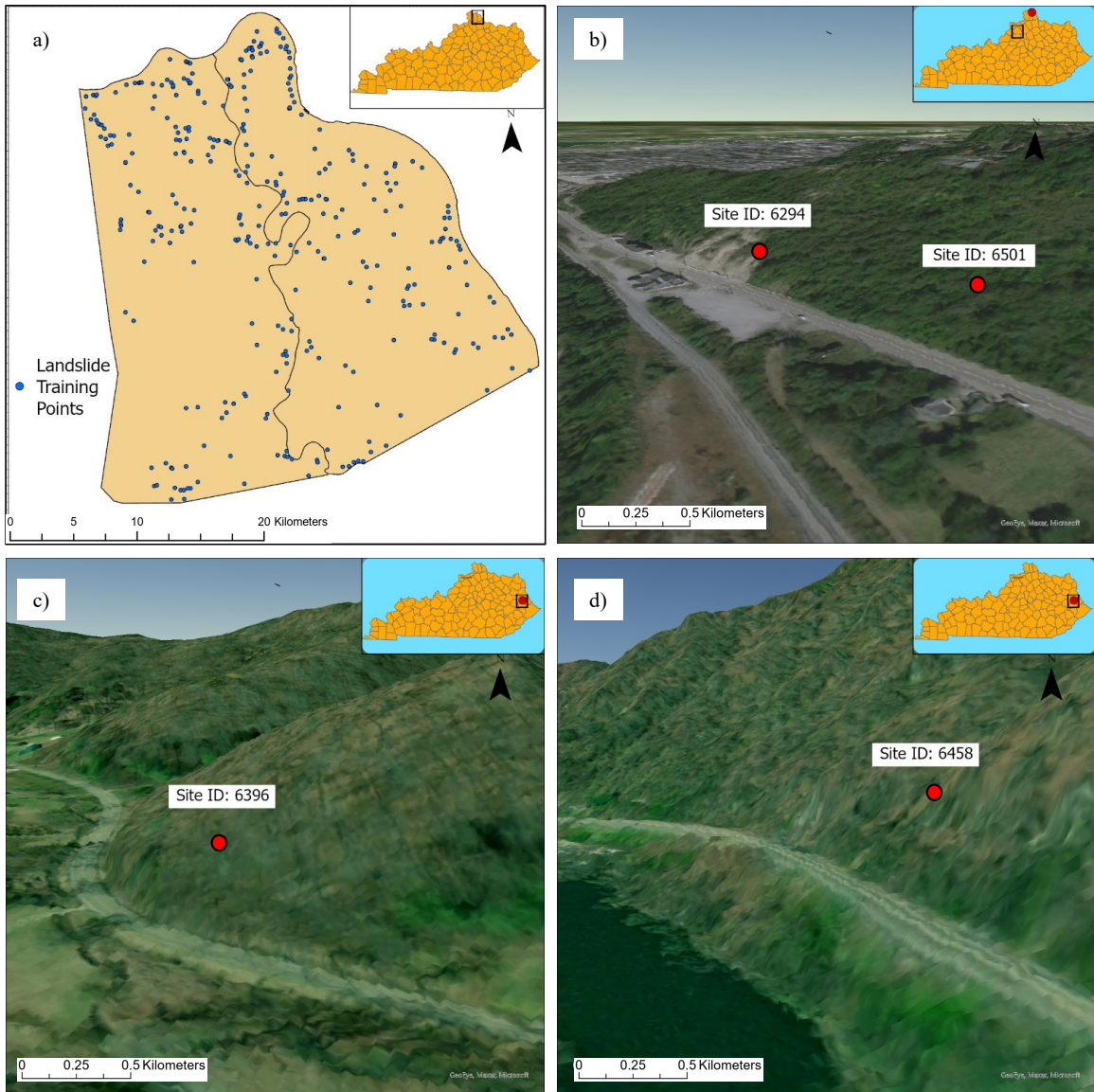


Figure 2-1. Location of the study sites: (a) landslide sites in northern Kentucky used for training the machine learning model for the LSM creation, (b) two sites in northern Kentucky (Site 6294 and Site 6501) used in validation of LHM, (c) validation site in Johnson County used in creating the LHM (Site 6396), (d) validation site in Pike County in eastern Kentucky (Site 6458) for creation of the LHM. Images were taken from DEM and Google Earth imagery using ArcGIS scene.

Figure 2-1 presents the locations of the study sites used in this study, relative to the State. Terrain information was obtained from the 1.5 m LiDAR DEM in the study area that was

flowed in 2011 and 2012. Sites 6501 and 6294 [Figure 2-1(b)] are in Campbell County and are about 180 m apart. The landslides occurred on 3/30/2018 and 2/04/2016, respectively. The slope angles in the vicinity of the landslide points range between 5 degrees to 38 degrees, with the slope surface partially covered with trees. The landslide runouts at these sites resulted in debris that blocked the road downslope of the sites. Site 6396 [Figure 2-1(c)] is in Johnson County, eastern Kentucky and was reported on 1/24/2017. The slope angles in the area ranged from 38 degrees to 58-degree along the roadway downslope of the site. Site 6458 [Figure 2-1(d)] is in Pike County, eastern Kentucky. The landslide was reported on 2/10/2018. The slope angle was approximately 63 degrees with sporadic rock outcropping with little to no overburden in some areas.

### **2.3 Landslide susceptibility map – machine learning analysis**

The general landslide susceptibility map (LSM) was created using machine learning technique for Campbell County in northern Kentucky. A Geographic Information System (GIS) model was used to prepare and extract features for the machine learning analysis such as geomorphic variables, soil physical variables, and land cover variables. These features are independent variables and are used to predict target variables, known as Landslide Conditioning Factors (LCFs). The LCFs in this study were extracted using GIS and compiled into database as training points for the machine learning analysis to predict landslide susceptibility. The logistic regression machine learning model was selected for this study based on the performance with a high prediction accuracy in the previous studies (Du et al. 2017; Kalantar et al. 2018; Kadavi et al. 2019; Nhu et al. 2020; Crawford et al. 2021).

The KGS Landslide Inventory listed 234 known landslide occurrences in Campbell and Kenton counties in northern Kentucky region, some of data entries include the reported failure dates. All 234 landslide sites were used to extract LCF features for the machine learning analysis. In training the model, 75 percent of the sites was used in training the model (176 points for each class; landslide and non-landslide) and 25 percent was used in testing the data for validation.

### **2.3.1 Variables and Statistics Compilation used in the Machine Learning Analysis**

Landslide occurrence is influenced by bedrock geology, geological structure, hillslope morphology, soil type and thickness, and hydrogeological conditions (Xu et al. 2018). We used 15 variables in the machine learning analysis from three distinct fields including soil property features, geomorphic features, and land cover information. Table 2-1 shows all used features with their definitions. The soil property features play important roles in determining soil strength and its behavior. Figure 2-2 shows the soil property features and their range of values extracted from the NRCS WSS that include percent sand, percent silt, percent clay, saturated hydraulic conductivity, available water capacity, one third bar water content, plasticity index, and liquid limit, all of which were used as features in the logistic regression analysis. Geomorphic features were used as LCF features in landslide susceptibility modelling in several studies (e.g., Reichenbach et al. 2018; Crawford et al. 2021). The geomorphic features used in this analysis were slope, aspect, curvature, elevation, roughness, and plan curvature, all derived from 1.5 m LiDAR DEM in ArcGIS and shown in Figure 2-3. The 1.5 m LiDAR DEM was taken from the Kentucky Elevation Data and Aerial Photography Program (KyFromAbove, 2021), and it is publicly accessible and open-source portal. In addition, land cover data from of 2016 National Land Cover Database (NLCD) product suite (Multi-Resolution, 2021) [Figure 2-2(i)] were used in the analysis.

Table 2-1. Features used in the logistic regression model.

<b>Geomorphic features</b>	<b>De finition</b>
Elevation	Vertical distance of a point above or below a reference surface, derived as a representation of the Earth's surface, derived as a representation of the Earth's surface (meters)
Slope	Gradient or steepness from each cell of an elevation raster (degrees)
Terrain roughness	Degree of terrain irregularity calculated as surface deviation from a smoothing window; scale of landscape features is important in choosing a smoothing-window size
Curvature	Second derivative value from each cell from an elevation raster (1/100 of a z-unit)
Plan curvature	Curvature of the surface perpendicular to the direction of maximum slope (1/100 of a z-unit)
Aspect	Compass direction of a downhill-facing slope, derived for each cell of an elevation raster.
<b>Soil Property Features</b>	<b>De finition</b>
Percent sand	Sand percentage in the soil layer.
Percent silt	Silt percentage in the soil layer.
Percent clay	Clay percentage in the soil layer.
Saturated hydraulic conductivity	The ease with which pores in a saturated soil transmit water (micrometers per second).
Available water capacity	Quantity of water that the soil can store for use by plants (centimeters of water per centimeter of soil layer)
One third bar water content	Amount of soil water retained at a tension of 1/3 bar (volumetric percentage of the whole soil)
Plasticity index	Numerical difference between the liquid limit and plastic limit of the soil. It is the range of water content in which a soil exhibits the characteristics of a plastic solid (percent).
Liquid limit	Water content, on a percent by weight basis, of the soil (passing #40 sieve) at which the soil changes from a plastic to a liquid state (percent)
<b>Land cover information</b>	<b>De finition</b>
Land cover data	Land cover characteristic (eg., grassland, shrub, deciduous forest)



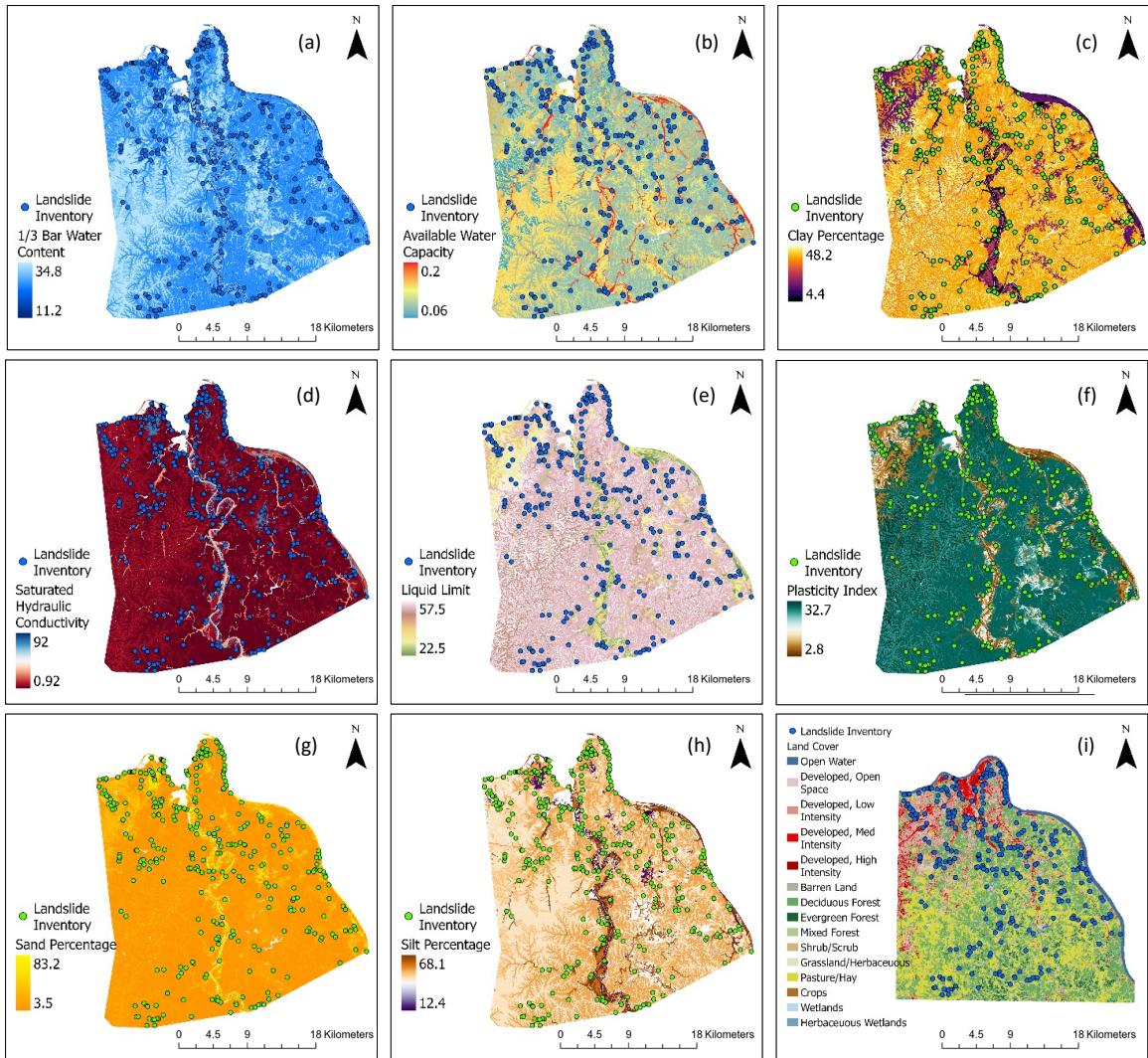


Figure 2-2. Soil property features from NRCS WSS: (a) one third bar water content, (b) available water capacity, (c) clay percentage, (d) saturated hydraulic conductivity, (e) liquid limit, (f) plasticity index, (g) sand percentage, (h) silt percentage, and (i) land.

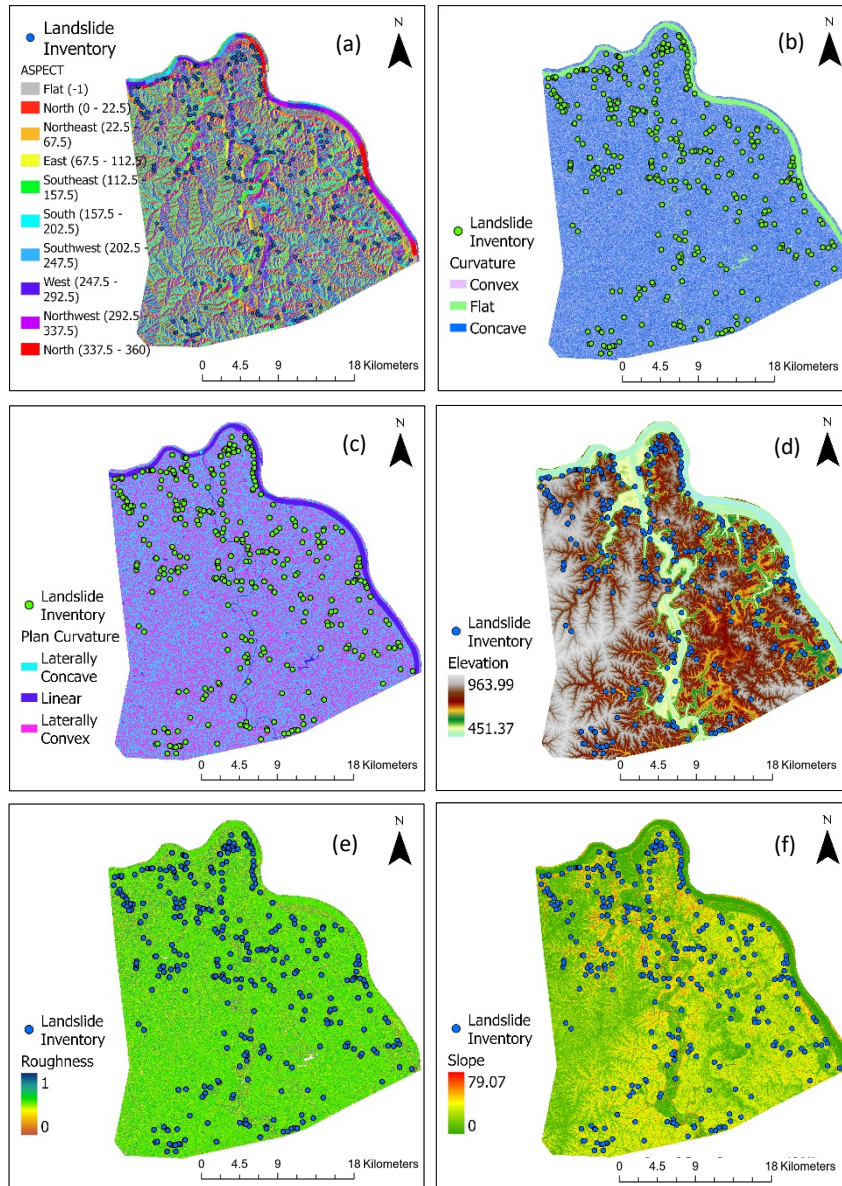


Figure 2-3. Geomorphic features derived from 1.5 m LiDAR DEM: (a) aspect, (b) curvature, (c) plan curvature, (d) elevation, (e) roughness, and (f) slope. All features were used in Logistic Regression model to create LSM in northern Kentucky.

The statistics used in the analysis were the mean, standard deviation, variance, skewness, and coefficient of variance as these statistics capture the variation of the distribution of the data in each buffer area for each LCF features. Descriptive statistics (e.g., mean, range, standard deviation) of elevation and slope are better predictors of the presence (or absence) of landslides than the same indices computed for the single DEM cells (Carrara et al. 1991, Alvioli et al. 2016, Reichenbach et al. 2018). The statistics used for each feature in the

buffered areas for the landslide and non-landslide were calculated in ArcGIS and shown in Table 2-2.

Table 2-2. Statistics for each of the features in a buffer area for logistic regression model.

Statistics	Formula
Mean	$Mean = \frac{\sum_{i=1}^N x_i}{N}$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - Mean)^2}$
Variance	$\sigma^2$
Coefficient of Variance	$CV = \sigma / Mean$
Skewness	$Sk = (3Mean - Md) / \sigma$

$x_i$  = grid cell value;  $Md$  = median value;  $N$  = the total number of samples in the training or testing dataset in each feature for the buffer area.

There were 15 independent variables with five statistics for each variable except for NLCD, which used only median value since the spatial resolution was 30 m. Therefore, a total of 71 features were used in the machine learning model as training points. There were some dependencies between the features such as elevation variance feature and elevation standard deviation feature. However, the model tended to start with more features that narrowed down to the optimal number of features to get the higher accuracy of the result as determined by statistical skill assessment method.

A circular buffer area of 45.7 m radius from the centroid point of the landslide extent was used as a mask layer to extract feature statistics. Therefore, it dictated the spatial resolution of the LSM (91 m). Figure 2-4 shows how the mask buffer areas used to extract each feature were aligned with the landslide extent in a close view. Most of the buffer masks fell within the extent of the landslide boundary (Figure 2-4). Timilsina et al. (2014) observed that a buffer polygon that represents most of the landslide extent is superior to a single point in accounting for variability in landslide characteristics. Crawford et al. (2021) evaluated the

performance of several buffer sizes for predicting landslides in eastern Kentucky using the MATLAB Classification Learner application (Mathworks Inc., 2021). These researchers found a buffer size of 45.7 m performed best in predicting landslide. This buffer size was used in this study to extract the features because the landslides in this study were in the same region as those investigated in Crawford et al. (2021).

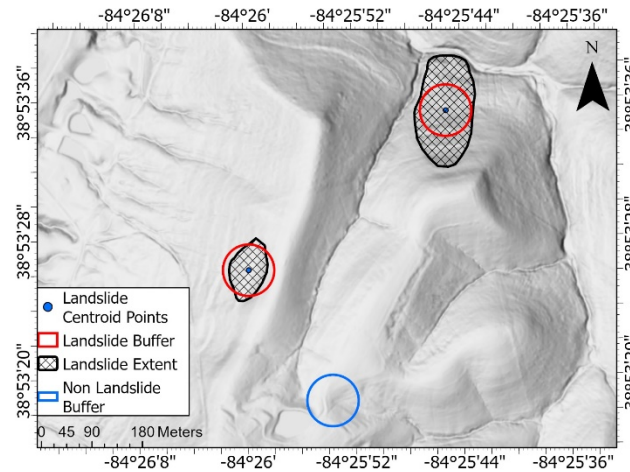


Figure 2-4. Landslide buffers (44.7m radius) around centroids of mapped landslides and a non-landslide, Campbell County, Kentucky.

For the data preparation, all 234 landslide occurrences in northern Kentucky were assigned the class value of 1 for landslide, and all 71 features for each landslide were extracted using the buffer mask area. An equal number of points that represented non-landslide areas were randomly selected and assigned the class of 0 for non-landslide. All 71 features for all 234 non-landslide areas were extracted the same way, using the same size buffer mask area as the landslide areas. Thus, the two classes; landslide=1 and non-landslide=0, had the same size and features and were extracted the same way. The non-landslide areas were manually inspected to confirm that there were no landslides in the non-landslide buffer area. An equal number of landslides (1) and non-landslides (0) was required for an equal class-distribution ratio, which helps to eliminate class bias (Gupta et al. 2019).

### 2.3.2 Logistic Regression Model for the Creation of the LSM

Logistic regression is a statistics-based linear model used to model the probability of the existence of a certain class. Logistic regression is also known in the literature as logit regression, maximum-entropy classification (MaxEnt) or the log-linear classifier. In this

model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function. The dependent variables or indicator variables, which were predicted using the logistic regression model, were value of 1 (landslide) or value of 0 (non-landslide). The probability of absence of landslide (0) or presence of landslide (1) was expressed as Eq. 1:

$$P = \frac{1}{1 + e^{-z}} \quad (1)$$

where  $P$  is the cumulative estimated output probability of a landslide occurrence (confined between 0 and 1),  $z$  was the weighed linear combination of independent variables (ranging from  $-\infty$  to  $+\infty$ ) that could be expressed as a Eq. 2 sum of constants (Kadavi et al. 2019):

$$z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \dots + \beta_n V_n \quad (2)$$

where  $p/(1-p)$  is the corresponding odds or the likelihood ratio,  $\beta_0$  is the constant intercept;  $V_i$  ( $i = 1, 2, 3, \dots, n$ ) are the independent variables (e.g., Curvature Std, Slope Variance) and  $\beta_i$  ( $i = 1, 2, 3, \dots, n$ ) are the coefficient estimates of the model.

The coefficients in Eq. 2 express the effects of the predictor independent variables on the relative risk of being a landslide or not a landslide (0 or 1). The relative risk increases or decreases with each value of the independent variable  $V_i$  (i.e., the rate of change) in log-odds as  $V$  changes (Crawford et al. 2021). Consequently, higher  $z$ -values suggest a  $P$ -value closer to 1, which indicates the presence of a landslide. Conversely, lower  $z$ -values suggest a  $P$ -value closer to 0, which indicates the absence of a landslide. The  $\beta_i$  coefficients of the model were estimated using a cost function as shown in the works of Lee and Liu (2003). The threshold value of  $P = 0.5$  in Eq.1 was used as a decision boundary such that a  $P$  value 0.5 or greater was classified as Class 1 (landslide), and a  $P$  value less than 0.5 was classified as Class 0 (non-landslide). We assumed that the variables were not normally distributed or did not have linear relationships (Nandi et al. 2016). The advantage of logistic regression is that the variables may be either continuous or discrete, and it is not necessary for these variables to have normal distributions (Du et al. 2017). Therefore, the logistic-regression

analysis was well suited for this study because the primary unknowns are the relationships among the variables (Crawford et al. 2021).

The initial result including all 71 features (independent variables) was developed in the logistic regression model. The optimal number of features were searched to produce the highest model accuracy using a univariate feature selection method, using all features in the logistic regression model. The univariate feature selection method works by selecting the features that produced the most accuracy of the model based on univariate statistical tests. The feature selection process is a preprocessing step to optimize the number of features to yield the best accuracy of the model. In this selection process, each feature to the indicator variable (1, 0) was compared to determine if there is any statistically significant relationship between them. This method is also known as analysis of variance (ANOVA), which analyzes the relationship between one feature and the indicator at a time without the consideration of the other features. The univariate selection method uses statistical tests to select those features that have the strongest relationship with the indicator variable (0 or 1) and yields test scores for each feature. All the test scores were compared, and the features with top scores were selected. Based on the univariate statistical test on all the training data, 18 features produced the highest accuracy and were subsequently used in the logistic regression model. Figure 2-5 shows the optimal number of features with their corresponding accuracy percentage (given as cross validation scores in the figure) using the univariate feature selection method. The accuracy for the model increased to optimal 77 percent at 18 features, and it dropped to an approximate accuracy value of 73% after 30 features (Figure 2-5). Table 2-3 shows the 18 features that produced the highest model accuracy using univariate selection method in the logistic regression model with their corresponding ranking scores.

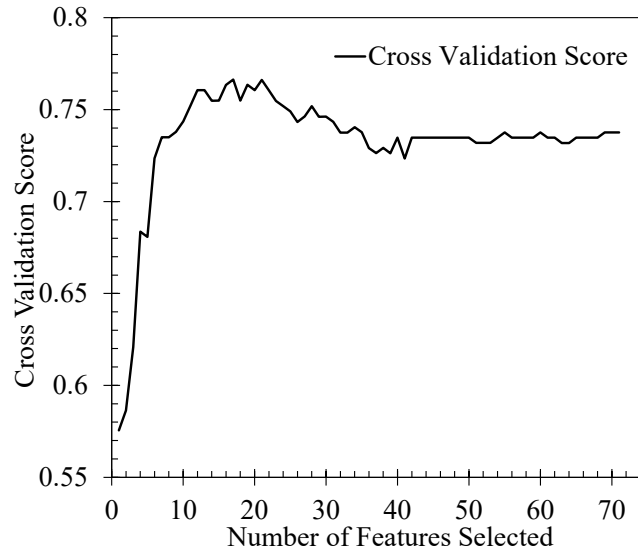


Figure 2-5. Optimum number of features using univariate feature selection method in logistic regression model.

Table 2-3. Features that yielded highest score of accuracy after univariate selection method in the logistic regression model.

Feature	Selection Scores	Feature	Selection Scores	Feature	Selection Scores
Elevation CV	14.50	Aspect Std	3.78	Aspect Variance	2.25
Elevation Variance	12.72	Slope CV	3.58	Sand Mean	2.18
Elevation Std	10.60	Curvature Std	3.40	Available Water - Content Mean	2.05
Slope Mean	7.76	Slope Variance	3.26	NLCD Median	1.85
Plan Curvature Std	4.05	Curvature Variance	2.88	Clay Std	1.28
Plan Curvature Variance	3.89	Slope Std	2.39	One Third Bar Water - Content Mean	1.25

### 2.3.3 LSM Model Performance and Validation

The logistic regression model for the LSM was validated using the Receiver Operating Characteristic (ROC) metric to evaluate the classifier output quality (Landslide = 1, Non-Landslide = 0). The ROC is a practical method to visualize the performance of the model. In the validation process, we used 25 percent of the landslide inventory points (59 landslide points and 59 non-landslide points) as test data. The area under the curve (AUC) of the

ROC provided the accuracy of the classifier on the test points. The ROC was created plotting specificity (false positive rate) on x-axis versus sensitivity on y-axis. Specificity and Sensitivity were defined as,

$$Specificity = \frac{TN}{(TN + FP)} \quad (3)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (4)$$

where TP is true positive i.e., the number of landslides (Class 1) predicted accurately; TN is true negative i.e., the number of non-landslides (Class 0) predicted correctly; FP is false positive - the number of non-landslides predicted as landslides whereas FN is false negative - the number of landslides predicted as non-landslides. The specificity indicates the proportion of the negative class that was correctly classified, and the sensitivity indicates the proportion of the positive class that was correctly classified. Figure 2-6 shows the AUC for the model, which was calculated in the test dataset as 0.841.

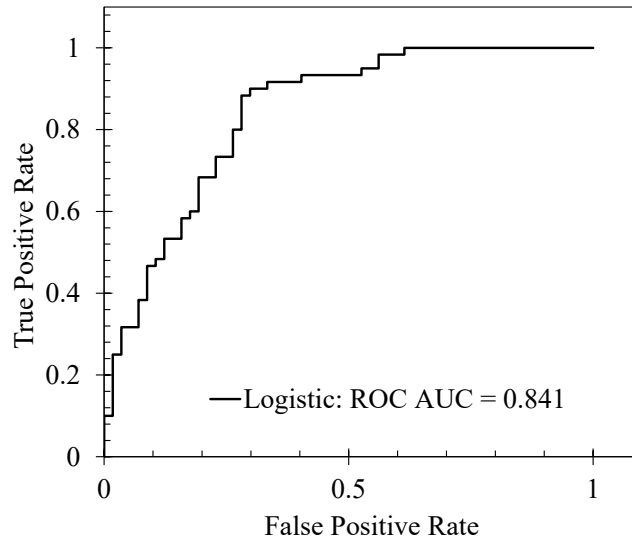


Figure 2-6. Results of Logistic Regression function. Model performance is indicated by the receiver operating characteristic curve, the area under the curve (ROC AUC) for the test data.

This machine learning model was used to map a landslide susceptibility on a large-scale area such as Campbell County. ArcGIS was used to extract the selected 18 features (Table 2-3) from the spatial data, and to enumerate them in the tables that needed to be processed



in machine learning model. The landslide susceptibility result from the machine learning model was inserted back into ArcGIS to produce the LSM. Figure 2-7 shows countywide LSM for Campbell County using the result of the machine learning model.

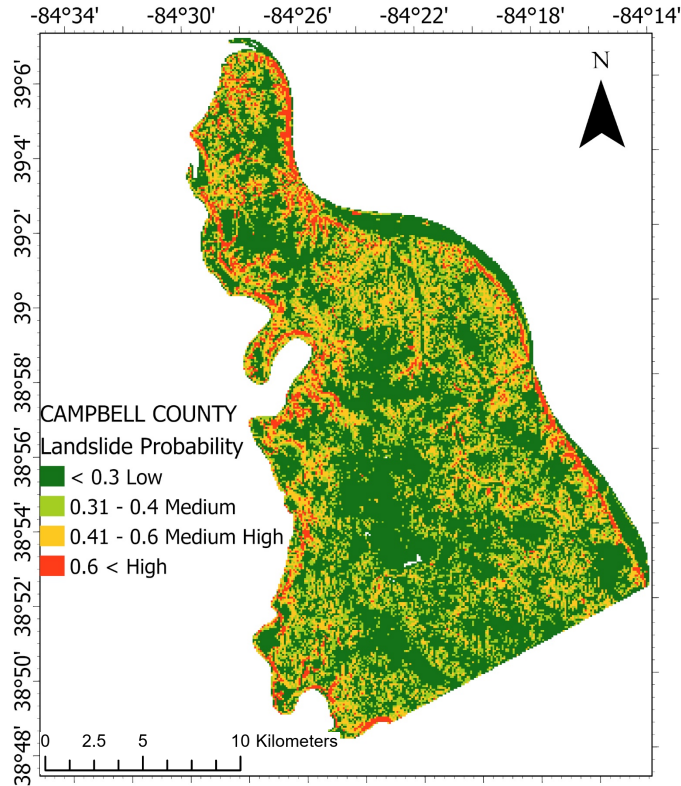


Figure 2-7. LSM for Campbell County, Kentucky using logistic regression model.

Crawford et al. (2021) developed a landslide susceptibility classification system based on standard deviation from the mean of landslide occurrence probability. We used a similar approach for this current study. The mean probability value of landslide occurrences for our study was 0.634 with a standard deviation of 0.186. One standard deviation from the mean was 0.448 and two standard deviations from the mean was 0.262. We assumed rounded values for break points. High susceptibility was values greater than mean (approximately 0.6), High to Moderate susceptibility was within one standard deviation from the mean (0.41 to 0.6), Moderate was within two standard deviations from the mean (0.31-0.4), and Low was values below two standard deviations from the mean (< 0.3). Table 2-4 shows the susceptibility classes for landslide occurrences, indicating 87 percent of the landslide extents in Campbell County were classified as high Moderate to High (Figure 2-7).

Table 2-4. Landslide occurrences modeled in LSM in Campbell County.

Total of Landslide occurrences		131
Susceptibility level	Number of Landslides	Percent of all landslides
Low (< 0.3)	6	4.6%
Moderate (0.31 and 0.4)	11	8.4%
High to Moderate (0.41 and 0.6)	36	27.5%
High (> 0.6)	78	59.5%

Campbell County has a total area of 413 sq.km of which 53.6 percent was classified as Low probability, 20.1 percent as Moderate probability, 19.9 percent as high to moderate probability, and 6.3 percent as high probability (Figure 2-7). Figure 2-8 shows the validation of the LSM using the actual landslide areal extents in a close view in which the extent of seven landslide occurrences in the area 3.5 km by 2 km are all within or crossing areas classified as high Moderate to High.

Because the statistical based LSM are created by prior landslide events, the LSM provides estimate probabilities of future landslide susceptibility in a spatial term. LSM describes the relative likelihood of future landslides occurrences based solely on statistical methods using slope variables such as the ones used in this study (Table 2-1) from the past landslide events. This implies that future landslide events will be more probable to occur under the conditions in which the past landslide events occurred. The mapped landslide susceptibility provides useful tools for avoiding or reducing potential damages from the landslides. In contrast, LHM predicts not only spatial component of the landslide susceptibility, but also temporal aspect of the landslide events, rendering the LHM time-aware dynamic map.

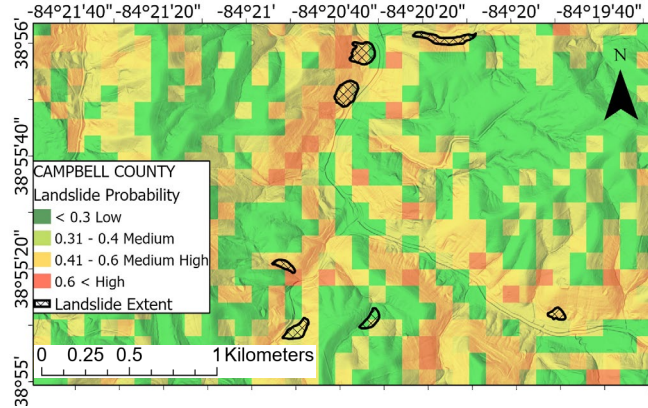


Figure 2-8. Zoomed view of LSM for Campbell County with landslide extents. Note, the cell size of the figure is 91 m.

## 2.4 Landslide Hazard Map – Factor of Safety Equation

Whereas the LSM provides an estimate of “where” a landslide is most likely to occur, the landslide hazard map (LHM) provides an estimate “when” a landslide is most likely within a region susceptible to landslides. We estimated the “when” landslides occurred using the limit equilibrium factor of safety equation presented by Lu and Godt (2008). Zhuo et al. (2019) found a correlation between seasonal variations in soil moisture and the occurrence of landslides. Traditional limit equilibrium factor of safety equations are based on static equilibrium involving resistance forces in equilibrium with the driving forces on a slope and do not count for the unsaturated soil conditions. Therefore, these equations produce factor of safety values that are constant regardless of the changing hydrologic conditions (i.e., volumetric water content and matric suction) of the soil. Lu and Godt (2008) suggested that the resisting forces in a hillslope system were functions of the hydrologic conditions in the slope and those conditions varied dynamically and with depth in the vadose zone. The implication of the Lu and Godt (2008) suggestion is the factor of safety of a hillslope will vary with dynamic variations and with depth of the hydrologic conditions. For this study, we evaluated the LHM produced by the factor of safety equation using four study sites described previously in this paper.

### 2.4.1 Factor Of Safety Equation – Creation Of Dynamic Landslide Hazard Map

The Lu and Godt (2008) equation presents the fact of safety for subaerial infinite slopes in the unsaturated saturated soil conditions. The Lu and Godt (2008) equation is given as the

sum of three distinct contributions to slope strength; the frictional strength component, the cohesion strength component, and the strength component derived from matric suction. The general form of the equation is given as:

$$FS = \frac{\tan(\phi')}{\tan(\beta)} + \frac{2c'}{\gamma H_{ss} \sin(2\beta)} + \frac{\sigma^s}{\gamma H_{ss}} [\tan(\beta) + \cot(\beta)] \tan(\phi') \quad (5)$$

where  $FS$  is the factor of safety;  $\phi'$  is the soil friction angle;  $c'$  is the effective soil cohesion;  $\beta$  is the slope angle;  $\gamma$  is the soil unit weight;  $H_{ss}$  is the depth to bedrock;  $\sigma^s$  is the suction stress, a characteristic function of the soil that describes the inter-particle stresses resulting from the wetting and drying of the soil. The suction stress is given as,

$$\sigma^s = S_e s \quad (6)$$

where  $S_e$  is the effective degree of saturation  $= (\theta - \theta_r) / (\theta_s - \theta_r)$ ;  $\theta$  is the volumetric water content;  $\theta_r$  is the residual volumetric water content;  $\theta_s$  is the saturated volumetric water content;  $s$  is the soil matric suction, which is defined in this study using the van Genuchten (1980) soil water characteristic curve (SWCC) model given as,

$$s = (u_a - u_w) = \frac{1}{\alpha} \left[ \left( \frac{\theta - \theta_r}{\theta_s - \theta_r} \right)^{\frac{1}{m}} - 1 \right]^{\frac{1}{n}} \quad (7)$$

where  $u_a$  is the pore air pressure;  $u_w$  is the pore water pressure;  $\alpha$  is a fitting parameter reflecting the air entry value;  $n$  is a fitting parameter related to the inflection point of the SWCC;  $m$  is fitting parameter related to the curvature of the SWCC near the residual point,  $m = (n-1)/n$ .

The shear strength parameters (i.e.,  $c'$  and  $\phi'$ ) for the sites used in the LHM model were acquired from various geotechnical reports obtain near the site locations from the Kentucky Transportation Cabinet Geotechnical Project Report Database (Kentucky Geological Survey, 2021) and from estimates made using data from the NRCS WSS database. The geotechnical reports indicated that the soils in the northern Kentucky area consisted predominantly of colluvium (i.e., silty clay). When not directly given, the soil friction

angles were estimated using the Wood (1990) equation and the plasticity index extracted from the WSS database given as,

$$\phi' = \sin^{-1} \left[ 0.35 - 0.1 \ln \left( \frac{PI}{100} \right) \right] \quad (8)$$

where  $PI$  is the plasticity index. The effective soil cohesion,  $c'$  for the colluvium was found to be negligible. Therefore, the second term in the Eq. 5 dropped out. Figure 2-9 shows the estimated map of the soil friction angles,  $\phi'$  derived from Eq. 8 and the plasticity index data in the vicinity of study sites 6294 and 6501. The soil friction angle,  $\phi'$  ranges between 27.6 degrees and 33.6 degrees with the lower end of this range dominating the study area in northern Kentucky (Figure 2-9).

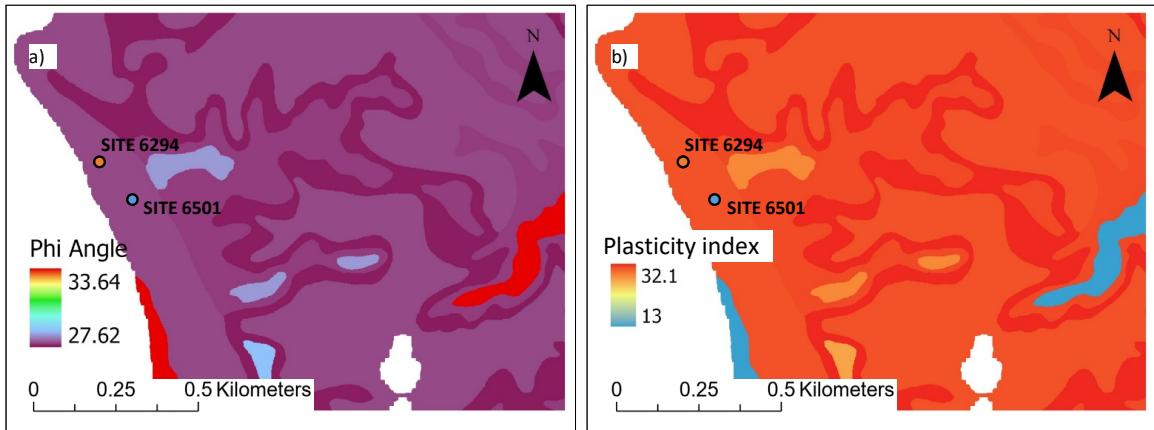


Figure 2-9. Soils data in northern Kentucky used for the creation of the LHM: (a) soil friction angles estimated for the study area in vicinity of Site 6294 and 6501 using the Wood (1990) equation; (b) plasticity index data taken from NRCS WSS.

The spatial resolution of the LHM was arbitrarily set as 15.2 m. This spatial resolution was chosen to detect landslide occurrences that fall between local-scale (grid size of 5 m or less) landslide occurrences and regional scale (grid size of 20 m -30 m) landslide events (Kakavas and Nikolakopoulos 2021). The slope angle map was created using the 1.5 m LiDAR DEM in ArcGIS. Therefore, the resolution of the slope map was 1.5 m. Consequently, the slope map was resampled to a 15.2 m cell grid using a bilinear resampling technique. The bilinear resampling technique performs a bilinear interpolation and determines the new value of a cell based on a weighted distance average of the four nearest input cell centers. The bilinear resampling technique is useful for continuous data

and will cause some smoothing of the data (ArcGIS, 2021). Figure 2-10 shows the resampled slope angles used in the Eq. 5, which ranges from 0 degree angle to 57 degree angle in the study site.

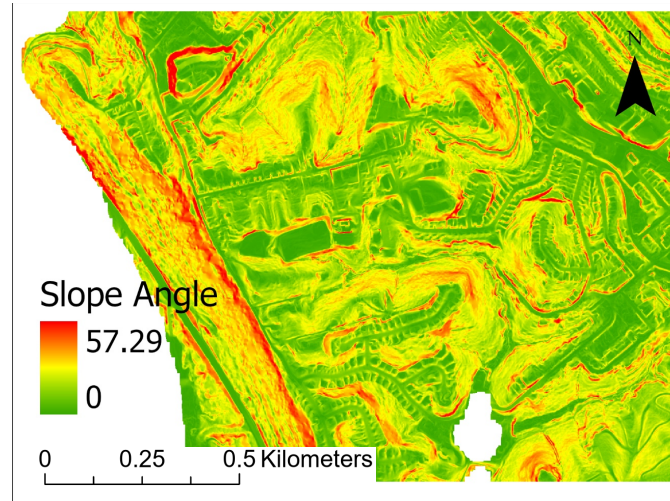


Figure 2-10. Resampled slope angle from 1.5 m to 15.2 m grid cells (taken from 1.5 m LiDAR DEM) for the study area in vicinity of Site 6294 and 6501 in northern Kentucky.

Depth to bedrock,  $H_{ss}$ , was assumed to be 4.5 m in the study region as a typical depth to the bedrock value taken from the geotechnical reports in the study area and used as the slip surface depth in Eq. 5. Bittelli et al. (2012) pointed out that the infinite slope model often assumes that the failure plane is coincident with the soil–bedrock interface. The unit weight of the soil,  $\gamma$  was assumed to be 18.8 kN/m<sup>3</sup>. A summary of the soil shear strength and slope parameters used for the study sites is given in Table 2-5.

Table 2-5. Soil properties used in the general factor of safety equation as presented in Eq. 5 at the study sites.

SITE ID	County	$\beta$ (deg)	$H_{ss}$ (m)	$\phi'$ (deg)	Failure Date
6294	Campbell	42.03	4.5	28.17	2/4/2016
6501	Campbell	23.1	4.5	28.17	3/30/2018
6396	Johnson	34.39	4.5	27.26	1/24/2017
6458	Pike	42.19	4.5	22.87	2/10/2018

All the parameters in Eq. 5 are constant except for  $\sigma^s$  = suction stress, which is a function of hydrologic parameters such as volumetric water content and matric suction. These parameters enable the LHM to be dynamic.

#### 2.4.2 Estimation of Matric Suction

As was previously discussed, soil matric suction was defined using the van Genuchten (1980) soil water characteristic curve (SWCC) model. The van Genuchten (1980) model parameters  $\theta_s$ ,  $\theta_r$ ,  $\alpha$ ,  $n$  for the study area were estimated using the Rosetta online tool (Rosetta, 2021). These parameters were estimated according to Schaap et al. (2001) pedotransfer function, which utilizes the percentages of sand, silt, clay and bulk density, all of which was extracted in ArcGIS from the NRCS WSS soil database. These data are shown in Table 2-6. Figure 2-11 represents the sand, silt, clay percentage used in estimating the hydrologic parameters for the Eq. 5. However, if the estimated  $\theta_s$  from the Rosetta tool was less than the highest value of  $\theta$  obtained from the satellite volumetric water content, then  $\theta_s$  was replaced by the highest value of  $\theta$ . It is observed from Figure 2-11 that the study area mostly consists of silt and clayey silt soil.

Table 2-6. Extraction of hydrologic parameter using the Rosetta online tool.

Input to Rosetta Online Tool				Output from Rosetta Online Tool				
Sand%	Silt%	Clay%	Bulk density (g/cm <sup>3</sup> )	$\theta_r$	$\theta_s$	$\alpha$ (1/cm)	n	m
26.8	48.7	24.5	1.67	0.09133	0.35042	0.00655	1.33572	0.25134
13.2	47.5	39.3	1.43	0.11740	0.43786	0.00575	1.34640	0.25728
12.1	44.2	43.6	1.55	0.12251	0.41273	0.00645	1.29802	0.22960
12.9	48.9	38.2	1.44	0.11608	0.43342	0.00557	1.35167	0.26017
13	44.2	42.7	1.55	0.12128	0.41092	0.00642	1.30076	0.23122
9.4	64.2	26.4	1.48	0.09860	0.40979	0.00420	1.42828	0.29986

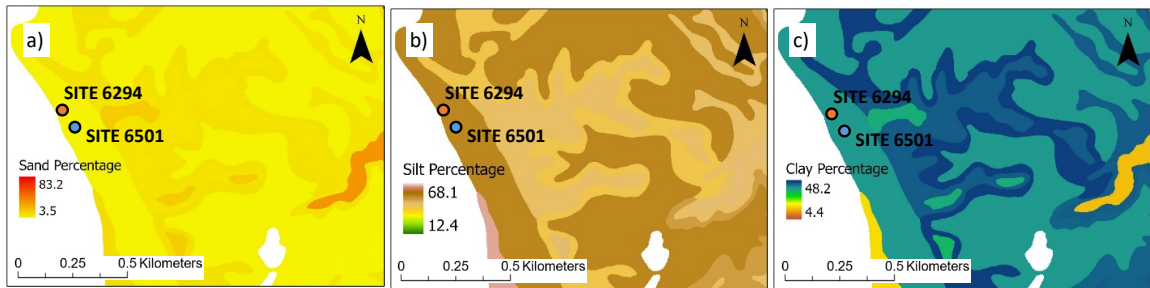


Figure 2-11. Soil data obtained from the NRCS WSS for the study area in vicinity of Site 6294 and Site 6501 in northern Kentucky: (a) sand percentage (b) silt percentage (c) clay percentage. The data were consequently used to estimate van Genuchten (1980) model parameters in Rosetta online tool.

#### 2.4.2.1 Volumetric Water Content Satellite Data Acquisition

Volumetric water content data from the National Aeronautics and Space Administration (NASA) Earth and Precipitation Satellite Missions were used for a time span of six weeks; three weeks prior to the reported landslide failure dates and three weeks post the landslide failure dates for each study area. Time-series data for satellite measured soil moisture were acquired from the NASA Soil Moisture Active Passive (SMAP) Earth satellite mission. The SMAP data were accessed and acquired using the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) tool. SMAP provides four data products one of which is the Level 4 (SMAP L4\_SM) product. This product is a model-derived value-added product obtained by merging SMAP observations with estimates from a land surface model in a data assimilation system. The land surface model component of the assimilation system is driven with observations-based meteorological forcing data,



including precipitation, which is the most important driver for soil moisture (Reichle et al. 2011). The model-derived product produces three-hourly estimates of surface and root zone soil moistures (to a depth of 100 cm) at a 9 km gridded spatial resolution with a data availability latency of 7 to 14 days (Chan et al. 2016). The SMAP L4\_SM product was used for this study. The multiple three-hourly soil moistures reported over a 24-hour period were averaged to provide a daily value of soil moisture. Figure 2-12 shows the SMAP data in the northern Kentucky area temporally and spatially. Figure 2-12(a) shows 9 km spatial resolution grids in Campbell County, over which the soil moisture data is uniform. The volumetric water content ranges from 0.42 to 0.51 in Campbell County [Figure 2-12(a)] from January 14, 2016 to February 25, 2016 during which the failure occurred. Figure 2-12(b) shows the variation of soil moisture change with time at Site 6294. The data were reported three weeks before and three weeks after the failure date.

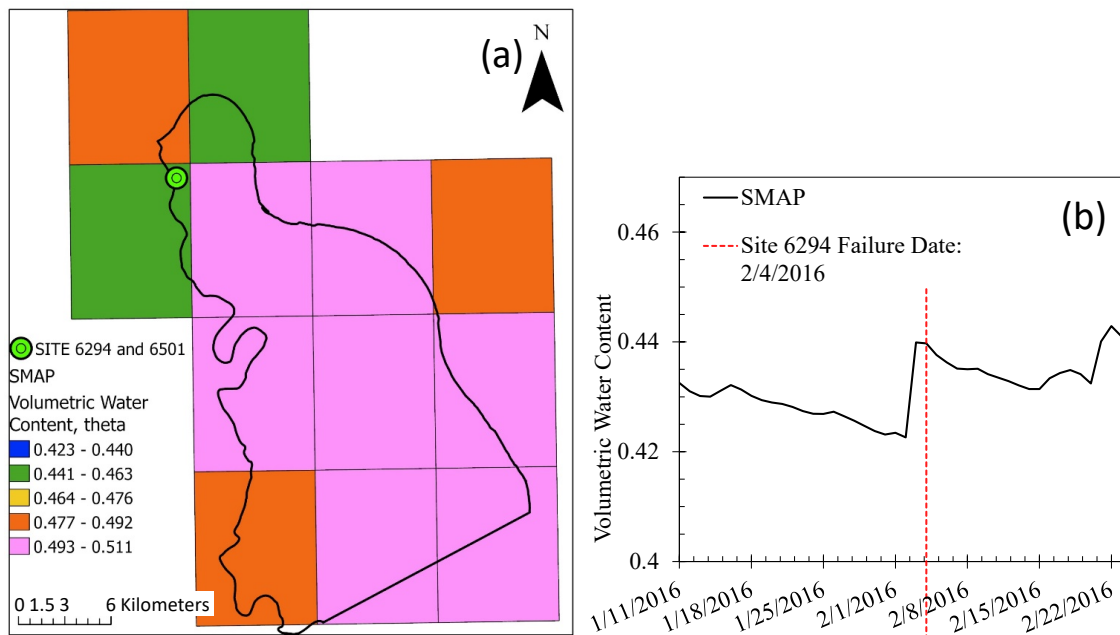


Figure 2-12. Temporal and Spatial SMAP data for the northern Kentucky area: (a) Spatial grid of volumetric water content from SMAP in Campbell County, Kentucky (b) volumetric water content at Site 6294 over six-week period during which the failure occurred.

### 2.4.3 Performance and Validation of Landslide Hazard Map

The soil and hydrologic parameters that remain constant ( $\beta, \phi', \theta_s, \theta_r, \alpha, n, m, H_{ss}$ ) and the temporally varying daily volumetric water content at each study site were used in the Eq. 5 to generate the daily factor of safety values for the given investigative period (i.e.,

three weeks prior to the landslide, and three weeks post the landslide). The factor of safety values were validated using the four study sites to assess whether the LHM correctly mapped the landslides spatially and temporally.

The LHM covered 300 m by 300 m area with a spatial resolution of 15.2 m at each study site. The validation was made by overlaying the known failure area with the created LHM temporally and spatially at each site. Table 2-7 shows the factor of safety values a week before the failure date, failure date, and a week after the failure date at each study site. Overall, the landslide hazard map indicated the areas where landslide occurred as having factor of safety values ranging from 1.15 -1.73. We observed that these factor of safety values exceed a value of 1.0 (i.e., the value at incipient failure). This exceedance is most likely due to geomorphic and geotechnical behavior that is too complicated to be captured by a simple subaerial infinite slope model. However, the general trend shows that the failures occurred as the factor of safety values reached their lowest levels.

Table 2-7. Factor of safety at the study sites at different times.

Site ID	Factor of Safety		
	7 days before	Failure Day	7 days after
6396	1.41	1.15	1.33
6458	3.13	1.27	1.02
6294	1.71	1.29	1.5
6501	2.17	1.73	1.58

#### 2.4.3.1 Validation of Model Performance at Sites in which the LHM Model Performed Optimally

Site 6396 in eastern Kentucky and Site 6294 in northern Kentucky show similar results that identified the landslides correctly temporally and spatially. The LHM derived from the Eq. 5 for Site 6396 and Site 6294 indicates that only areas around these sites showed a low factor of safety on the date the actual failures occurred. The LHM was symbolized with different colors to illustrate the areas with different factor of safety values by the break points based on factor of safety values; values lower than the factor of safety value on the failure date (Table 2-7) for each site (red), values between this value and 1.75 (orange), values between 1.751 and 2.5 (yellow), values between 2.51 and 3.5 (light green), and values  $\geq 3.51$  (dark green).

Figure 2-13 shows the LHM the factor of safety evolution of the Site 6396 for the dates a week before the failure [Figure 2-13(a)], the failure date [Figure 2-13(b)] and a week after the failure [Figure 2-13(c)]. As seen in Figure 2-13(b), only the area around the failure location was correctly identified with the red color grid indicating the low factor safety of 1.15 on the day it failed, and everywhere else, the factor of safety is greater than 1.15.

Also, the LHM for a week before the failure [Figure 2-13(a)] shows that the factor of safety is more than 1.15 everywhere in the map, and the same is shown in Figure 2-13(c) a week after the failure capturing the evolution of the landslide area. Figure 2-13(d) illustrates the factor of safety values at Site 6396 over a time span of six weeks covering three weeks prior and three weeks after the failure. The factor of safety reaches 1.15 on the failure date which was the lowest value during the time span [Figure 2-13(d)] at the Site 6396.

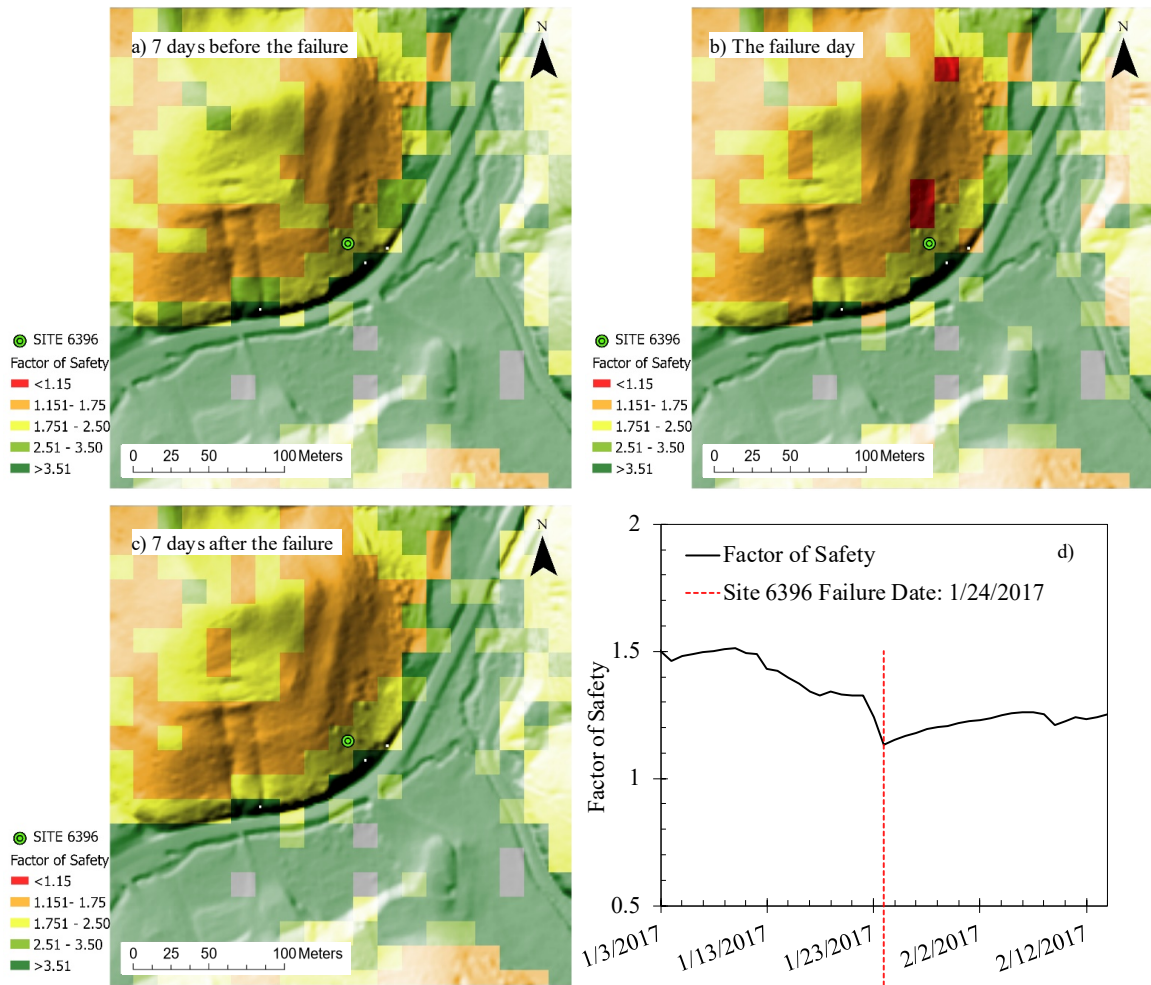


Figure 2-13. The LHM for Site 6396: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety at Site 6396 between January 3, 2017 and February 12, 2017 during which the failure happened.

Figure 2-14 validates the LHM showing that only the location where the landslide at Site 6294 happened indicates the lowest factor of safety in red area on the failure date and it matches the low factor of safety in the time graph spanning six weeks [Figure (2-14.d)]. Similar to the Site 6396, Figure 2-14 illustrates the evolution of the factor of safety values leading to the landslide, the landslide and the after the landslide.

Figure 2-13(d) and Figure 2-14(d) illustrate that the factor of safety value lowers to 1.15 and 1.29 (Table 2-7) when the failure happened, and they correctly show the failure time in the period of six weeks. The Figure 2-14(d) also shows that the factor of safety can reach less than 1.29 after the failure date, for the Eq. 5 accounts for the soil moisture, and the soil

moisture can increase during and post the failure. It is observed from the Figure 2-13(d) and Figure 2-14(d) that the failure happened after a threshold point of 1.15 and 1.29 in factor safety values for the Site 6396 and Site 6294, respectively. Furthermore, the factor of safety value at Site 6294 even lowered to 0.5 as the soil moisture increased to the highest after 10 days of the failure date. It is inferred that the lowest factor of safety does not mean the failure happens at this point in time, rather it is the certain factor of safety threshold, which was calculated as 1.15 and 1.29 for the Site 6396 and 6294, the soil has to reach for the failure to occur. Seed at al. (1989) hypothesized that during and post slope failure, the pore water pressure build-up in granular materials under undrained conditions goes up reducing the strength of the soil which corroborates the behavior of the Figure 2-14(d).

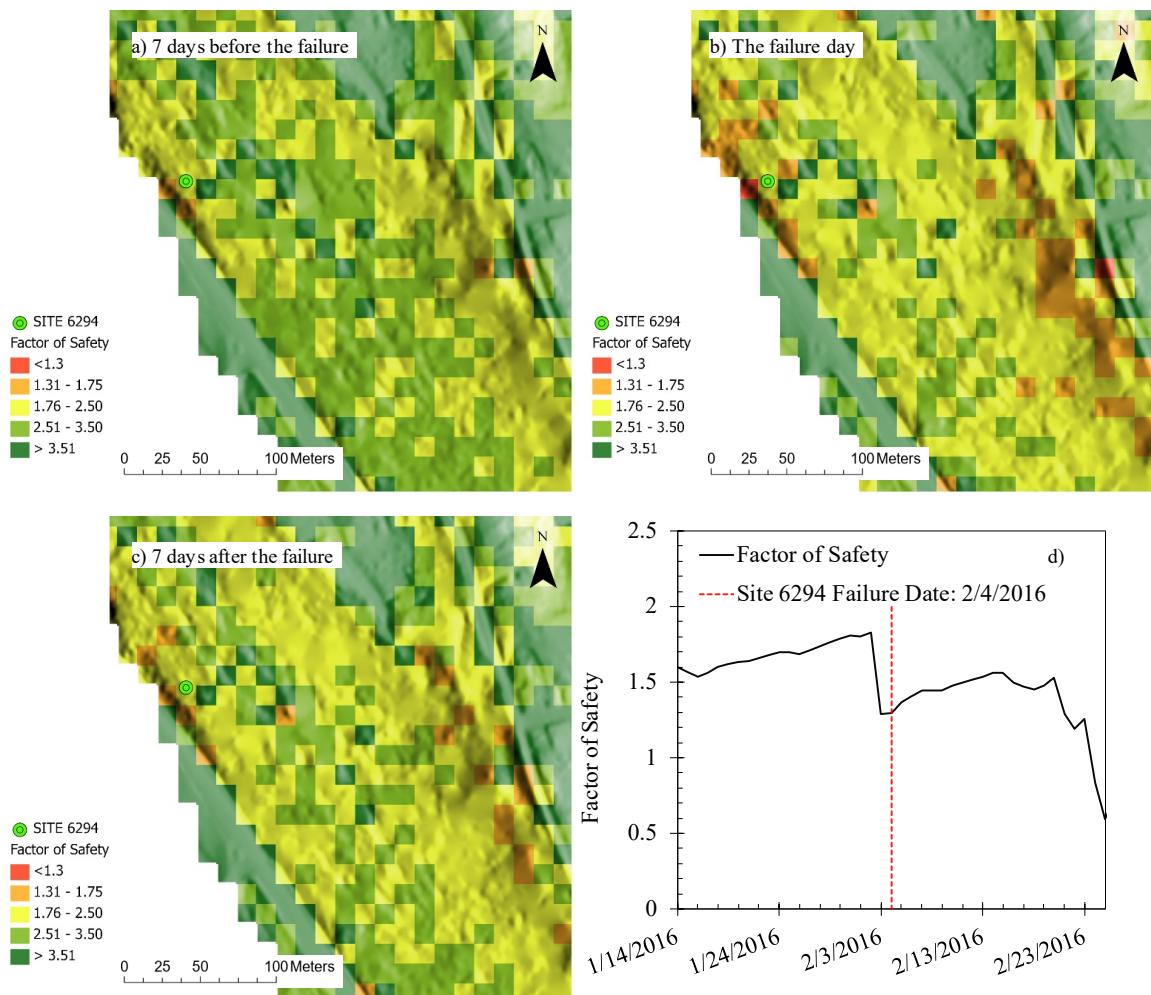


Figure 2-14. The LHM for Site 6294: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between January 14, 2016 and February 23, 2016 during which the failure happened.

#### **2.4.3.2 Validation of Model Performance at Sites in which the LHM Model Performed Slightly Less than Optimal**

Figure 2-15 shows that the factor of safety was 1.73 at the Site 6501 on the failure date, and four red grids near where the landslide at the Site 6294 occurred show the factor of safety less than 1.3 [Figure 2-15(b)]. However, these areas failed in 2016 and remediated and stabilized since they were next to the highway, so the slope angles and the soil properties are most likely not the same any longer. Therefore, the red grids are not representative of the conditions at the Site 6294 for the time shown in Figure 2-15(b). However, Figure 2-15(c) shows that the LHM for a week after the failure shows the lower factor of safety around the failure area at Site 6501, for the soil moisture increased post the failure date. The red areas upslope of the failure area [Figure 2-15(c)] might have been failed, but since the red areas are in the wooded area up the hill, they were not verified. Figure 2-14 (d) illustrates the factor of safety values at the Site 6501 over the time span of six weeks, and it shows that the factor of safety lowered from 1.73 on the failure date to 1.26 five days after the failure due to the post soil moisture increase.

Figure 2-16 correctly shows the failure location with factor of safety of 1.27 on the failure date. However, Figure 2-16(b) shows more areas with similar or less factor of safety values than the Site 6458. Some of the areas have very steep angles (45 degree and more); therefore, assumed to be sloped rock surface. The repose angle of sand is 45 degrees (Glover, 1999) at the max in the nature, and the areas that had more than 45 degrees of angles were assumed to be rock surface areas in which the factor of safety values become null because the Eq. 5 is not applicable to rocks.

Figure 2-16(c) indicates lower factor of safety values for the map area than that of the failure date [Figure 2-16(b)], and we argue that the sloped rock surface area is existent in the area, the actual internal angle in the area is stronger than the ones used in the Eq. 5 to create the map, the actual depth to the bedrock shallower than 4.5 m, all of which could potentially contribute to the red areas of the map [Figure 2-16(b) and 2-16(c)]. The friction angle of the soil was assumed to be uniform in the study, but the soil strength of slope materials is spatially heterogeneous, as the slope material is produced by a natural process (Chowdhury et al. 2010). In addition, it is observed from Figure 2-16 (d) that the factor of

safety value kept lowering after the failure date indicating the soil moisture kept increasing for the three weeks after the failure.

Overall, the factor of safety map aligns correctly with the study sites temporally and spatially, with some additional areas indicating the false landslide warnings such as areas away from the landslide location at the study sites in red color [Figure 2-16(b)]. We argue that the reasons can be the area is composed of rock surface or outcropping having more than 45 degrees, the heterogeneity of the soil; friction angle of the soil the higher than estimated friction angle, and the lower depth to the bedrock than the estimated depth used in creation of the map. The LHM for each study site indicates that using the Eq.(5), good insights can be gained into the likelihood of landslide temporally and spatially based on the soil moisture evolution and coarse soil properties.

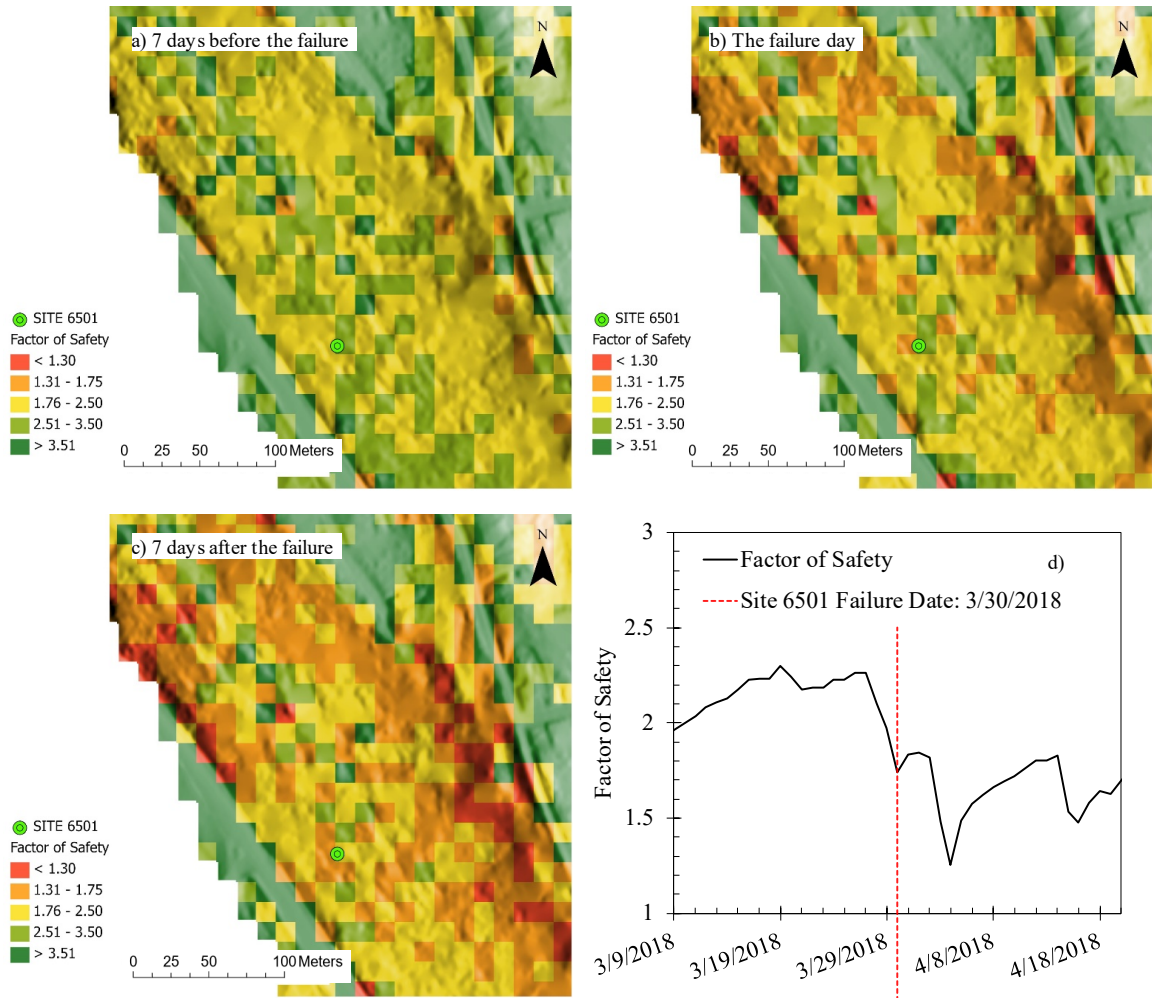


Figure 2-15. The LHM for Site 6501: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between March 9, 2018 and April 18, 2018 during which the failure happened.



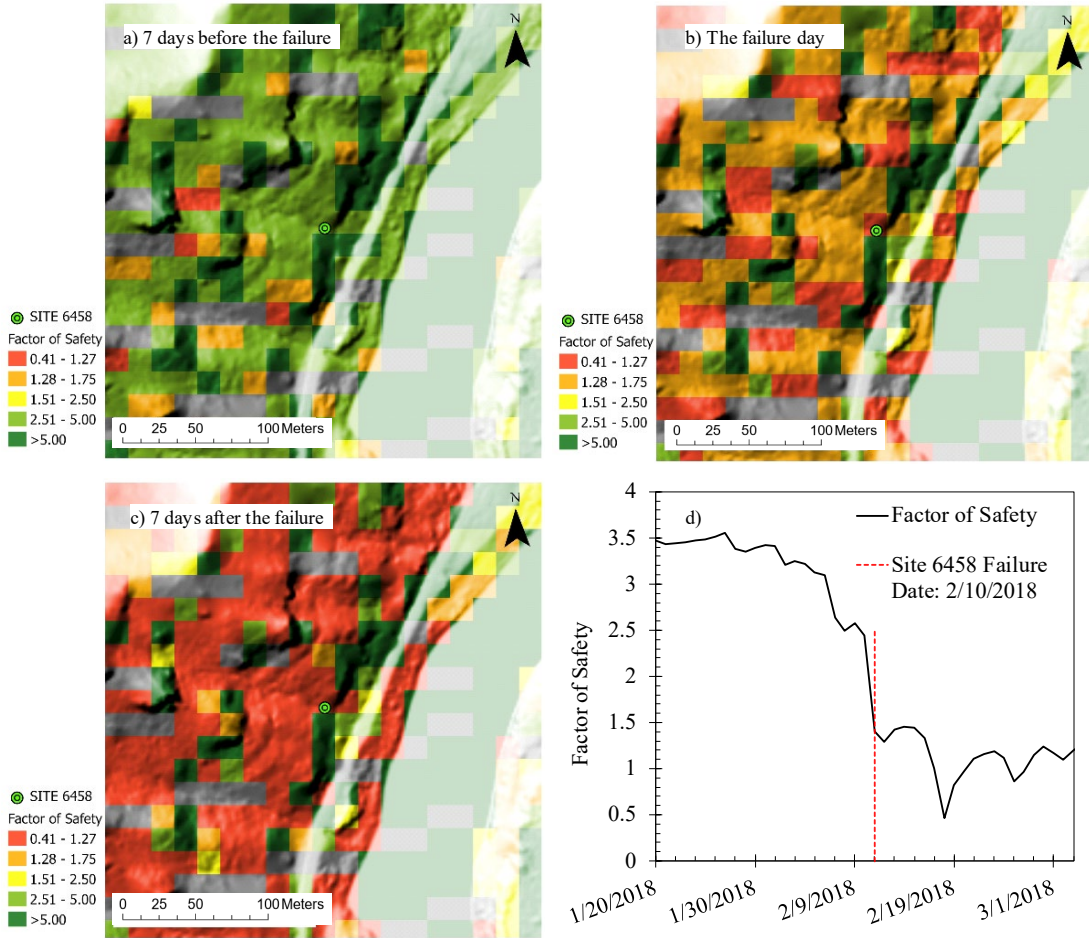


Figure 2-16. The LHM for Site 6458: (a) a week before the failure date; (b) on the failure date; (c) a week after the failure date; and (d) the factor of safety between January 20, 2018 and March 1, 2018 during which the failure happened.

#### 2.4.3.3 Comparison of the landslide hazard map to the landslide susceptibility map

Direct comparison between the LHM at Site 6501 on the date its lowest factor of safety occurred (4/4/2018) and the LSM were conducted to assess how these two models map landslides compared to the one another. In addition, direct comparison between the LHM at Site 6501 on 3/20/2018, and the LSM were carried out the same exact way to see the change in the landslide occurrence. For the comparison to be made, a spatial resolution of the maps had to be the same. Therefore, resampling of LSM with 91.4 m spatial resolution was done converting it to 15.2 m spatial resolution to match the resolution of the LHM. The resampled spatial resolution of LSM matches that of the LHM (Figure 2-7) 15.2 m allowing the direct comparison of the values quantitatively.

The quantitative comparison between the probability map and factor of safety map was carried out by setting the factor of safety value less than 1.75 to be likely to be a landslide because our study sites (Table 2-7) show that the factor of safety ranged from 1.15 to 1.75 on the failure dates and by setting probability value of 0.40 and higher to be a landslide in LSM because 87 percent (Table 2-4) of actual landslides in inventory belonged to this category.

#### 2.4.3.4 Comparison of LHM at Site 6501 on 4/4/2018 and LSM.

Figure 2-17 shows the comparison of the maps visually and Table 2-8 compares the maps quantitatively, grid-by-grid in the confusion matrix indicating that the LSM models 46 percent more probable locations of landslide occurrences (3805 occurrences versus 2610 occurrences) than that modeled by the LHM. A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. In this study, it is used to compare the model classes (landslide or non-landslide) on grid to grid. Table 2-8 shows that the 1633 grids on the maps are classified as landslides by both LHM and LSM, and 4726 grids are classified as non-landslides by both of the models out of 9508 total grids on the map (Figure 2-17).

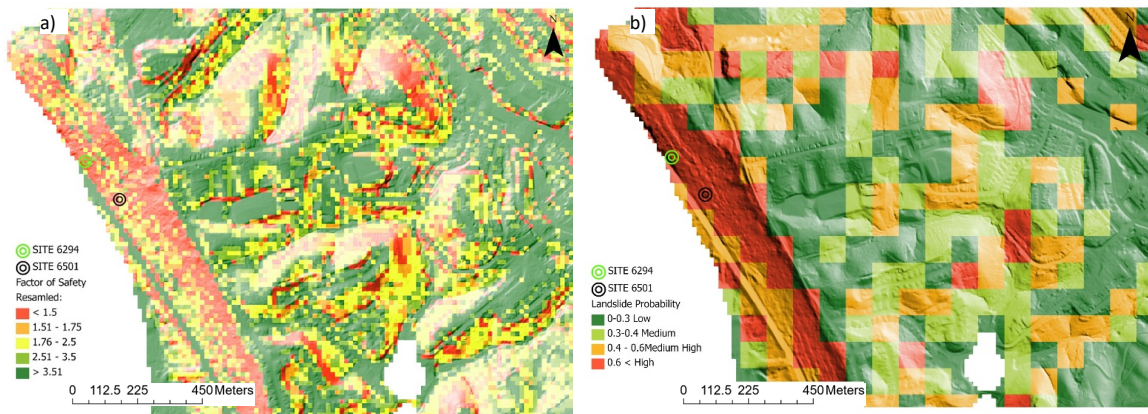


Figure 2-17. Comparison of (a) the lowest value of LHM in 5 years at Site 6501 on 4/4/2018 and (b) the LSM.

Table 2-8. Confusion matrix of LHM (4/4/2018) vs LSM in part of northern Kentucky site area (Figure 2-17).

		LSM	
		Landslide	Non-Landslide
LHM	Landslide	1633	977
	Non-Landslide	2172	4726

**2.4.3.5 Comparison of LHM at Site 6501 on 3/20/2018 and LSM.**

Figure 2-18 shows the comparison of the LHM at Site 6501 on 3/20/2018, two weeks before the lowest factor of safety date (4/4/2018), and LSM visually. Table 2-9 compares the maps quantitatively grid by grid in confusion matrix the same way that LHM on 4/4/2018 was compared to LSM. Table 2-9 indicates that the LSM models 48 times more probably locations of landslide occurrences (3805 occurrences vs 77 occurrences) than that modeled by the LHM. Furthermore, it shows that 57 grids on the maps are classified as landslides by both LHM and LSM, and 5683 grids are classified as non-landslides by both of the models out of 9508 total grids on the map (Figure 2-18) for 3/20/2018.

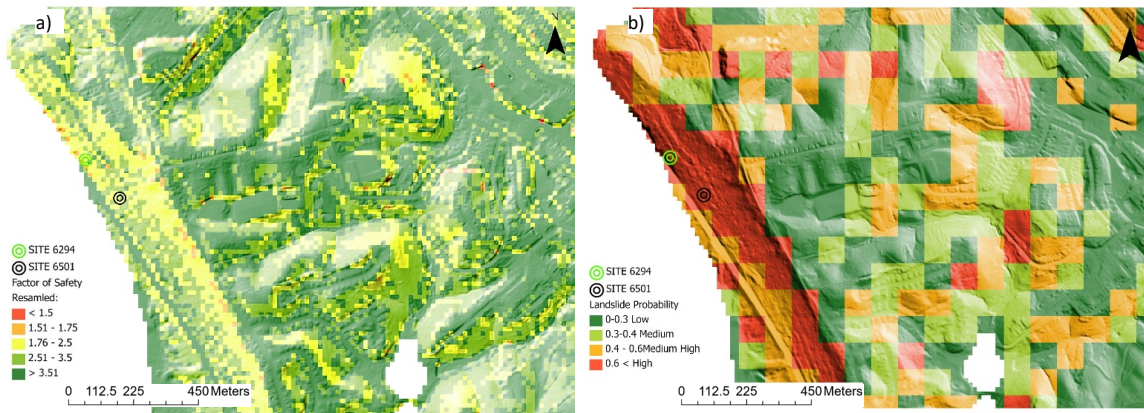


Figure 2-18. Comparison of (a) LHM at Site 6501 on 3/20/2018 and (b) the LSM.

Table 2-9. Confusion matrix of LHM (3/20/2018) vs LSM in part of northern Kentucky site area (Figure 2-18).

		LSM	
		Landslide	Non-Landslide
LHM	Landslide	57	20
	Non-Landslide	3748	5683

These comparisons corroborate the main statement that the LHM derived from the factor of safety map shows when and where landslides are more likely to occur and gives us more insights into the evolution of a given state of factor of safety of soil for any time period and compared well to the LSM derived from the machine learning model (logistic regression), which was based on the actual landslide failures from the past.

## 2.5 Discussion

During the last decade, there has been an increase in spatial and temporal data available to the public, such as high resolution DEMs and national soil property databases, which can be used for probabilistic and deterministic slope stability analysis. These available data can be prepared in Geographic Information System (GIS) for a regional-scale landslide susceptibility analysis. There is a need to better understand, identify and assess landslide susceptibility and develop landslide hazard map for communities that are in the close vicinity of the landslide zones. Furthermore, the public can use the LHM in their decision-making process to minimize the potential risk of injuries and damages caused by landslide.

The limitation for this study includes that the soil properties used in the Eq. 5 to create the LHM;  $\phi'$ ,  $H_{ss}$ ,  $c'$ , and hydrologic parameters;  $\theta_s$ ,  $\theta_r$ ,  $\alpha$ ,  $n$  were all estimated using soil database from WSS which was made in 1:20000 scale (nrcs.usda.gov) not representing the soil variance in small scale. The internal friction angles of sand for Site 6458 and 6396 in eastern Kentucky area were taken from the geotechnical reports for the project located within 3.2 km radius of the study sites, so the actual friction angle and the variance of the soil strength in the study site was not verified. However, the friction angles used in the location produced LHMs that matched the location and the time of the failure sites. The soil moisture data from SMAP has a spatial resolution of nine km, and the uniform soil moisture data for the whole nine km grid was used in the Eq. 5 for the creation of the LHM

for part of northern Kentucky area in the vicinity of Site 6294 and 6501, assuming that there was no variation. Also, the depth of the bedrock was taken from the average depth from the geotechnical engineering reports in northern Kentucky area, so the actual depth to the bedrock at the sites was not in-situ measurement. Uncertainty in model parameter evaluation has been recognized as an important cause of mismatch between simulated and observed distributions of landslide occurrence (Burton et al., 1998). Therefore, when performing susceptibility analysis with physically based models, spatial variability and uncertainties in ground conditions must be considered (Bittelli et al. 2012).

LSM reveals the probability of landslide occurrences whereas LHM shows the temporal and spatial evolution of landslide occurrences. The general procedure used in this study can be generalized with a certain degree of confidence to assess the stability of the slope using satellite and remote sensing data that are available publicly without any measured in-situ soil properties. As this study demonstrated that the land hazard map at close-to-real time can be created using Eq. 5 (Lu and Godt 2008), using the available data, and the factor of safety threshold of 1.15 to 1.75 in this study correctly predicted the landslide spatially and temporally with a few additional areas indicating the false landslide warnings because of the uncertainty of the estimated values of the soil and the overburden soil at the area as previously shown in this paper. This paper shows the use of soil moisture data, the elevation data, and the soil property data, all accessible publicly gives a good insight about the assessment of the factor of safety against the landslide.

## **2.6 Conclusion**

In this study, we used the infinite slope factor of safety equation (Lu and Godt, 2008) utilizing satellite soil moisture data, high resolution LiDAR digital elevation map (DEM), and soil strength data derived from soil database from USDA to produce the LHM, and we validated the results using the four study sites with known landslides and the failure dates. The result of the LHM correctly indicated low factor of safety values at the study sites temporally and spatially. However, some additional areas with low factor of safety values at Site 6548 and 6501 where the occurrence of the landslide was not verified on the dates the landslide occurred was identified as a landslide in the LHM, and we argue that the area at the Site 6548 was composed of sporadic rock surfaces where it had more than 45 degree

slope angles, and the uncertainty of the estimated friction angle, the depth to the bedrock and the soil moisture data, all of which contributed to the false indication of the landslide areas in the LHM for those sites. Over a long period of factor of safety graph of the sites, it indicated that use of the Eq. 5 could assess the landslide susceptibility in an area temporally and spatially instead of regional statistically produced LSM which was insensitive to a variation in the soil moisture. LSM was created using the logistic regression machine learning model using the variables that included six geomorphic variables of slope, aspect, curvature, elevation, roughness, and plan curvature, all extracted from 1.5 m LiDAR DEM, eight variables of clay percent, silt percent, sand percent, saturated hydraulic conductivity, available water capacity, one third bar pressure water content, plasticity index, liquid limit, all extracted from the physical property of soil database from (WSS), and a variable of 2016 National Land Cover Database (NLCD) product suite. The result of LSM produced the AUC of 0.841. The comparisons of the LHM and LSM were done quantitatively using the confusion matrix (Table 2-8 and Table 2-9), which showed that even on the date with the highest soil moisture content, the LHM still modeled less occurrence of landslides and verified correctly against the study sites. The validation of LHM for each study site indicated that using the infinite slope factor of safety equation (Lu and Godt, 2008), we can model the likelihood of landslide temporally and spatially based on the soil moisture evolution of the area, the soil properties and DEM data all available publicly rather than statistically produced LSM.

---

### **3 Using Multi-Temporal UAV Images to Detect Land Movement and to Estimate Soil Moisture Data Using Machine Learning Model.**

#### **3.1 Introduction**

Land movement and soil moisture of a slope are good indicators of the slope stability, and used as a forecasting of the slope failures and landslides. Landslides are common and dangerous natural hazards that occur worldwide, often causing severe direct impacts on human lives, public and private properties, and lifelines (Klose et al. 2014). Mapping and displacement monitoring of unstable slopes is a crucial issue for the prevention and assessment of hazards (Rossi et al. 2018). Landslide forecasting is an early warning system that helps to predict a slope failure spatially and temporally. Different approaches exist to predict the location and the time of the slope failures, depending on which parameter is adopted to indicate probable imminent failure. The most reliable (and most commonly used) parameters for forecasting the time of failure are the slope displacement and its derivatives; velocity, and acceleration (Intrieri et al. 2019). These kinematic parameters are directly related with the stability conditions of the moving mass, and there are a number of methods and approaches to monitor them in real time (Intrieri et al. 2019, Lacasse and Nadim 2009, Raspini 2018). Land movement monitoring can be carried out with various techniques and approaches. They are mainly divided into three categories; in-ground monitoring instruments, geodetic, and remote sensing techniques. In-ground instruments include inclinometer, extensometers, and piezometers, and strain gages. Geodetic techniques are based on the survey of a network of selected points in the landslide body and displacements computation using a total station, a GPS survey equipment that gives very high accuracies. Remote sensing techniques include space-borne, air-borne platforms that have on-board passive or active sensors. Remote sensing techniques are effective tools to rapidly obtain spatially distributed information on landslide kinematics (Delacourt et al. 2007), and can be operational from spaceborne, airborne, and ground-based platforms (Rossi et al. 2018). Synthetic Aperture Radar (SAR) (space-borne) active sensor can observe the Earth's surface at any time of the day or night, regardless of weather and environmental conditions. SAR has the advantage of operating at wavelengths not impeded by cloud cover or lack of illumination (European Space, 2021). Several studies including works of Metternicht et al. (1998), Tofani et al. (2013), and Scaioni et al. (2014) used SAR

data successfully to detect the change in land movement in sub-centimeter scale using Interferometric Synthetic Aperture Radar (InSAR) techniques. InSAR techniques have become widely used and broadly recognized tools for landslide mapping and monitoring. Differential Interferometry (DInSAR) is a technique that exploits the interferometric phase difference between combinations of SAR images with the orbital information to detect a point displacement in a line of sight direction. The primary goal of such techniques is to derive an interferogram that expresses the phase difference for each image pixel between two passages of the satellite on the same area. By using multiple SAR images and the interferograms derived, it is possible to retrieve the temporal evolution of ground displacements over time (Solari et al. 2020).

Unmanned aerial vehicle (UAV) is used in an air-borne remote sensing technology to capture aerial images with high resolution and high efficiency. UAV-based remote sensing has the following advantages: real-time applicability, flexible survey planning, high resolution, low cost, and it can collect information in dangerous environments without risk (Changchun et al. 2010, Rossi et al. 2018). UAV can be mounted with many different type of sensor such as Light Detection and Ranging (LiDAR), hyperspectral, thermal sensor to analyze the surface information and its topography. The land movement was detected using UAV in studies including the works of Yang et al. (2020), Turk (2018), and Lucieer et al. (2014), and these studies demonstrated the effectiveness of UAV images using precise orthorectification, image co-registration, and image correlation technique. Co-registration of Optically Sensed

Images and Correlation COSI-Corr (Ayoub et al. 2009) technique was used in these studies successfully demonstrating the land movement can be detected using optical truecolor imagery (RGB image) taken by UAV.

Use of multi-temporal images collected by UAV in extraction of soil moisture data was shown in the works of Lu et al. (2020), and Ge et al. (2019) employing machine learning models to extract soil moisture from visible and hyperspectral images. Antecedent soil moisture conditions play a crucial role in the initiation of landslides (Zhuo et al. 2019). It is important to use the actual soil moisture data instead of rainfall data for developing a limit equilibrium factor of safety to assess the stability of a slope. The antecedent soil



moisture and evapotranspiration govern the actual soil moisture regime and thus govern the factor of safety. Zanetti et al. (2015) used RGB images and artificial neural networks to estimate soil moisture. Dos Santos et al. (2014) established different linear models using RGB, HSV (Hue, Saturation, Value), and the digital number of a panchromatic image to estimate soil moisture in each type of soil (Lu et al. 2020).

In this paper, we aimed to demonstrate the use of UAV images to detect the land movement, to extract soil parameters in the area where landslide occurred. These findings can be used to evaluate and assess the current condition of slope stability as means of forecasting. UAV with an optical digital camera was used to capture 20 multi-temporal RGB images over three month period over an area, where landslide occurred and remediated in 1995. The UAV images were processed and ortho-rectified employing digital photogrammetric algorithms in PIX4D (Pix4D S.A., Lausanne, Switzerland) software. Once the images were ortho-rectified, they were co-registered pixel to pixel so that the images can be used in the image-correlation algorithm (COSI-Corr) to detect any land movement. Furthermore, the statistics from the images were extracted and input into machine learning model to extract soil parameters. This work attempted to demonstrate the process flow that can use UAV images to obtain land detection and soil parameter to gain a good insight into the close to near-time state of slope stability.

### **3.2 Study area**

A site in Garrard County, Kentucky was chosen as a study site due to its documented landslide history and its close distance to our location. Garrard County is located in central Kentucky area, and it is characterized by steep hills, and deep valleys. This area is defined as the Hills of the Bluegrass physiographic region of Kentucky, and the area is composed of Upper Ordovician calcareous siltstone and shale of the Garrard and Clays Ferry Formations. Elevation ranges from 157 m in the northern part near the Kentucky River to 424 m in the southern part adjacent to Rockcastle County. The Hills of the Bluegrass belong to the Eden-Nicholson-Lowell soil series (Thompson and Poindexter 2005). The area is mostly composed of silty sand underlain by gray shale with some interbedded limestone.

The slope failure at the site was first reported on July 7, 1995, and the preliminary investigation was done on the same date. According to the engineering memorandum, a

head scarp appeared in the southbound lane of KY 39 that affected about 55 m of the roadway. The approximate area of the landslide was 0.011 sq. km. The embankment was about 20 m in height and the slope transitioned from a 2:1 to an approximate 12:1 or flatter near the bottom. Figure 3-1 shows the study area that was flown over by UAV and the extent of the landslide. The slope area was heavily vegetated with tall grasses and shrubs. The access to the slope was very limited which was one of the advantages of using UAV in inaccessible areas. The slope angles in the area ranged from 10 degrees to 36-degree slope angles along the downslope of the roadway.



Figure 3-1. Aerial view of the study site. Landslide occurred and remediated in 1995, Garrard County, Kentucky.

Images were taken between 1100 and 1600 (Eastern Time, USA) throughout the observation period for three months. Land cover type is classified as pasture/hay from National Land Cover Data (NLCD) 2016 map (Multi-Resolution, 2016).

### 3.3 UAV deployment and operation

The UAV used for this study was DJI Phantom 4 Advanced quadcopter for its low cost and ease of use with smart flying camera that was able to intelligently avoid obstacles during flights. Its camera is RGB camera that offers image quality with greater clarity, lower noise, and high resolution. Table 3-1 shows the specification of the UAV used in the study. The camera parameters were consistent for each UAV flights, so the image processing can be done with the same settings. The flight altitude was set at 62 m because the sub-centimeter spatial resolution of the images were desired to detect any land movement at sub-pixel

scale at the study site. The area for each image was about 54 m by 82 m with 80 percent overlapping between the adjacent images rendering a spatial resolution of 5 cm. Figure 3-2 shows the UAV and its flight path that was set in an automatic flight mission app used for photogrammetric data acquisition for each flight. Figure 3-2.a shows the UAV used, and its controller component.

Table 3-1. Technical specification of the UAV

Detail	Parameters	Value
Drone	Model	DJI Phantom 4 Advanced
	Weight	1368 g
	Dimensions	350 mm diagonally
	Max Speed	31 kmph
Camera	Camera	FC6310_8.8
	Image sensor	1" CMOS effective Pixel : 20M
	Lens	FOV 84° 8.8 mm/24 mm (35 mm format equivalent)
	Photo size	5472 x 3078
	Aperture	f/2.8
	ISO	3200
	Shutter speed	8-1/8000 s

### 3.3.1 Flight path

Automatic flight and easy-to-use remote controls for small UAVs enable convenient one-man operation of UAVs flying over field on-demand. Acquiring high-resolution aerial imagery requires autonomous flight on a pre-defined flight route, particularly when flying over wide fields. The forward and side overlap for the autonomous flight are highly critical for successfully constructing RGB orthoimages or reflectance map (Guan et al. 2019). The flight plan was carried out using 80 percent frontal overlap and 60 percent lateral overlap with the spatial resolution of 0.05 m. Consequently, based on the resolution and specification parameters of the RGB camera, an above ground level (AGL) altitude for the UAV was set to 62 m to acquire images. In order to ensure the consistent frontal and side overlap and keeping the resolution as desired, PIX4Dcapture app was used to acquire all

the images. PIX4Dcapture is a flight planning and image acquisition app developed for Android and iOS mobile operating system by Pix4D. Figure 3-2.c shows the aerial view of the flight plan used in acquiring the images. An average of 90 images were taken for each flight, and they were processed in image processing software later.

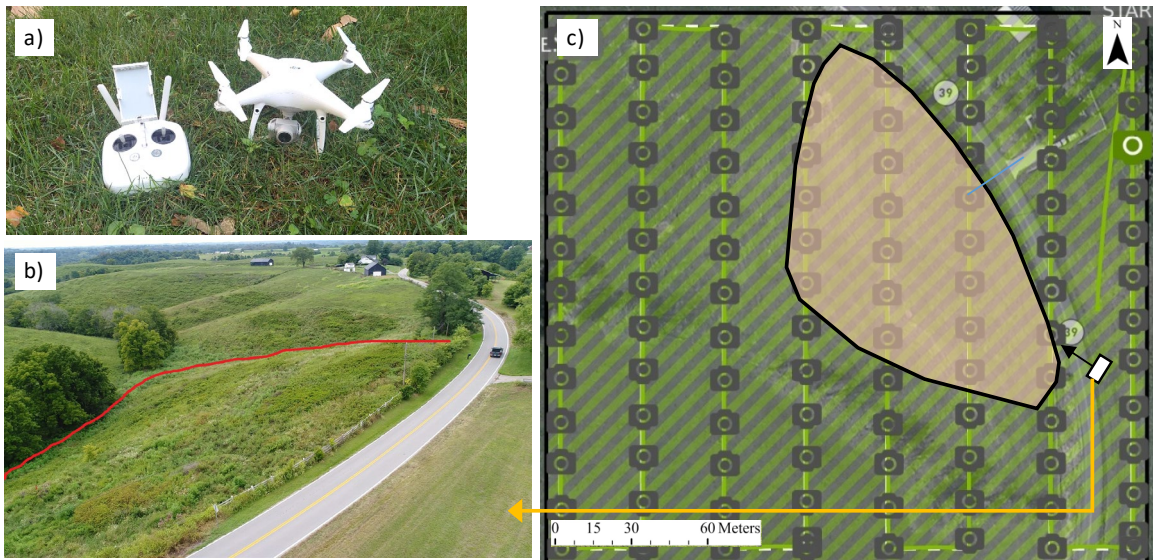


Figure 3-2. UAV and its flight path used in the study area (a) Phantom 4 advanced UAV equipped with a digital camera to collect images (b) study site from oblique angle from the look point shown in (c) and the landslide boundary extent in red line marker (c) aerial view of flight path for the UAV set in flight planning mobile app.

### 3.4 Image processing:

The acquired images can not be used in landslide detection or soil parameter extraction analysis until they are processed such that all pixels are in an accurate (x, y) position on the ground. Photogrammetry is a discipline, developed over many decades, for processing imagery to generate accurately georeferenced orthorectified images. Orthorectification is a process that removes the geometric distortions introduced during image capture and produces an image product that has planimetric geometry, like a map. Orthorectified imagery, also known as orthoimagery, is precisely registered to a ground coordinate system and the image scale is constant throughout the entire image (L3Harris Geospatial, 2021). The orthorectified images are processed to apply corrections for optical distortions from the sensor system, and apparent changes in the position of ground objects caused by the perspective of the sensor view angle and ground terrain (ArcGIS manual, <http://learn.arcgis.com/en/arcgis-imagery-book>). Common production workflows employ

structure from motion (SfM) and multi-view stereopsis (MVS) techniques to build digital surface model and to produce ortho-rectified images. All image processing was done in photogrammetry software Pix4D mapper (Pix4D, 2021). Pix4D mapper searches for matching points by analyzing all images using SIFT (Lowe 2004) feature matching technique with an improved version of the binary descriptors proposed in Strecha et al. (2011), which are very powerful to match keypoints quickly and accurately. Figure 3-3 shows one of the matching keypoint on the ground from different images in one flight. Those matching points as well as approximate values of the image position and orientation provided by the UAV autopilot were used in a bundle block adjustment (Triggs 1999) to reconstruct the exact position and orientation of the camera for every acquired image. Based on this reconstruction, the matching points were verified and their 3D coordinates calculated using GPS measurement from the UAV during flight. Those 3D points were interpolated to form a triangulated irregular network in order to obtain a DEM that was used to project every image pixel and to calculate the geo-referenced ortho-rectified images (Strecha et al. 2012).

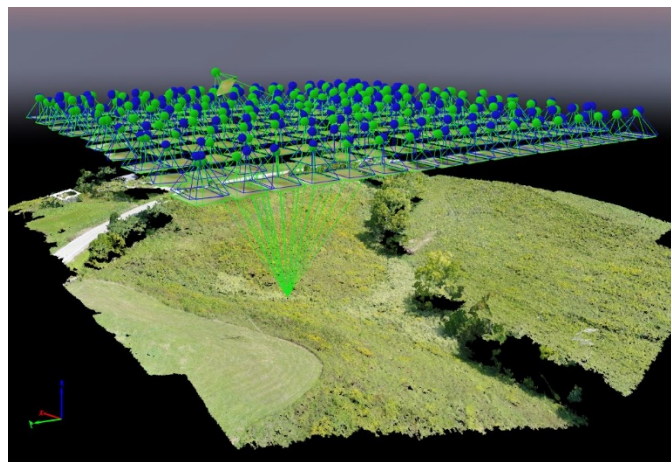


Figure 3-3. Matching point on the ground from different overlapping images in SIFT photogrammetry technique processed in Pix4D mapper.

Once the orthorectified images were processed, they were co-registered to have the same geometry so that every pixel in the images matched correctly.

Image co-registration is a process of geometrically aligning two or more images to integrate or fuse corresponding pixels that represent the same objects. Typically, the geometric relationship between images is obtained through a number of tie points that are matching

points in each image. Fully automatic tie point generation is possible through area-based matching technique. This technique compares the gray scale values of patches of two or more images and tries to find conjugate image locations based on similarity in those gray scale value patterns. The results of area-based matching largely depend upon the quality of the approximate relationship between the images. This is determined through traditional or pseudo rational polynomial coefficients (RPC) map information, or by using three or more tie points (L3HARRIS Geospatial, 2021.). Automatic co-registration process was carried out in ENVI 5.5 (Exelis Visual Information Solutions, Boulder, Colorado) remote sensing software.

Figure 3-4 shows the co-registration of two images using automatic registration tool. Once the multi-temporal images were orthorectified and co-registered, they were analyzed further to detect a land movement.

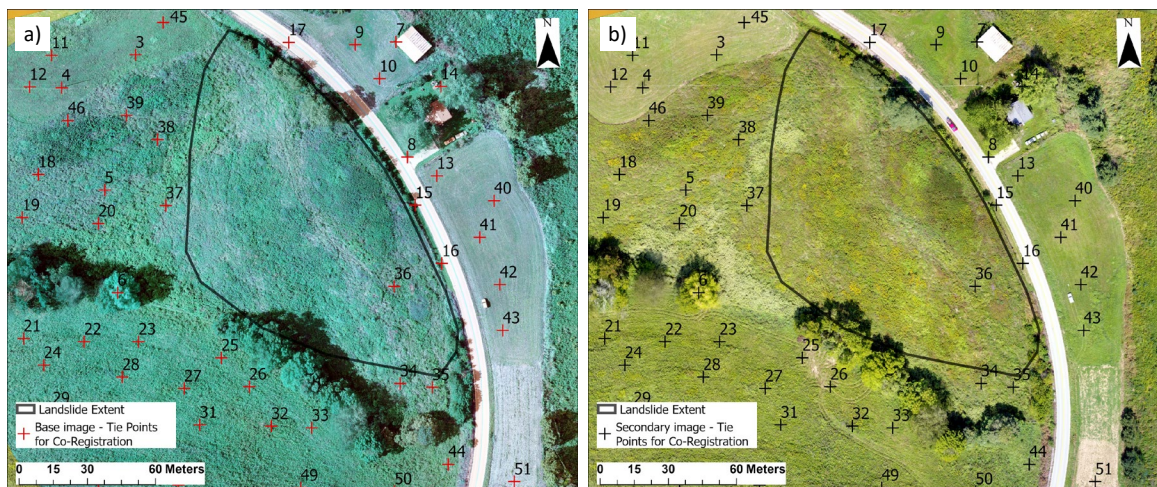


Figure 3-4. Coregistration of images were performed using area based matching technique in ENVI (a) tie points in the base image (8/11/2020) to which (b) tie points in secondary image (9/11/2020) were co-registered.

### 3.5 Research Methods

#### 3.5.1 Landslide movement detection

The co-registered images were further processed to detect any land movement employing image-correlation technique developed by LePrince et al. (2007) and Ayoub et al. (2009). The method is referred to as COSI-Corr: Co-registration of Optically Sensed Images and Correlation (Ayoub et al., 2009). This technique produces an automatic pixel-wise change

detection between two orthorectified and precisely co-registered optical images and was originally designed to measure displacements from seismic activity (Leprince et al. 2007). It correlates two single-band images of any optical remotely sensed images. Therefore, red band was used to detect the landslide movement detection. COSI-Corr uses an image kernel to compute the correlation between the two images (Lucieer et al. 2014). There are two correlators in COSI-Corr; frequency and statistical. The frequency correlator is Fourier based and is more accurate than the statistical one. However, this method is very sensitive to noise. We used statistical correlator due to the high noise in the images from seasonal color change. The statistical correlator maximizes the absolute value of the correlation coefficient and is coarser but more robust than the frequency one. The displacement algorithm in COSI-Corr requires a number of initial settings (Ayoub et al. 2009): (1) window size – the size in pixels of the patches that will be correlated in x and y direction; (2) step – determines the step in x and y direction in pixels between two sliding windows; (3) search range – sets the maximum distance in the x and y direction in pixels where the displacements to measure are to be searched (Lucieer et al. 2014). The window size for this analysis was set to 32x32 pixels, step 16x16 pixels, and search range 10x10 pixels.

We chose three pair images, each taken one-month apart for the period of 8/11/2020 to 11/10/2020 such that images taken on 8/11/20 and 9/11/20 formed a pair, an another pair of images taken 9/11/20 and 10/9/20, and the last pair of images taken 10/9/20 and 11/10/20 were correlated to detect a horizontal land movement in COSI-Corr. In addition to the horizontal movement detection, the vertical movement detection between 8/6/20 and 9/11/20 was carried out using SAR images using D-InSAR (Differential Interferometry) technique that exploits single interferometric SAR pair acquired by an SAR sensor. The result from D-InSAR analysis produced the displacement map in the direction of line of sight (LOS) of the satellite. This map does not capture the horizontal movement of land. In many scientific and commercial studies including works of Parker et al. (2017), Alshammari et al. (2018), and Yang et al. (2018), LOS measurements were converted into the vertical direction by projecting the data using the sensor incidence angle. This procedure neglected the horizontal components of motion that would also be mapped into the LOS. The assumption of a purely vertical motion field was mentioned in most of these studies (Fuhrmann and Garthwaite 2019).

### 3.5.2 Results of land displacement analysis

The result of correlation of the images using COSI-Corr produces three layers for each image pair; displacement layers in east-west (E/W) and north-south (N/S) directions and the signal/noise ratio (SNR). Positive values in the E/W and N/S layers represent that the slope moves east and north, respectively. Negative values indicate movements to the west and south. SNR values range from 0 to 1 higher values indicating high confidence in estimating displacements. The vector magnitude of the land movement was calculated using Euclidian distance;  $Dist = \sqrt{(E/W)^2 + (N/S)^2}$ . Figure 3-5 shows the horizontal land movement at the study site for each pair of images in Euclidian distance and the direction of the land movement in arrows. The displacement map break points in the color legend of Figure 3-5 was created using standard deviations from the mean. A rectangular area of 17 m x 37 m on the slope where the slope angle was the highest was selected as an area to get metrics of the displacement rather than the entire slope embankment. Table 3-2 shows that mean displacement for a pair dataset (8/11/20 and 9/11/20) has the biggest movement within the rectangular area, and the direction of the mean movement indicates that it pointed in the direction that was 90 degrees rotated in clockwise direction compared to the slope direction. The directional movement arrows in Figure 3-5.a show this movement. The mean Euclidean distance for this pair within the rectangular box was 9.23 cm. This movement was considerably higher than those of the other two pairs. Figure 3-5.b and Figure 3-5.c show the movements for the pair image of 10/09/20 - 9/11/20, and 11/10/20-10/09/20. The mean Euclidian distances were 3.03 cm and 0.23 cm, respectively. The movement direction of latter two pairs appeared to be aligned with the direction of the slope compared to the first pair. In overall area shown in Figure 3-5, the biggest movement in each pair was 66.2 cm to 66.9 cm, and all appeared to be located around the area where there were presence of trees.

Authors such as Lucieer et al. (2014), Yang et al. (2020) found that shadow differences in solar zenith angles in the optical image pair influenced the images and produced unreliable results using COSI-Corr technique. Thus, in this study, the shadow of the trees appeared to be incorrectly producing the biggest displacement in the image pairs. Figure 3-6 shows the



total accumulated mean displacement in E/W and N/S direction from 8/11/20 to 11/10/20 within the box area.

Furthermore, Sentinel-1 SAR image pair from 8/06/2020 and 9/11/2020 was analyzed using DinSAR technique mentioned in Section 3.2 of this paper. The result of analysis shows that 4 cm of subsidence in the line of sight direction was produced within the rectangular box area for this image pair. Figure 3-5.d shows the subsidence map created using DinSAR in the study area with 15 m spatial resolution. This result only shows the subsidence in LOS with no regards to the horizontal displacement.

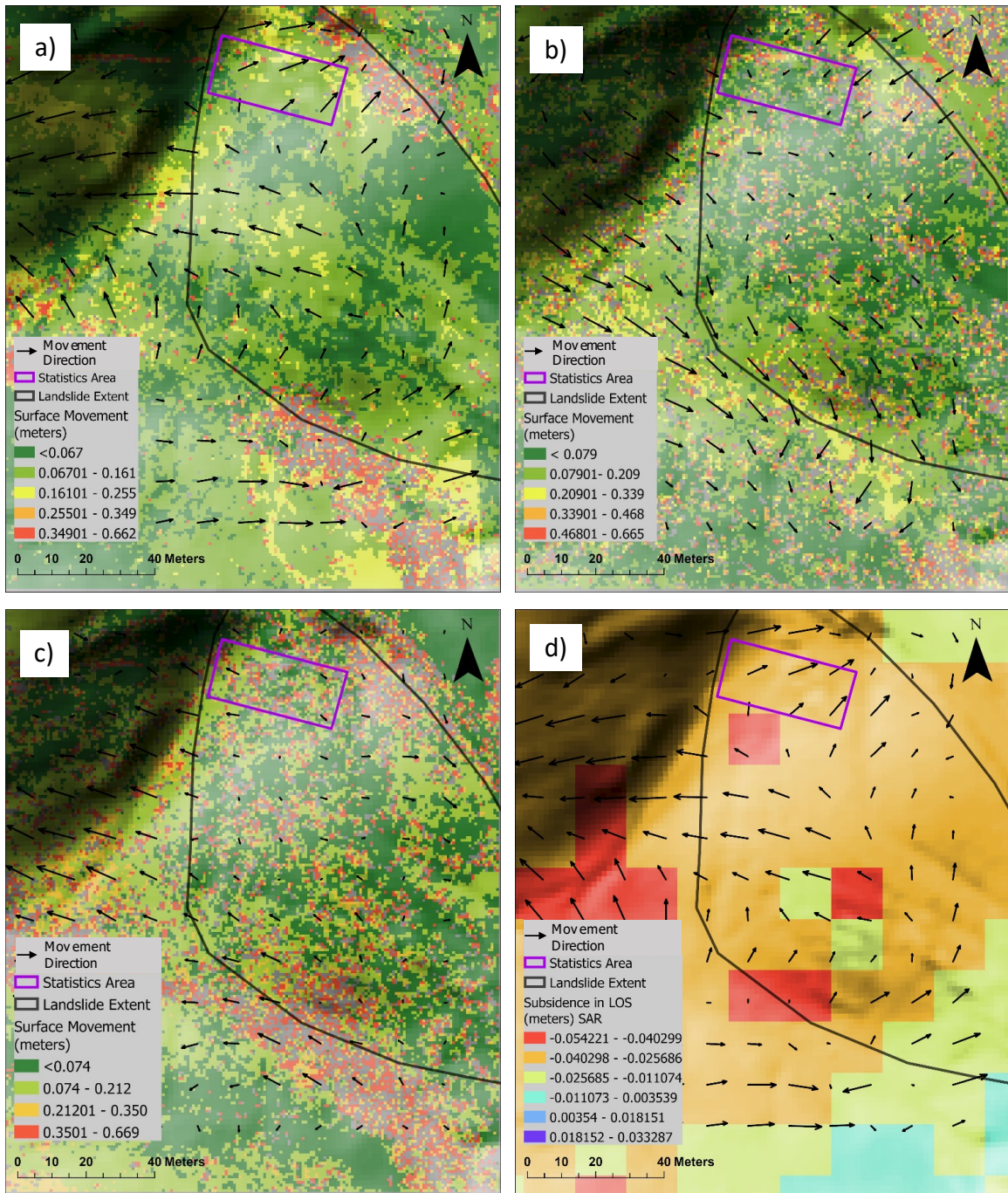


Figure 3-5. Horizontal displacement map of the study site for the pair images from (a) 8/11/20 and 9/11/20, (b) 9/11/20 and 10/9/20, (c) 10/9/20 and 11/10/20 using COSI-Corr technique. (d) Vertical displacement (in LOS direction of SAR satellite) of the study site between 8/6/20 and 9/11/20 using DinSAR technique, Garrard County, Kentucky.

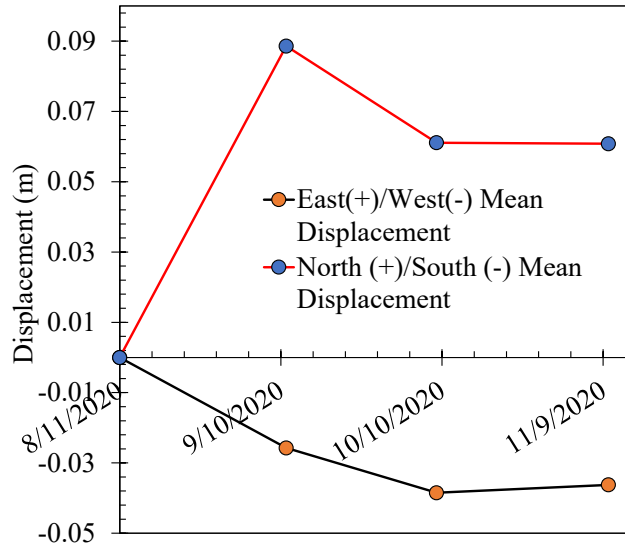


Figure 3-6. Accumulated mean displacement in E (+)/W (-) and N (+)/S (-) direction within the metrics area for the time span of 8/11/20 and 11/10/2020.

Table 3-2. The mean displacement in E (+)/W (-), N (+)/ S (-), and Euclidean distance for each pair of the images.

Dates	East (+)/West(-) Mean (meter)	North(+)/South(-) Mean (meter)	Euclidean distance (abs) (meter)
8/11/2020	0	0	0
9/11/2020	-0.0257	0.0886	0.0923
10/9/2020	-0.0128	-0.0275	0.0303
11/10/2020	0.0023	-0.0003	0.0023

### 3.5.3 Extraction of soil parameters from optical images using logistic regression machine learning model.

Soil parameters such as soil moisture are important parameter in slope stability as shown in the works of Lu and Godt (2008), Zhuo et al. (2019), Wicki et al. (2020), and Guzzetti et al. (2020), linking it to the triggering of the landslides. In this study, using all multi-temporal, ortho-rectified and co-registered images, an attempt was made to extract soil parameters that have an importance in slope stability. The target soil parameters we desired to estimate using the UAV images are presented in Table 3-3. Machine learning analysis

was carried out to get the correlation between the soil parameters and the RGB (Red = Band 1, Green = Band 2, Blue = Band3) band of the images. A RGB source is a light source which emits at the same time red, green and blue light. A combination of these colors makes up a true color image. A wide range of colors can be obtained by mixing different amounts of red, green and blue light. A wavelength of spectral RGB band ranges between; red light (635-700 nm), green light (520-560 nm), and blue light (450-490 nm). In digital imaging, the brightness values of the bands are scaled to 8-bit (0-255) values that represent a surface reflectance. Figure 3-7 illustrates the band differences in the image captured on 9/11/2020, and its histogram. As seen in Figure 3-7(d) through (f), all bands are scaled to digital numbers (DN), and the different band has a different distribution of the DN. The mean DN for each band in Figure 3-7 is band 1 = 95.1, band 2 = 159.7 , band 3 = 144.2.

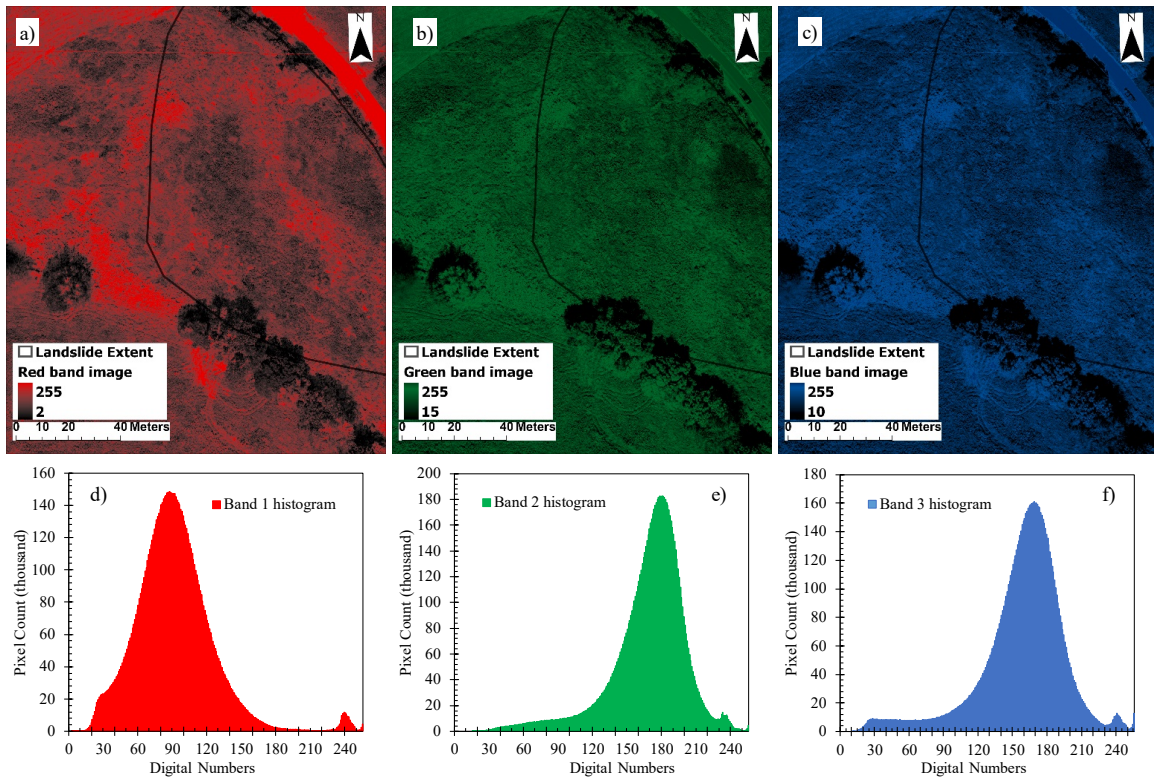


Figure 3-7. Spectral bands of digital images (9/11/2020) and their histograms of digital numbers (0-255) for surface reflectance. (a) Red (Band 1) (b) green (band 2) (c) blue (band 3) (d) histogram of band 1 (e) histogram of band 2 (f)

Linear regression machine learning model was used to correlate the soil parameters with the brightness variables from the images. The values for 12 target variables (Table 3-3) were taken from two satellite data sources, which were the Global Land Data Assimilation

System (GLDAS), and NASA Soil Moisture Active Passive (SMAP). These target variable were used to train the model to predict the target variables. Table 3-3 shows each of the target variables, and their spatial resolutions. Temporal resolution of all data was 3 hours.

Table 3-3. Soil parameters extracted from UAV images using machine learning model.

Target Variables	Description	Spatial resolution
GLDAS_NOAH025_3H_2_1_SoilMoi10_40cm_inst	Area-Averaged of Soil moisture content (10 - 40 cm underground)	0.25 degree
GLDAS_NOAH025_3H_2_1_SoilMoi0_10cm_inst	Area-Averaged of Soil moisture content (0 - 10 cm underground)	0.25 degree
GLDAS_NOAH025_3H_2_1_RootMoist_inst	Root zone soil moisture (0-100cm underground)	0.25 degree
GLDAS_NOAH025_3H_2_1_SoilMoi40_100cm_inst	Area-Averaged of Soil moisture content (40 - 100 cm underground)	0.25 degree
SPL4SMGP_005_Geophysical_Data_sm_surface_5	Top layer soil moisture (0-5cm) (m <sup>3</sup> /m <sup>3</sup> )	9km
SPL4SMGP_005_Geophysical_Data_sm_surface_wetness_5	Top layer soil wetness (0-5cm)	9km
SPL4SMGP_005_Geophysical_Data_heat_flux_ground_5	Downward ground heat flux into layer 1 of soil heat diffusion model.	9km
SPL4SMGP_005_Geophysical_Data_land_evapotranspiration_flux_5	Evapotranspiration from land (excluding areas of open water and permanent ice)	9km
SPL4SMGP_005_Geophysical_Data_sm_rootzone	Root zone soil moisture (0-100 cm) (m <sup>3</sup> /m <sup>3</sup> )	9km
GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst	Area-Averaged of Soil moisture content (100 - 200 cm underground)	0.25 degree
SPL4SMGP_005_Geophysical_Data_sm_rootzone_pctl_5	Root zone soil moisture (0-100 cm; percentile units)	9km
SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5	Vegetation "greenness" or fraction of transpiring leaves averaged over the land area (excluding areas of open water and permanent ice) of the grid cell	9km

### 3.5.3.1 Variables and Statistics compilation used in machine learning analysis

We used three bands from the images as independent variables and their statistics in the training process to model the target variables (Table 3-3). The statistics used for each band in the analysis were the mean, standard deviation, Eigenvalue for each band, and covariance between the bands as these statistics capture the variation of the distribution of the data in each band. The mask area that was used to get the statistics from each band of the images was the same image extent as the one used in the land displacement analysis (Figure 3-5). Figure 3-8 shows the mean and standard deviation of all three bands in comparison over three month period at the study site. The band 2 and band 3 statistics appeared to have a similar trend (Figure 3-8).

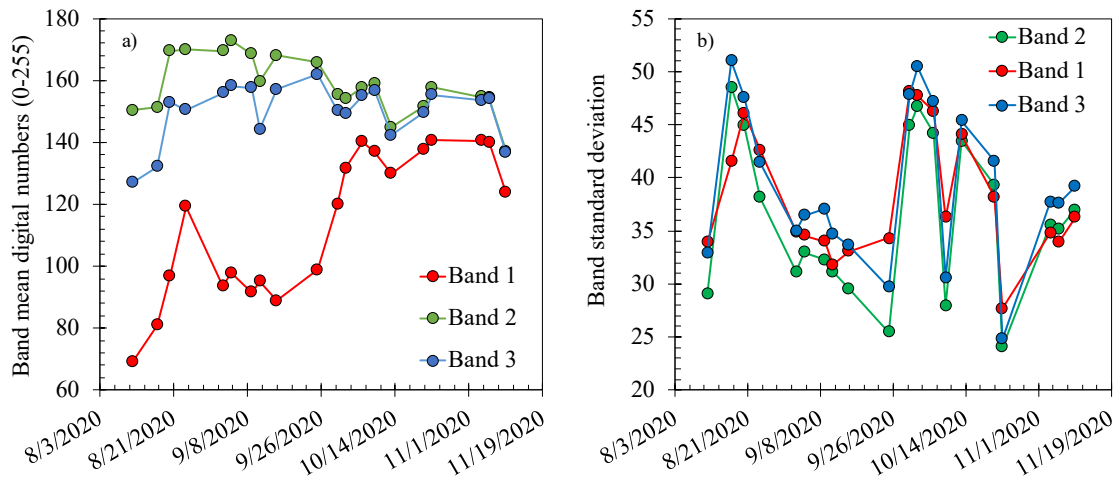


Figure 3-8. Mean and standard deviation for the multi-temporal images (a) mean band values and (b) standard deviation values for each image over three month period at the study site.

Alizamir et al. (2020) successfully used a machine learning model with various statistical combination of variables to predict soil temperature at different depths. We had three band variables with three statistics for each band, and three covariance between pairs of bands (band1 and band2, band1 and band3, band2 and band3) making the total number of features to be 12 to be used as training points in the machine learning model. The statistics for each image in the study area were calculated using the formulas in Table 3-4.

Table 3-4. Statistics for each band of the images for linear regression model.

Statistics	Formula
Mean	$Mean = \frac{\sum_{i=1}^N x_i}{N}$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - Mean)^2}$
Covariance	$COV_{b_1, b_2} = \frac{\sum (b_{1_i} - \bar{b}_1)(b_{2_i} - \bar{b}_2)}{N - 1}$
Eigenvalue	$Av = w = \lambda v$

$x_i$  = grid cell value;  $N$  = the total number of grid cells;  $b_{1_i}, b_{2_i}$  ( $i = 1, 2, 3, \dots, n$ ) are the band cell values in band 1 and band 2;  $\bar{b}_1, \bar{b}_2$  are the mean value of the bands;  $A = n \times n$  matrix of pixel;  $\lambda$  = Eigenvalue,  $v$  = Eigenvector such that matrix multiplication of  $\lambda$  and  $v$  is the same as just multiplying the vector by a constant as follows:

$$\lambda v = w = Av = A_1 v_1 + A_2 v_2 + \dots + A_i v_i = \sum_{j=1}^n A_j v_j \quad (9)$$

An eigenvalue/eigenvector decomposition of the covariance matrix reveals the principal directions of variation between images in the collection. This has applications in image coding, image classification, object recognition. All statistics were calculated in ENVI software for each band in the study area.

For the data preparation, a set of all 12 features shown in Table 3-5 as predictor or independent variables was extracted for each image in ENVI, and enumerated in a table that was ready to be put into linear regression model. The values for the target variables (Table 3-3) for each collection date of the images were obtained from GLDAS, and SMAP, and added to a table as the target values. Once the predictor variables and the target variables for each image were in the table, they were used as training points in linear regression machine learning model.



Table 3-5. Features used to model soil parameters in linear regression model.

band1_mean	band1_band2_covarian ce	band2_std	band3_mean
band1_std	band1_band3_covarian ce	band2_Eigenvalue	band3_std
band1_Eigenval ue	band2_mean	band2_band3_covarian ce	band3_Eigenval ue

### 3.5.3.2 Linear regression model for predicting soil parameters

Linear regression is a method of modelling a target value from predictor values. In linear regression model, the target value is expected to be a linear combination of the predictor features. In mathematical notation,

$$\bar{y}(w, x) = w_0 + w_1x_1 + \dots + w_ix_i \quad (10)$$

where,  $\bar{y}$  = target value;  $w_0$  = constant intercept;  $w_i (i=1,2,3,\dots,n)$  = coefficient estimates of the model;  $x_i (i=1,2,3,\dots,n)$  = predictor variables (e.g, band1 mean, band2 Eigenvalue) (Table 3-5). The coefficients express the effects of the predictor variables on the target variable value. Linear regression model fits a linear model with coefficients  $w_i$  to minimize the residual sum of squares between the observed targets such as soil moisture data from SMAP in the dataset, and the targets predicted ( $\bar{y}$ ) by the linear approximation. Mathematically it solves a problem of the form:

$$\underset{w}{\text{minimize}} \frac{1}{n} \sum_{i=1}^n (\bar{y} - y)^2 \quad (11)$$

where,  $\bar{y}$  = predicted value or target value ;  $y$  = observed or actual target value;  $n$  = all data points. The model coefficients are optimized in Eq. 11 using Gradient Descent method explained in the work of Ruder (2016).

### 3.5.3.3 Model Performance and validation

Normally in machine learning analysis, all dataset is divided into a training set (typically 75 percent) and a test set (25 percent) to evaluate the performance of the trained model

against the test set. However, in this study, we had only 20 points for the acquisition dates of the images, so the entire dataset was used as a training set, and the trained model was tested using the entire dataset, which was the same as the train test score. Table 3-6 lists the constant intercept; ( $w_0$ ), coefficients;  $w_i (i=1,2,3,\dots,n)$ , that were calculated in the linear regression model, and the predictor variables;  $x_i (i=1,2,3,\dots,n)$ . Plugging these coefficients and independent variables in Eq.10, the predicted target values can be estimated.

Table 3-6. Coefficients to estimate soil parameters from Linear Regression model using Eq. 10.

<b>Coeff.</b>	<b>Pred. Var</b>	<b>SPL4SMGP_005_Geophysical_Data_sm_rootzone_pctl_5</b>	<b>GLDAS_NOAH025_3H_2_1_SoilMoisture_100_200cm_inst</b>	<b>SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5</b>	<b>SPL4SMGP_005_Geophysical_Data_sm_rootzone</b>
<b>w<sub>0</sub></b>	intercept	1.05	293.65	0.09	0.37
<b>w<sub>1</sub></b>	(x1) band1_mean	5.85	-17.44	-0.04	-0.02
<b>w<sub>2</sub></b>	(x2) band1_std	221.02	-265.57	0.03	0.21
<b>w<sub>3</sub></b>	(x3) band1_Eigen value	-1573.24	1718.54	0.86	-2.61
<b>w<sub>4</sub></b>	(x4) band1_band2_covariance	423.22	-136.18	0.01	0.67
<b>w<sub>5</sub></b>	(x5) band1_band3_covariance	112.26	-388.96	-0.45	0.48
<b>w<sub>6</sub></b>	(x6) band2_mean	83.11	-17.82	0.16	0.01
<b>w<sub>7</sub></b>	(x7) band2_std	-202.44	25.57	-0.13	-0.16
<b>w<sub>8</sub></b>	(x8) band2_Eigen value	-13.16	34.94	-0.04	0.02
<b>w<sub>9</sub></b>	(x9) band2_band3_covariance	1046.42	-1209.44	-0.32	1.5
<b>w<sub>10</sub></b>	(x10) band3_mean	-66.14	20.98	-0.12	0.01
<b>w<sub>11</sub></b>	(x11) band3_std	30.14	187.52	0.02	-0.03
<b>w<sub>12</sub></b>	(x12) band3_Eigen value	-37.02	-13.53	-0.03	-0.02

Coeff = coefficient estimates; Pred. Var = predictor variables.

R-squared ( $R^2$ ) value was selected as evaluation metrics for the assessment of the model estimating the target value.  $R^2$  is a statistical measure of how close the dataset is fitted to the regression line in the model. It represents the proportion of variance of actual value that is explained by the independent variables in the model. It provides an indication of goodness of fit and therefore a measure of how well target variables are likely to be predicted by the model (scikit-learn 2021). Table 3-7 shows the predicted target variables that had  $R^2$  values more than 0.8.

Table 3-7. Target variables that yielded more than 0.8  $R^2$  score in the linear regression model provided input from the variables from UAV images.

Selected Target Variables > 0.8 ( $R^2$ -value)	$R^2$ - Value
SPL4SMGP_005_Geophysical_Data_sm_rootzone	0.823
GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst	0.906
SPL4SMGP_005_Geophysical_Data_sm_rootzone_pct1_5	0.961
SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5	0.988

The result from the machine learning analysis indicates that using combination of band statistics of multi-temporal optical images, we can model with a certain degree of confidence, soil moisture (SM) in units of  $m^3/m^3$ , and SM percentile at root zone (0-100 cm), SM in unit of  $kg\ m^2$  at 100-200 cm, and vegetation greenness fraction (Table 3-6). Figure 3-9 shows the result of the linear regression model with the aforementioned predicted target values and the actual values. The soil moisture plays important role in slope stability. The extraction of SM using UAV can help assess the state of slope stability using SM threshold values, or limit equilibrium equation (Lu and Godt 2008) which yields a factor of safety of the slope accounting for the SM content.

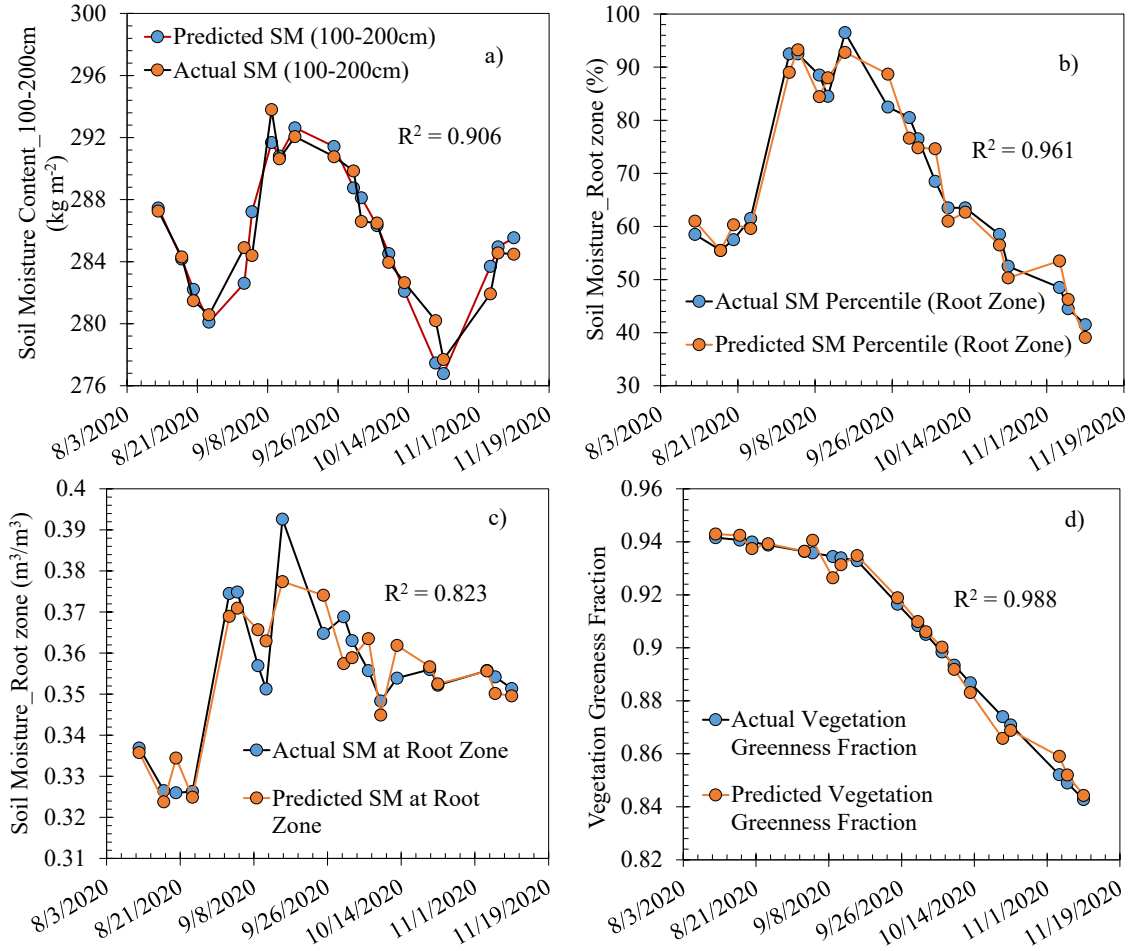


Figure 3-9. Predicted target values using linear regression model, actual target values from Satellite data (GLDAS, SMAP), and  $R^2$  score. (a) soil moisture percentile at root zone (b) Soil moisture at root zone (c) Vegetation Greenness fraction (d) Soil Moisture Content at 100-200 cm at study site, Garrard County, Kentucky.

### 3.6 Discussion

In this paper, we examined the workflow of using UAV with optical digital camera to gain an understanding of slope dynamic and stability employing a land surface detection and extraction of soil moisture content, both of which were presented in Section 5. The flights were taken between 1100 and 1600 (Eastern Time, USA) on days with less than 20 percent clouds for the reason that visible light-electromagnetic radiation is transmitted in the sun-surface-sensor (UAV) route, and the transmission process is affected by the absorption and scattering of atmospheric molecules, water vapor, aerosol, and other atmospheric components. The visible light-electromagnetic radiation captured by UAV includes the

surface reflection information and records the interference of the atmosphere in the surface reflection information (Lu et al. 2020, Shoshany et al. 2019).

This workflow of using UAV to detect a land movement and to extract soil parameters can be generalized for a slope with low vegetation. Areas with high vegetation have a shadow effect on movement detection process (COSI-Corr), producing unreliable results. Weather conditions with scattered clouds and resulting sharp and dark shadows makes the process more difficult. Differences in illumination conditions and in vegetation condition (e.g. lush spring grass versus short winter grass) between the acquisition dates can hamper the COSI-Corr algorithm to identify matching features between images acquired under different lighting conditions (Lucieer et al. 2014). The large land movement areas represented by red color in Figure 3-5 were highly associated with the shadow of trees or bushes in the study area.

The use of ground control points (GCPs) and UAV camera GPS locations produces georeferenced ortho-rectified images. GCPs improve the absolute and relative accuracy of photogrammetry. Although we used no GCPs for this project, Pix4D mapper searched for matching points by analyzing all images using SIFT (Lowe 2004) feature matching technique to ortho-rectify the images. Without GCPs, images were co-registered exploiting fully automatic tie point generation possible through area-based matching technique in ENVI. In addition, seasonal change in vegetation affects the reflectance impacting the displacement computation since the algorithm searches for similar surface patterns in a specified search radius. This seasonal reflectance change has an impact on these patterns and hence an impact on the displacement vector computation (Lucieer et al. 2014).

There was no in-ground measurement for land movement detection nor soil moisture measurement to validate the results. Soil parameters obtained from GLDAS and SMAP have a spatial resolution of 0.25 degree and 9 km, respectively. They were used as actual or observed values to train the model, and assumed to uniform across the spatial resolution area, neglecting the variance in soil parameters in the area.

Machine learning models normally use a number of training points to train a good working model, and a portion of the points are used as test points to validate the performance of the model. However, we had only 20 points for image acquisition dates and all points were

used as training points. Therefore, the model is highly specific to the study area of which the features were taken. Despite all the limitation aforementioned above, we demonstrated feasibility of the workflow that can detect land movement (Figure 3-5), and extract soil parameters (Table 3-6) from UAV flights with methods presented earlier section of this paper. Further studies are needed to verify and validate the performance of the workflow. This paper aimed to introduce the workflow of using UAV to gain an insight into slope dynamic and stability.

### **3.7 Conclusion.**

In this study, we used a UAV equipped with optical digital camera to collect multi-temporal images over the area where a landslide occurred and repaired in the past. The images were processed using matching points by analyzing all images employing SIFT (Lowe 2004) and bundle block adjustment (Triggs 1999) techniques to create ortho-rectified images. The ortho-rectified images were co-registered in ENVI, geometrically aligning corresponding pixels, and representing the same objects. We selected three pairs of the images (8/11/20 and 9/11/20, 9/11/20 and 10/9/20, and 10/9/20 and 11/10/20) to see if there was any land movement employing COSI-Corr, image correlation technique presented in Ayoub (2009). The result showed that there was a mean movement of 9.23 cm, 3.03 cm, and 0.23 cm in the area 17 m x 37 m where the metrics were taken on the slope for each pair in chronological order. In addition, SAR images from 8/6/20 and 9/11/2020 were analyzed exploiting DinSAR technique to get the subsidence in LOS, and the result showed 4 cm depression in the metrics area. Furthermore, using multi-temporal images, we estimated the soil parameters; soil moisture in units of  $m^3/m^3$  ( $R^2=0.823$ ), soil moisture in percentile ( $R^2=0.961$ ) at root zone, soil moisture in unit of  $kg\ m^2$  ( $R^2=0.906$ ) at 100-200 cm, and vegetation greenness fraction ( $R^2=0.988$ ) employing linear regression machine learning model. Statistical combination of variables from all 20 images were used as training points in the linear regression model, and the target variable values were taken from satellite data (GLDAS, SMAP). Our results indicate that using UAV equipped with an optical digital camera, we can estimate land surface movement, and extract soil parameters such as soil moisture data using the technique presented in this paper. The land movement and soil moisture data can be used to gain significant information about the

active landslide, and the stability of a slope. The further study is needed to validate and verify the performance of this workflow.



---

## 4 Conclusions

Using satellite soil moisture data, remote sensing data including high resolution LiDAR digital elevation map (DEM), and soil strength data derived from soil database from USDA, close-to-real time landslide hazard map (LHM) was produced at the local scale. The final multi-temporal landslide hazard map was validated temporally and spatially using four study sites of known landslide locations and failure dates. The resulting product correctly indicated low factor of safety values at the sites on the dates the landslide occurred. Landside susceptibility map (LSM) was created using the logistic regression machine learning model utilizing the variables that included six geomorphic variables extracted from 1.5 m LiDAR DEM, eight variables extracted from the physical property of soil database from (WSS), and a land cover variable from 2016 National Land Cover Database (NLCD). The receiver operating characteristic curve (ROC), area under the curve (AUC), were used for the accuracy of the model which yielded a success rate of 0.84. The comparisons of the LHM and LSM were done quantitatively using the confusion matrix (Table 2-8 and Table 2-9), which showed that even on the date with the highest soil moisture content, the LHM still modeled less occurrence of landslides and verified correctly against the study sites. The validation of LHM for each study site indicated that using the infinite slope factor of safety equation (Lu and Godt, 2008), we can model the likelihood of landslide temporally and spatially based on the soil moisture evolution of the area, the soil properties and DEM data all available publicly rather than statistically produced LSM.

Land movement was detected using multi-temporal images collected by UAV equipped with a digital camera over the area where a landslide occurred and repaired in the past. The collected images were ortho-rectified, and co-registered using photogrammetric and area-based matching techniques. Image correlation technique (COSI-Corr) was applied to three pairs of images that were one month apart to detect a horizontal land movement. The result showed a mean movement of 9.23 cm, 3.03 cm, and 0.23 cm in the metrics area for each pair in chronological order. Furthermore, using the multi-temporal images, we estimated the soil parameters; soil moisture in units of  $\text{m}^3/\text{m}^3$  ( $R^2=0.823$ ), soil moisture in percentile ( $R^2=0.961$ ) at root zone, soil moisture in unit of  $\text{kg m}^2$  ( $R^2=0.906$ ) at 100-200 cm, and vegetation greenness fraction ( $R^2=0.988$ ) employing linear regression machine learning

model. Statistical combination of variables from all multi-temporal images were used as training points in the linear regression model, and the target variable values were taken from satellite data (GLDAS, SMAP). The land movement and soil moisture data can be used to gain significant information about the active landslide, and the stability of a slope.

---

# **Appendix A**

**Python Codes For Landslide Susceptibility Map  
Using Logistic Regression Model.**

## Classifier Model for Landslide Susceptability Using **Logistic Regression** with **RFE** (Univariate Selection Tool and Recursive Feature Elimination) as Feature Selection Tool

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import warnings

from sklearn.decomposition import PCA
from sklearn.feature_selection import RFE
from sklearn.feature_selection import RFECV
from sklearn.feature_selection import SelectKBest, chi2
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, f1_score
from sklearn.metrics import f1_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
import matplotlib.pyplot as plt

%matplotlib notebook

np.set_printoptions(precision=3)
pd.set_option('display.float_format', lambda x: '%.3f' % x)
warnings.filterwarnings('ignore')
np.random.seed(8)
%matplotlib inline

def generate_accuracy_and_heatmap(model, x, y):
    cm = confusion_matrix(y,model.predict(x))
    sns.heatmap(cm,annot=True,fmt="d")
    ac = accuracy_score(y,model.predict(x))
    f_score = f1_score(y,model.predict(x))
    print('Accuracy is: ', ac)
```

```

print('F1 score is: ', f_score)
print ("\n")
print (pd.crosstab(pd.Series(model.predict(x), name='Predicted'),
                  pd.Series(y['Predictor'],name='Actual'))))

return

# My data
df =pd.read_csv('FINAL_CAMPB_KENTON_CO.csv')

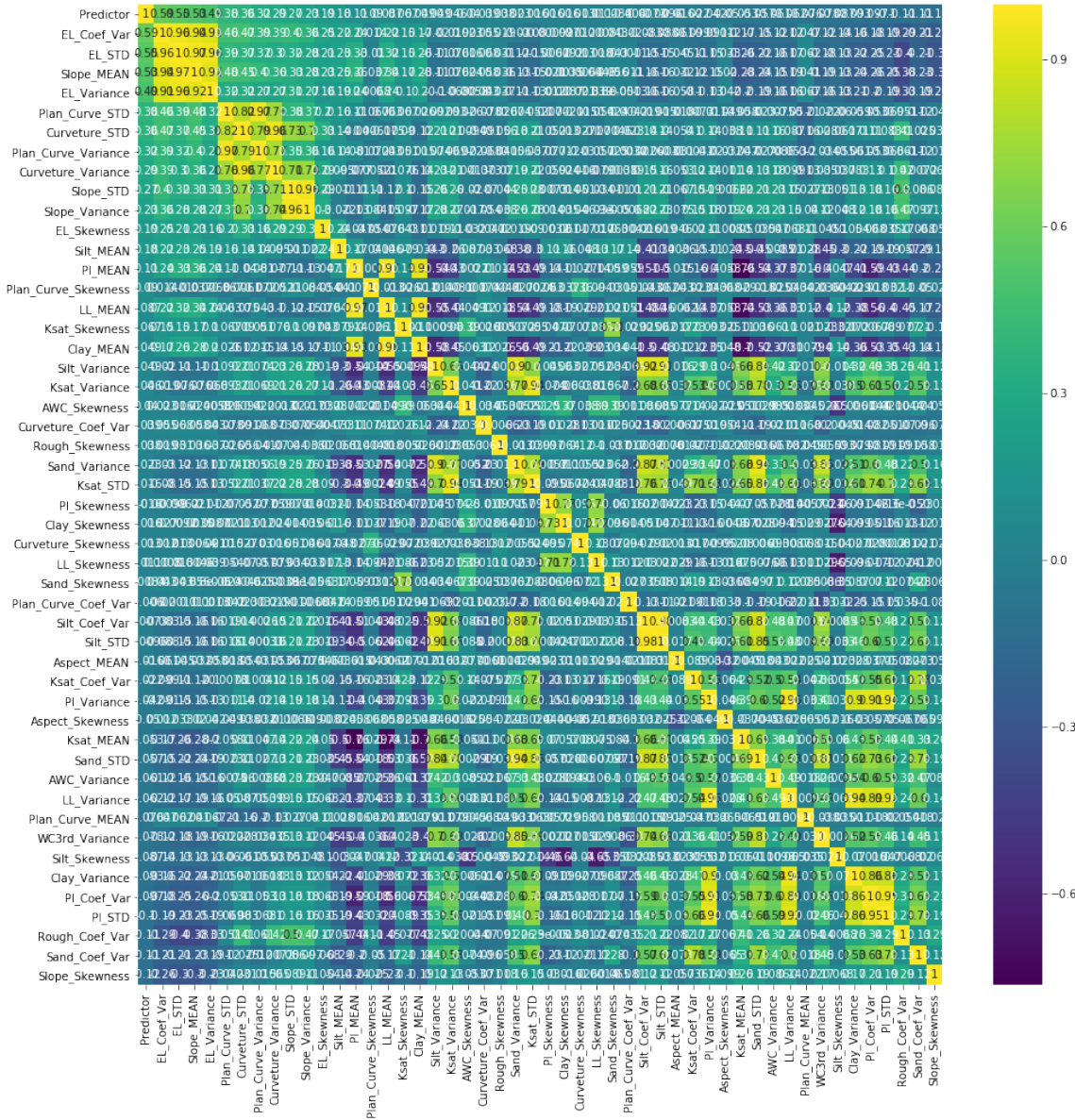
print('The number of landslides: 1 and Not Landslides: 0: \n', df.Predictor.value_counts())

# We see that it is not inbalanced , it has a probability of 50% chance

numerical_feature_columns = list(df._get_numeric_data().columns) # creating
a list to see the header of the columns
numerical_feature_columns

target = 'Predictor'
k = 50 #number of variables for heatmap
cols = df[numerical_feature_columns].corr().nlargest(k, target)[target].index
correl = df[cols].corr()
plt.figure(figsize=(16,16))
sns.heatmap(correl, annot=True, cmap = 'viridis' ) # cmap color options: '
inferno', , 'cividis','YlGnBu','magma','plasma'

```



```
# Separating the features from the labels
```

```
X1 = df.loc[:, df.columns != target] # Condition statement nested in this line of code.
```

```
Y1 = df.loc[:, df.columns == target]
```

```
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler()
```

```
X = scaler.fit_transform(X1)
```

```

# we must apply the scaling to the test set that we computed for the tra
ining set
print('The shape of the features',X.shape)
print('The shape of the labels', Y1.shape)

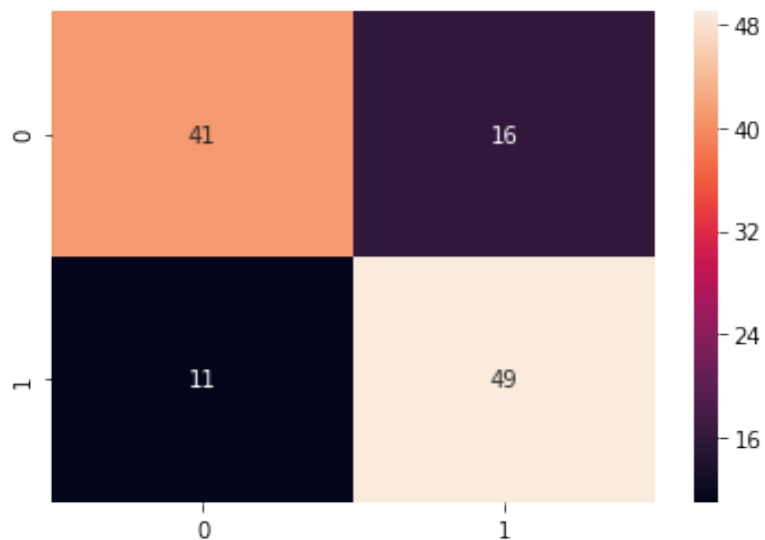
# Splitting the train set and the test set
x_train, x_test, y_train, y_test = train_test_split(X, Y1, random_state=
0) # the default is set to 0.25 test set, also now
# the data is np array no longer the panda dataframe

C = [0.01,0.05,0.1, 0.5, 1, 2, 5, 8, 10 ,15, 20, 30, 40, 50]
# after trying different regularization, it shows that the C = 0.01 give
s the best result.

clf_lr = LogisticRegression(C = 0.0005)
lr_baseline_model = clf_lr.fit(x_train,y_train)

generate_accuracy_and_heatmap(lr_baseline_model, x_test, y_test)

```



## Univariate feature selection Tool

```

select_feature = SelectKBest(chi2, k=19).fit(x_train, y_train)

# Creating a panda dataframe from the List , X1 column, and
selected_features_df = pd.DataFrame({'Feature':list(X1.columns),
                                     'Scores':select_feature.scores_})
selected_features_df.sort_values(by='Scores', ascending=False)

```

Feature	Scores	
23	EL_Coef_Var	14.495
22	EL_Variance	12.716
21	EL_STD	10.595
61	Slope_MEAN	7.758
20	EL_MEAN	5.145
42	Plan_Curve_STD	4.048
43	Plan_Curve_Variance	3.885
1	Aspect_STD	3.782
64	Slope_Coef_Var	3.583
16	Curvature_STD	3.399
63	Slope_Variance	3.257
17	Curvature_Variance	2.881



<b>Feature</b>	<b>Scores</b>	
<b>62</b>	Slope_STD	2.393
<b>2</b>	Aspect_Variance	2.247
<b>51</b>	Sand_MEAN	2.175
<b>5</b>	AWC_MEAN	2.050
<b>35</b>	NLCD_MEDIAN	1.848
<b>11</b>	Clay_STD	1.283
<b>66</b>	WC3rd_MEAN	1.254
<b>67</b>	WC3rd_STD	1.170
<b>3</b>	Aspect_Coef_Var	1.155
<b>31</b>	LL_STD	1.073
<b>48</b>	Rough_Variance	1.040
<b>13</b>	Clay_Coef_Var	0.868
<b>58</b>	Silt_Variance	0.854
<b>33</b>	LL_Coef_Var	0.806
<b>47</b>	Rough_STD	0.686

<b>Feature</b>	<b>Scores</b>	
<b>15</b>	Curvature_MEAN	0.648
<b>8</b>	AWC_Coef_Var	0.637
<b>69</b>	WC3rd_Coef_Var	0.634
...	...	...
<b>37</b>	PI_STD	0.277
<b>25</b>	Ksat_MEAN	0.275
<b>26</b>	Ksat_STD	0.271
<b>45</b>	Plan_Curve_Skewness	0.250
<b>56</b>	Silt_MEAN	0.243
<b>54</b>	Sand_Coef_Var	0.198
<b>68</b>	WC3rd_Variance	0.196
<b>30</b>	LL_MEAN	0.131
<b>29</b>	Ksat_Skewness	0.117
<b>28</b>	Ksat_Coef_Var	0.113
<b>52</b>	Sand_STD	0.070

<b>Feature</b>	<b>Scores</b>	
<b>32</b>	LL_Variance	0.067
<b>41</b>	Plan_Curve_MEAN	0.060
<b>4</b>	Aspect_Skewness	0.053
<b>57</b>	Silt_STD	0.043
<b>9</b>	AWC_Skewness	0.042
<b>60</b>	Silt_Skewness	0.036
<b>10</b>	Clay_MEAN	0.035
<b>14</b>	Clay_Skewness	0.028
<b>0</b>	Aspect_MEAN	0.025
<b>59</b>	Silt_Coef_Var	0.023
<b>7</b>	AWC_Variance	0.022
<b>34</b>	LL_Skewness	0.020
<b>40</b>	PI_Skewness	0.011
<b>50</b>	Rough_Skewness	0.006
<b>18</b>	Curvature_Coef_Var	0.002

	<b>Feature</b>	<b>Scores</b>
<b>44</b>	Plan_Curve_Coef_Var	0.002
<b>19</b>	Curvature_Skewness	0.001
<b>55</b>	Sand_Skewness	0.001
<b>38</b>	PI_Variance	0.000

```
x_train_chi = select_feature.transform(x_train) # reducing the features
to the k number of features.
```

```
x_test_chi = select_feature.transform(x_test)
```

In [43]:

```
select_feature.get_support(indices = True)
```

Out[43]:

```
array([ 1,  2,  5, 11, 16, 17, 20, 21, 22, 23, 35, 42, 43, 51, 61, 62, 6
3,
      64, 66], dtype=int64)
```

In [44]:

```
select_feature_list = list(select_feature.get_support(indices = True))
```

```
select_feature_list
```

Out[44]:

```
[1, 2, 5, 11, 16, 17, 20, 21, 22, 23, 35, 42, 43, 51, 61, 62, 63, 64, 66
]
```

In [45]:

```
# to see which features are being used.
```

```
selected_features_df1 = selected_features_df.iloc[select_feature_list]['
Feature']
```

```
selected_features_df1.to_csv('Selected_feaures.csv')
```

In [47]:

```
lr_chi_model = clf_lr.fit(x_train_chi,y_train)
```

In [48]:

```
generate_accuracy_and_heatmap(lr_chi_model, x_test_chi, y_test)
```

```
Accuracy is: 0.7863247863247863
```

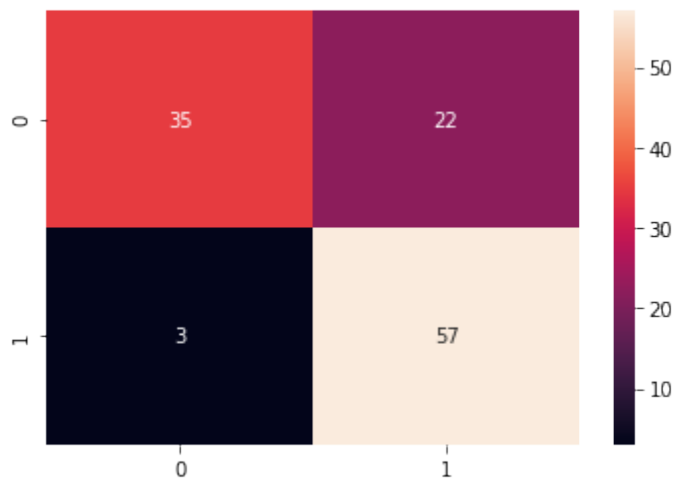
```
F1 score is: 0.8201438848920861
```

```
Actual      1
```

```
Predicted
```

```
0          12
```

```
1          19
```



```
# Cross Validation scores:
```

```
from sklearn.model_selection import cross_val_score
```

```
cv_scores = cross_val_score(lr_chi_model, X, Y1, cv =5)
```

```
print('Cross-validation scores (3-fold):', cv_scores)
```

```
print('Mean cross-validation score (3-fold): {:.3f}'
```

```
      .format(np.mean(cv_scores)))
```

```
Cross-validation scores (3-fold): [0.777 0.798 0.691 0.691 0.707]
```

```
Mean cross-validation score (3-fold): 0.733
```

**Equation for the selected features.**

In [50]:

```
coef = lr_chi_model.coef_.tolist()
class_ = lr_chi_model.classes_
print(lr_chi_model.intercept_)
print(coef)
coef_list = list(coef)
[-0.023]
[[-0.09635920203764367, -0.05312688523528351, -0.06627828655493466, -0.0
3371384619115812, 0.07069903909893288, 0.04155803063697834, -0.142537961
47726086, 0.14688932481330982, 0.11348556424732094, 0.17650182480232998,
-0.07040267728098656, 0.0840502094330824, 0.056347475249267816, -0.03497
305484503876, 0.1308785299334852, 0.062354221671975046, 0.04932229482041
728, -0.08241909801742367, -0.08388179482495318]]
```

## RFE with cross validation

In [20]:

```
rfevcv = RFECV(estimator=clf_lr, step=1, cv=5, scoring='accuracy')
rfevcv = rfevcv.fit(x_train, y_train)
print('Optimal number of features :', rfevcv.n_features_) # takes the ra
nking#1 data
print('Best features :', X1.columns[rfevcv.support_])
Optimal number of features : 17
Best features : Index(['Aspect_STD', 'Aspect_Variance', 'AWC_MEAN', 'Cur
vature_MEAN',
                    'Curvature_STD', 'EL_MEAN', 'EL_STD', 'EL_Variance', 'EL_Coef_Var
',
                    'NLCD_MEDIAN', 'Plan_Curve_STD', 'Plan_Curve_Variance', 'Slope_ME
AN',
                    'Slope_STD', 'Slope_Variance', 'Slope_Coef_Var', 'WC3rd_MEAN'],
                    dtype='object')
```

In [21]:

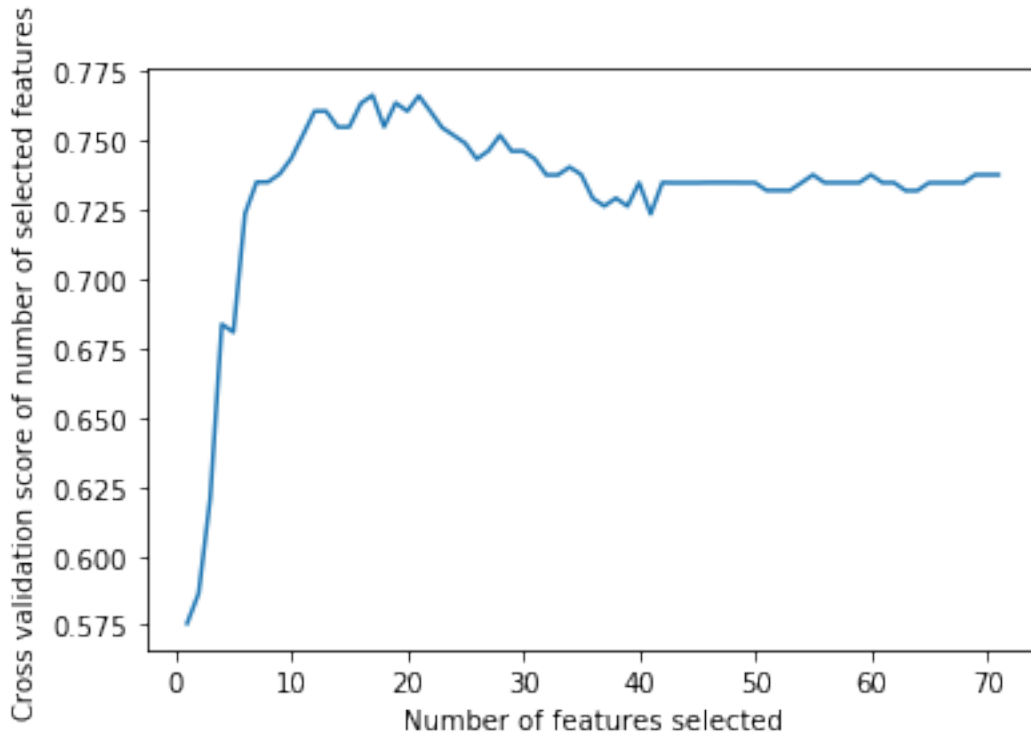
```
important_features = list(X1.columns[rfevcv.support_])
important_features
```

Out [21]:

```
['Aspect_STD',  
 'Aspect_Variance',  
 'AWC_MEAN',  
 'Curvature_MEAN',  
 'Curvature_STD',  
 'EL_MEAN',  
 'EL_STD',  
 'EL_Variance',  
 'EL_Coef_Var',  
 'NLCD_MEDIAN',  
 'Plan_Curve_STD',  
 'Plan_Curve_Variance',  
 'Slope_MEAN',  
 'Slope_STD',  
 'Slope_Variance',  
 'Slope_Coef_Var',  
 'WC3rd_MEAN']
```

In [23]:

```
plt.figure()  
plt.xlabel("Number of features selected")  
plt.ylabel("Cross validation score of number of selected features")  
plt.plot(range(1, len(rfecv.grid_scores_) + 1), rfecv.grid_scores_)  
plt.show()
```



In [25]:

```
# To graph everything in excel the score and rate need to be copied to t
xt file.
```

```
optimalFeature_df = pd.DataFrame({'Number of Features Selected':list(ra
nge(1, len(rfecv.grid_scores_) + 1)),
                                'Cross Validation Score':list(rfecv
.grid_scores_)})
optimalFeature_df.to_excel("C:\\Users\\bda227\\Documents\\RESEARCH\\Pap
er1_landslideMapping\\tables_OptimalFeature.xlsx")
```

In [65]:

```
x_test.shape
```

Out[65]:

```
(117, 71)
```

In [66]:

```
x_train_rfecv = rfecv.transform(x_train) # Data with 17 features from t
he selection model
x_test_rfecv = rfecv.transform(x_test)
x_test_rfecv.shape
```

Out[66]:



```
(117, 18)
```

```
In [67]:
```

```
lr_rfecv_model = clf_lr.fit(x_train_rfecv, y_train)
```

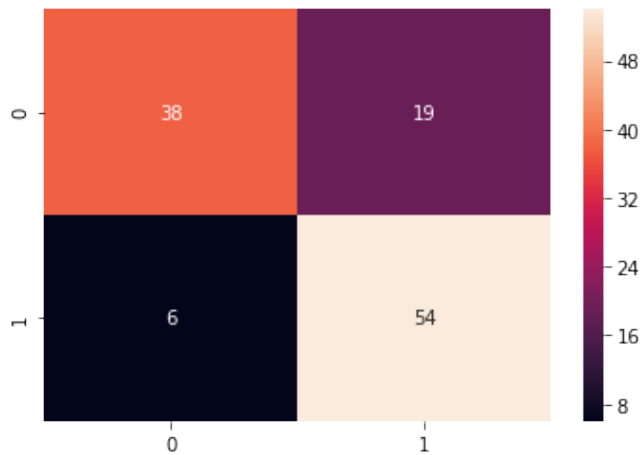
```
In [68]:
```

```
generate_accuracy_and_heatmap(lr_rfecv_model, x_test_rfecv, y_test)
```

```
Accuracy is: 0.7863247863247863
```

```
F1 score is: 0.8120300751879699
```

```
Actual      1
Predicted
0           13
1           18
```



```
# Cross Validation scores:
```

```
cv_scores = cross_val_score(lr_rfecv_model, X, Y1, cv =5)
```

```
print('Cross-validation scores (5-fold):', cv_scores)
```

```
print('Mean cross-validation score (5-fold): {:.3f}'
```

```
      .format(np.mean(cv_scores)))
```

```
Cross-validation scores (5-fold): [0.777 0.798 0.691 0.691 0.707]
```

```
Mean cross-validation score (5-fold): 0.733
```

## Evaluation

We can check precision, recall, f1-score using classification report!

In [71]:

```
predictions = lr_rfecv_model.predict(x_test_rfecv)
```

In [72]:

```
from sklearn.metrics import classification_report
```

In [73]:

```
print(classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.86	0.67	0.75	57
1	0.74	0.90	0.81	60
accuracy			0.79	117
macro avg	0.80	0.78	0.78	117
weighted avg	0.80	0.79	0.78	117

In [74]:

```
print('Coefficients:\n', clf_lr.coef_)
```

```
print('Intercepts:\n', clf_lr.intercept_)
```

Coefficients:

```
[[-0.093 -0.051 -0.063 -0.053  0.074 -0.136  0.152  0.116  0.182 -0.066  
  0.088  0.058  0.136  0.066  0.051 -0.08  -0.076 -0.049]]
```

Intercepts:

```
[-0.01]
```

---

# **Appendix B**

**Python Codes To Estimate Soil Parameters From  
Multi-Temporal UAV Images.**

## Regression Model to correlate image statistics to soil parameters using **Linear Regression**

```
%%javascript
```

```
IPython.OutputArea.auto_scroll_threshold = 9999;
```

In [1]:

```
import os
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import warnings

from sklearn.decomposition import PCA
from sklearn.feature_selection import RFE
from sklearn.feature_selection import RFECV
from sklearn.feature_selection import SelectKBest, chi2
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
from sklearn.model_selection import cross_val_score
plt.style.use('ggplot')
```

In [25]:

```
from sklearn import feature_selection
```

In [2]:

```
np.set_printoptions(precision=3)
pd.set_option('display.float_format', lambda x: '%.3f' % x)
warnings.filterwarnings('ignore')
np.random.seed(8)
%matplotlib inline
```

In [3]:

```
# see the plots interactively
```

```
%matplotlib notebook
```

```
In [4]:
```

```
os.chdir('C:\\Users\\bda227\\Documents\\ENVI\\Paper2\\python_codes') # to change your directory to the location you want.
```

```
In [4]:
```

```
os.getcwd() # to see the working directory
```

```
Out [4]:
```

```
'C:\\Users\\bda227\\Documents\\ENVI\\Paper2\\python_codes'
```

```
In [5]:
```

```
# My data
```

```
df =pd.read_excel('All_satellite_droneData.xlsx')
```

```
In [6]:
```

```
df.shape
```

```
Out [6]:
```

```
(20, 52)
```

```
In [7]:
```

```
df.info()
```

```
# everything looks good! no missing data.
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 20 entries, 0 to 19
```

```
Data columns (total 52 columns):
```

```
Date 20 non-null da
```

```
tetime64[ns]
```

```
Date_Time 20 non-null da
```

```
tetime64[ns]
```

```
band1_mean 20 non-null fl
```

```
oat64
```

```
band1_std 20 non-null fl
```

```
oat64
```

```
band1_Eigenvalue 20 non-null fl
```

```
oat64
```

band1_band2_covariance oat64	20 non-null fl
band1_band3_covariance oat64	20 non-null fl
band2_mean oat64	20 non-null fl
band2_std oat64	20 non-null fl
band2_Eigenvalue oat64	20 non-null fl
band2_band3_covariance oat64	20 non-null fl
band3_mean oat64	20 non-null fl
band3_std oat64	20 non-null fl
band3_Eigenvalue oat64	20 non-null fl
SPL4SMGP_005_Geophysical_Data_sm_rootzone loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_AvgSurfT_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_CanopInt_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_Evap_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_GWS_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_Qg_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_Qh_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_Qle_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_SoilMoist_P_tavg loat64	20 non-null f

GLDAS_CLSM025_DA1_D_2_2_SoilMoist_RZ_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_SoilMoist_S_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_TVeg_tavg loat64	20 non-null f
GLDAS_CLSM025_DA1_D_2_2_TWS_tavg loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_AvgSurfT_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_Evap_tavg loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_Qg_tavg oat64	20 non-null fl
GLDAS_NOAH025_3H_2_1_Qh_tavg oat64	20 non-null fl
GLDAS_NOAH025_3H_2_1_Qle_tavg oat64	20 non-null fl
GLDAS_NOAH025_3H_2_1_RootMoist_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilMoi0_10cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilMoi10_40cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilMoi40_100cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilTMP0_10cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilTMP10_40cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilTMP40_100cm_inst loat64	20 non-null f
GLDAS_NOAH025_3H_2_1_SoilTMP100_200cm_inst loat64	20 non-null f

GLDAS_NOAH025_3H_2_1_Tveg_tavg loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_heat_flux_ground_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_heat_flux_sensible_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_land_evapotranspiration_flux_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_sm_rootzone_pct1_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_sm_surface_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_sm_surface_wetness_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_soil_temp_layer4_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_surface_temp_5 loat64	20 non-null f
SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5 loat64	20 non-null f

dtypes: datetime64[ns](2), float64(50)

memory usage: 8.2 KB

## Function to create a graph

In [66]:

```
# without scaling the data. Original
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
def PlotPolly(model, independent_variable, dependent_variabble, Name,power):
    x_new = np.linspace(independent_variable.min(), independent_variable.max()
, 20)
    y_new = model(x_new)
```



```

y_hat = np.asanyarray(model(independent_variable))
y_test = np.asanyarray(dependent_variabble)

R2_score = r2_score(y_test, y_hat)
# print("R2-score: %.3f" % r2_score(y_hat , y_test) )

MSE = mean_squared_error(y_test, y_hat)

plt.figure()

plt.plot(independent_variable, dependent_variabble, '.', x_new, y_new, '-')
)
# plt.title('Polynomial Fit with Matplotlib for q0 x e_eoc VS e_ini')
plt.title('Polymomial Regression(Power={})\n\
          R^2 = {:.3f}, MeanSqrd Error:{:.3f} '.format(power, R2_score, MSE)
)

ax = plt.gca()
ax.set_facecolor((0.898, 0.898, 0.898))

fig = plt.gcf()
plt.xlabel(Name)
plt.ylabel(y)

plt.show()
plt.close()

```

## Feature Selection

In [8]:

```

y_list = ['SPL4SMGP_005_Geophysical_Data_sm_rootzone', 'GLDAS_NOAH025_3H_2_1_RootMoist_inst', 'GLDAS_NOAH025_3H_2_1_SoilMoi0_10cm_inst',
          'GLDAS_NOAH025_3H_2_1_SoilMoi10_40cm_inst', 'GLDAS_NOAH025_3H_2_1_SoilMoi40_100cm_inst', 'GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst',
          'SPL4SMGP_005_Geophysical_Data_heat_flux_ground_5', 'SPL4SMGP_005_Geophysical_Data_land_evapotranspiration_flux_5',
          'SPL4SMGP_005_Geophysical_Data_sm_rootzone_pct1_5', 'SPL4SMGP_005_Geophysical_Data_sm_surface_5',
          'SPL4SMGP_005_Geophysical_Data_sm_surface_wetness_5', 'SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5',

```

```
'SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5']
```

## Using only all the variables in Linear Regression Model

In [15]:

```
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
```

## Normalizing the X

In [13]:

```
# Scaling it using min max scaler
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X_norm = scaler.fit_transform(X)
```

## Using correlation Features

## Using f\_regression

In [126]:

```
# configure to select all features

for i in range(1,13):
    fs = SelectKBest(score_func=f_regression, k=i)
    # Learn relationship from training data
    fs.fit(X, y)
    # transform train input data
    X_train_fs = fs.transform(X)
    # transform test input data

    # fit the model
    model = LinearRegression()
    model.fit(X_train_fs, y)
```

```

# evaluate the model
yhat = model.predict(X_train_fs)

# evaluate predictions
mae = mean_absolute_error(y, yhat)

print('MAE: %.3f' % mae)

# R2 score
R2_score = r2_score(y, yhat)

print("R2-score: %.3f and when selected features : %.f" % (r2_score(y, yha
t), i))
MAE: 0.008
R2-score: 0.586 and when selected features : 1
MAE: 0.007
R2-score: 0.661 and when selected features : 2
MAE: 0.007
R2-score: 0.713 and when selected features : 3
MAE: 0.007
R2-score: 0.716 and when selected features : 4
MAE: 0.007
R2-score: 0.717 and when selected features : 5
MAE: 0.007
R2-score: 0.724 and when selected features : 6
MAE: 0.006
R2-score: 0.734 and when selected features : 7
MAE: 0.006
R2-score: 0.740 and when selected features : 8
MAE: 0.006
R2-score: 0.809 and when selected features : 9
MAE: 0.005
R2-score: 0.819 and when selected features : 10
MAE: 0.005
R2-score: 0.819 and when selected features : 11
MAE: 0.005

```

R2-score: 0.823 and when selected features : 12

In [119]:

```
def select_features1(X, y):
    # configure to select all features
    fs = SelectKBest(score_func=mutual_info_regression, k= 5)
    # Learn relationship from training data
    fs.fit(X, y)
    # transform train input data
    X_train_fs = fs.transform(X)
    return X_train_fs, fs

# feature selection
X_train_fs, fs = select_features1(X, y)
# fit the model
model = LinearRegression()
model.fit(X_train_fs, y)
# evaluate the model
yhat = model.predict(X_train_fs)
# evaluate predictions
mae = mean_absolute_error(y, yhat)
print('MAE: %.3f' % mae)
# R2 score
R2_score = r2_score(y, yhat)
print("R2-score: %.3f" % r2_score(y, yhat))
MAE: 0.007
R2-score: 0.691
```

## Using RFE (Recursive Feature Elimination)

In [217]:

```
# evaluate RFE for regression
from numpy import mean
```

```

from numpy import std
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.feature_selection import RFE
from sklearn.tree import DecisionTreeRegressor
from sklearn.pipeline import Pipeline

list_target_corr = []
for k in y_list:
    y = df[k]

    # to formulate the graph for the number of features Vs the R2 values.
    x_value = []

    y_value = []
    for i in range(1,13):
        # create pipeline
        rfe = RFE(estimator=LinearRegression(), n_features_to_select=i)
        model = LinearRegression()
        pipeline = Pipeline(steps=[('s',rfe),('m',model)])

        # fit the model on all available data
        pipeline.fit(X_norm, y)

        # make a prediction for one example

        yhat = pipeline.predict(X_norm)
        # print('Predicted: %.3f' % (yhat))
        R2_score = r2_score(y, yhat)
        if R2_score >= 0.8:
            print("R2-score: %.3f" % r2_score(y, yhat),f': {k} , and feature s
elected : {i}')

```

```

        list_target_corr.append(k)

    x_value.append(i)
    y_value.append(R2_score)

# To plot the image
    if R2_score >= 0.8:
        plt.plot(x_value, y_value, label = k)
        plt.plot(x_value, y_value, 'o', alpha=0.8, color = 'r')
        # thisaxis.plot(X_test, y_test, 'o', Label='Test Value', alpha=0.8)
        plt.xlabel('number of features')
        plt.ylabel('R2 values')
        plt.title('RFE graph with number of features vs R2 scores')
        plt.legend()
        plt.tight_layout(pad=0.4, w_pad=0.5, h_pad=1.0)
        list_target_corr.append(k)

print(f'The names of the target variables:\n {set(list_target_corr)} and the l
length id {len(set(list_target_corr))}')

## evaluate model
# cv = RepeatedKFold(n_splits=5, n_repeats=2, random_state=1)
# n_scores = cross_val_score(pipeline, X, y, scoring='neg_mean_absolute_error'
, cv=cv, n_jobs=-1, error_score='raise')
## report performance
# print('MAE: %.3f (%.3f)' % (mean(n_scores), std(n_scores)))
# R2_score = r2_score(y, yhat)
# print("R2-score: %.3f" % r2_score(y, yhat))
# n_scores

R2-score: 0.802 : SPL4SMGP_005_Geophysical_Data_sm_rootzone , and feature sele
cted : 7

```

R2-score: 0.809 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone , and feature selected : 8

R2-score: 0.810 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone , and feature selected : 9

R2-score: 0.811 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone , and feature selected : 10

R2-score: 0.822 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone , and feature selected : 11

R2-score: 0.823 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone , and feature selected : 12

R2-score: 0.830 : GLDAS\_NOAH025\_3H\_2\_1\_SoilMoi100\_200cm\_inst , and feature selected : 10

R2-score: 0.831 : GLDAS\_NOAH025\_3H\_2\_1\_SoilMoi100\_200cm\_inst , and feature selected : 11

R2-score: 0.906 : GLDAS\_NOAH025\_3H\_2\_1\_SoilMoi100\_200cm\_inst , and feature selected : 12

R2-score: 0.832 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5 , and feature selected : 8

R2-score: 0.960 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5 , and feature selected : 9

R2-score: 0.960 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5 , and feature selected : 10

R2-score: 0.961 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5 , and feature selected : 11

R2-score: 0.961 : SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5 , and feature selected : 12

R2-score: 0.839 : SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5 , and feature selected : 9

R2-score: 0.843 : SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5 , and feature selected : 10

R2-score: 0.847 : SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5 , and feature selected : 11

R2-score: 0.856 : SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5 , and feature selected : 12

R2-score: 0.958 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 6

R2-score: 0.968 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 7

R2-score: 0.968 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 8

R2-score: 0.979 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 9

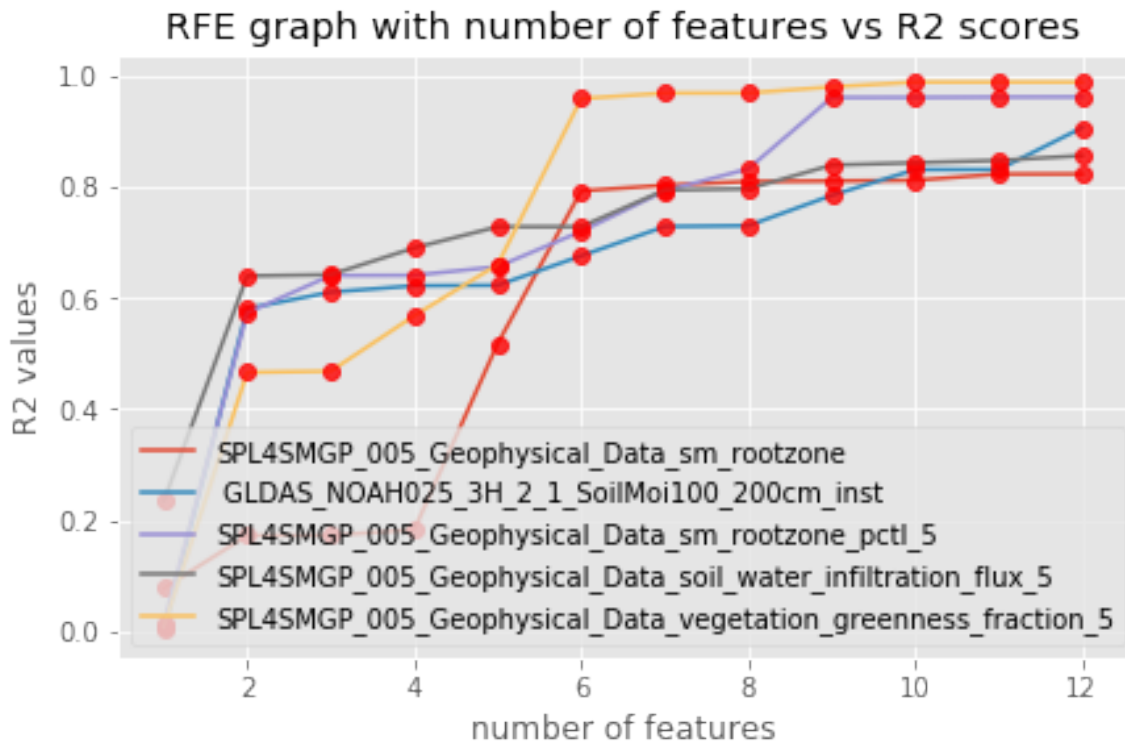
R2-score: 0.988 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 10

R2-score: 0.988 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 11

R2-score: 0.988 : SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5 , and feature selected : 12

The names of the target variables:

{'SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5', ' GLDAS\_NOAH025\_3H\_2\_1\_SoilMoi100\_200cm\_inst', 'SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5', 'SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5', 'SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone'} and the length id 5



**IMPORTANT!!**



The following dataframe is the one we want to use to correlate the data.

```
In [10]:  
  
# Originally We had the following soil parameters, however I decided to drop t  
he ones that give us less than 0.8 R2 value  
  
# After filtering through the list\  
  
print('Original selected Target variables from Satellites')  
  
for k in y_list:  
    print(f'{k}')  
  
print('\n')  
  
y_list_final = ['SPL4SMGP_005_Geophysical_Data_sm_rootzone_pctl_5', ' GLDAS_NO  
AH025_3H_2_1_SoilMoi100_200cm_inst',  
                'SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5'  
,  
                'SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5'  
, 'SPL4SMGP_005_Geophysical_Data_sm_rootzone']  
  
for f in y_list_final:  
    print(f'Final selected target variable that has more than 0.8 R2-Value: {f  
}')  
  
df_used = df[['SPL4SMGP_005_Geophysical_Data_sm_rootzone_pctl_5', ' GLDAS_NOAH  
025_3H_2_1_SoilMoi100_200cm_inst',  
              'SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5',  
              'SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5',  
              'SPL4SMGP_005_Geophysical_Data_sm_rootzone']]  
  
Original selected Target variables from Satellites  
SPL4SMGP_005_Geophysical_Data_sm_rootzone  
GLDAS_NOAH025_3H_2_1_RootMoist_inst  
GLDAS_NOAH025_3H_2_1_SoilMoi0_10cm_inst  
GLDAS_NOAH025_3H_2_1_SoilMoi10_40cm_inst  
GLDAS_NOAH025_3H_2_1_SoilMoi40_100cm_inst
```

```

GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst
SPL4SMGP_005_Geophysical_Data_heat_flux_ground_5
SPL4SMGP_005_Geophysical_Data_land_evapotranspiration_flux_5
SPL4SMGP_005_Geophysical_Data_sm_rootzone_pct1_5
SPL4SMGP_005_Geophysical_Data_sm_surface_5
SPL4SMGP_005_Geophysical_Data_sm_surface_wetness_5
SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5
SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5

```

Final selected target variable that has more than 0.8 R2-Value: SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone\_pct1\_5

Final selected target variable that has more than 0.8 R2-Value: GLDAS\_NOAH025\_3H\_2\_1\_SoilMoi100\_200cm\_inst

Final selected target variable that has more than 0.8 R2-Value: SPL4SMGP\_005\_Geophysical\_Data\_soil\_water\_infiltration\_flux\_5

Final selected target variable that has more than 0.8 R2-Value: SPL4SMGP\_005\_Geophysical\_Data\_vegetation\_greenness\_fraction\_5

Final selected target variable that has more than 0.8 R2-Value: SPL4SMGP\_005\_Geophysical\_Data\_sm\_rootzone

In [26]:

```

data_head = X.columns
data_head

```

Out [26]:

```

Index(['band1_mean', 'band1_std', 'band1_Eigenvalue', 'band1_band2_covariance'
,
      'band1_band3_covariance', 'band2_mean', 'band2_std', 'band2_Eigenvalue'
,
      'band2_band3_covariance', 'band3_mean', 'band3_std',
      'band3_Eigenvalue'],
      dtype='object')

```

## Predicted soil parameters Vs Actual soil Values

In [33]:

```

model = LinearRegression()

for f in y_list_final:
    y = df[f]
    # fit the model

    model.fit(X_norm, y)
    # evaluate the model
    yhat = model.predict(X_norm)
    # evaluate predictions
    mae = mean_absolute_error(y, yhat)
    print('MAE: %.3f' % mae)
    # R2 score
    R2_score = r2_score(y, yhat)
    print("R2-score: %.3f" % r2_score(y, yhat))
    # creating new column for the dataframe with y_hat data.
    new_column_name = f + 'yhat'
    df_used[new_column_name] = yhat
    # creating new column for the coefficients
    # Plotting the predicted Vs the actual
    # Predicted values
    plt.plot(df['Date'], yhat)
    plt.plot(df['Date'], yhat, '^', markersize = 10,
             label='Predicted', alpha=0.8, color = 'g')
    # Actual Values
    plt.plot(df['Date'], y)
    plt.plot(df['Date'], y, 'o', label='True Value', alpha=0.8, color = 'r')
    # thisaxis.plot(X_test, y_test, 'o', label='Test Value', alpha=0.8)
    plt.xlabel('Date')
    plt.ylabel('Target value')
    plt.title('Linear Regression \n\

```

```

Train $R^2 = {:.3f}$'.format(R2_score))

plt.title(f)

plt.legend()

plt.tight_layout(pad=0.4, w_pad=0.5, h_pad=1.0)

plt.show()

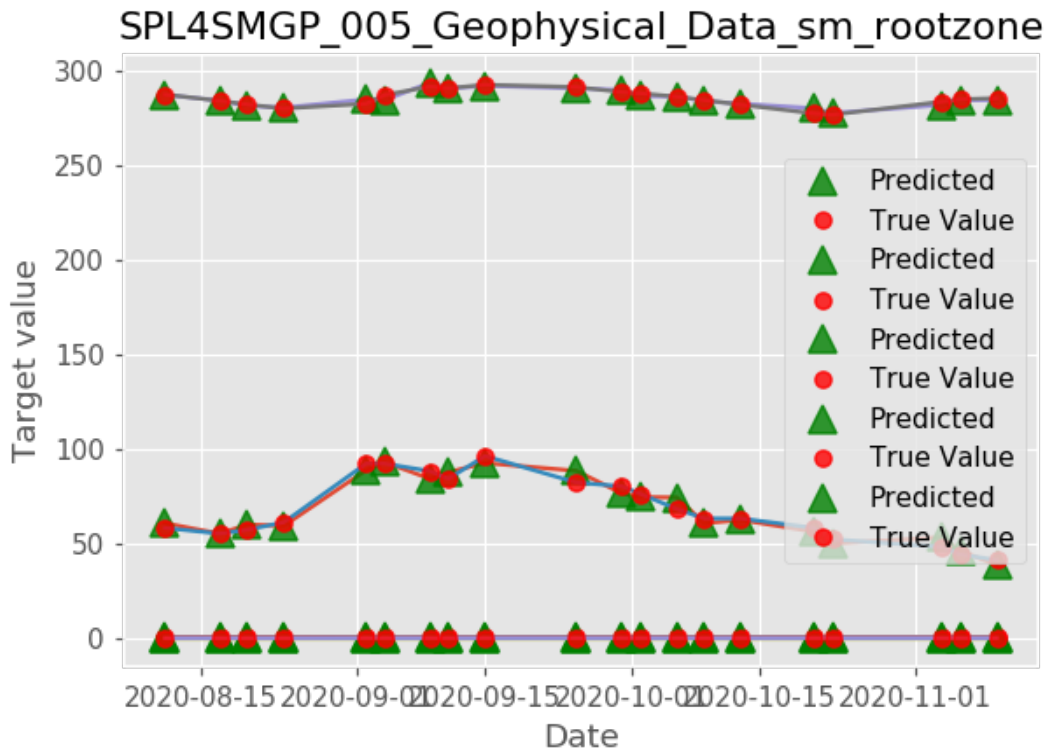
print(f"{f} is {model.coef_}")

# to see if the yhat values are appended to this

```

MAE: 2.865

R2-score: 0.961



```

SPL4SMGP_005_Geophysical_Data_sm_rootzone_pct1_5 is [ 5.846 221.024 -1573
.241 423.222 112.26 83.11 -202.441
-13.155 1046.425 -66.14 30.136 -37.017]

```

MAE: 1.046

R2-score: 0.906

```

GLDAS_NOAH025_3H_2_1_SoilMoi100_200cm_inst is [ -17.437 -265.57 1718.541
-136.183 -388.959 -17.821 25.572

```

```
34.941 -1209.445 20.979 187.523 -13.553]
```

```
MAE: 0.000
```

```
R2-score: 0.856
```

```
SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5 is [ 4.108e-06 1  
.536e-05 -4.551e-04 7.703e-05 1.529e-04 7.472e-06  
-2.683e-05 1.101e-05 3.194e-04 -5.976e-06 -8.779e-05 3.375e-06]
```

```
MAE: 0.003
```

```
R2-score: 0.988
```

```
SPL4SMGP_005_Geophysical_Data_vegetation_greenness_fraction_5 is [-0.035 0.02  
5 0.862 0.006 -0.445 0.162 -0.134 -0.038 -0.323 -0.118  
0.024 -0.025]
```

```
MAE: 0.005
```

```
R2-score: 0.823
```

```
SPL4SMGP_005_Geophysical_Data_sm_rootzone is [-0.021 0.214 -2.609 0.666 0.4  
75 0.01 -0.159 0.017 1.499 0.005  
-0.033 -0.019]
```

```
In [18]:
```

```
df_used.drop(columns = ['SPL4SMGP_005_Geophysical_Data_soil_water_infiltration  
_flux_5yhat', 'SPL4SMGP_005_Geophysical_Data_soil_water_infiltration_flux_5'],  
inplace=True)
```

Exportng the predicted data VS Actual data and plot the graph in excel for the paper later.

```
In [232]:
```

```
df_used.to_excel('Correlated_data_predicted_VS_actual.xlsx', index = False)
```

```
In [ ]:
```

```
In [162]:
```

```
df.columns[1]
```

```
Out [162]:
```

```
'Date_Time'
```

```
In [180]:
```

```
rfe = RFE(estimator=LinearRegression(), n_features_to_select=7)
```

```

model = LinearRegression()
pipeline = Pipeline(steps=[('s',rfe),('m',model)])

# fit the model on all available data
selector = pipeline.fit(X_norm, y)
# make a prediction for one example

yhat = selector.predict(X_norm)
# print('Predicted: %.3f' % (yhat))
R2_score = r2_score(y, yhat)
print("R2-score: %.3f" % r2_score(y, yhat))

for i in range(X.shape[1]):
    if rfe.support_[i] == True:
        print(f'The selected features are {df.columns[i+2]}')

# print('Column: %d, Selected %s, Rank: %.3f' % (i, rfe.support_[i], rfe.r
anking_[i]))
R2-score: 0.802
The selected features are band1_std
The selected features are band1_Eigenvalue
The selected features are band1_band2_covariance
The selected features are band1_band3_covariance
The selected features are band2_std
The selected features are band2_band3_covariance
The selected features are band3_std

```

END OF THE DRONE DATA ANALYSIS

---

## References

- Alizamir, M., Kisi, O., Ahmed, A.N., Mert, C., Fai, C.M., Kim, S., Kim, N.W. and El-Shafie, A. (2020). Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *Plos one*, 15(4), p.e0231055.
- Alshammari, L., Large, D.J., Boyd, D.S., Sowter, A., Anderson, R., Andersen, R. and Marsh, S. (2018). Long-term peatland condition assessment via surface motion monitoring using the ISBAS DInSAR technique over the Flow Country, Scotland. *Remote Sensing*, 10(7), p.1103.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F. and Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, 9(11), pp.3975-3991.
- ArcGIS, 2021. ArcGIS Pro help. Electronic document, available at: <https://pro.arcgis.com/en/pro-app/2.7/help/main/welcome-to-the-arcgis-pro-app-help.htm>
- Ayoub, F., Leprince, S. and Avouac, J.P. (2009). Co-registration and correlation of aerial photographs for ground deformation measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), pp.551-560.
- Bittelli, M., Valentino, R., Salvatorelli, F. and Pisa, P. R. (2012). Monitoring soil-water and displacement conditions leading to landslide occurrence in partially saturated clays. *Geomorphology*, 173, 161-173. <https://doi.org/10.1016/j.geomorph.2012.06.006>.
- Burton, A., Arkell, T.J. and Bathurst, J.C. (1998). Field variability of landslide model parameters. *Environmental Geology*, 35, 100-114.
- Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A. and Kerr, Y. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4994-5007.
- Changchun, L., Li, S., Hai-bo, W. and Tianjie, L. (2010). The research on unmanned aerial vehicle remote sensing and its applications. In 2010 2nd International Conference on Advanced Computer Control (Vol. 2, pp. 644-647). IEEE.
- Chapella, H., Haneberg, W.C., Crawford M.M. and Shakoor, A. (2019). Landslide inventory and susceptibility models, Prestonsburg 7.5-min quadrangle, Kentucky, USA. In: Shakoor A, Cato K (eds) IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018—Volume 1, 217–226. [https://doi.org/10.1007/978-3-319-93124-1\\_26](https://doi.org/10.1007/978-3-319-93124-1_26)
- Chowdhury, R., Flentje, P. and Bhattacharya, G. (2010). *Geotechnical Slope Analysis*. CRC, New York, 738.
- Colliander, A., Fisher, J. B., Halverson, G., Merlin, O., Misra, S., Bindlish, R. and Yueh, S. (2017). Spatial downscaling of SMAP soil moisture using MODIS land surface temperature and NDVI during SMAPVEX15. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 2107-2111.

- Crawford, M.M. (2014). Kentucky Geological Survey landslide inventory: From design to application. Kentucky Geological Survey Information Circular 31, Series 12. <https://doi.org/10.13023/kgs.ic31.12>
- Crawford, M.M. and Andrews W. (2012). Assessing Early Stages of a Landslide Inventory, Digital Mapping Techniques '10—Workshop Proceedings, Sacramento, California, May 16–19, 2010: U.S. Geological Survey Open-File Report 2012–1171, 170 p., available only online at <http://pubs.usgs.gov/of/2012/1171/>
- Crawford, M.M. and Bryson, L.S. (2017). Assessment of active landslides using field electrical measurements. *Engineering Geology*, 233, 146–159. <https://doi.org/10.1016/j.enggeo.2017.11.012>
- Crawford, M.M., Dortch, J.M., Koch, H.J., Killen, A.A., Zhu, J., Zhu, Y., Bryson, L.S. and Haneberg, W.C. (2021). Using landslide-inventory mapping for a combined bagged-trees and logistic-regression approach to determining landslide susceptibility in eastern Kentucky, United States. *Quarterly Journal of Engineering Geology and Hydrogeology*, <http://dx.doi.org/10.1144/qjegh2020-177>.
- Crozier, M. J. (1997). The climate-landslide couple: a southern hemisphere perspective. Rapid mass movement as a source of climatic evidence for the Holocene, 333-354.
- Delacourt, C., Allemand, P., Berthier, E., Raucoules, D., Casson, B., Grandjean, P., Pambrun, C. and Varel, E. (2007). Remote-sensing techniques for analysing landslide kinematics: a review. *Bulletin de la Société Géologique de France*, 178(2), pp.89-100.
- dos Santos, J.F., Silva, H.R., Pinto, F.A. and Assis, I.R.D. (2016). Use of digital images to estimate soil moisture. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 20(12), pp.1051-1056.
- Du, G. L., Zhang, Y. S., Iqbal, J., Yang, Z. H. and Yao, X. (2017). Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *Journal of Mountain Science*, 14(2), 249-268.
- EUROPEAN SPACE AGENCY, 2021. Copernicus Open Access Hub. Electronic document, available at: <https://scihub.copernicus.eu/>
- Fuhrmann, T. and Garthwaite, M.C. (2019). Resolving three-dimensional surface motion with InSAR: Constraints from multi-geometry data fusion. *Remote Sensing*, 11(3), p.241.
- Ge, X., Wang, J., Ding, J., Cao, X., Zhang, Z., Liu, J. and Li, X. (2019). Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *PeerJ*, 7, p.e6926.
- Glover, T. J. and Young, M. M. (1999). *Pocket PC Reference*. Sequoia Publishing.



- Guan, S., Fukami, K., Matsunaka, H., Okami, M., Tanaka, R., Nakano, H., Sakai, T., Nakano, K., Ohdan, H. and Takahashi, K. (2019). Assessing correlation of high-resolution NDVI with fertilizer application level and yield of rice and wheat crops using small UAVs. *Remote Sensing*, 11(2), p.112.
- Gupta, V., Vij, Y. and Gupta, C. (2019). *Toward Analysis of Requirement Prioritization Based Regression Testing Techniques*. System Performance and Management Analytics, Springer, Singapore, 97-101.
- Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Science Reviews*, 200, 102973.
- Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1-2), 42-66. <https://doi.org/10.1016/j.earscirev.2012.02.001>
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72, 272–299. <http://dx.doi.org/10.1016/j.geomorph.2005.06.002>.
- Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd Ed. Chapter 5, John Wiley and Sons, New York, NY, 160-164.
- Huang, Y. and Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena* 165, 520–529. <https://doi.org/10.1016/j.catena.2018.03.003>.
- Intrieri, E., Carlà, T. and Gigli, G. (2019). Forecasting the time of failure of landslides at slope-scale: A literature review. *Earth-science reviews*, 193, pp.333-349.
- Kadavi, P. R., Lee, C. W. and Lee, S. (2019). Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. *Environmental Earth Sciences*, 78(4), 116.
- Kakavas, M.P. and Nikolakopoulos, K.G. (2021). Digital Elevation Models of Rockfalls and Landslides: A Review and Meta-Analysis. *Geosciences*, 11(6), 256.
- Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A. and Mansor, S. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk*, 9(1), 49-69.
- Kentucky Geological Survey, 2019. Landslide Inventory Map. Electronic document, available at: [https://kgs.uky.edu/kgsmap/helpfiles/landslide\\_help.shtml](https://kgs.uky.edu/kgsmap/helpfiles/landslide_help.shtml)
- Kentucky Geological Survey, 2021. Kentucky Transportation Center Geotechnical Data. Electronic document, available at: <https://kgs.uky.edu/kgsmap/kytcLinks.asp>
- Klose, M., Highland, L., Damm, B. and Terhorst, B. (2014). Estimation of direct landslide costs in industrialized countries: Challenges, concepts, and case study. *Landslide Science for a Safer Geoenvironment*, Springer, Cham, 661-667.

- KyFromAbove, 2021. Kentucky's Elevation Data & Aerial Photography Program. Electronic document, available at: <https://kyfromabove.ky.gov>
- L3HARRIS Geospatial, 2021. L3HARRIS Technology Training. Electronic document, available at: <https://www.l3harrisgeospatial.com/Learn/Training>
- Lacasse, S. and Nadim, F. (2009). Landslide risk assessment and mitigation strategy. In *Landslides—disaster risk reduction* (pp. 31-61). Springer, Berlin, Heidelberg.
- Lee, W.S. and Liu, B. (2003). Learning with positive and unlabeled examples using weighted logistic regression. *International Conference on Machine Learning*, Vol. 3, 448-455.
- Leprince, S., Barbot, S., Ayoub, F. and Avouac, J.P. (2007). Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), pp.1529-1558.
- Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), pp.91-110.
- Lu, F., Sun, Y. and Hou, F. (2020). Using UAV Visible Images to Estimate the Soil Moisture of Steppe. *Water*, 12(9), p.2334.
- Lu, N. and Godt, J. (2008). Infinite slope stability under steady unsaturated seepage conditions. *Water Resources Research*, 44(11), W11404.
- Lu, N. and Likos, W. J. (2006). Suction stress characteristic curve for unsaturated soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 132(2), 131-142.
- Lucieer, A., Jong, S.M.D. and Turner, D. (2014). Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. *Progress in Physical Geography*, 38(1), pp.97-116.
- Mathworks Inc., 2021. Classification Learner. Electronic document, available at: <https://www.mathworks.com/help/stats/classificationlearner-app.html>
- McDowell, R.C. (1986). *Geology of Kentucky—A text to accompany the geologic map of Kentucky*. U.S. Geological Survey Professional Paper 1151-H.
- McGrain, Preston. (1983). *The geologic story of Kentucky*. Vol. 8. Kentucky Geological Survey, University of Kentucky.
- Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. and Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 103225. <https://doi.org/10.1016/j.earscirev.2020.103225>.
- Metternicht, G., Hurni, L. and Gogu, R. (2005). Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote sensing of Environment*, 98(2-3), pp.284-303.

- Monsieurs, E., Dewitte, O. and Demoulin, A. (2019). A susceptibility-based rainfall threshold approach for landslide occurrence. *Natural Hazards and Earth System Sciences*, 19(4), 775-789. <https://doi.org/10.5194/nhess-19-775-2019>, 2019.
- Multi-Resolution Land Characteristics (MRLC) Consortium, 2016. The National Land Cover Database (NLCD). Electronic document, available at: <https://www.mrlc.gov/>
- Nandi, A., Mandal, A., Wilson, M. and Smith, D. (2016). Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environmental Earth Sciences*, 75(6), 465.
- Nhu, V. H., Shirzadi, A., Shahabi, H., Singh, S. K., Al-Ansari, N., Clague, J. J. and Ahmad, B. B. (2020). Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *International journal of environmental research and public health*, 17(8), 2749. <https://doi.org/10.3390/ijerph17082749>
- Parker, A.L., Filmer, M.S. and Featherstone, W.E. (2017). First results from Sentinel-1A InSAR over Australia: application to the Perth Basin. *Remote Sensing*, 9(3), p.299.
- Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H., and Bui, D. T. (2019). Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena*, 175, 203-218.
- PIX4D, 2021. Photogrammetry knowledge. Electronic document, available at: <https://support.pix4d.com/hc/en-us/articles/360000235026-Scientific-papers-about-UAV-and-drone-mapping>
- Raspini, F., Bianchini, S., Ciampalini, A., Del Soldato, M., Solari, L., Novali, F., Del Conte, S., Rucci, A., Ferretti, A. and Casagli, N. (2018). Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific reports*, 8(1), pp.1-11.
- Ray, R. L., Jacobs, J. M. and de Alba, P. (2010). Impacts of unsaturated zone soil moisture and groundwater table on slope instability. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(10), 1448-1458.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60-91.
- Reichle, R., De Lannoy, G., Crow, W.T., Koster, R., Kimball, J. (2011). Using SMOS observations in the development of the SMAP level 4 surface and root-zone soil moisture project. *International Geoscience and Remote Sensing Symposium Proceedings*. August 1-5, 2011.
- Rosetta, 2021. Estimate unsaturated soil hydraulic parameters from soil characterization data. Electronic document, available at: <https://www.handbook60.org/rosetta/>

- Rossi, G., Tanteri, L., Tofani, V., Vannocci, P., Moretti, S., and Casagli, N. (2018). Multitemporal UAV surveys for landslide mapping and characterization. *Landslides*, 15(5), 1045-1052.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
- Scaioni, M., Longoni, L., Melillo, V. and Papini, M. (2014). Remote sensing for landslide investigations: an overview of recent achievements and perspectives. *Remote Sensing*, 6(10), pp.9600-9652.
- Schaap, M. G., Leij, F. J., and Van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251(3-4), 163-176. doi: 10.1016/S0022-1694(01)00466-8.
- scikit-learn, 2021. Machine Learning in Python. Electronic document, available at: <https://scikit-learn.org/stable/>
- Seed, H.B., Seed, R.B., Harder, L.F. and Jong, H.L. (1989). Re-evaluation of the slide in the lower San Fernando dam in the earthquake of February 9, 1971. Report no. UCB/EERC-88/04, Earthquake Engineering Research Center, University of California, Berkeley, CA.
- Segoni, S., Piciullo, L., and Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15, 1483–1501. <https://doi.org/10.1007/s10346-018-0966-4>
- Segoni, S., Tofani, V., Rosi, A., Catani, F. and Casagli, N. (2018). Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Frontiers in Earth Science*, 6, 85.
- Shoshany, M., Spond, H. and Bar, D.E. (2019). Overcast versus clear-sky remote sensing: Comparing surface reflectance estimates. *International Journal of Remote Sensing*, 40(17), pp.6737-6751.
- Solari, L., Del Soldato, M., Raspini, F., Barra, A., Bianchini, S., Confuorto, P., Casagli, N. and Crosetto, M. (2020). Review of satellite interferometry for landslide detection in Italy. *Remote Sensing*, 12(8), p.1351.
- Strecha, C., Bronstein, A., Bronstein, M. and Fua, P. (2011). LDAHash: Improved matching with smaller descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 34(1), pp.66-78.
- Strecha, C., Küng, O. and Fua, P. (2012). Automatic mapping from ultra-light UAV imagery (No. CONF).
- Thompson, R.L. and Poindexter, D.B. (2005). Host specificity of American mistletoe (*Phoradendron leucarpum*, Viscaceae) in Garrard County, Kentucky. *Journal of the Kentucky Academy of Science*, 66(1), pp.40-43.
- Timilsina, M., Bhandary, N. P., Dahal, R. K. and Yatabe, R. (2014). Distribution probability of large-scale landslides in central Nepal. *Geomorphology*, 226, 236-248. <https://doi.org/10.1016/j.geomorph.2014.05.031>

- Tofani, V., Segoni, S., Agostini, A., Catani, F. and Casagli, N. (2013). Use of remote sensing for landslide studies in Europe. *Natural Hazards and Earth System Sciences*, 13(2), pp.299-309.
- Triggs, B., McLauchlan, P.F., Hartley, R.I. and Fitzgibbon, A.W. (1999). Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms* (pp. 298-372). Springer, Berlin, Heidelberg.
- Türk, T. (2018). Determination of mass movements in slow-motion landslides by the Cosi-Corr method. *Geomatics, Natural Hazards and Risk*, 9(1), pp.325-336.
- van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5), 892-898.
- Wicki, A., Lehmann, P., Hauck, C., Seneviratne, S. I., Waldner, P., and Stähli, M. (2020). Assessing the potential of soil moisture measurements for regional landslide early warning. *Landslides*, 17(8), 1881-1896.
- Wood, D.M (1990). *Soil Behavior and Critical State Soil Mechanics*. Cambridge University Press, Cambridge.
- Xu, C., Sun, Q., and Yang, X. (2018). A study of the factors influencing the occurrence of landslides in the Wushan area. *Environmental Earth Sciences*, 77(11), 1-8. <https://doi.org/10.1007/s12665-018-7584-2>
- Yang, M., Yang, T., Zhang, L., Lin, J., Qin, X. and Liao, M. (2018). Spatio-temporal characterization of a reclamation settlement in the Shanghai coastal area with time series analyses of X-, C-, and L-band SAR datasets. *Remote Sensing*, 10(2), p.329.
- Yang, W., Wang, Y., Wang, Y., Ma, C., and Ma, Y. (2020). Retrospective deformation of the Baige landslide using optical remote sensing images. *Landslides*, 17(3), 659-668.
- Zanetti, S.S., Cecílio, R.A., Alves, E.G., Silva, V.H. and Sousa, E.F. (2015). Estimation of the moisture content of tropical soils using colour images and artificial neural networks. *Catena*, 135, pp.100-106.
- Zhuo, L., Dai, Q., Han, D., Chen, N., Zhao, B. and Berti, M. (2019). Evaluation of remotely sensed soil moisture for landslide hazard assessment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1), 162-173.

---

**Vita**

BATMYAGMAR DASHBOLD

EDUCATION

**Master of Science:** Fall 2019 – present, Geotechnical Engineering, University of Kentucky, KY.

**Bachelor of Science:** May 2014, Civil Engineering, University of Kentucky, KY.

**Associate of Science:** December 2010, General Engineering, Santa Ana College, CA.

PROFESSIONAL EXPERIENCE

**Water Resource Engineer, Lexington, KY** April 2017- August 2019

**Assistant Dam Safety Engineer, Raleigh, NC** March 2016 – February 2017

**Assistant Dam Safety Engineer I, Frankfort, KY** November 2014 – March 2016