




2023

## HISTORICAL AND FORECASTED KENTUCKY SPECIFIC SLOPE STABILITY ANALYSES USING REMOTELY RETRIEVED HYDROLOGIC AND GEOMORPHOLOGIC DATA

Daniel M. Francis

University of Kentucky, [danny.francis@uky.edu](mailto:danny.francis@uky.edu)

Author ORCID Identifier:

 <https://orcid.org/0009-0001-0598-143X>

Digital Object Identifier: <https://doi.org/10.13023/etd.2023.324>

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

### Recommended Citation

Francis, Daniel M., "HISTORICAL AND FORECASTED KENTUCKY SPECIFIC SLOPE STABILITY ANALYSES USING REMOTELY RETRIEVED HYDROLOGIC AND GEOMORPHOLOGIC DATA" (2023). *Theses and Dissertations--Civil Engineering*. 136.  
[https://uknowledge.uky.edu/ce\\_etds/136](https://uknowledge.uky.edu/ce_etds/136)

This Doctoral Dissertation 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.

Daniel M. Francis, Student

Dr. L. Sebastian Bryson, Major Professor

Dr. Mei Chen, Director of Graduate Studies

HISTORICAL AND FORECASTED KENTUCKY SPECIFIC SLOPE  
STABILITY ANALYSES USING REMOTELY RETRIEVED  
HYDROLOGIC AND GEOMORPHOLOGIC DATA

---

DISSERTATION

---

A dissertation submitted in fulfillment of the  
requirements for the degree of Doctor of Philosophy in the  
College of Arts and Sciences at the University of Kentucky

By

Daniel Marvin Francis

Lexington, Kentucky

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

Lexington, Kentucky

2023

Copyright © Daniel Marvin Francis 2023

# ABSTRACT OF DISSERTATION

## HISTORICAL AND FORECASTED KENTUCKY SPECIFIC SLOPE STABILITY ANALYSES USING REMOTELY RETRIEVED HYDROLOGIC AND GEOMORPHOLOGIC DATA

Hazard analyses of rainfall-induced landslides have typically been observed to experience a lack of inclusion of measurements of soil moisture within a given soil layer at a site of interest. Soil moisture is a hydromechanical variable capable of both strength gains and reductions within soil systems. However, in situ monitoring of soil moisture at every site of interest is an unfeasible goal. Therefore, spatiotemporal estimates of soil moisture that are representative of in-situ conditions are required for use in subsequent landslide hazard analyses.

This study brings together various techniques for the acquisition, modeling, and forecasting of spatiotemporal retrievals of soil moisture across areas of Eastern Kentucky for use in hazard analyses. These techniques include: A novel approach for determination of satellite-based soil moisture retrieval correction factors for use in acquisition of low orbit-based soil moisture retrievals in site-specific analyses, unique spatiotemporal modeling of soil moisture at various depths within the soil layer through assimilation of satellite-based and land surface modeled soil moisture estimates, and the development of a novel workflow to effectively provide 7-day forecasts of soil moisture for use in subsequent forecasting of landslide hazards.

Soil moisture retrieved through the previous approaches was implemented within landslide hazard and susceptibility analyses across known rainfall-induced landslides within Eastern Kentucky. Investigated analyses were conducted through a coupling of spatial soil moisture retrievals with that of site-specific geomorphologic data. These analyses proved capable in the detection of incipient failure conditions indicative of landslide occurrence over these known investigated slides. These soil moisture-based analyses show that inclusion of soil moisture, as hydromechanical variable, yields a more capable hazard analysis approach. Additionally, these analyses serve as a means to gain a better understanding of the coupled hydro-mechanical behavior associated with the initiation of rainfall-induced landslides.

KEYWORDS: Rainfall-induced, landslides, Land Information System, machine learning, susceptibility, soil moisture

---

Daniel M. Francis

---

July 3, 2023

---

---

HISTORICAL AND FORECASTED KENTUCKY SPECIFIC SLOPE STABILITY  
ANALYSES USING REMOTELY RETRIEVED HYDROLOGIC AND  
GEOMORPHOLOGIC DATA

By

Daniel Marvin Francis

---

---

L. Sebastian Bryson

Director of Dissertation

---

Mei Chen

Director of Graduate Studies

---

July 3, 2023

---

## ACKNOWLEDGEMENTS

Thanks to my committee members, Dr. L. Sebastian Bryson, Dr. Michael Kalinski, Dr. Reginald Souleyrette, and Dr. Edward Woolery who provided thoughtful advice during the development of this research. I would like to give a special thank you to the director of my dissertation, Dr. Bryson. Dr. Bryson has always shown himself to be extremely knowledgeable, patient, understanding, eager, and overall, an exceptional mentor throughout this process. Dr. Bryson continually pushed me to be able to find answers to my questions on my own, rather than simply providing an easy answer. However, if I was ever truly baffled by a problem, he would do everything in his power to work with me to be able to find a solution together. I would not have benefitted nearly as much as I have from the process of pursuing this degree if it were not for Dr. Bryson.

I would also like to thank following individuals for their invaluable help in understanding the development of existing techniques utilized herein: Dr. Doug Baldwin of SCS Global Services, Dr. Dalia Kirschbaum of the NASA Earth Sciences Division, and Mr. Thomas Stanley of NASA Landslide Research Group.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES .....	viii
CHAPTER 1: Introduction .....	1
1.1    Problem Statement.....	1
1.2    Soil Moisture Hydromechanical Influences.....	2
1.3    Conceptual Overview of Study.....	2
1.4    Objectives .....	4
1.5    Contents of Dissertation.....	5
CHAPTER 2: Proposed Methodology for Site Specific Soil Moisture Obtainment Utilizing Coarse Satellite-Based Data.....	7
2.1    INTRODUCTION .....	7
2.2    SATELLITE DATA ACQUISITION.....	9
2.2.1    L4_SM RETRIEVALS .....	9
2.2.2    2.2 L3SMP_E RETRIEVALS .....	9
2.3    GROUND BASED SOIL MOISTURE SITES .....	10
2.4    EXISTING METHODS FOR SITE-SPECIFIC SOIL MOISTURE .....	12
2.4.1    STATISTICAL DOWNSCALING APPROACHES.....	12
2.4.2    SITE-SPECIFIC ASSIMILATION APPROACH .....	18
2.5    SITE-SPECIFIC SOIL MOISTURE ESTIMATES THROUGH REGRESSION .....	21
2.5.1    SITE-SPECIFIC L4_SM CORRECTION FACTORS .....	22
2.5.2    MULTIVARIATE LINEAR REGRESSION USING SOIL TEXTURE DATA..	23
2.5.3    SITE-SPECIFIC CORRECTED L4_SM VS IN-SITU DATA .....	25
2.6    COMPARATIVE ANALYSIS OF SITE-SPECIFIC METHODS.....	28
2.7    CONCLUSIONS.....	32



CHAPTER 3: Rainfall-Induced Landslide Hazard Analyses using Spatiotemporal Retrievals of Soil Moisture and Geomorphologic Data .....	34
3.1    INTRODUCTION .....	34
3.2    SPATIAL GEOMORPHOLOGIC DATA.....	36
3.3    ARTIFICIAL NEURAL NETWORK PREDICTION OF FRICTION ANGLES .....	37
3.3.1    USE OF EMPIRICAL RELATIONSHIP TO ESTIMATE FRICTION ANGLE .	37
3.3.2    ARTIFICIAL NEURAL NETWORK PREDICTION OF FRICTION ANGLES	38
3.4    SPATIAL SOIL MOISTURE RETRIEVALS .....	40
3.4.1    OVERARCHING LIS FRAMEWORK.....	41
3.4.2    LIS SATELLITE-BASED DATA ASSIMILATION PREPROCESSING .....	43
3.4.3    LIS NOAH 3.6 LAND SURFACE MODEL RUNS .....	45
3.4.4    LIS NOAH 3.6 MODEL RUNS .....	46
3.5    HYDROLOGIC BEHAVIOR VIA SPATIAL LIS DATA .....	49
3.6    SLOPE STABILITY ANALYSES.....	52
3.6.1    LANDSLIDE STUDY AREA .....	52
3.6.2    EASTERN KENTUCKY RAINFALL EVENT.....	53
3.6.3    INFINITE SLOPE STABILITY .....	55
3.6.4    LIS NOAH 3.6 BASED STABILITY ANALYSES.....	55
3.7    CONCLUSIONS.....	61
CHAPTER 4: Coupled Landslide Analyses Through Dynamic Susceptibility and Forecastable Hazard Analysis .....	63
4.1    INTRODUCTION .....	63
4.2    LIS-BASED SPATIAL SOIL MOISTURE RETRIEVALS .....	65
4.2.1    LIS EXECUTABLES AND WORKFLOW .....	66
4.2.2    LIS LAND SURFACE MODEL AND ASSIMILATED PRODUCT.....	68
4.2.3    LIS NOAH 3.6 ASSIMILATION RUNS .....	70
4.3    SPATIAL RETRIEVALS OF GEOMORPHOLOGIC DATA .....	72

4.3.1	SPATIAL ESTIMATES OF DEPTH TO BEDROCK .....	73
4.3.2	SPATIAL ESTIMATES OF FRICTION ANGLE .....	74
4.4	MACHINE LEARNING-BASED LANDSLIDE SUSCEPTIBILITY .....	74
4.4.1	NDVI ACQUISITION FOR LANDSLIDE SUSCEPTIBILITY .....	74
4.4.2	LOGISTIC REGRESSION SUSCEPTIBILITY MODEL .....	75
4.4.3	TEMPORAL VARIATION OF SUSCEPTIBILITY .....	77
4.4.4	VALIDATION OF SUSCEPTIBILITY MODEL .....	77
4.5	HISTORIC LANDSLIDE HAZARD ANALYSES .....	79
4.5.1	HYDROLOGIC VARIABLES VIA LIS-BASED MOISTURE RETRIEVALS..	79
4.5.2	INFINITE SLOPE STABILITY .....	81
4.5.3	LIS-MOISTURE BASED LANDSLIDE HAZARD ANALYSES .....	81
4.6	FORECASTING OF LANDSLIDE HAZARD ANALYSES .....	83
4.6.1	LSTM OVERARCHING FRAMEWORK .....	84
4.6.2	LSTM MODEL TRAINING .....	86
4.6.3	LSTM-BASED FORECASTED HAZARDS .....	89
4.7	CONCLUSIONS.....	91
CHAPTER 5: Conclusions .....		95
APPENDIX.....		97
REFERENCES .....		108
CURRICULUM VITAE .....		117

## LIST OF TABLES

Table 2-1: Commonwealth of Kentucky In-Situ Ground Station Locations .....	11
Table 2-2: Valid Ranges and Scale Factors for MODIS Data .....	14
Table 2-3: $R^2$ Values from Data Imputation for MODIS Time-Series Data .....	16
Table 2-4: Average Optimized Hyperparameter Values/Criterion used for RFR Training over In-Situ Sites .....	18
Table 2-5: Optimized L4_SM Offset Factors from 10 Analyzed In-Situ Sites .....	23
Table 2-6: Independent Variables used in Multivariate Regression Analysis for MF and AF Factors.....	24
Table 2-7: Excel Optimized Offset Factors Compared to Multivariate Regression Predicted Offset Factors.....	25
Table 3-1: Soil and Slope Characteristic Data Sources .....	36
Table 3-2: LDT Preprocessed Physical Parameter Information .....	42
Table 4-1: LDT Preprocessed Physical Parameter Information .....	67
Table 4-2: Soil Geomorphologic and Slope Characteristic Data Sources .....	72

## LIST OF FIGURES

Figure 1-1: Overall Aim and Chapter Objectives of Work Discussed.....	5
Figure 2-1: Location of In-Situ Data Sites Within Commonwealth of Kentucky .....	12
Figure 2-2: 9 km x 9 km EASE2.0 Grid Over Doe Run In-Situ KGS Station (red marker).....	14
Figure 2-3: Generalized Workflow for Scale Matching MODIS data.....	15
Figure 2-4: General Workflow for Downscaling L4_SM Data Using RFR Model.....	17
Figure 2-5: Generalized Workflow for EnKF Assimilation .....	19
Figure 2-6: 9 km L4_SM (blue) vs Doe Run In-Situ Data (orange), Excel Optimized Regressed L4_SM (grey), and L4_SM Regressed Using Correction Factors Predicted by Multivariate Regression (yellow) .....	26
Figure 2-7: 9 km L4_SM (blue) vs Caldwell County In-Situ Data (orange), Excel Optimized Regressed L4_SM (grey), and L4_SM Regressed Using Correction Factors Predicted by Multivariate Regression (yellow) .....	27
Figure 2-8: 9 km L4_SM (blue) vs Barren County In-Situ Data (orange), SEE Downscaled L4_SM Data (grey), and SEE Downscaled Linear Regressed L4_SM Data (yellow) .....	29
Figure 2-9: Comparison of Roberts Bend In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR .....	30
Figure 2-10: 9 km L4_SM RMSE vs Obtained Site-Specific Moisture RMSE .....	31
Figure 3-1: Plasticity Index vs Liquid Limit Retrieved via WSS and Kentucky Borehole Explorations .....	38
Figure 3-2: Multilayer Perceptron Neural Network Architecture, after (da Silva et al., 2018) .....	39
Figure 3-3: MLP Neural Network Predicted vs Mesri Friction Angles at 20 Retained Test Sites	40
Figure 3-4: Generalized Flowchart for LIS Framework .....	41
Figure 3-5: LIS Domain Over Kentucky .....	41
Figure 3-6: (A) Missing L3SMP_E (shown as areas of black) Over Domain Due to LIS Satellite Data Quality Checks and (B) Missing L3SMP_E (shown as areas of black) Over Domain After Adjustment to LIS Satellite Data Quality Checks .....	45
Figure 3-7: Ground-Based In-Situ Sites Across Kentucky .....	46
Figure 3-8: Comparison of OL and DA LIS NOAA 3.6 Estimated Moisture to Mesonet Barren County In-Situ Moisture at: (A) 100 mm, (B) 300 mm, (C) 600 mm, and (D) 1000 mm	47

Figure 3-9: Comparison of OL and DA LIS NOAH 3.6 Estimated Moisture to Mesonet Fayette County In-Situ Moisture at: (A) 100 mm, (B) 300 mm, (C) 600 mm, and (D) 1000 mm	48
Figure 3-10: Mesonet Barren County Measurements of: (A) In-Situ vs LIS Estimates of VWC, (B) In-Situ vs LIS-based Matric Suction, and (C) In-Situ vs LIS-based Suction Stress...	51
Figure 3-11: Landslide Study Area: (A) Overview of Kentucky with Locations of Eastern Kentucky Rain Induced Landslides Enclosed in Black Box and (B) Exploded View of Eastern Kentucky Rain Induced Landslide Region .....	52
Figure 3-12: Rainfall (mm/day) Over Landslide Study Area from 4/1/2022 to 10/31/2022 .....	53
Figure 3-13: Average of LIS Estimates of Volumetric Water Content from Landslide Area over:(A) Full LIS Landslide Period Run (01/01/21 - 10/01/22) and (B) Snapshot of LIS Landslide Period Run Over Period Including Rain Event (07/10/22 - 08/9/22).....	54
Figure 3-14: Hill Shade map of Landslide Area with 3 x 3 km Landslide Focus Areas A and B .	56
Figure 3-15: Stability Analyses from Focus Area A at 1000 mm Depth on: (A) 07/22/2022, (B) 07/28/2022, (C) 07/29/2022, and (D) 07/30/2022 .....	57
Figure 3-16: Stability Analyses from Focus Area B at 1000 mm Depth on: (A) 07/22/2022, (B) 07/28/2022, (C) 07/29/2022, and (D) 07/30/2022 .....	58
Figure 4-1: LIS Domain Over Eastern Kentucky .....	66
Figure 4-2: Ground Based Validation Sites within LIS Domain .....	69
Figure 4-3: Comparison of Assimilated SMAP and NOAH 3.6 Estimated Moisture to Mesonet Breathitt County In-Situ Moisture at: (A) 100 mm, (B) 200 mm, (C) 500 mm, and (D) 1000 mm .....	71
Figure 4-4: (A) Locations of KYTC Boreholes with DTB Data Across Domain and (B) Novel DTB Map Developed Through Kriging of DTB Data from 3000 Boreholes .....	73
Figure 4-5: Confusion Matrix for Trained Logistic Regression Susceptibility Classifier .....	76
Figure 4-6: Susceptibility Over Harlan County, Kentucky Using NDVI from: (A) 02/10/2023 to 02/18/2023 and (B) 03/30/2023 to 04/07/2023 .....	77
Figure 4-7: Susceptibility Over: (A) Landslide 10757, Occurred 03/17/2021, (B) Landslide 9677, Occurred 06/17/2019, (C) Landslide 8596, Reported 01/04/2019, and (D) Landslide 6456, Occurred 02/10/2018 .....	78
Figure 4-8: Infinite Slope Stability Analysis at: (A) Landslide 10757, Occurred 03/17/2021, (B) Landslide 8596, Reported 01/04/2019, (C) Landslide 9677, Occurred 06/17/2019, and (D) Landslide 6456, Occurred 02/10/2018 .....	82
Figure 4-9: Generalized Framework of Long Short Term Memory Model.....	84
Figure 4-10: Stacked LSTM Configuration.....	85

Figure 4-11: RMSE of 7 Day Forecasted LSTM Data Compared to Amount of Training Data ...	87
Figure 4-12: Behavior of Soil Moisture Over: (A) 365 Days, (B) 730 Days, and (C) 1825 Days of Training Data and (D) Example of LSTM-based 7 Day Forecast Using 1825 Days of Training Data .....	88
Figure 4-13: (A) 100 m Susceptibility Analysis for Landslide 10793 (Occurred 01/02/2022), (B) Pseudo-Forecasted Hazard Analysis for Landslide 10793, (C) 100 m Susceptibility Analysis for Landslide 10794 (Occurred 01/03/2022), and (D) Pseudo-Forecasted Hazard Analysis for Landslide 10794 .....	90
Figure A1: (A) 9 km L4_SM corrections at Adair County site, (B) 9 km L4_SM corrections at Barren County site, (C) 9 km L4_SM corrections at Caldwell County site, (D) 9 km L4_SM corrections at Doe Run site, (E) 9 km L4_SM corrections at Fayette County site, (F) 9 km L4_SM corrections at Madison County site, (G) 9 km L4_SM corrections at Mammoth Cave site, (H) 9 km L4_SM corrections at Roberts Bend site, (I) 9 km L4_SM corrections at Shelby County site, and (J) 9 km L4_SM corrections at Versailles site ....	98
Figure A2: Comparison of Adair County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR .....	99
Figure A3: Comparison of Barren County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	100
Figure A4: Comparison of Caldwell County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	101
Figure A5: Comparison of Doe Run In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	102
Figure A6: Comparison of Fayette County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	103
Figure A7: Comparison of Madison County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	104
Figure A8: Comparison of Mammoth Cave In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	105
Figure A9: Comparison of Shelby County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR ....	106

Figure A10: Comparison of Versailles In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR.... 107

# CHAPTER 1: Introduction

## 1.1 PROBLEM STATEMENT

Rainfall-induced landslides are spatial geohazards which pose a severe threat to residential and civil infrastructure systems. Every year, rainfall induced landslides are associated with significant economic loss as well as losses of human life (Petley, 2012). Typically, these landslides occur due to rapid increases of moisture within the soil due to these rainfall events. This rapid increase in soil moisture reduces the overall strength of the soil as well as increases existing stresses within the soil (Hong et al., 2007; Ray and Jacobs, 2007). Additionally, as global variability increases, so do climatological influences that lead to the occurrence of these severe rainfall events. This observable increase of these rainfall events leads to overall increases of the likelihood of the occurrences of these rainfall induced landslides (Sidle and Bogaard 2016, Swain et al., 2020). With this knowledge of potential increase in these severe geohazards comes the desire for a means to accurately monitor landslide prone areas for potential risks of occurrence. Existing systems for this monitoring typically only consider site conditions and levels of potential precipitation. While surface soil moisture (0 – 5 cm) levels respond rapidly to precipitation, deeper seated moisture typically does not respond as quickly (Akbar et al., 2018). Subsequently, measurements of potential precipitation over an area are unlikely to capture in-situ soil moisture conditions in these deeper layers at the time of analysis. Therefore, a lack of inclusion of soil moisture, considered a driving factor in the occurrence of landslides, has been observed within existing landslide hazard analysis systems.

This study primarily focuses upon the inclusion of soil moisture measurements within landslide hazard analyses. These measurements are retrieved either directly via low Earth orbit satellite, or through direct assimilations of satellite-based moisture measurements. The inclusion of these moisture measurements in landslide hazard analyses provides two overarching benefits as compared to existing landslide analysis systems. The first benefit being that these analyses now consider the driving factor of landslide occurrence, soil moisture. The second benefit is that, as soil moisture varies with time, this implies that these analyses will also vary with time. Therefore, these analyses are no longer static analyses, as observed with certain existing systems, but are instead capable of temporal variations. Considering these benefits, this study develops a means to conduct landslide hazard analyses observed to vary spatially and temporally, as with soil moisture measurements, which are capable of detecting incipient conditions indicative of potential landslide occurrences.



## 1.2 SOIL MOISTURE HYDROMECHANICAL INFLUENCES

The study presented herein conducts investigations into strength gains and losses within a soil layer dependent upon variations of soil moisture. These investigations are investigated through relations of suction stress ( $\sigma^s$ ) to that of soil moisture variations. Suction stress is a force within a soil layer composed of physiochemical, cementation, surface tension, and negative pore-water pressure forces subjected to a soil through exposure to moisture (Lu and Likos, 2006). Suction stress provides tensile forces within the soil layer, leading to an overall increase in the strength of that layer. Additionally, suction stress can be observed to reduce the overall effective stress that a soil system is exposed to. This generalized effective stress for soils, proposed by Lu and Likos (2004, 2006), is given as the total stress minus the pore air pressure minus the suction stress (i.e.,  $\sigma' = (\sigma - u_a) - \sigma^s$ ). As the soil approaches saturated conditions, suction stress values decrease leading to a reduction in the overall strength gain previously provided at values of lesser saturation. Therefore, suction stress, and related strength gains, can be observed as inversely related to soil moisture. As soil moisture increases, suction stress decreases. This decrease leads to an overall increase of effective stress and decrease of strength of the soil layer.

As a function of soil moisture, suction stress effectively becomes a function of the effective degree of saturation ( $S_e$ ) and the matric suction ( $s$ ) within the analyzed soil. Matric suction, the difference between pore air and pore water pressures (i.e.,  $u_a - u_w$ ), is maintained as a strong precursor to rainfall-induced landslides. Particularly, matric suction is a good indicator of reductions of shear strength within a soil. These shear strength reductions are non-linearly related to decreasing levels of matric suction. Therefore, as matric suction decreases with increasing soil moisture, so does the available shear strength within the soil (Bittelli et al., 2012; Mauri et al., 2022). This reduction in strength leads to an increased likelihood of landslide occurrence in direct relation to increases of soil moisture.

## 1.3 CONCEPTUAL OVERVIEW OF STUDY

This study develops a Kentucky specific hydrologic based landslide hazard analysis regime through inclusion of soil moisture measurements over analyzed areas. Herein, soil moisture is assumed to be the primary factor in the overall occurrence of a rainfall-induced landslide. Therefore, the primary aim of this study became the acquisition of spatially and temporally representative measurements of soil moisture. Additionally, it was observed that spatially measurements of requisite geomorphologic parameters, such as friction angle and depth to

bedrock, were seen to be lacking. As this study progressed, efforts were made to address this lack of spatial geomorphic data as well as spatiotemporal retrievals of soil moisture.

The first aim of this study was the determination of the efficacy of soil moisture measurements retrieved via satellite-based platforms. Direct comparisons were conducted between soil moisture retrieved via the NASA Soil Moisture Active Passive (SMAP) earth observing satellite and that of in-situ measurements. In general, it was observed that SMAP estimates captured seasonality, but either over or underpredicted as compared to in-situ measurements. This disparity between in-situ and SMAP data was observed to be due to the resolution of SMAP retrievals. SMAP retrievals are readily available for retrieval at a gridded resolution of 9 km. This resolution makes this product a powerful tool for global analyses. However, a product more representative of site conditions is desired for use in site specific analyses, such as those of landslide hazards. Therefore, efforts were made to improve the efficacy of these SMAP based retrievals for use in site-based analyses. These efforts investigated both downscaling (i.e., the improvement of resolution) and reductions of biases (i.e., errors) through assimilation approaches. Both downscaled and bias-reduced SMAP performed better in comparison to in-situ measurements as compared to native resolution SMAP. However, these approaches were observed to be both intensive and tedious to conduct. Therefore, a multivariate linear regression approach was developed herein which was capable of yielding SMAP measurements observed to better align with that of in-situ data. This approach conducted no downscaling or bias-reduction through assimilations and relied upon soil textures at a given site, which are readily available via web databases for retrieval. The herein developed regression approach provides a far less intensive approach that is still capable of improving the utility of SMAP-based retrievals at a site-based scale.

The previously investigated SMAP soil moisture measurements are a powerful tool in hydrologic analyses. However, these measurements are only available at the surface layer (0 – 50 mm) and at the root zone layer (1000 mm). As landslides can occur at various depths within a soil layer, spatial measurements of soil moisture at varying depths were desired. To obtain these measurements, the Land Information System (LIS) framework was utilized. LIS allows for assimilation of Land Surface Modeled and satellite-based soil moisture products. The LIS framework was modified through this study to obtain spatial retrievals of soil moisture at a resolution of 1 km across Kentucky as a function of modeled and satellite-based moisture products. LIS-based moisture estimates were then used to conduct hazard analyses over a known severe landslide event in Eastern Kentucky. However, these analyses required spatial retrievals of

friction angle, which was observed to be unavailable across the area of interest. To remedy this, an Artificial Neural Network (ANN) machine learning workflow was developed that determined relationships between spatial measurements of soil texture and friction angle. Estimates of friction angle were then used in conjunction with LIS-based soil moisture estimates at varying depths to efficiently identify incipient conditions over the known severe landslide event.

With a means to obtain spatial estimates of soil moisture at varying depths, the efforts of this study shifted to the improvement of existing landslide susceptibility and hazard analysis approaches. Current susceptibility analyses typically account only for static geomorphologic data, therefore becoming static tools themselves. Current hazard analyses consider both static geomorphic data as well as dynamic precipitation data but are capable of providing warnings of potential landslide occurrences within near real-time of the occurrence. To address these observed shortcomings, a dynamic landslide susceptibility was first developed. This susceptibility system was developed upon the inclusion of measurements of vegetation, known to have influences on landslide occurrences (Guo et al., 2020; Phillips et al., 2021), over analyzed sites. As vegetation measurements vary spatially and temporally, so do the developed susceptibility analyses. With dynamic susceptibility, a forecastable landslide hazard approach was then developed. This approach was developed through construction of a Long Short-Term Memory (LSTM) forecasting machine learning workflow capable of forecasting 7-day soil moisture measurements at a site of interest. As landslide hazard analyses were conducted throughout this study as a function of soil moisture, the ability to forecast soil moisture effectively yields forecastable landslide hazard analyses. The developed dynamic susceptibility and forecastable hazard analyses are capable of providing greater warnings of potential landslide occurrences. This extended warning period provided by these systems will undoubtedly provide powerful tools for planners and designers in mountainous landslide prone regions around the world.

#### **1.4 OBJECTIVES**

The overall objectives, as well as the order in which these objectives were approached throughout this work, are shown in the following figure, Figure 1-1.

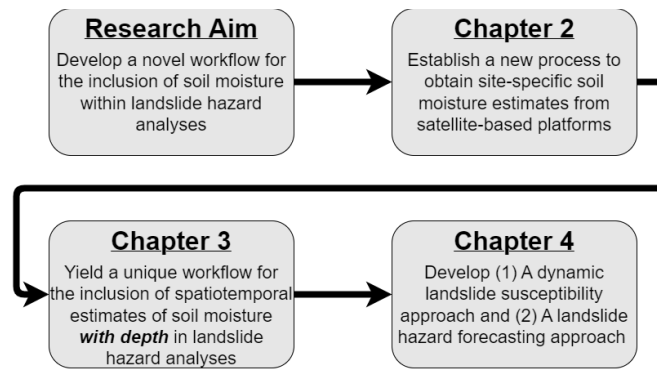


Figure 1-1: Overall Aim and Chapter Objectives of Work Discussed

As shown in Figure 1-1, the overall aim of this research was the development of a novel workflow for the inclusion of soil moisture within landslide hazard analyses. The aim of this research was achieved through the work conducted throughout Chapters 2-4. These chapters investigated novel objectives which aided in the completion of the overall research aim. The novel objectives for each chapter herein are given in Figure 1-1.

## 1.5 CONTENTS OF DISSERTATION

Chapter 1 – Introduction: This chapter consists of the Problem Statement (what issues this research addresses), a generalized conceptual overview of the conducted research presented herein, study objectives, and contents of the Dissertation.

Chapters 2 – 4 consist of published studies and the contents within are verbatim.

- Chapter 2 presents investigations into existing workflows for the downscaling or refinement of SMAP satellite-based soil moisture retrievals. Downscaling made efforts to improve the resolution of SMAP retrievals from a resolution of 9 km to a finer resolution of 1 km. Refinement efforts made use of bias-reduction and modeling techniques to improve accuracy of the native 9 km product as compared to in-situ data. In contrast to these existing workflows, a novel multivariate regression approach was developed which associated soil textures at a site to correction factors applied to 9 km SMAP retrievals. It was observed that the developed regression approach was less tedious while just as capable as existing approaches. This chapter was submitted to Environmental Earth Sciences Journal in 2022.

*Francis, D. M., and Bryson, L. S. (2023). "Proposed Methodology for Site Specific Soil Moisture Obtainment Utilizing Coarse Satellite-Based Data." Environmental Earth Sciences, <https://doi.org/10.1007/s12665-023-11057-0>.*

- Chapter 3 presents a unique methodology for spatiotemporal retrieval of soil moisture with depth through assimilation of modeled and satellite-based soil moisture products. These retrievals were first validated against multiple in-situ based soil moisture monitoring sites. Once validated, these moisture retrievals were then used in conjunction with a spatiotemporal landslide hazard analysis over a known severe landslide event. Through implementation of these spatiotemporal retrievals of soil moisture, incipient conditions of a majority of investigated known landslides were able to be detected. This chapter was submitted to the AGU Water Resources Journal in 2023.

*Francis, D. M., and Bryson, L. S. (2023). "Rainfall-Induced Landslide Hazard Analyses using Spatiotemporal Retrievals of Soil Moisture and Geomorphologic Data." Water Resources Research Journal. SUBMITTED.*

- Chapter 4 presents the novel workflow for dynamic landslide susceptibility and forecastable landslide hazard analyses. The dynamic susceptibility system was built upon the inclusion of temporally varying vegetation measurements, which have been observed to be lacking in existing Kentucky susceptibility systems. Landslide hazard forecasting was developed through the inclusion of forecasted soil moisture measurements. These measurements were obtained through a predictive analytics machine learning workflow developed as a soil moisture forecasting tool. As hazard analyses were conducted as a function of soil moisture, this allowed for forecasting of these analyses through inclusion of forecasted moisture. The dynamic susceptibility and forecasted hazard analysis systems both performed satisfactorily at capturing temporally varied incipient conditions indicative of landslide occurrences.

*Francis, D. M., and Bryson, L. S. (2023). "Coupled Landslide Analyses Through Dynamic Susceptibility and Forecastable Hazard Analysis." Journal of Natural Hazards. SUBMITTED.*

Chapter 5 – Conclusions: This chapter summarizes the overall methods and findings of the research presented in Chapters 2 – 4.

## CHAPTER 2: Proposed Methodology for Site Specific Soil Moisture Obtainment Utilizing Coarse Satellite-Based Data

### 2.1 INTRODUCTION

Soil Moisture (SM) in the climate system integrates the combined influence of the atmosphere, land surface, and soil (Quiring et al., 2016). Soil moisture data is frequently used in drought monitoring, climate forecasting, and natural hazard analyses. However, acquisition of comprehensive soil moisture data at sites of interest (e.g., agricultural sites exposed to drought or sites prone to the occurrence of hazards, such as landslides) can be a costly and tedious endeavor. Additionally, ground-based soil moisture networks are too sparsely distributed to produce soil moisture maps at local, regional, or global scales (Vinnikov et al., 1999, Yuan et al., 2020). Fortunately, remotely sensed (e.g., satellite-based) estimates of soil moisture data can be used in lieu of physical acquisitions of soil moisture from ground-based networks. Due to its sensitivity to subsurface SM and relative insensitivity to vegetation, low-frequency passive microwave remote sensing has been established as the primary tool for retrievals of SM on a global scale (Schmugge et al., 1986). The National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) satellite mission was designed to retrieve soil moisture estimates at a 9 km x 9 km resolution (Miralles et al., 2010). This resolution is generally too coarse for most site-specific analyses. Therefore, a means to obtain site-specific soil moisture estimates from the SMAP satellite-based moisture dataset is desired. The study discussed herein provides a comparative analysis upon the efficacy of methods used to remotely obtain these site-specific moisture estimates.

This study investigates various methods to obtain site-specific soil moisture estimates from the NASA SMAP Earth satellite mission. The 9 km SMAP products investigated for use in obtaining these site-specific estimates are the L3SMP\_E enhanced surface (0-5 cm) soil moisture and Level 4 (L4\_SM) Root Zone Soil Moisture (RZSM) (0-100 cm of the soil column) products. To utilize these products at a more site-specific scale, this study investigates downscaling or assimilation techniques that are commonly used by members of the geoscience and remote sensing communities. These approaches are that of Random Forest downscaling, Soil Evaporative Efficiency (SEE) downscaling, and an Ensemble Kalman Filter (EnKF) based assimilation routine. With respect to Random Forest and SEE downscaling, these methods are utilized to improve the resolution of the SMAP product from the native resolution to a finer resolution. Through this study, these downscaling approaches were implemented to improve the resolution of

SMAP from 9 km to 1 km. EnKF based assimilation does not improve resolution of the SMAP product, but instead improves the accuracy of the product as compared to in-situ sites within an analysis area. Through this study, EnKF assimilation was used to improve the accuracy of L3SMP\_E estimates. These improved estimates were then coupled with an infiltration model to estimate root zone moisture at sites of interest.

In addition to investigating existing methodologies currently used to obtain site-specific soil moisture from satellite-based data, this study proposes a new approach for site-specific analyses. The proposed approach developed a multivariate regression analysis which characterized relationships between site-specific soil texture data and SMAP L4\_SM root zone soil moisture correction factors. As the intent of this study is to provide a means to obtain site-specific estimates remotely, these texture data were also obtained remotely via Web Soil Survey (WSS). This approach was conducted over various in-situ sites across the Commonwealth of Kentucky. These sites served as control sites where various L4\_SM correction factors and associated soil texture data were obtained. These data were then incorporated into the developed multivariate regression analysis. The proposed approach makes no efforts to improve the resolution of the L4\_SM data, but instead aims to improve the accuracy of the data as compared to site-specific locations. Once validated, this approach allows for L4\_SM data to be obtained over a site of interest and then corrected to that site based on easily obtained/estimated soil texture parameters.

The intent of this study was that of a twofold approach. The first was that of the proposal of the multivariate regression methodology used to obtain site-specific data from coarse resolution SMAP L4\_SM estimates through inclusion of site-specific soil information. The second was that of the associated comparative analysis between existing methods and the herein proposed approach. It has been observed that, while effective, the existing methods (Random Forest downscaling, SEE downscaling, and EnKF assimilation) typically require intensive data preprocessing and understanding of the associated approach to be effectively conducted over sites of interest. The intent behind the proposed approach, multivariate regressive analysis, is to provide a means to quickly, yet accurately, obtain site-specific soil moisture data from the readily available SMAP soil moisture datasets. The study presented herein shows that the proposed multivariate regression approach is far less intensive, yet still yields site-specific moisture estimates comparable to that of associated downscaling or assimilated approaches.

## **2.2 SATELLITE DATA ACQUISITION**

This study made use of two remotely sensed soil moisture products available via the SMAP low Earth orbit satellite mission. The first soil moisture product, used for downscaling and proposed regression methods, is that of 9 km Root Zone Soil Moisture (L4\_SM). The second product, used for EnKF assimilation, is that of 9 km Enhanced Surface Soil Moisture (L3SMP\_E).

### **2.2.1 L4\_SM RETRIEVALS**

The SMAP Level 4 soil moisture (SMAP L4\_SM) product is obtained through the merging (i.e., assimilation) of SMAP soil moisture estimates with estimates from a Catchment land surface model, effectively constructing a model-derived and value-added product. The Catchment LSM describes the transfer of soil moisture vertically between the surface and root zone reservoirs and is driven by observation-based surface meteorological forcing data, which includes precipitation (Reichle et al., 2014). The model-derived product produces 3-hourly estimates of surface and root zone soil moistures (to a depth of 100 cm, referred to as root zone throughout this study) at a gridded resolution of 9 km. The L4\_SM product is available with a latency of 7 to 14 days (Chan et al., 2016). To create a daily (i.e., 24 hour) SMAP L4\_SM product during this study, the 3-hourly estimates were simply averaged across a given day. The NASA EarthData Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) tool was used to obtain the SMAP L4\_SM data used for this current study.

### **2.2.2 L3SMP\_E RETRIEVALS**

The SMAP Level 3 (SM\_L3) products provide global daily composites of the half orbit-based Level 2 instantaneous fields, with SMAP Level 2 products being geophysical retrievals of soil moisture on a fixed Earth grid. The widely used SMAP L3 soil moisture datasets are that of the 36-km radiometer product (SM\_L3) and the 9-km enhanced radiometer product (L3SMP\_E). This work made use of the 9-km enhanced L3 product, L3SMP\_E. The L3SMP\_E product is an enhanced Level 3 radiometer passive soil moisture product, which provides daily composite estimates of surface soil moisture (0-5 cm) retrieved from the SMAP a.m. (i.e., from the 6:00 a.m. descending half orbits) and p.m. (i.e., from the 6:00 p.m. ascending half orbits) surface brightness temperature data sets, respectively. The average data latency for the 9-km L3SMP\_E is 1-3 days for either a.m. or p.m. half orbit (ascending and descending) fields (Xu, 2020). As with the L4\_SM product, L3SMP\_E data used for this current study were accessed and acquired using the AppEEARS tool.



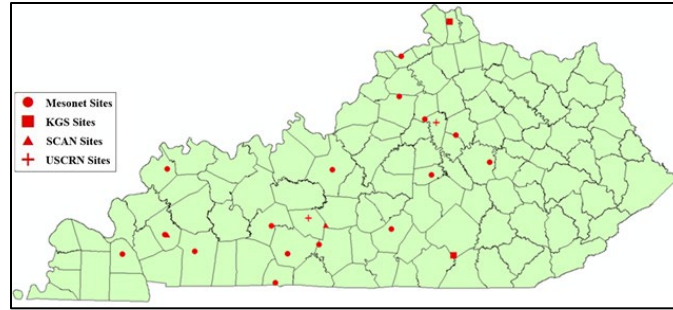
### **2.3 GROUND BASED SOIL MOISTURE SITES**

Through this study, the investigated methods for obtaining site-specific soil moisture were conducted over a number of in-situ sites across the Commonwealth of Kentucky. Doing so allowed for the direct comparison of site-specific in-situ data to that of the obtained site-specific satellite-based moisture estimates. These in-situ sites monitor soil moisture in the vertical soil column ranging from surface moisture (0-5 cm) to root zone depths (100 cm). The sites are maintained by various agencies, which are as follows: (1) The Kentucky Mesonet Network, (2) Stations implemented and maintained by the Kentucky Geological Survey (KGS) (Crawford and Bryson, 2018; Crawford et al., 2019), Soil Climate Analysis Network (SCAN) stations, and (4) U.S. Climate Reference Network (USCRN) stations. These in-situ networks were used to obtain Root Zone Soil Moisture (RZSM) and surface moisture data at varying locations across the commonwealth of Kentucky. The geographic locations (i.e., the latitudes/longitudes) of each of these stations can be seen in Table 2-1.

Table 2-1: Commonwealth of Kentucky In-Situ Ground Station Locations

Network	Station Name	Latitude	Longitude
Mesonet	Adair County	37.14	-85.29
	Barren County	37.01	-86.11
	Boyle County	37.62	-84.82
	Butler County	37.18	-86.65
	Caldwell County	37.10	-87.86
	Carroll County	38.69	-85.14
	Christian County	36.95	-87.52
	Fayette County	37.97	-84.53
	Franklin County	38.12	-84.88
	Hardin County	37.68	-85.95
	Madison County	37.72	-84.15
	Marshall County	36.92	-88.34
	Shelby County	38.33	-85.17
	Simpson County	36.67	-86.61
	Union County	37.69	-87.84
	Warren County	36.93	-86.47
KGS	Doe Run	38.99	-84.57
	Roberts Bend	36.89	-84.59
SCAN	Mammoth Cave	37.18	-86.03
	Princeton #1	37.10	-87.83
USCRN	Bowling Green	37.25	-86.23
	Versailles	38.09	-84.75

Additionally, Figure 2-1 shows the spatial distribution of the various in-situ monitoring stations spread across the Commonwealth of Kentucky, where the operating agency of the site is indicated symbolically through the associated legend.



*Figure 2-1: Location of In-Situ Data Sites Within Commonwealth of Kentucky*

## 2.4 EXISTING METHODS FOR SITE-SPECIFIC SOIL MOISTURE

As discussed, this study first investigated three site-specific methods observed to be prevalent within the geoscience and remote sensing communities. These methods are that of Random Forest statistical downscaling, Soil Evaporative Efficiency statistical downscaling, and Ensemble Kalman Filter (EnKF) data assimilation. Through this study, the downscaling methods were used to refine SMAP L4\_SM data from the native resolution of 9 km to a more site-specific scale of 1 km. The EnKF approach was used to estimate and remove biases between SMAP L3SMP\_E and in-situ surface moisture measurements. The bias-reduced L3SMP\_E estimates were then coupled with the SMAR infiltration model to model root zone moisture at the site(s) of interest. Through this study, the EnKF/SMAR approach made no improvements upon the resolution, but instead was made to be site-specific through the inclusion of site-specific soil texture data within the infiltration model.

### 2.4.1 STATISTICAL DOWNSCALING APPROACHES

There are typically two types of downscaling approaches that can be utilized: Dynamical and Statistical. Dynamical downscaling makes use of either Regional Climate Models (RCMs) or Numerical Weather Models (NWMs). These models (RCMs and NWMs) provide high-resolution precipitation and other climate variables by modeling the physical processes of the coupled land-atmosphere system. However, these models are usually quite taxing on the system and require large computational resources. In contrast, statistical downscaling methods aim to model statistical relationships observed between small and large-scale covariates. In statistical downscaling, auxiliary data (such as vegetation indices, land surface temperature, elevation, cloud cover properties, etc.) are utilized. A low-resolution image (e.g., 9 km L4\_SM) is enhanced to a finer resolution using another higher-resolution product via a regression approach (Sharifi et al., 2019).

Through this study, the high-resolution product used to downscale 9 km L4\_SM to 1 km resolution was that of the Moderate Resolution Imaging Spectroradiometer (MODIS) products. For this research, the daily MODIS Aqua (MYD) and Terra (MOD) products were retrieved: 1 km Level 3 Day/Night LST (MYD11A1 and MOD11A1, version 6), and 1 km Level 2 Albedo (MYDTBGA and MODTBGA, version 6). The following 500 m 8-day products were retrieved: Level 4 Evapotranspiration (MYD16A2 and MOD16A2, version 6) and Level 4 LAI (MYD15A2H and MOD15A2H, version 6). The following 500 m 16-day products were retrieved: Level 3 EVI (MYD13A1 and MOD13A1, version 6) and Level 3 NDVI (MYD13A1 and MOD13A1, version 6). The change in LST ( $\Delta LST$ ) was not retrieved via MODIS and was instead calculated manually as follows:

$$\Delta LST = DLST - NLST \quad (2-1)$$

where,  $\Delta LST$  = the change in LST,  $DLST$  = Day LST, and  $NLST$  = Night LST. For the 8-day and 16-day products, it was assumed that the value given was a constant value that could be applied to the entire interval. For example, given an NDVI value over a 16-day period, it was assumed that NDVI value would remain the same over the entire 16 days. This enabled the creation of daily datasets, rather than 8 and 16-day datasets. MODIS Aqua and Terra data can be retrieved via the AppEEARS tool.

#### **2.4.1.1 PREPROCESSING OF STATISTICAL DOWNSCALING DATA**

All SMAP Level 2-4 products employ the Equal-Area Scalable Earth2.0 (EASE2.0) Grid (also referred to as WGS 1984 Cylindrical Equal Area) projection (Brodzik et al., 2012). The L4\_SM product, as used throughout this study, is provided at a gridded spatial resolution of 9 km across these EASE2.0 grids (Reichle et al., 2016). Figure 2-2 shows a 9 km EASE2.0 grid, from which L4\_SM data was obtained, over the Doe Run KGS in-situ site. The need for downscaling of L4\_SM can be seen through this figure. The red marker (Doe Run in-situ site) is a site-specific measurement while the L4\_SM estimate is constant over the entire 9 km grid. To provide better comparisons between L4\_SM and in-situ data, L4\_SM must be downscaled to a resolution near site scale, such as 1 km or finer.

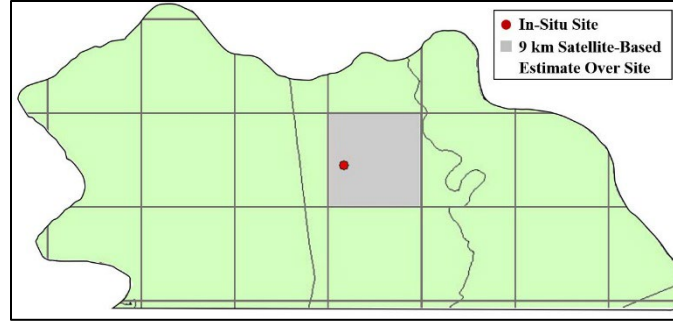


Figure 2-2: 9 km x 9 km EASE2.0 Grid Over Doe Run In-Situ KGS Station (red marker)

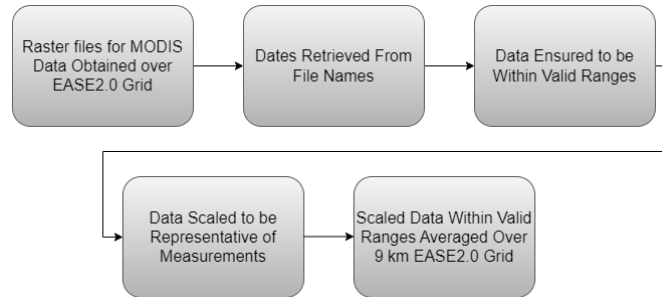
However, before downscaling of L4\_SM data can be conducted, data used in the downscaling process (e.g., MODIS data through this study) must first be scale matched to that of L4\_SM data (i.e., upsampled to 9 km). Through this study, scale matching was completed within a Python coding environment. The first step of the scale matching process was to obtain rasters of MODIS data (e.g., 1 km LST data) across a given SMAP 9 km EASE2.0 grid. The MODIS rasters were then ingested into the Python environment, where areal averaging was used to construct 9 km MODIS data. However, before areal averaging was conducted, a validity check upon the MODIS data was conducted. Each MODIS data set has a valid range of data whereas any data outside that range is excluded. Once validity was confirmed, the data was scaled to convert the data from satellite measurements (such as surface brightness) to that of applicable values. The valid ranges of data as well as the applied scale factors can be seen in Table 2-2.

Table 2-2: Valid Ranges and Scale Factors for MODIS Data

MODIS Product	Valid Range	Scale Factor	Units after Scale Factor Applied
Albedo	-100 to 5000	0.0001	Unitless
ET	-32767 to 32700	0.1	kg/m <sup>2</sup>
EVI	-2000 to 10000	0.0001	Unitless
LAI	0 to 100	0.1	m <sup>2</sup> /m <sup>2</sup>
LST	7500 to 65535	0.02	Kelvin
NDVI	-2000 to 10000	0.0001	Unitless

Scale matching is a crucial first step in both downscaling routines investigated through this study. For machine learning, the machine model is first trained with all obtained 9 km MODIS data over a given site as the input variables and L4\_SM over a site as the response variable. For Soil Evaporative Efficiency (SEE), there is coarse resolution (9 km) SEE as well as fine resolution (1

km in this study) SEE which both are then used in the subsequent SEE downscaling routine. SEE does not make use of all MODIS data, but instead only utilizes that of NDVI and LST retrievals. Due to this need of scale-matched data, extensive preprocessing of data is required prior to downscaling efforts being carried out. The generalized preprocessing workflow completed during this study to obtain scale matched MODIS data is shown in Figure 2-3.



*Figure 2-3: Generalized Workflow for Scale Matching MODIS data*

Additional preprocessing was required for individual MODIS datasets. Through this study, it was observed that both native resolution (500 m or 1 km) and upscaled resolution (9 km) MODIS daily LST (day and night) and 8-day evapotranspiration data had large amounts of missing time-series data. This was likely caused by excessively cloudy days and the fact that cloud cover is a common problem for visible and infrared sensing (i.e., LST sensing) (Wan, 1999). Researchers and the data scientist community are constantly working on addressing the problem of dealing with missing data within a dataset. The most accepted way to handle this problem is through missing data imputation, which is nothing more than the estimation of plausible values to substitute the missing ones (Jadhav et al., 2019). Through this study, data imputation was conducted using an 8 term Sum of Sine curve fitting program included within the MATLAB curve fitting toolbox. This imputation was seen to match the pattern of the LST and ET data well. Additionally, quantitative measurements of the data imputation (i.e.,  $R^2$  values for the imputed product) for both upscaled (9 km) and native (1 km) resolution MODIS data can be seen in Table 2-3.

Table 2-3:  $R^2$  Values from Data Imputation for MODIS Time-Series Data

Site	Upscaled (9 km) Resolution $R^2$			Native (1 km) Resolution $R^2$		
	ET	Day LST	Night LST	ET	Day LST	Night LST
Adair	0.9511	0.7896	0.7736	0.9239	0.8581	0.7817
Barren	0.9589	0.7586	0.7826	0.935	0.8199	0.769
Caldwell	0.9689	0.7839	0.7592	0.9369	0.7408	0.7895
Doe Run	0.96	0.842	0.7494	0.9219	0.8311	0.7504
Fayette	0.9468	0.821	0.7859	0	0.7582	0.7926
Madison	0.9537	0.7605	0.7781	0.9147	0.8134	0.7718
Mammoth	0.9489	0.7845	0.7515	0.8878	0.7836	0.772
Roberts Bend	0.9254	0.6909	0.732	0.914	0.6996	0.7587
Shelby	0.9549	0.7677	0.7656	0.9282	0.7835	0.8056
Versailles	0.9554	0.798	0.7512	0.9146	0.8243	0.7695

#### 2.4.1.2 RANDOM FOREST MACHINE LEARNING

Random forest regressive (RFR) machine learning was selected as one of two routines for downscaling L4\_SM data during this study. However, Random Forest is not the only available machine learning approach. This selection was based on work presented by Im et al (2016), which saw success in using RFR to downscale soil moisture measurements from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) satellite mission. The downscaling conducted by Im et al saw the resolution of AMSR-E estimates improved from 36 km to a much finer resolution of 1 km. RFR is an example of a tree-based regression ensemble which consists of a large number of decision trees (for example, 1000) based on random samples of training data (i.e., MODIS data in the case of this study). A random forest is random in two ways: (1) each tree is built upon a random subset of the training data, and (2) each branch within a given tree is created based on a random subset of candidate variables. The overall prediction of a random forest is the average of the predictions from each individual decision tree (Grömping, 2009). The following figure, Figure 2-4, shows a generalized workflow for the implementation of Random Forest downscaling.

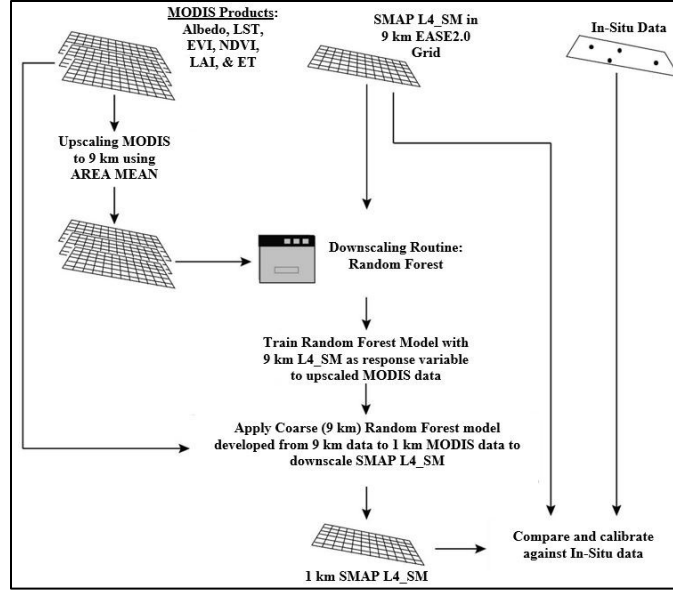


Figure 2-4: General Workflow for Downscaling L4\_SM Data Using RFR Model

Additionally, as Random Forest is a machine learning model, the selection of hyperparameters is an important step in the model construction. In the scope of machine learning, hyperparameters are parameters that are used to control the learning process of the model. Additionally, every machine learning system has hyperparameters, and the most basic task in machine learning is to automatically set these hyperparameters to optimize performance. Hyperparameter optimization has several important uses in that it can: (1) Reduce the effort required to conduct machine learning, (2) Improve the performance of the machine learning algorithms, and (3) Improve the reproducibility and fairness of scientific studies (Hutter et al., 2019). For this study, the technique chosen for hyperparameter optimization was that of RandomizedSearchCV. This optimization technique was chosen based on previous research that saw success in using RandomizedSearchCV to optimize the hyperparameters in an XGBoosting routine (Priscilla et al., 2020; Tiwari et al., 2021). As this study investigated downscaling over individual sites, each downscaling approach was constructed over an individual site of interest. This implies that hyperparameters could vary across sites due to the nature of using RandomizedSearchCV individually at a site. The optimized hyperparameters through this study were as follows: (1) **Max depth** of the tree, (2) Mean Squared Error (**MSE**) or Mean Absolute Error (**MAE**), (3) **Max features** considered when deciding upon the best split in a tree, (4) **Min Samples Split** (minimum number of samples required to split a node), (5) Bootstrapping **True** or **False**, (6) **Min Samples Leaf** (the minimum number of samples required to be at a leaf node), and (7) **Number of Estimators** (number of trees in the random forest). The average of each optimized



hyperparameter (across the analyzed sites), in addition to the possible ranges RandomizedSearchCV investigated, are shown in Table 2-4.

*Table 2-4: Average Optimized Hyperparameter Values/Criterion used for RFR Training over In-Situ Sites*

<b>Hyperparameter</b>	<b>Value/Criterion</b>	<b>Range/Alternative</b>
Max Depth	7	1 to 10
MSE or MAE	MSE	MSE or MAE
Max Features	5	1 to 8
Min Samples Split	151	2 to 200
Bootstrapping	False	True or False
Min Samples Leaf	107	100 to 400
Number of Estimators	333	100 to 500

#### **2.4.1.3 SOIL EVAPORATIVE EFFICIENCY (SEE)**

Soil Evaporative Efficiency (SEE) was chosen as the second investigated 9 km L4\_SM statistical downscaling routine during this study. Soil evaporative efficiency is defined as the ratio of the actual to potential soil evaporation and the downscaling efforts discussed herein make use of the assumed relationship between soil evaporation and soil moisture (Merlin et al., 2008).

Additionally, this work follows work conducted by Colliander et al. (2017) which made efforts to downscale SMAP Level 2 Radiometer surface soil moisture (0 – 5 cm) estimates from 36 km to that of 1 km during the SMAP Validation Experiment 2015 (SMAPVEX15). Although the downscaled datasets differ between this work and that of Colliander et al, the utilization of the SEE relationships for use in downscaling of L4\_SM data remains consistent. SEE downscaling did not require near the amount auxiliary (i.e., MODIS) data as that of the RFR statistical downscaling method discussed previously. For SEE downscaling, the required auxiliary MODIS data was that of Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI). SEE did, however, require scale matching of these auxiliary products to that of 9 km L4\_SM in the same fashion as that of RFR preprocessing. Merlin et al and Colliander et al provide in-depth discussions of the associated relationships between 9 km and 1 km LST/NDVI and soil moisture measurements assumed through SEE downscaling.

#### **2.4.2 SITE-SPECIFIC ASSIMILATION APPROACH**

The third existing approach investigated through this study to obtain site-specific soil moisture estimates was that of an Ensemble Kalman Filter (EnKF) assimilation approach. This approach has frequently been observed to be used in published research to estimate and remove biases

between satellite and surface-based measurements. Although surface soil moisture data (such as L3SMP\_E) from different sources (satellite retrievals, in-situ measurements, and land surface model integrations of meteorological forcing data) have been shown to contain consistent and useful information in their seasonal cycles and anomaly signals, they typically exhibit vastly different mean values and variabilities. These biases pose a severe obstacle to exploiting the useful information contained in satellite retrievals through data assimilation (Reichle, 2004). In Metrology, these biases, also called “systematic errors”, are simply measurement errors. Additionally, these biases include both additive and multiplicative errors. Additive errors, such as offset errors, do not change with the measured values. This is to say that additive errors remain constant. In contrast to this, multiplicative errors, such as gain errors, change linearly with the measured values, depending on the input values. Biases (i.e., systematic errors) can be classified as: (1) Instrumental errors (e.g., calibration errors in the system or model), (2) Assumption errors (e.g., errors in assumptions made whilst constructing the model), (3) Environmental errors (e.g., RFI interference), (4) Dynamic Errors (e.g., precipitation events, vegetation effects, etc.), and (5) Static errors (e.g., errors in soil texture information) (Lee et al., 2017). The generalized workflow for EnKF assimilation through this study can be seen in Figure 2-5.

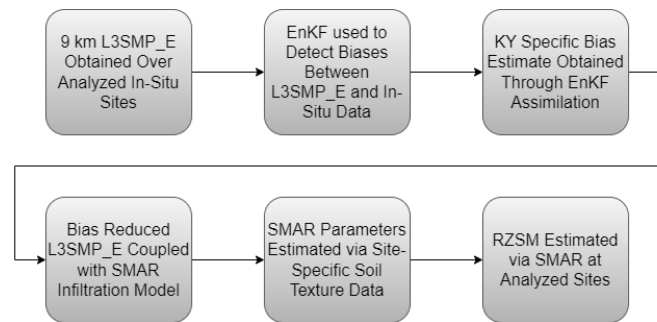


Figure 2-5: Generalized Workflow for EnKF Assimilation

#### 2.4.2.1 BIAS REDUCTION THROUGH ENKF ASSIMILATION

Baldwin et al (2017) made use of an Ensemble Kalman Filter (EnKF), a Monte-Carlo variant of the Kalman Filter (Evensen, 2003), to estimate bias between AMSR-E surface moisture estimates and in-situ measurements over various ecoregions throughout the contiguous United States. The work discussed herein uses the workflow established by Baldwin et al to estimate bias between SMAP L3SMP\_E surface moisture (0-5 cm) estimates and in-situ surface moisture measurements at 20 sites (shown in Table 2-1) within the Commonwealth of Kentucky. This assimilation approach effectively produced a Kentucky specific bias parameter. This parameter could then be

applied to L3SMP\_E retrievals, to yield “bias-reduced” L3SMP\_E estimates across Kentucky to move the satellite estimates nearer to that of in-situ readings. These bias-reduced estimates were then coupled with the SMAR infiltration model, discussed in the following section, to estimate root zone moisture at sites of interest. Therefore, this approach makes no efforts to refine the resolution of satellite-based data, as did the discussed downscaling approaches. Instead, this method compares satellite-data to in-situ data and then allows for estimation of root zone moisture through inclusion of site-specific texture parameters.

#### **2.4.2.2 ESTIMATING SOIL MOISTURE WITH THE SMAR INFILTRATION MODEL**

As discussed previously, the bias-reduced L3SMP\_E estimates were coupled with the Soil Moisture Analytical Relationship (SMAR) infiltration model (Manfreda et al., 2014). The SMAR model is based on existing hydrological models which are divided into several layers. However, Manfreda et al. assumed that the soil could be divided into two layers: The surface layer (0-5 cm), and the root zone layer (60 – 150 cm). Between these two layers, the most relevant water mass exchange is represented primarily through that of infiltration. The challenge faced by Manfreda et al. was to define a soil water balance equation which modelled this infiltration as function of the soil moisture in the surface layer, rather than that of a function of precipitation. This would then allow the derivation of a function of the soil moisture in the root zone layer as a function of the soil moisture estimated/measured within the surface layer. The movement of soil moisture from the surface layer to that of the root zone layer would be considered to be significant only when the soil moisture within the surface layer exceeds that of the field capacity of the surface layer.

Through this study, the bias-reduced L3SMP\_E estimates were utilized as the forcing parameters for the SMAR infiltration model. The coupling of these surface estimates with the SMAR model allowed for estimation of root zone soil moisture specific to sites of interest due to inclusion of site-specific soil texture parameters. Through this study, these soil parameters were as follows: %Sand, %Silt, and %Clay at both the surface (0-5 cm) and root zone (0-100 cm) levels. As discussed, these parameters were obtained remotely via Web Soil Survey (WSS). WSS is a web application operated by the USDA National Resources Conservation Service (NRCS) which allows for easy extraction of soil survey information at user defined sites. These parameters were then used with regression equations, established by Baldwin et al, to estimate parameters required by the SMAR infiltration model. However, utilization of the Baldwin et al regression equations yielded SMAR parameters which were observed to overestimate root zone moisture at analyzed Kentucky in-situ sites. Due to this overestimation, SMAR estimates were optimized through a simple additive offset with range [-1,1]. Optimization was conducted using Excel GRG nonlinear

solver to reduce the squared difference between optimized SMAR and in-situ measurements. Once these offsets were obtained at all investigated sites, a fitting parameter was developed for use at potential future sites not investigated herein. This parameter, shown in Equation 2-2, was derived through multivariate analyses which related available texture data to that of optimized fitting parameters seen to yield a better alignment between SMAR and root zone measurements.

$$\begin{aligned} \delta_{KY} = & 130.81 - 174.78(Sand_{RZ}) - 174.47(Silt_{RZ}) - 174.80(Clay_{RZ}) \\ & + 43.58(Sand_{Surf}) + 43.53(Silt_{Surf}) + 43.70(Clay_{Surf}) \end{aligned} \quad (2-2)$$

where  $\delta_{KY}$  = SMAR modification parameter for use at Kentucky sites;  $Sand_{RZ}$  = % Sand at root zone (100 cm);  $Silt_{RZ}$  = % Silt at root zone (100 cm);  $Clay_{RZ}$  = % Clay at root zone (100 cm);  $Sand_{surf}$  = % Sand at surface (5 cm);  $Silt_{surf}$  = % Silt at surface (5 cm);  $Clay_{surf}$  = % Clay at surface (5 cm).

## 2.5 SITE-SPECIFIC SOIL MOISTURE ESTIMATES THROUGH REGRESSION

In addition to investigating existing methods to obtain site-specific soil moisture estimates, this study proposes a new approach for site-specific analysis. The proposed approach developed a multivariate linear regression analysis which characterizes relationships between site-specific soil texture data and L4\_SM correction factors. This approach again made use of the following texture data: %Sand, %Silt, and %Clay at both the surface (0-5 cm) and root zone (0-100 cm) levels. As the intent of this study is to provide a means to obtain site-specific estimates remotely, these texture data were obtained remotely via Web Soil Survey (WSS). Through this approach, the L4\_SM correction factors are intended to be applied to 9 km L4\_SM data for a given site of interest. This effectively yields a 9 km product, that is now specific to the analyzed site. This approach was conducted over various in-situ sites across the Commonwealth of Kentucky. These sites effectively served as control sites where various L4\_SM correction factors and associated soil texture data were obtained. These data were then incorporated into a multivariate regression analysis. As with the previously discussed EnKF coupled SMAR assimilation routine, this proposed approach makes no efforts to improve the resolution of the L4\_SM data. This approach instead aims to improve the accuracy of the data as compared to site-specific locations through a computationally straightforward approach. Once constructed, this approach allows for L4\_SM data to be obtained over a site of interest and then corrected to that site based on easily obtained/estimated site soil texture parameters.

### 2.5.1 SITE-SPECIFIC L4\_SM CORRECTION FACTORS

This approach uses the application of simple multiplicative and additive factors to correct 9 km L4\_SM to a site of interest. Herein, these factors will simply be referred to as L4\_SM correction factors. These correction factors manipulate the 9 km L4\_SM data such that a better alignment with that of in-situ data is obtained. The application of these correction factors is as follows:

$$SM_{cal} = (L4\_SM \cdot MF) + AF \quad (2-3)$$

where  $SM_{cal}$  = site specific SMAP L4\_SM data;  $L4\_SM$  = 9 km L4\_SM data;  $MF$  = multiplicative factor (ranging from 0.0 – 1.0); and  $AF$  = additive factor (ranging from -1.0 – 1.0).

To ensure the highest feasible accuracy whilst conducting site specific L4\_SM corrections, least squares optimization was conducted using the Microsoft Excel GRG nonlinear solver. The ranges of the discussed offset factors in Equation 2-3 were used with said nonlinear solver to minimize the squared differences between the corresponding in-situ site and corrected 9 km L4\_SM data. The intent behind the optimization of these correction factors, and the general use of said correction factors, is as follows: (1) Acquire 9 km L4\_SM data over multiple in-situ sites in user defined study area, (2) Perform site-specific L4\_SM corrections at each site using Equation 2-3 and Excel optimization, (3) Construct multivariate regression equations using Excel optimized  $MF$ ,  $AF$ , and site soil texture data, and (4) Use constructed equations to allow for these offset factors to be calculated at sites where no in-situ data is available, but soil texture information is (either through physical obtainment or estimation). This method is meant to not only effectively correct 9 km L4\_SM data to that of site specific in-situ moisture data, but also allow for these corrections to be applied to L4\_SM data where no in-situ moisture data is available. The optimized offset factors, found using Excel optimization, can be seen in Table 2-5. A discussion of the multivariate linear regression can be found in the following section.

Table 2-5: Optimized L4\_SM Offset Factors from 10 Analyzed In-Situ Sites

Station Name	$MF$	$AF$
Adair County	0.142	0.306
Barren County	0.519	0.214
Caldwell County	0.177	0.333
Fayette County	0.073	0.403
Madison County	0.284	0.212
Shelby County	0.378	0.430
Doe Run	0.387	0.271
Roberts Bend	0.634	0.065
Mammoth Cave	0.219	0.404
Versailles	0.147	0.365

## 2.5.2 MULTIVARIATE LINEAR REGRESSION USING SOIL TEXTURE DATA

The main objective for this proposed approach was to yield a means to readily obtain site-specific soil moisture estimates from the 9 km SMAP L4\_SM product. To do so, multivariate linear regression was conducted using the Excel optimized offset factors ( $MF$  and  $AF$ ) and soil texture data at each analyzed in-situ site. The soil texture used throughout the multivariate regression was that of % Sand, % Silt, and % Clay at both the surface (0-5 cm) and root zone (0-100 cm) levels within the soil column. Additionally, % Organic Matter was investigated throughout this regression work but was observed to be statistically insignificant for this analysis. The soil texture data used throughout this study was acquired via WSS. WSS data was used in conjunction with ArcMap and the ArcMap addon, SoilViewer, to analyze soil survey information and to acquire the required soil texture data at each of the analyzed sites in Table 2-5.

The multivariate linear regression discussed herein was developed in a Python coding environment. Within this environment, multivariate regression equations were constructed individually for both  $MF$  and  $AF$  offset factors. However, before these equations were constructed, it was necessary to determine which of the soil texture variables would be used for the regression. To determine this, backwards elimination was used. Backwards elimination investigates the p-values that are associated with the various independent variables (soil texture data) and dependent variables ( $MF$  and  $AF$ ). For this study, a p-value of 0.05 was chosen. This selection was based on the knowledge that a p-value of 0.05 has, for decades, been conventionally accepted as the threshold to discriminate significant from non-significant results (Di Leo and

Sardanelli, 2020). Through backwards elimination, it was seen that %Sand, %Silt, and %Clay at both levels (surface and root zone) had p-values less than 0.05 while %Organic Matter at both levels had p-values greater than 0.05. Therefore, based on backwards elimination with a p-value of 0.05, %Sand, %Silt, and %Clay at surface and root zone were chosen as the independent variables for the multivariate regression analysis. The values for these independent variables can be seen in Table 2-6.

Table 2-6: Independent Variables used in Multivariate Regression Analysis for MF and AF Factors

Station Name	$Sand_{RZ}$	$Silt_{RZ}$	$Clay_{RZ}$	$Sand_{Surf}$	$Silt_{Surf}$	$Clay_{Surf}$
Adair County	0.677	0.197	0.125	0.457	0.373	0.17
Barren County	0.149	0.391	0.46	0.22	0.59	0.19
Caldwell County	0.08	0.539	0.381	0.045	0.725	0.23
Fayette County	0.039	0.543	0.418	0.05	0.62	0.33
Madison County	0.083	0.665	0.252	0.1	0.7	0.2
Shelby County	0.122	0.45	0.428	0.267	0.538	0.195
Doe Run	0.308	0.289	0.402	0.688	0.162	0.15
Roberts Bend	0.213	0.54	0.248	0.291	0.534	0.175
Mammoth Cave	0.075	0.653	0.272	0.08	0.71	0.21
Versailles	0.062	0.671	0.267	0.066	0.763	0.171

Development of the multivariate regression equations for both  $MF$  and  $AF$  was conducted using a 70-30 train-test split fashion. In machine learning, this implies that 70% of the data being used to build the model is set aside to train the model, while 30% of the data is kept from the model during training and is used to gauge model performance after model training is completed. With 10 sites being used during this regression analysis, seven sites became training sites while three sites were withheld to be used as testing sites (i.e., to compare model predicted vs expected values). The multivariate regression equations for  $MF$  and  $AF$  can be seen in Equations 2-4 and 2-5. Additionally, the  $R^2$  values for the constructed  $MF$  and  $AF$  equations were seen to be 0.907 and 0.987, respectively.

$$\begin{aligned}
 MF = & -267.4 + 357.23(Sand_{RZ}) + 356.44(Silt_{RZ}) + 357.57(Clay_{RZ}) \\
 & - 89.23(Sand_{Surf}) - 89.37(Silt_{Surf}) - 88.8(Clay_{Surf})
 \end{aligned}
 \tag{2-4}$$

$$AF = 172.75 - 230.42(Sand_{RZ}) - 229.77(Silt_{RZ}) - 230.37(Clay_{RZ}) + 57.67(Sand_{Srf}) + 57.59(Silt_{Srf}) + 57.49(Clay_{Srf}) \quad (2-5)$$

where  $MF$  = Multiplicative factor;  $AF$  = Additive factor;  $Sand_{RZ}$  = %Sand at root zone (100 cm);  $Silt_{RZ}$  = %Silt at root zone (100 cm);  $Clay_{RZ}$  = %Clay at root zone (100 cm);  $Sand_{Srf}$  = %Sand at surface (5 cm);  $Silt_{Srf}$  = %Silt at surface (5 cm);  $Clay_{Srf}$  = %Clay at surface (5 cm).

The predicted values for  $MF$  and  $AF$  via the following multivariate linear regression equations can be seen in Table 2-7.

Table 2-7: Excel Optimized Offset Factors Compared to Multivariate Regression Predicted Offset Factors

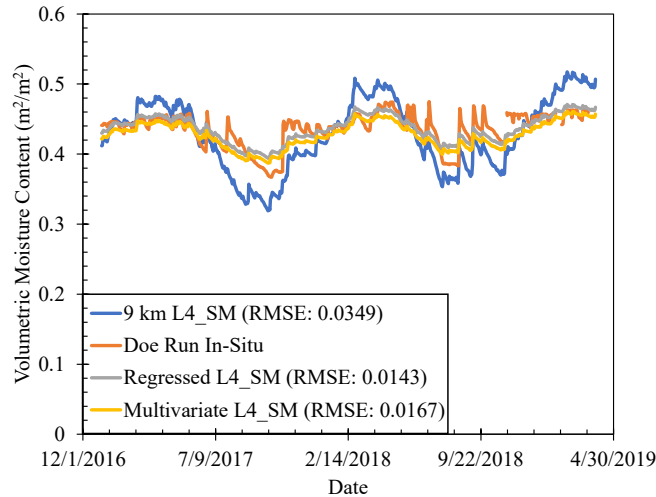
Station Name	Excel Optimized		Multivariate Regression	
	$MF$	$AF$	$MF$	$AF$
Adair County	0.142	0.306	0.152	0.308
Barren County	0.519	0.214	0.451	0.199
Caldwell County	0.177	0.333	0.306	0.274
Fayette County	0.073	0.403	0.372	0.268
Madison County	0.284	0.212	0.153	0.357
Shelby County	0.378	0.430	0.402	0.239
Doe Run	0.387	0.271	0.194	0.399
Roberts Bend	0.634	0.065	0.619	0.062
Mammoth Cave	0.219	0.404	0.172	0.348
Versailles	0.147	0.365	0.132	0.362

### 2.5.3 SITE-SPECIFIC CORRECTED L4\_SM VS IN-SITU DATA

The site-specific corrections of 9 km L4\_SM data to that of ground-based in-situ data were meant to provide a simplistic means to obtain L4\_SM data for use at a finer scale than that of 9 km. Two examples of these L4\_SM corrections can be seen in the following figures. Figure 2-6 shows 9 km L4\_SM data (shown in blue) corrections at the Doe Run in-situ site (shown in orange). The plot shown in grey is that of L4\_SM corrections conducted using the Excel optimization technique to determine  $MF$  and  $AF$  (i.e., directly correcting L4\_SM to that of in-situ data). The plot shown in yellow is that of L4\_SM corrections using the associated multivariate regression equations (Equations 2-4 & 2-5) to calculate  $MF$  and  $AF$  (i.e., regression using excel optimized correction factors and soil texture data from the site of interest). As seen in Figure 2-6, the Root Mean Squared Error (RMSE) of the 9 km L4\_SM, Excel regressed L4\_SM,



and Multivariate regressed L4\_SM are 0.0349, 0.0143, and 0.0167, respectively. The Excel regressed L4\_SM had the lowest RMSE when compared to that of in-situ data. This is to be expected due to the nature of directly optimizing this data to that of in-situ data. However, the multivariate regressed L4\_SM data only experienced a slightly higher RMSE, which is promising for the proposed approach of using multivariate regression to correct L4\_SM data over sites where in-situ data is not readily available.



*Figure 2-6: 9 km L4\_SM (blue) vs Doe Run In-Situ Data (orange), Excel Optimized Regressed L4\_SM (grey), and L4\_SM Regressed Using Correction Factors Predicted by Multivariate Regression (yellow)*

Site-specific corrections of L4\_SM data at the Caldwell County in-situ site can be seen in Figure 2-7. As with Figure 2-6, 9 km L4\_SM data is plotted in blue, in-situ data in orange, Excel optimized data in grey, and data obtained via multivariate regression in yellow. The same trend was observed in RMSE values as was seen at the Doe Run site. Both correction methods experienced significant decreases in the RMSE value, with the Excel optimized version experiencing the lowest RMSE. However, an interesting note to make about this site, and several other in-situ sites in this study, is the behavior of the recorded in-situ data and the subsequent effect on the correction of L4\_SM data. The in-situ data can be observed to experience abrupt and considerable changes in soil moisture over time. An example of this can be seen near the beginning of the time-series plot of the in-situ data where the soil moisture abruptly increases from  $0.35 \text{ m}^3/\text{m}^3$  to nearly  $0.5 \text{ m}^3/\text{m}^3$ . This trend in the in-situ data, which was observed to be prevalent in much of the recorded root zone moisture across Kentucky, made it difficult to accurately scale the L4\_SM data to match the in-situ data. This in turn caused the corrected L4\_SM data to become flattened due to Excel optimization attempting to best fit the corrected

L4\_SM data between the large spikes observed in the in-situ data. However, The site-specific corrected L4\_SM data does tend to capture the overall seasonal trends of the in-situ data (i.e., general wetting and drying trends) as well as experience a decrease in RMSE values as compared to that of native 9 km L4\_SM data.

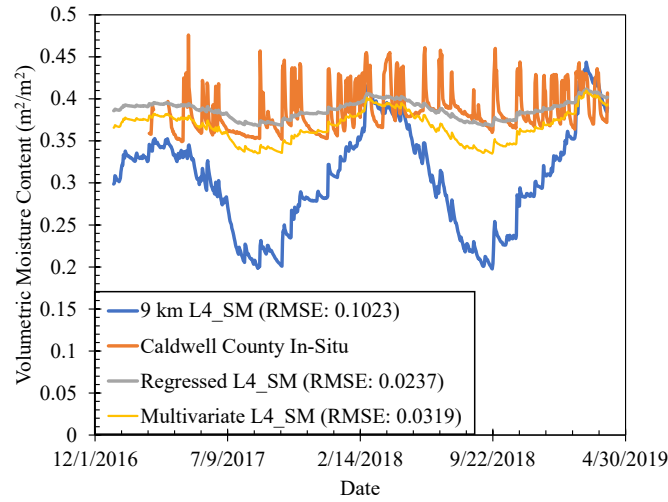


Figure 2-7: 9 km L4\_SM (blue) vs Caldwell County In-Situ Data (orange), Excel Optimized Regressed L4\_SM (grey), and L4\_SM Regressed Using Correction Factors Predicted by Multivariate Regression (yellow)

The proposed multivariate regression approach is intended for use in obtaining site-specific 9 km L4\_SM data. This approach does first require in-situ data throughout an area of study to be readily available. This in-situ data is necessary for conducting the initial direct corrections of L4\_SM data to that of ground-based data. These corrections are again carried out using the discussed Excel optimization routine, which employs simple correction factors (i.e.,  $MF$  and  $AF$  offset factors) to better align L4\_SM estimates with in-situ measurements. With optimized correction factors at various in-situ sites known, the overarching goal of this proposed approach moves to the estimation of site-specific L4\_SM data over sites with no in-situ data availability. Through the implementation of multivariate regression, based on the inclusion of optimized correction factors and respective soil texture information, this goal can be achieved. The multivariate regression allows for the creation of correction factor estimation equations (such as Equations 2-4 & 2-5). Implementation of these equations over sites of interest, where no in-situ moisture data is readily available, yields correction factors unique to those sites. These correction factors can then be applied to 9 km L4\_SM data, effectively yielding L4\_SM data specific to these sites. Site specific data corrections over remaining investigated sites can be seen in Figure A1 in the appendix of this paper. This proposed approach was observed to be both relatively

simple and computationally efficient as compared to existing methods used to obtain site-specific data.

## **2.6 COMPARATIVE ANALYSIS OF SITE-SPECIFIC METHODS**

The intent of this study was that of a twofold approach. The first intent was the investigation into the efficacy of existing approaches to obtain site-specific soil moisture estimates. These approaches, through this study, were that of Random Forest downscaling, SEE downscaling, and EnKF assimilation. The investigated downscaling approaches saw the resolution of 9 km L4\_SM improved to a finer resolution of 1 km. The EnKF assimilation approach made no improvements to the resolution of L4\_SM. EnKF assimilation instead improved accuracy of estimates through the inclusion of site-specific parameters within the coupled SMAR infiltration model. The efficacy of these methods was gauged both qualitatively and quantitatively. The efforts and data preprocessing required to conduct a given approach were considered through qualitative analysis. The accuracy of the site-specific estimates as compared to in-situ data were examined through quantitative analysis. The second intent of this study was to propose a less intensive approach to obtain site-specific soil moisture estimates. The proposed approach was that of a multivariate regression analysis which characterized relationships between L4\_SM correction factors and site-specific soil texture data.

The efficacy of the herein proposed multivariate approach was gauged through direct comparisons to results obtained using the aforementioned approaches. An example of this comparison can be seen in Figure 2-9. However, the herein proposed approach, as with the EnKF assimilation approach, made no improvements upon the resolution of 9 km L4\_SM data. This approach improves accuracy of L4\_SM data as compared to in-situ data through the inclusion of site-specific correction factors and soil texture parameters. Due to this lack of improvement within the resolution of 9 km L4\_SM, linear regression was coupled with SEE downscaling. Effectively, SEE downscaling (9 km to 1 km) was conducted and directly corrected to the in-situ site using the linear regressive approach shown in Equation 2-3. SEE was chosen as the downscaling approach to be coupled with linear regressive direct corrections due to the observed performance of SEE downscaling. SEE downscaled data did not experience a considerable increase in RMSE as compared to that of 9 km L4\_SM data. An example of SEE downscaling and SEE downscaled directly corrected (SEE/LR) estimates can be seen in Figure 2-8.

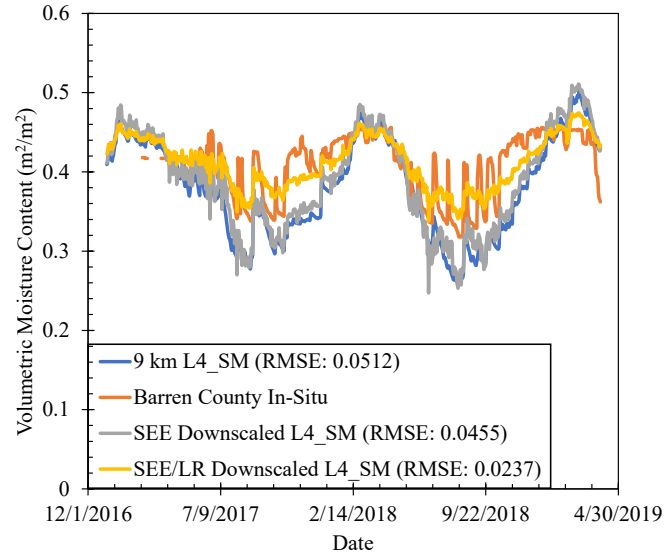


Figure 2-8: 9 km L4\_SM (blue) vs Barren County In-Situ Data (orange), SEE Downscaled L4\_SM Data (grey), and SEE Downscaled Linear Regressed L4\_SM Data (yellow)

As seen, SEE downscaled data experiences little improvement as compared to in-situ data than that of 9 km L4\_SM data. This lack of improvement in downscaling of L4\_SM data led to the coupling of SEE downscaling with that of direct corrections to that of in-situ data. Therefore, the SEE/LR downscaled data, as shown in Figure 2-8, becomes a 1 km product that is corrected to that of an investigated site. However, as seen in Figure 2-9C, site-specific data obtained through multivariate regression performs comparably with that of linearly regressed SEE downscaled data. This implies that, while the proposed multivariate approach makes no improvement to resolution, the improvement to accuracy of site-specific estimates outweighs the need for downscaling with respect to this proposed approach.

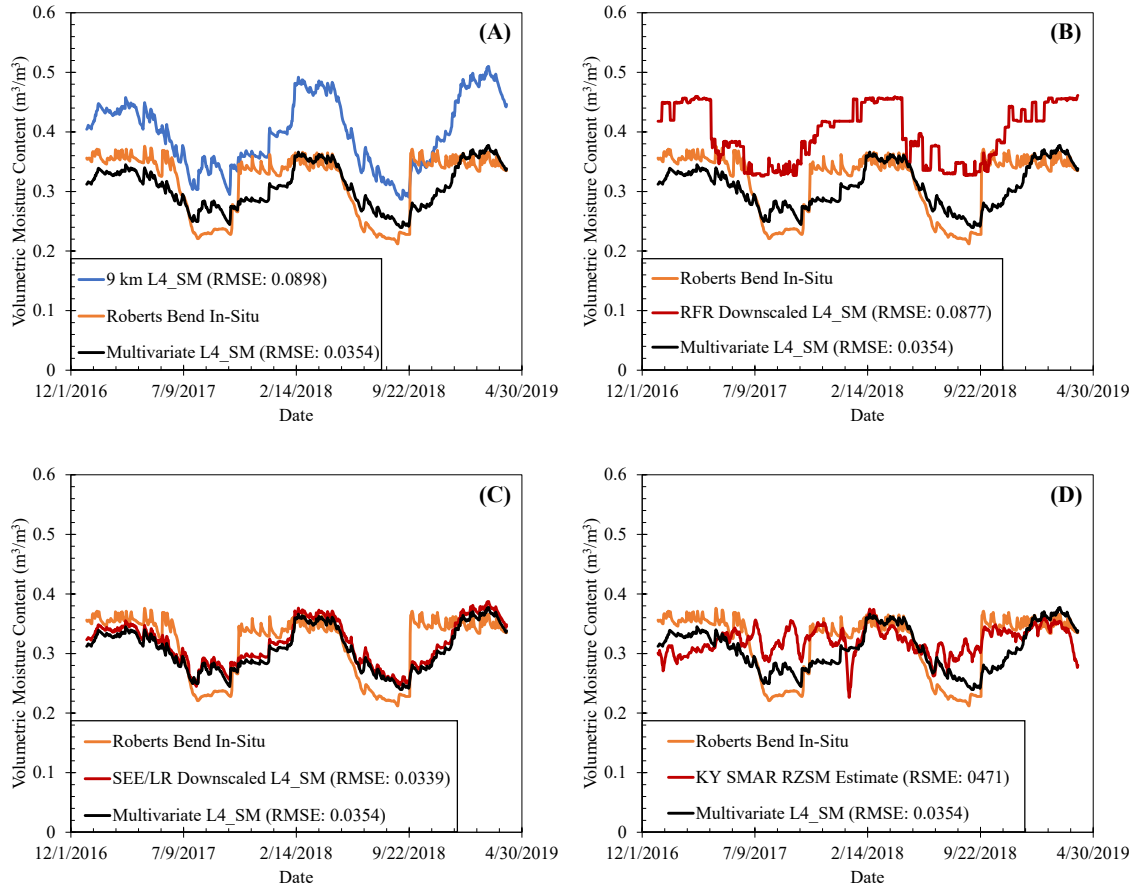


Figure 2-9: Comparison of Roberts Bend In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

The herein investigated approaches, existing and proposed, were observed to experience significant decreases in the RMSE values as compared against 9 km L4\_SM estimates. However, it was not the intent of this study to purely analyze various approaches and return associated RMSE values. The intent of this study was instead to present a less intensive approach to obtain site-specific soil moisture that consistently performs comparably to, or better than, that of existing approaches. The existing approaches, with respect to those discussed within this study, are rather labor/knowledge intensive. The downscaling approaches (RFR and SEE) require intensive preprocessing of auxiliary data (e.g., upscaling and imputation of MODIS data). The EnKF approach requires an intensive understanding of the Ensemble Kalman assimilation approach as well as data preprocessing. The herein proposed multivariate approach requires only in-situ data availability across an area of interest (e.g., statewide) and access to remotely accessible datasets such as SMAP L4\_SM and Web Soil Survey (or another means of soil texture obtainment). Through this study, the proposed multivariate approach was observed to be a far less intensive

approach that yielded accurate estimates. This can be seen through Figure 2-9, which shows the proposed approach to consistently perform well as compared to in-situ data. Comparisons of this proposed approach were also observed to perform well at remaining investigated sites. These comparisons can be seen in Figures A2-10 in the appendix of this paper.

The RMSE values for the respective site-specific soil moisture methods are shown in Figure 2-10. These RMSE values consider the error of each method as compared to that of in-situ data at each analyzed site. Figure 2-10 is formatted as follows: RMSE values for a given site-specific method are presented along the Y-axis while the RMSE values for native L4\_SM data over the analyzed site are presented along the X-axis. Figure 2-10 makes use of a line of unity, which is represented as a dashed line from the origin of the plot to the upper extent of the plot. The line of unity is useful for quickly analyzing improvements/decreases in the RMSE values. If an approach, for example RFR downscaling, plots along the line of unity, then the RMSE value for the site-specific estimate is very close to the RMSE value of the 9 km L4\_SM data. If a plot occurs above the line, then that indicates that the RMSE is higher than the RMSE of the 9 km L4\_SM data. Therefore, it is expected that site-specific data that experienced improved RMSE values would plot *below* the line of unity.

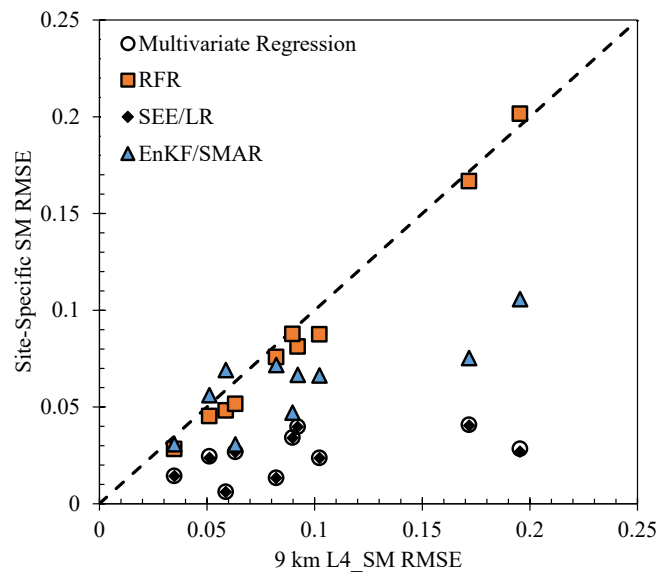


Figure 2-10: 9 km L4\_SM RMSE vs Obtained Site-Specific Moisture RMSE

From Figure 2-10, it can be observed that the investigated approaches for obtaining site-specific data generally experienced decreases in RMSE values. It is worth noting that the Random Forest approach plotted close to that of the line of unity. As discussed, this indicates that the RMSE value of the RFR downscaled product was very similar to that of the native 9 km L4\_SM product.

Aside from this, the remaining approaches generally plotted below the line of unity, indicating improvements to the RMSE values as compared to the RMSE of the 9 km product. Of these approaches, those that made use of regression (SEE coupled LR and proposed multivariate regression) were observed to experience the largest improvements in RMSE as compared to L4\_SM data. This can be observed through Figure 2-10 whereas these approaches plotted well below the line of unity as compared to other approaches. However, it is expected that these approaches perform well in this analysis due to the nature of conducting the approach (i.e., inclusion of direct corrections to in-situ data). With that being said, the approach which made use of no direct correction techniques and still saw favorable results was that of EnKF/SMAR. This approach made use of bias reduction over 20 in-situ sites throughout the Commonwealth of Kentucky. However, the routine did not specifically focus on any individual site, as the SEE/LR and multivariate regression routines did. Therefore, EnKF/SMAR was observed to be non-site specific, but still be able to be utilized at a site of interest through the inclusion of soil texture and site data from the site within the SMAR model. The EnKF/SMAR approach does perform well, but still depends on extensive knowledge of Ensemble Kalman filtering as well as intensive data preprocessing and obtainment. As compared to the investigated existing downscaling and assimilation approaches, the proposed multivariate regression is that of a non-intensive approach capable of yielding exceptional estimates of site-specific soil moisture.

## **2.7 CONCLUSIONS**

The initial intent behind this study was the investigation into existing approaches that utilize coarse satellite-based soil moisture data to obtain site-specific moisture estimates. The existing approaches investigated through this study were as follows: (1) Random Forest Downscaling, (2) Soil Evaporative Efficiency Downscaling, and (3) Ensemble Kalman Filter (EnKF) assimilation coupled with the SMAR infiltration model. The investigated downscaling approaches were utilized to refine the resolution of L4\_SM data from a coarse resolution of 9 km to a finer resolution of 1 km. Through this study, the EnKF coupled SMAR approach made no improvements upon the resolution of 9 km L4\_SM data. The EnKF approach was instead used to estimate and remove biases between 9 km L3SMP\_E surface moisture estimates and that of in-situ surface measurements. These bias reduced estimates were then coupled with the SMAR infiltration model to estimate root zone moisture at the various sites of interest. Therefore, these three existing approaches provided site-specific estimates of root zone moisture through either downscaling or assimilation approaches. Additionally, this study proposes a less intensive approach for use in obtaining site-specific soil moisture estimates. The herein proposed approach

utilizes multivariate regression to characterize relationships between site-specific L4\_SM correction factors and soil texture data. The proposed approach makes no improvement upon the resolution of the L4\_SM data. The multivariate approach instead improves the accuracy of the 9 km L4\_SM data as compared to a specific site through the inclusion of remotely obtainable site-specific parameters.

The subsequent intent of this study, after initial investigation of the various existing and proposed approaches to obtain site-specific moisture, was to conduct a comparative analysis upon the investigated approaches. It is worth noting that a modification was made to the SEE downscaling routine during this study that factored into the comparative analysis. As the proposed multivariate approach makes no improvement upon the resolution of L4\_SM data, a coupling of SEE downscaling and linear regression (shown in Equation 2-3) from the multivariate regression approach was conducted. This allowed for L4\_SM data to be both downscaled and directly corrected to an analyzed site. This coupled approach (downscaled and corrected) was considered in place of pure SEE downscaling through this study. It was observed that all investigated approaches experienced slight to moderate improvements in the accuracy of the respective soil moisture estimates. This was observed through the reduction of RMSE values of the site-specific soil moisture estimates as compared to that of in-situ data. Random Forest downscaling saw the lowest improvement of RMSE values, with these values being only marginally better than that of 9 km L4\_SM estimates. EnKF/SMAR saw moderate improvements of RMSE as compared to L4\_SM data. However, EnKF/SMAR is easily the most foreboding of the approaches investigated through this study, requiring knowledge of EnKF assimilation and extensive preprocessing of data. The methods which saw the greatest improvement of RMSE values were those which made use of regression: SEE coupled linear regression and the proposed multivariate regression approach. The coupled SEE/linear regression approach does first require extensive preprocessing of auxiliary data for initial downscaling to be conducted. The overarching intent of this study was to investigate existing methods to obtain site-specific moisture and to then propose a less intensive approach observed to yield comparable results to these existing approaches. As seen through this study, the proposed multivariate regression approach is one that is far less intensive than the investigated existing methods. The multivariate approach was observed to yield site-specific moisture estimates comparable to in-situ measurements. Therefore, the proposed multivariate regression approach becomes a readily available means to obtain site-specific moisture estimates through inclusion of remotely obtainable parameters.



## CHAPTER 3: Rainfall-Induced Landslide Hazard Analyses using Spatiotemporal Retrievals of Soil Moisture and Geomorphologic Data

### 3.1 INTRODUCTION

Rainfall-induced landslides are geohazards that pose a severe threat to residential and civil infrastructure. Every year, rainfall-induced landslides are associated with significant economic loss and loss of human life. As rainfall extremes increase globally, due to increased global warming and climate variability, the likelihood of rainfall-induced landslides increases as well. (Petley 2012; Sidle and Bogaard 2016, Swain et al., 2020). The occurrence of these geohazards is dependent upon the interactions that occur between both static and dynamic factors. Static factors do not change over time and include geomorphologic parameters such as slope angle, soil properties, and lithologic factors. Landslide susceptibility analyses are performed using static factors. Landslide susceptibility analyses describes the spatial distribution for the probability of landslide occurrence in each area according to the geographical setting (Stanley and Kirschbaum 2017; Crawford et al. 2021; Dashbold et al. 2023). Dynamic factors, such as climatological parameters, experience spatiotemporal variability (i.e., vary both over time and location). Dynamic factors in the context of rainfall-induced landslides refer to changes in the hydrologic regime (i.e., soil moisture, soil suction) driven by changes in precipitation (rainfall, evapotranspiration). Rapid changes in the hydrologic regime during heavy rainfall events are often precursors for potential landslides (Jiang et al. 2017; Piciullo et al. 2018). Landslide hazard analyses are performed using dynamic factors. Landslide hazard analyses give the probability that a landslide will occur in a given area and at a given time (Segoni et al. 2018; Monsieurs et al. 2019; Dashbold et al. 2023). Although susceptibility analyses provide indications of “where” landslides will occur, these types of studies do not give any indication of “when” or under “what” conditions landslides will occur. Landslide hazard analyses provide the “where”, “when”, and “what” for landslide occurrences. This foreknowledge is key to developing early warning systems or general preparation measures that can reduce the societal and economic impact of geohazards. Knowledge from hazard analyses is also vital to providing a better understanding of the coupled hydro-mechanical behavior associated with the initiation of rainfall-induced landslides.

This paper presents a novel approach to landslide hazard analyses using spatiotemporal retrievals of requisite data. These analyses require spatial estimates of temporally varying (i.e., dynamic) hydrologic parameters such as soil moisture and suction. It is not feasible to obtain dynamic spatial measures of soil moisture and suction from ground-based measurements. Ground-based

measurements represents data at discrete locations. To develop a landslide hazard platform (i.e., a dynamic mapping system) for a county or a region using ground-based measurements alone would require tedious and costly implementation of measurement stations across the study area. Spatiotemporal estimates of soil moisture can be retrieved through the National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) earth observing satellite mission. Soil suction can then be estimated as a function of SMAP soil moisture data. However, SMAP estimates are available at a resolution of 9 km. While the resolution of 9 km is not overly coarse, a finer resolution product is desired for use in applications such as landslide hazard analyses (Francis and Bryson, 2022). Land surface modeling (LSM) workflows provide a means to obtain the desired fine resolution estimates of dynamic soil moisture and suction data. For this study, a workflow was developed which saw the implementation of the Land Information System (LIS) framework, available through NASA. LIS was used herein to obtain requisite fine resolution spatiotemporal estimates of soil moisture and suction across the given study area. LIS is a flexible land surface modeling framework that allows users to conduct land surface modeling and observation assimilations at various spatial resolutions (Kumar et al., 2006; Peters-Lidard et al., 2007). Through LIS, these spatial estimates of soil moisture were obtained through implementation of the NOAH 3.6 land surface model (LSM). For use in the subsequent analyses, these LIS-based retrievals of soil moisture across Kentucky were retrieved at a fine resolution of 0.01 degrees (~1 km).

Fundamental to performing a landslide hazard analyses is the underlying stability analyses for the hillslopes. This study used a planar limit equilibrium approach for stability analyses. A requisite variable of these stability analyses was that of friction angle, a shear strength parameter of a given soil. However, spatial measurements of friction angle were not available across the designated study area. To remedy this lack of data, an Artificial Neural Network (ANN) machine learning workflow was developed. This ANN workflow yielded spatial estimates of friction angle over landslide-prone areas previously observed to lack this data availability. Herein, the developed ANN workflow proved an effective means to obtain desired spatial geomorphologic data requisite for subsequent stability analyses.

This study made use of the aforementioned spatial static and dynamic landslide parameters to conduct slope stability analyses over known rainfall-induced landslides in Kentucky. Namely, the focus was upon a severe landslide event which occurred between 07/25/2022 and 07/30/2022 in Eastern Kentucky. Over this 5-day period, severe thunderstorms developed over the area which brought about heavy rain which yielded 600% of the normal rainfall totals typically seen at this

time of year over this area of Kentucky (NOAA, 2022). This high influx of rainfall led to severe flash flooding and over 500 small to moderate sized translational landslides and debris flows. In conjunction with the rainfall-induced landslide event, the aim was to investigate the hydro-mechanical role of soil moisture and suction as precursors for known rainfall-induced landslide hazards. The study presented herein shows that spatiotemporal estimates of soil moisture and suction can be effectively used within hazard analyses to detect incipient failure conditions in the occurrence of rainfall-induced landslides.

### 3.2 SPATIAL GEOMORPHOLOGIC DATA

Herein, slope stability analyses were conducted in order to investigate the role of soil moisture and suction as precursors indicative of incipient conditions for rainfall-induced landslide occurrences. To support these analyses, a variety of static geomorphologic parameters were acquired for investigation into spatial landslide detection. Spatial estimates of these required geomorphologic data used for these analyses were retrieved from Web Soil Survey (WSS) (<https://websoilsurvey.nrcs.usda.gov/app/>). WSS is an online platform operated by the USDA National Resources Conservation Service (NRCS), that provides digitized soil survey maps across the contiguous United States at very fine resolutions (100 – 300 m). WSS data was utilized in conjunction with the ArcMap addon, SoilViewer. SoilViewer allowed for the acquisition of required data estimates at a given desired depth. Additionally, slope angles at sites of interest were able to be retrieved through analysis of Digital Elevation Maps (DEMs) that are readily available through ArcMap. Table 3-1 shows the various soil and slope characteristic data that was obtained throughout this study as well as the source from which the data was obtained. The data acquired and generated in this study can be found at the UKnowledge Civil Engineering data repository, <https://doi.org/10.13023/wzxy-w420>

*Table 3-1: Soil and Slope Characteristic Data Sources*

Characteristic Data	Obtained Via
% Sand	WSS ( <a href="https://websoilsurvey.nrcs.usda.gov/app/">https://websoilsurvey.nrcs.usda.gov/app/</a> )
% Silt	
% Clay	
Bulk Density of Soil (g/cm <sup>3</sup> )	
Depth to Bedrock (m)	
Slope Angle (degree)	ArcMap DEM Esri. "Terrain: Slope in Degrees" [basemap]. No Scale

### 3.3 ARTIFICIAL NEURAL NETWORK PREDICTION OF FRICTION ANGLES

A key physical parameter of the associated slope stability analyses was friction angle ( $\phi'$ ). However, there were no readily available means of spatial acquisition for friction angle in a soil. To remedy this lack of data availability, an Artificial Neural Network (ANN) machine learning workflow was developed. The workflow included site specific data obtained from the WSS database and an empirical relationship for estimation of the friction angle from basic soil data. This workflow was implemented to obtain spatial estimates of friction angle at sites where slope stability analyses were conducted.

#### 3.3.1 USE OF EMPIRICAL RELATIONSHIP TO ESTIMATE FRICTION ANGLE

Machine learning workflows require a dataset to be trained upon. However, no spatial friction angle dataset was presently available. To address this requirement, a dataset of empirically estimated friction angles was first developed. The empirical approach implemented in this study to estimate friction angles at sites of interest was that given by Mesri et al. (1996), which relates the plasticity index of a clayey soil to the friction angle. The Mesri et al. (1996) equation is given as:

$$\phi' = -0.00001PI^3 + 0.0033PI^2 - 0.3554PI + 36.194 \quad (3-1)$$

where,  $\phi'$  = Friction Angle in degrees; and  $PI$  = Plasticity Index (%)

Using the PI values provided in the WSS database in conjunction with the Mesri et al. (1996) relationship provided spatial estimates of friction angles at sites of interest. However, the implicit assumption with the Mesri et al. (1996) relationship is that the soil is clayey. Therefore, a twofold investigation was conducted into both the PI values that were retrieved via WSS as well as PI values retrieved via borehole data throughout Kentucky (<https://kgs.uky.edu/kgsmmap/kytclinks.asp>). This investigation was conducted to confirm the soil types and characteristics indicated by WSS data aligned with data retrieved via the borehole data. The results of this investigation can be seen in Figure 3-1 which shows WSS data compared to borehole data retrieved near WSS study sites.

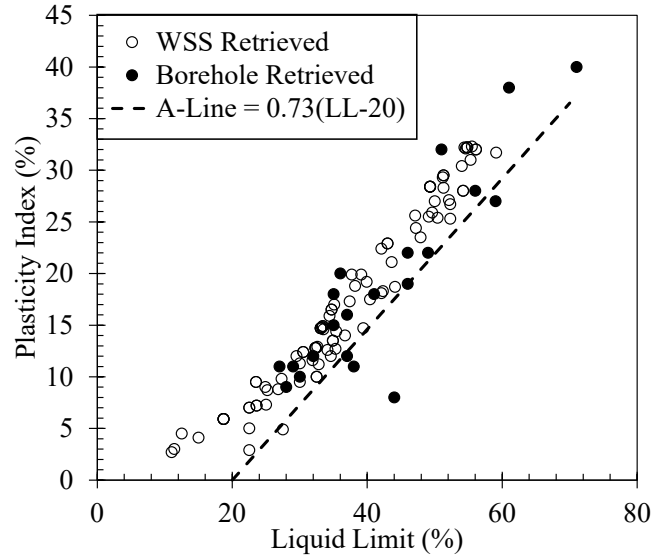


Figure 3-1: Plasticity Index vs Liquid Limit Retrieved via WSS and Kentucky Borehole Explorations

Figure 3-1 shows PI data retrieved via WSS and borehole explorations across Kentucky. As seen in Figure 3-1, nearly all retrieved values plot above the A-line. Soils that plot above the A-line are assumed to be clayey whereas soils that plot below the A-line are assumed to be silty.

Additionally, WSS retrievals are observed to align well with borehole obtained measurements of PI. Therefore, WSS data was assumed to be representative of the composition of the soil at investigated sites. Thus, the use of the described empirical approach to estimate friction angles in a clayey soil was assumed to be valid for this study. It must be noted, however, that the estimation of friction angles through this approach inherently biases the estimated values to that of the given empirical approach. This bias, and its inherent implications, is understood and noted herein.

### 3.3.2 ARTIFICIAL NEURAL NETWORK PREDICTION OF FRICTION ANGLES

This study not only investigates slope stability analyses, but also implements machine learning workflows. These workflows, which utilized an Artificial Neural Network (ANN) architecture, were implemented to estimate sparse parameters not readily available for spatial obtainment. Through this study an ANN was developed to predict friction angle values from soil texture information (% Sand, % Silt, and % Clay). The overarching ANN architecture was selected based upon 1: Previous research (Das et al., 2008; Nguyen et al., 2020) that has seen success through ANN prediction of various soil index properties and 2: The ability of this architecture to model highly nonlinear functions (e.g., relationships between soil texture) and to generalize when given new, unseen data (Gardner and Dorling, 1998). The intent behind development of the ANN workflow for this study was to specifically investigate potential relationships between soil texture

information and friction angles. Through ANN machine learning, this relationship was observed to be present. Based on this observation, the ANN workflow was used herein to estimate friction angles for use in subsequent stability analyses.

Artificial neural networks are machine learning models, within a branch of artificial intelligence. For this study, a Multilayer Perceptron (MLP) artificial neural network architecture was utilized (da Silva et al., 2018). The architecture of the MLP neural network used is shown in Figure 3-2.

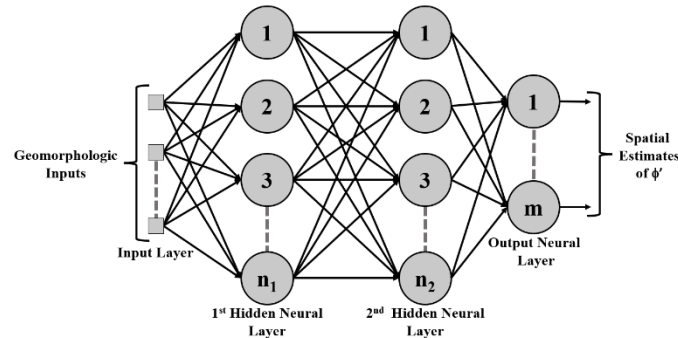


Figure 3-2: Multilayer Perceptron Neural Network Architecture, after (da Silva et al., 2018)

MLP neural networks typically consist of one hidden neural layer, but networks with two hidden layers can represent functions with any kind of shape (Panchal et al., 2011). Therefore, this study made use of a two-layer MLP where the number of neurons within each of the two hidden layers is referred to as the network architecture. This study saw success in implementation of MLP architecture of ANN for use in prediction of friction angles through soil texture information.

The number of neurons required for this model were optimized using the RandomizedSearchCV (RSCV) Method. The RSCV method tests various values for the parameters of a given model and returns those which yield the highest score of the model (i.e., the parameters which yield the best predictions compared to true values). RSCV requires a large amount of data to test the accuracy of the model as optimization occurs. This study used an 80-20 test/train split. In this split, 80% of the input data (i.e., soil texture and estimated friction angles at 80 sites) was used to train the model during the optimization process. To test the optimization, 20% of the data was retained to test the accuracy of the trained model per optimization. Through RSCV optimization, and the associated test/train split, it was found that roughly 1100 neurons per hidden layer yielded the best performance of the MLP neural network. Figure 3-3 shows the MLP neural network predicted friction angles at the 20 retained test sites with the optimized number of neurons.

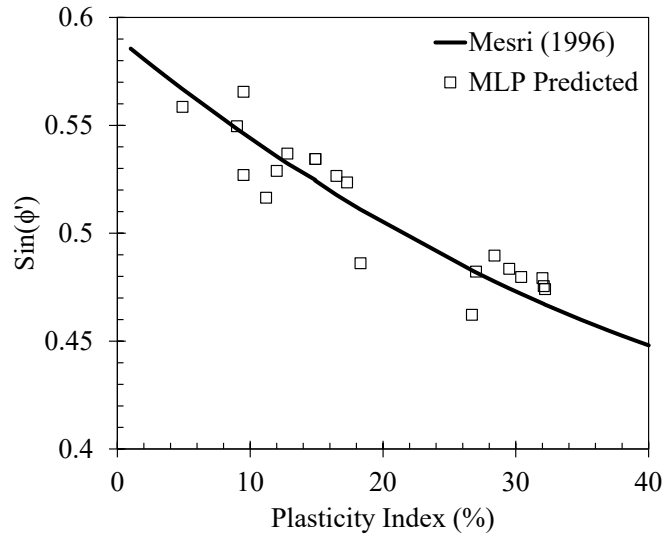


Figure 3-3: MLP Neural Network Predicted vs Mesri Friction Angles at 20 Retained Test Sites

As can be seen in Figure 3-3, the MLP Artificial Neural Network performed very well at predicting friction angles when compared to the Mesri et al. (1996) equation (Equation 3-1) at the respective test site (via 80/20 train/test split). Additionally, an RMSE value of 0.86 was achieved between the predicted and input friction angle values. This observed success led to the use of the MLP ANN model in prediction of friction angle based on soil texture information over landslide sites analyzed in the subsequent slope stability analyses.

### 3.4 SPATIAL SOIL MOISTURE RETRIEVALS

Shallow landslides, similar to those investigated by the study, are primarily caused by prolonged or heavy rainfall leading to an influx of moisture upon already saturated hillslopes (Hong et al., 2007). Herein, the study implicitly assumed soil moisture to be the controlling factor in landslide occurrence. Soil moisture information was retrieved over Kentucky using the Land Information System (LIS) framework. LIS is a flexible land surface modeling and data assimilation system that has been made available to the public. LIS has been structured such that the public can reuse modeling tools, data resources, and assimilation techniques to recreate or construct new models over individual areas of interest (Kumar et al., 2006; Peters-Lidard et al., 2007). These soil moisture retrievals were used in conjunction with previously discussed geomorphologic parameters in subsequent slope stability analyses.

### 3.4.1 OVERARCHING LIS FRAMEWORK

The LIS framework is composed of three subcomponents: (1) The Land Data Toolkit (LDT) executable, (2) The Land Information System (LIS) core executable, and (3) The Land surface Verification Toolkit (LVT) executable. These executables are designed for construction and implementation within a Linux-based operating system and are readily available to the public via the NASA-LIS GitHub repository (<https://github.com/NASA-LIS/LISF>). Figure 3-4 shows the generalized workflow for the implementation of the LIS framework utilized herein.

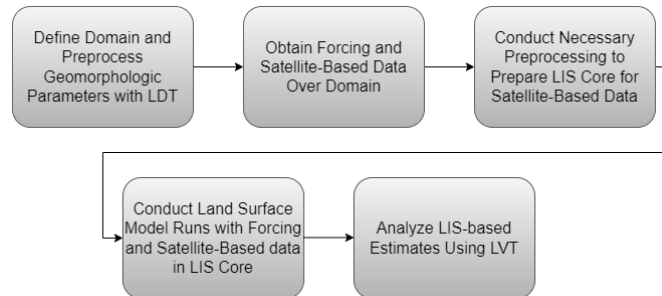


Figure 3-4: Generalized Flowchart for LIS Framework

#### 3.4.1.1 LAND DATA TOOLKIT (LDT) EXECUTABLE

The Land Data Toolkit serves as a preprocessor to the Land Surface Models (LSMs) implemented within the LIS executable. The first step of the LDT executable was the identification of the domain over which modeling would be conducted. Herein, a 0.01-degree (~1 km) resolution domain was constructed such that the entirety of Kentucky was considered. Figure 3-5 shows the domain over Kentucky.

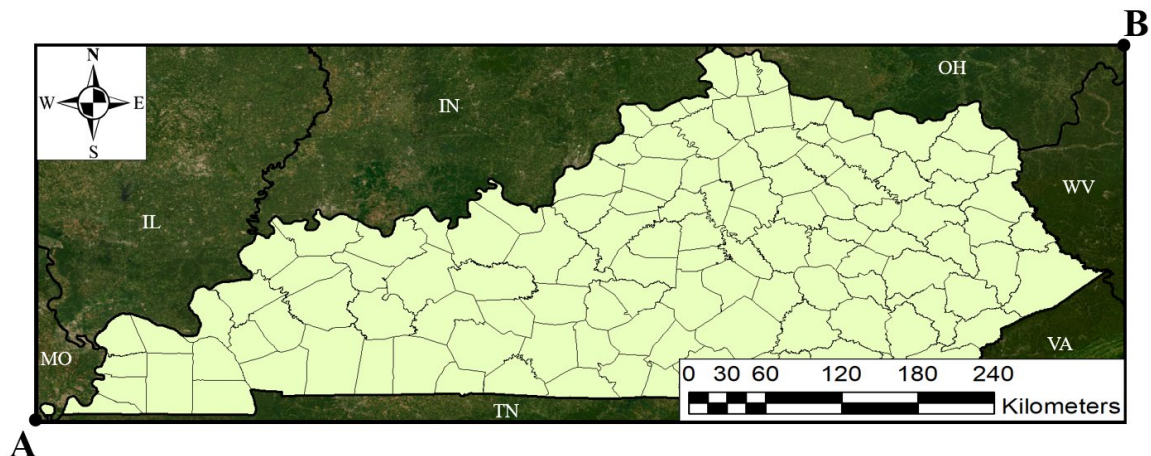


Figure 3-5: LIS Domain Over Kentucky

The domain shown in Figure 3-5 extends from Point A (36.45, -89.60) to Point B (39.10, -81.85). The primary role of the LDT is to provide physical parameters over the domain for use in the



LSM. The key physical parameters needed for LSMs include: (1) Land cover/vegetation, (2) Land/water masks, (3) soil texture information, and (4) topography (Arsenault et al., 2018). The LDT preprocesses these physical parameters for use in subsequent LIS LSM model runs. The sources and resolutions of these physical parameters are shown in Table 3-2.

*Table 3-2: LDT Preprocessed Physical Parameter Information*

<b>Geomorphic Parameter</b>	<b>Source</b>	<b>Available Resolution</b>
Land Cover/Vegetation	USGS	30 arcsecond (~0.01 degree)
Land/Water Masks	MODIS	
Soil Texture	STATSGO_FAO	
Topography	SRTM	

The land cover and vegetation classification was obtained via the USGS Earth Resources Observation and Science (EROS) Center (<https://www.usgs.gov/centers/eros>). Land or water mask information was obtained via the Moderate Resolution Imaging Spectroradiometer instrument aboard the Aqua and Terra Earth Observing Satellite (EOS) Earth missions. The soil texture map implemented within LIS is that of a global hybrid soil texture map produced by the National Center for Atmospheric Research (NCAR). The soil texture is a hybrid of maps derived from that of the NRCS State Soil Geographic (STATSGO) Data Base and a soil map of the world, produced by the Food and Agriculture Organization (FAO) of the United Nations Educational, Scientific, and Cultural Organization (UNESCO). The STASGO\_FAO texture map is constructed as a 16-category (i.e., 16 soil texture categories) two-horizon soil system. The first horizon extends from 0 – 30 cm, while the second horizon extends from 30 – 100 cm. The final physical parameter, topography, was obtained via the NASA Shuttle Radar Topography Mission (SRTM) (SRTMGL30 V021, 2023). All preprocessed geomorphologic parameters shown in Table 3-2 are available at a fine resolution of 30 arcseconds (~0.01 degree). Additionally, the geomorphologic data shown in Table 3-2 were only used within LIS for spatial soil moisture estimation. This is to say that these data were *not* used for the subsequent slope stability analyses. Instead, the data discussed in Sections 2 and 3 were implemented for these analyses.

#### **3.4.1.2 LAND INFORMATION SYSTEM (LIS) EXECUTABLE**

The Land Information System is the core executable of the LIS framework. The LIS executable conducts the user selected Land Surface Model (LSM) runs. These LSMs require static geomorphologic data (provided by LDT) *and* dynamic climatological forcing data. Therefore, the LIS executable also becomes the processor for the required climatological data. Forcing data via

the North American Land Data Assimilation System (NLDAS-2) was utilized herein. NLDAS-2 data is available over the contiguous United States at 1/8-degree spatial resolution. With forcing climatological and geomorphologic data available, the LIS core executable effectively becomes an LSM. The NOAH 3.6 LSM, discussed in section 4.3.1, was chosen for implementation through the LIS core. When conducting LSM runs, the LIS executable can conduct either Open-Loop (OL) or Data-Assimilation (DA) runs. In OL modeling, the LSM estimates are only a function of geomorphologic and forcing data. In DA modeling, LSM (NOAH 3.6 herein) estimated values are assimilated with observations (satellite-based moisture retrievals herein). The Ensemble Kalman Filter (EnKF) technique was used for assimilation. EnKF has been used successfully in the assimilation of soil moisture data by several researchers. (Kumar et al., 2008; Fertig et al., 2009; Baldwin et al., 2017).

#### **3.4.1.3 LAND SURFACE VERIFICATION TOOLKIT (LVT) EXECUTABLE**

The Land surface Verification Toolkit is the final executable of the LIS framework modeling process. The LVT conducts no preprocessing of input data, as does LDT, nor does it conduct modeling workflows as conducted within the LIS core executable. The LVT instead serves as a means to conduct verification of LIS estimated data, soil moisture in this case, as compared to either in-situ or satellite-based measurements. The LVT was designed to provide an integrated environment whereas systematic evaluations of LSM estimated data could be conducted. Additionally, the LVT provides a range of verification approaches and analysis capabilities for use in further validation of LSM estimates and/or assumptions (Kumar et al., 2012). The LVT was primarily utilized as a means for data retrieval for use in comparative analyses against in-situ data and use within subsequent stability analyses.

#### **3.4.2 LIS SATELLITE-BASED DATA ASSIMILATION PREPROCESSING**

The intent herein was the investigation into the implicit assumption that soil moisture is a controlling factor of landslide occurrence. This assumption requires spatial measurements of soil moisture at varying depths across the analyzed domain. These spatial estimates were retrieved via LIS through assimilation of satellite-based and LSM-based soil moisture estimates. Herein, NASA 9 km Soil Moisture Active Passive (SMAP) L3SMP\_E surface moisture retrievals were investigated for assimilation. The L3SMP\_E product is an enhanced Level 3 radiometer passive soil moisture product, providing daily composite estimates of surface soil moisture (0-5 cm). These composites are retrieved via the SMAP surface brightness temperature A.M. and P.M. descending and ascending, respectively, half orbit datasets with a data latency of 1-3 days (Xu, 2020).

However, before these retrievals could be assimilated within the LIS framework, changes to the LIS source code were required due to observed quality control checks within the LIS assimilation framework. These initial quality control checks allowed for only a percentage of the defined domain to be populated with L3SMP\_E retrievals. The L3SMP\_E retrievals not meeting these quality checks can be seen in Figure 3-6A, which shows a large portion of the domain over Kentucky to be lacking L3SMP\_E data. Due to this lack of data, assimilation of L3SMP\_E data would be unavailable in these areas. This required changes in the quality control flags within the LIS source code. The control flags that were found to be problematic were the vegetation water content (VWC) control flag and the overall retrieval quality flag (RQF). The VWC flag accounts for potential issues in transmission of signals from the soil through the vegetative cover. As VWC increases, so does the vegetative cover, increasing the difficulty to receive signals from the underlying soil. The RQF is a binary flag that indicates whether the retrieval does (i.e., 0) or does not (i.e., 1) have the recommended quality. The VWC (code lines 795 and 1067) and RQF (code lines 800 and 1072) control flags can be found within the NASA\_SMAPsm.F90 source code via the NASA-LIS GitHub repository. These flags were initially set to ignore L3SMP\_E retrievals that had a  $VWC > 5 \text{ kg/m}^2$ , which recommends retrieval at recommended quality, and an  $RQF = 0$ , which indicates the retrieval has recommended quality. Through this research, these flags were changed to allow retrievals with a  $VWC \leq 30 \text{ kg/m}^2$  and a RQF of either 0 or 1. The L3SMP\_E retrievals allowed by these adjusted quality checks can be seen in Figure 3-6B, which shows a vast majority of the domain to now be populated. However, the implication of this change was that the L3SMP\_E retrievals allowed by the adjusted quality checks may be of uncertain quality rather than the recommended quality. This implication was noted as acceptable to enable assimilation of L3SMP\_E retrievals to aid in spatial retrievals of soil moisture for subsequent slope stability analyses.

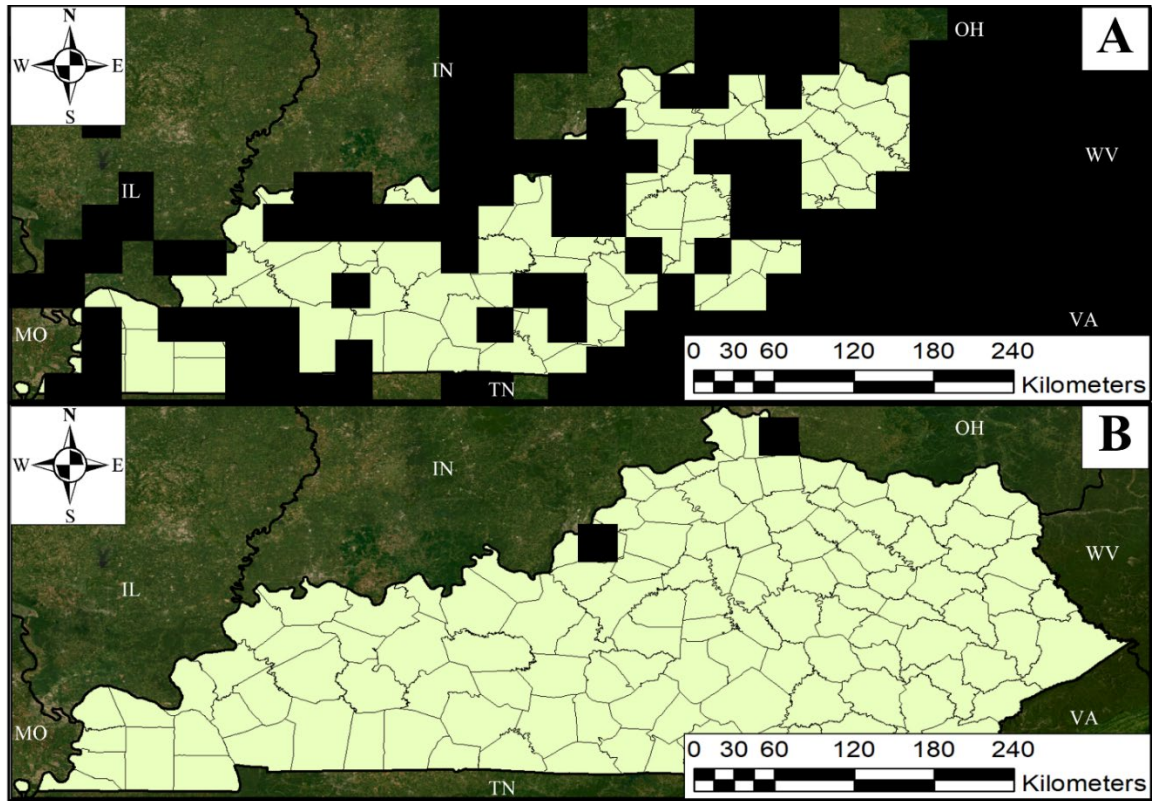


Figure 3-6: (A) Missing L3SMP\_E (shown as areas of black) Over Domain Due to LIS Satellite Data Quality Checks and (B) Missing L3SMP\_E (shown as areas of black) Over Domain After Adjustment to LIS Satellite Data Quality Checks

### 3.4.3 LIS NOAH 3.6 LAND SURFACE MODEL RUNS

The study retrieved spatial and temporal estimates of soil moisture across the defined domain through implementation of the NOAH 3.6 LSM. These estimates, through LIS assimilation, became functions of both climatological forcing data and satellite-based observations.

#### 3.4.3.1 NOAH 3.6 LSM

The NOAH 3.6 Land Surface Model (LSM) was used to obtain spatial soil moisture estimates. The NOAH 3.6 model was developed through efforts by the National Center for Atmospheric Research (NCAR). For the study herein, the model was developed through the LIS framework with default layers and at a fine resolution of 0.01-degree x 0.01-degree (~1 km x ~1 km). The NOAH 3.6 LSM has a four-layer configuration. By default, these four layers are at increasing depths of 100 mm, 300 mm, 600 mm, and 1000 mm. The NOAH 3.6 model provides estimates of soil moisture at these layers through computation of soil moisture in the unsaturated zones through implementation of Richard's equation. Additionally, the model consists of both a diurnally dependent Penman potential evaporation approach as well as a primitive canopy model

(Shrestha et al., 2020). The NOAH 3.6 model requires both static geomorphologic data (via LDT) as well as dynamic climatological data (via NLDAS-2).

### 3.4.3.2 GROUND BASED SOIL MOISTURE SITES

Validation of the LIS framework was conducted through direct comparisons of LIS estimated moisture to the soil moisture measured by existing in-situ stations throughout Kentucky. These in-situ soil monitoring stations are maintained by the following agencies: (1) Stations implemented and maintained by the Kentucky Geological Survey (KGS) (Crawford and Bryson, 2018; Crawford et al., 2019), (2) Soil Climate Analysis Network (SCAN) stations maintained by the United States Department of Agriculture (USDA), (3) U.S. Climate Reference Network (USCRN) stations maintained by the Atmospheric Turbulence and Diffusion Division (ATDD) of NOAA, and (4) the Kentucky Mesonet maintained by Western Kentucky University (WKU). The in-situ stations provide soil moisture monitoring depths of 102 mm, 203 mm, 508 mm, and 1016 mm. These depths were comparable to those available through NOAH 3.6 LSM. The locations of these stations throughout Kentucky are shown in Figure 3-7.

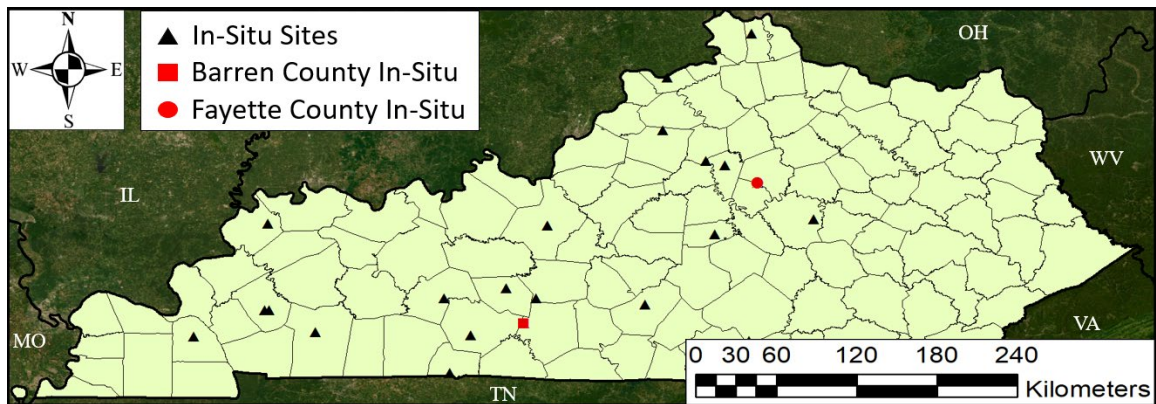


Figure 3-7: Ground-Based In-Situ Sites Across Kentucky

### 3.4.4 LIS NOAH 3.6 MODEL RUNS

Two distinct types of LIS modeling were investigated herein: Open-Loop (OL) and Data-Assimilation (DA). OL models soil moisture solely as a function of LDT preprocessed data and climatological forcing data. DA models soil moisture in the same fashion as OL, but also assimilates observations (L3SMP\_E herein) to effectively yield an assimilated product. The NOAH 3.6 model was run over three unique periods. The first period, solely an OL period, modeled soil moisture from 01/01/2016 – 10/01/2022. This period was selected to provide ample LIS modeled data to construct a modeled cumulative density function (CDF) for use in later CDF matching. The second period, the validation period, modeled soil moisture from 01/01/2017 –

01/01/2019. This period was selected due to observed availability of in-situ moisture data for use in validations of modeled soil moisture. The final period, the landslide period, modeled soil moisture from 01/01/2021 – 10/01/2022. This period was selected to temporally encompass the associated severe landslide occurrences in 07/2022. Both the validation and landslide periods experienced OL and DA modeling. Within LIS, DA modeling requires a rescaling of L3SMP\_E to that of modeled data. LIS performs this rescaling through CDF matching, where CDFs (from LIS and observation-based data) are used to bias-correct and reduce differences in observed and modeled data (Arsenault et al., 2018). Figure 3-8 shows LIS NOAH 3.6 OL and DA model runs over the validation period compared to in-situ data at the Kentucky Mesonet Barren County In-Situ site. The location of this site is shown in Figure 3-7.

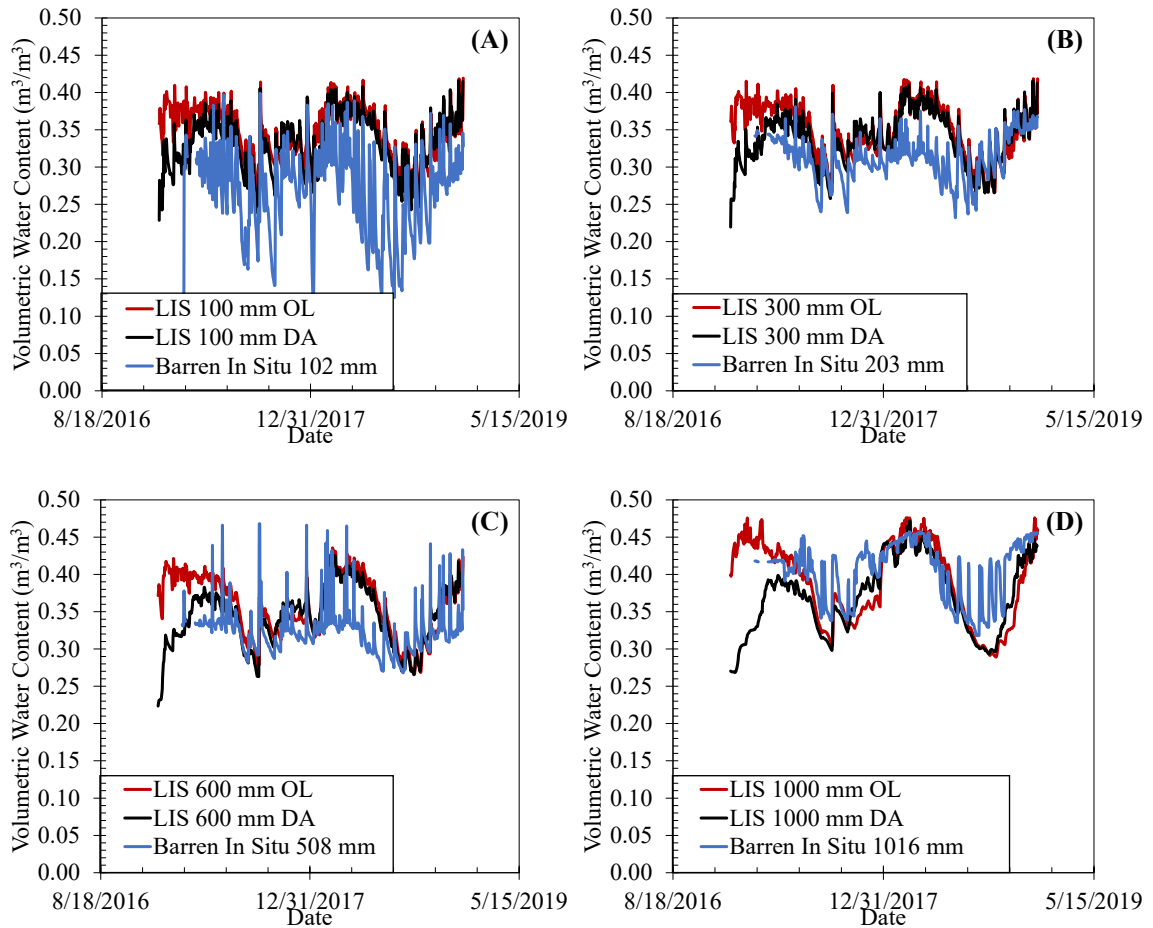


Figure 3-8: Comparison of OL and DA LIS NOAH 3.6 Estimated Moisture to Mesonet Barren County In-Situ Moisture at: (A) 100 mm, (B) 300 mm, (C) 600 mm, and (D) 1000 mm

As seen in Figure 3-8, LIS soil moisture performed well compared to in-situ data. It can be observed that the DA soil moisture performs better than OL modeled data. This can be observed in Figures 3-8A-8D whereas the DA-based soil moisture performs better at capturing the initial



conditions of the in-situ soil moisture. This is likely an artifact of the OL model spin-up, an adjustment process whereas the model moves from unusual to equilibrium conditions (Rahman and Lu, 2015), being a function of only forcing data. The DA spin-up period is a function of both forcing data and soil moisture observations, likely leading to DA estimates better capturing initial conditions of the period.

LIS estimates of soil moisture were generally observed to perform well compared to in-situ data. However, it was observed that estimates would occasionally lack in performance at deeper depths (e.g., depths > 300 mm). An example of this disparity can be seen in Figure 3-9 which shows OL and DA against in-situ data at the Kentucky Mesonet Fayette County In-Situ site (Figure 3-7).

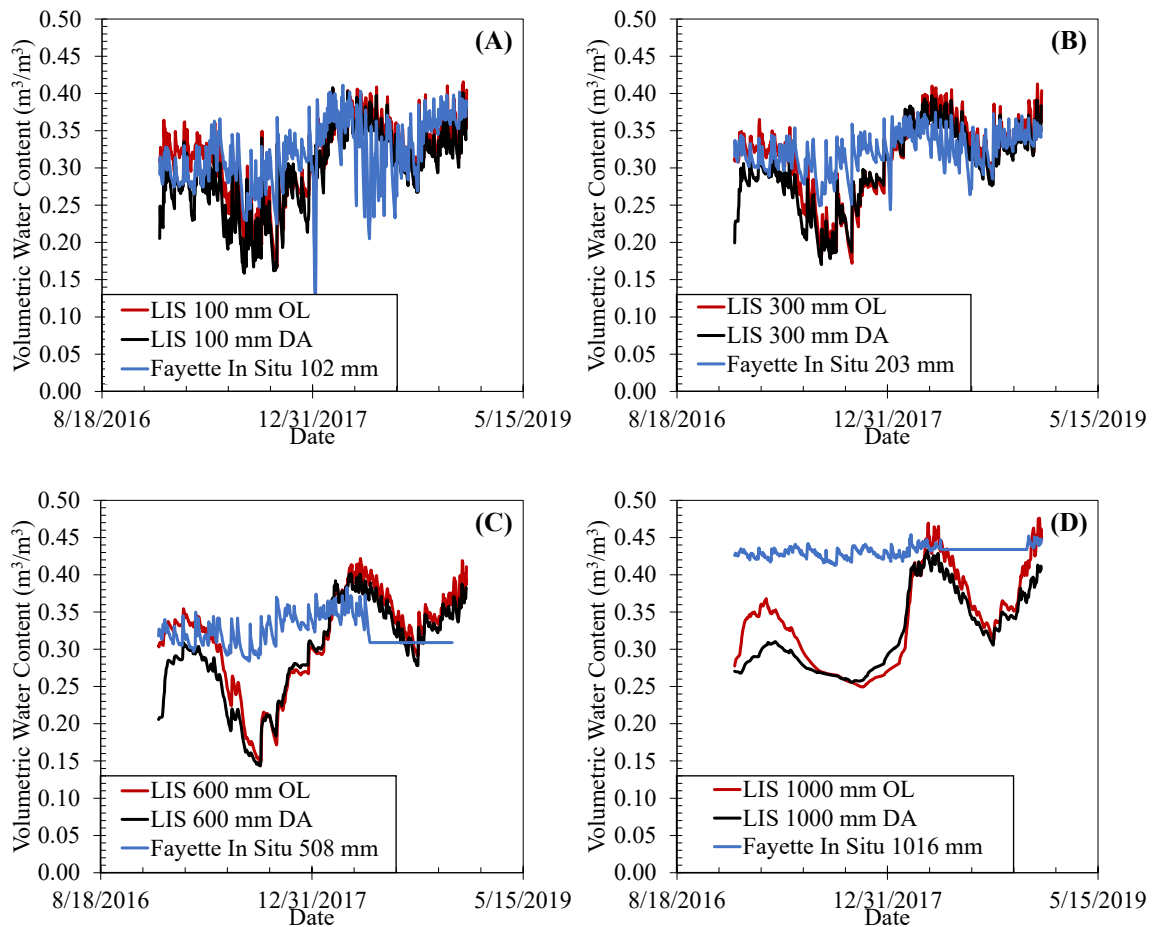


Figure 3-9: Comparison of OL and DA LIS NOAH 3.6 Estimated Moisture to Mesonet Fayette County In-Situ Moisture at: (A) 100 mm, (B) 300 mm, (C) 600 mm, and (D) 1000 mm

As seen in Figures 3-9A and 9B, the LIS estimates at 100- and 300-mm depths tend to perform very well as compared to the in-situ measured data at the same depths. However, Figures 3-9C and 3-9D reveal an obvious disparity at 600- and 1000-mm depths. There are two speculations as

to the cause of this disparity. The first speculation pertains to the soil texture map preprocessed within LDT. The STATSGO\_FAO soil texture map is defined as a two-horizon system, where the first and second horizons extends from 0 – 300 mm and 300 – 1000 mm, respectively. Within this system, the predominant soil texture class within the respective 0.01-degree grid cell is defined as the soil texture for the given layer. Therefore, it is speculated that a disparity exists between the second horizon (300 mm – 1000 mm) of the soil texture map and that of in-situ soil conditions. As LSMs rely on soil texture for modeling, this disparity would likely lead to the observed differences between estimated soil moisture values and in-situ values at these deeper depths within the soil layer. The second speculation pertains more upon the in-situ measuring device. As seen in Figures 3-9C and 3-9D, there is an approximately 6-month period that remains at constant moisture value. There are constant fluxes in a soil environment (e.g., rainfall, evapotranspiration, horizontal movement of moisture, etc.). Therefore, soil moisture remaining constant for an extended period is an unlikely phenomenon. This leads to speculation that the deeper seated in-situ monitoring devices were either installed improperly or were not functioning properly in terms of monitoring soil moisture levels. Regardless, LIS estimates were observed to perform well at a majority of ground-based verification sites. This performance is quite adequate for the implementation of spatial retrievals of the NOAA 3.6 L3SMP\_E assimilated soil moisture product within subsequent stability analyses.

### 3.5 HYDROLOGIC BEHAVIOR VIA SPATIAL LIS DATA

The overarching goal was the investigation into the feasibility of detection of incipient failure conditions at known landslide sites through fine resolution spatial site data. These investigations were conducted through the Lu and Godt (2008) infinite slope equation. Suction stress is a key variable in that of the associated infinite slope stability analyses due to impact of suction stress upon that of a soil system. Suction stress provides tensile forces within the soil, which yields an overall effect that is observed to increase the factor of safety within a stability analysis.

Additionally, suction stress can be observed to reduce the overall effective stress upon a soil system, which in turn increases the limit equilibrium factor of safety. The suction stress is defined as:

$$\sigma^s = S_e s \quad (3-2)$$

where,  $S_e$  = effective degree of saturation =  $(\theta - \theta_r)/(\theta_s - \theta_r)$ ;  $\theta$  = volumetric water content (VWC) (estimated through LIS NOAA 3.6 LSM);  $\theta_r$  = residual VWC;  $\theta_s$  = saturated VWC;  $s$  =



matric suction =  $(u_a - u_w)$ ;  $u_a$  = pore air pressure;  $u_w$  = pore water pressure; and  $\sigma^s$  = suction stress (kPa).

The implicit assumption herein is that soil moisture is a controlling factor of landslide occurrences, suction stress becomes a controlling factor as it becomes inversely related to soil moisture. This is due to the fact that, as soil moisture is seen to increase (i.e., the soil is nearing saturated conditions), the suction stress is seen to decrease. This decrease in suction stress carries the implication of less tensile forces and more effective stress being applied to the soils. This phenomenon could be the physical mechanism that is seen to trigger many shallow landslides when the respective hillslopes are subjected to intense periods of precipitation (Lu and Godt, 2008).

For calculation of suction stress, the effective degree of saturation variable ( $S_e$ ) was calculated through soil moisture estimates via LIS NOAH 3.6. However, data for matric suction was not readily available where slope stability analyses were conducted. To address this lack of data availability, the soil water characteristic curve (SWCC) model established by van Genuchten (1980) was manipulated for use. The manipulated van Genuchten (1980) SWCC equation for matric suction is as follows:

$$s = \frac{1}{\alpha} \left[ \left( \frac{1}{S_e} \right)^{1/m} - 1 \right]^{1/n} \quad (3-3)$$

where,  $s$  = matric suction;  $S_e$  = effective degree of saturation; and  $\alpha, n, m$  = fitting parameters that reflect the air entry value, the slope at the inflection point of the SWCC, and the curvature of the SWCC near the residual point, respectively.

The  $\alpha, n$  and  $m$  fitting parameters were obtained using the online tool, Handbook60 (<https://www.handbook60.org/rosetta/>). Handbook60 uses Schaap et al. (2001) pedotransfer functions (PTFs) to estimate the desired fitting parameters. The inputs for these PTFs are the % Sand, % Silt, % Clay, and Bulk Density at each analysis point. These data were retrieved via the WSS tool. The outputs retrieved via the Handbook60 Rosetta tool were estimates of  $\theta_r$ ,  $\theta_s$ ,  $\alpha$  and  $n$ . However, the Handbook60 Rosetta estimate of saturated VWC ( $\theta_s$ ) was not utilized. Instead,  $\theta_s$  was estimated as the maximum VWC retrieved via LIS NOAH 3.6 soil moisture over a given analysis period. Additionally, Handbook60 Rosetta did not provide an estimate for the  $m$

fitting parameter. For this, the Mualem (1976) approximation,  $m = 1 - 1/n$ , was used. As these fitting parameters were estimated as a function of very fine resolution (~100-300 m) WSS data, this implies these parameters to be at the same very fine resolution. This in turn implies that matric suction and suction stress can also be estimated at these very fine resolutions. Figure 3-10 shows the relationship between soil moisture measured or estimated at the Barren County In-Situ site as compared to matric suction and suction stresses.

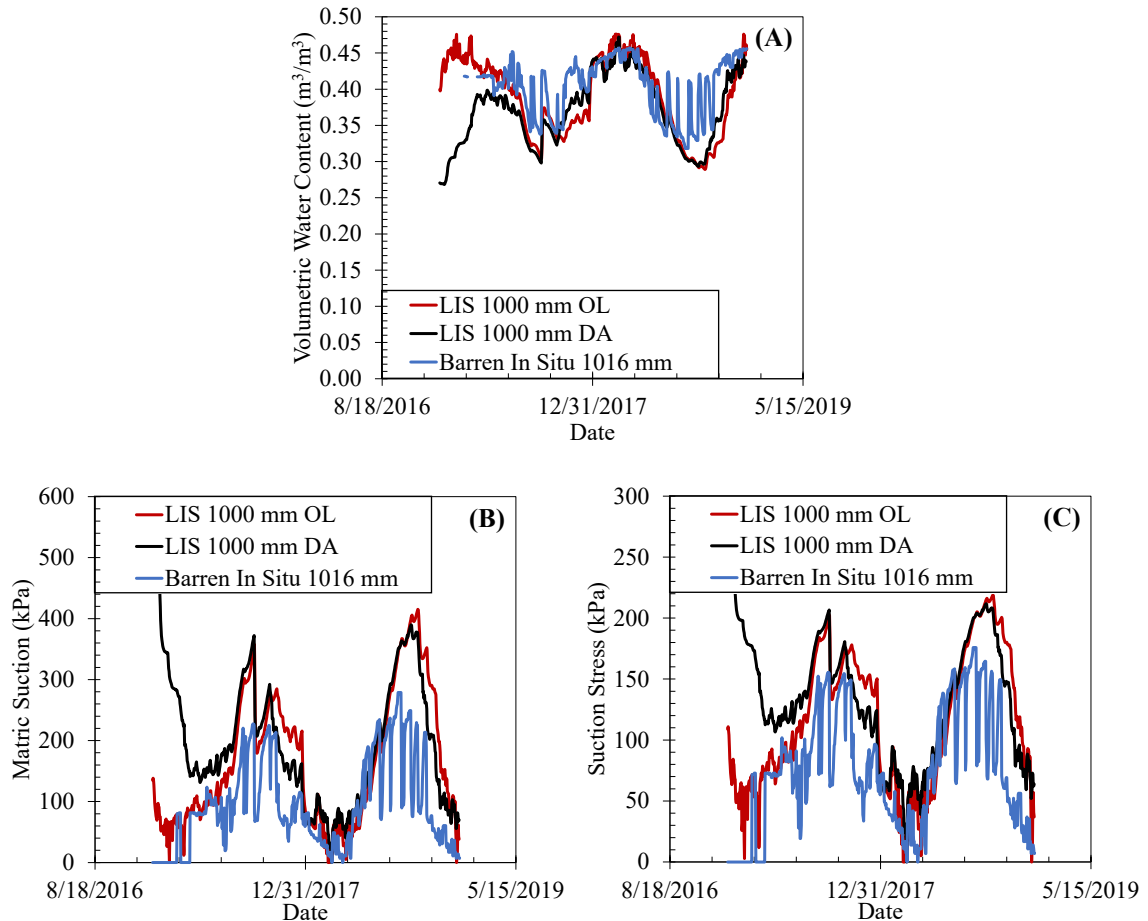


Figure 3-10: Mesonet Barren County Measurements of: (A) In-Situ vs LIS Estimates of VWC, (B) In-Situ vs LIS-based Matric Suction, and (C) In-Situ vs LIS-based Suction Stress

It can be observed that as soil moisture in Figure 3-10A decreases, the matric suction and suction stresses, Figures 3-10B and 3-10C, increase. The opposite can be observed to be true when soil moisture increase. This again shows that matric suction and suction stress are inversely related to that of soil moisture. Suction stress is again a key variable in the associated infinite slope stability analyses. Therefore, retrieval of this parameter at fine spatial resolutions enabled the subsequent slope stability analyses to be conducted at very fine resolutions as well.

### 3.6 SLOPE STABILITY ANALYSES

The analyses were conducted over an area of Eastern Kentucky subjected to an extreme amount of rainfall over a very short period (07/25/2022 – 07/30/2022) of time. Due to this short period of intense rainfall, multiple landslides were detected and recorded throughout the area. Therefore, the approximate dates, rather than exact dates, of occurrences of these slides are known. The spatiotemporally retrieved geomorphic and hydrologic data were used in conjunction with slope stability analyses to detect the occurrence of these known slides.

#### 3.6.1 LANDSLIDE STUDY AREA

The area of study throughout this work was focused within the Eastern region of Kentucky. This area is shown in Figure 3-11.

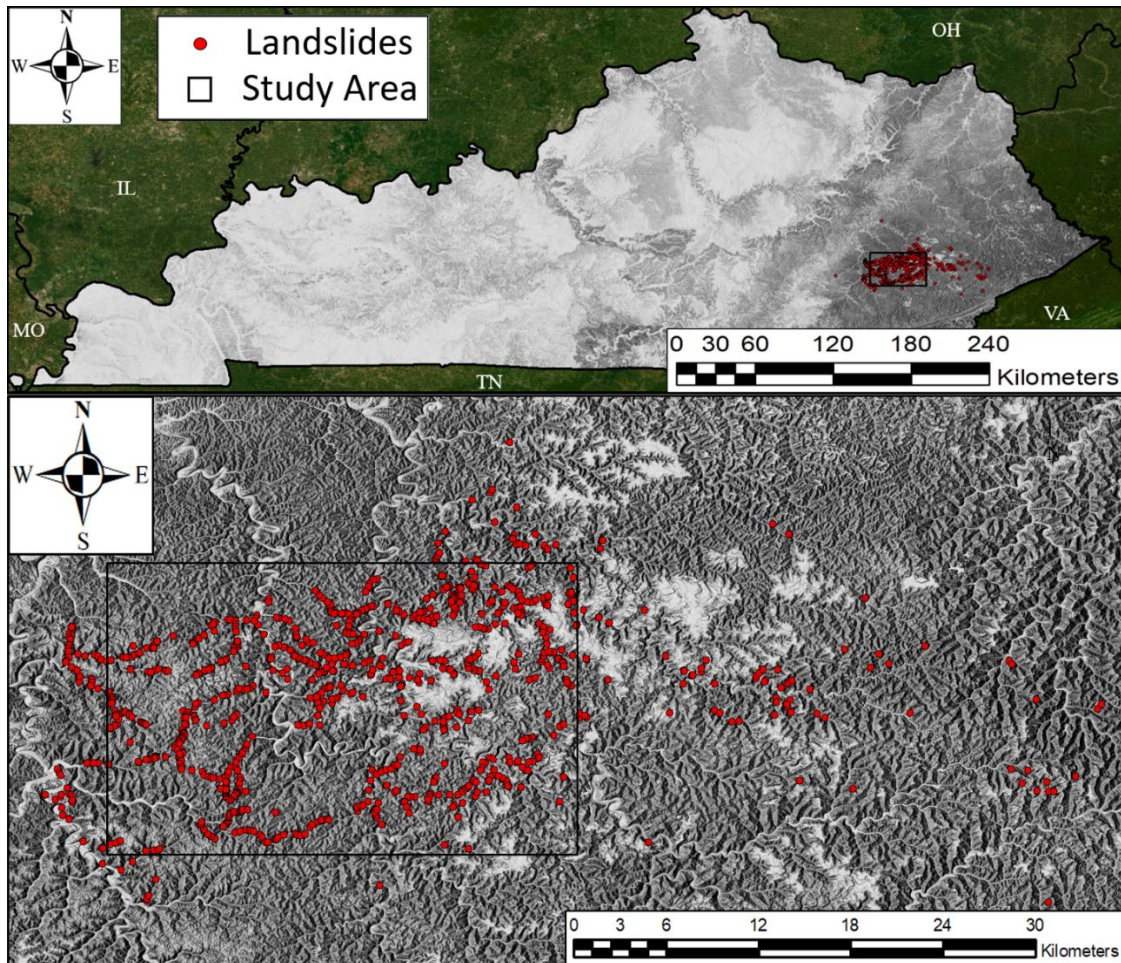


Figure 3-11: Landslide Study Area: (A) Overview of Kentucky with Locations of Eastern Kentucky Rain Induced Landslides Enclosed in Black Box and (B) Exploded View of Eastern Kentucky Rain Induced Landslide Region

Figure 3-11A shows the overall location of rainfall-induced landslides. It is important to note that these landslides were observed after the rainfall had ceased. Therefore, an exact date of

occurrence is not known. Instead, a range of dates for occurrence, 07/25/2022 to 07/30/2022 is known. Figure 3-11B shows an exploded view of this region, whereas the landslides observed to have occurred are indicated as individual points. Through the associated rainfall event, over 500 small to moderate sized translational landslides and debris flows were observed. The study area for these slides is indicated in Figure 3-11 by the bounded area which encompasses a majority of occurred landslides. Additionally, the points that indicate an observed landslide simply indicate the runout of the slide or flow. This is to say that an exact point of origin for the landslide is also unknown. The implications of this will be discussed within the following slope stability analyses.

### 3.6.2 EASTERN KENTUCKY RAINFALL EVENT

The landslides investigated herein were caused by a heavy rainfall event. Figure 3-12 shows the daily rates of rainfall for a 6-month period over the investigated area with the dates of occurrence for the investigated landslides indicated by the red-shaded box. These rainfall rates were retrieved via the NASA Integrated Multi-satellite Retrievals for GPM (IMERG) Global Precipitation Measurement (GPM) mission.

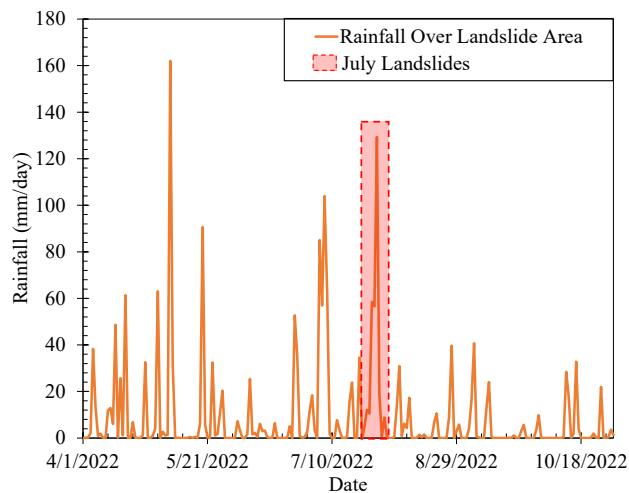


Figure 3-12: Rainfall (mm/day) Over Landslide Study Area from 4/1/2022 to 10/31/2022

From Figure 3-12, it can be observed that a large amount of rain fell over the period when the reported landslides occurred, with rates reaching approximately 140 mm/day. A large rain event also occurred on approximately 5/6/2022, with no reported landslides. However, it can be observed that a significant rain event occurred prior to the rainfall associated with the occurrence of the investigated landslides. It is hypothesized that this earlier rain event (~7/5/2022) led to higher antecedent soil moisture in the area with the second primary event (07/25/2022-07/30/2022) being more severe and leading to rapid saturation in the area, thus causing landslide

occurrences. While this primary event was not continuous (i.e., it did not rain continually over the period), the associated storms did cause rainfall rates seen to exceed 100 mm/hour over the affected area. Additionally, radar-based rainfall estimates indicate that upwards of 350 – 400 mm of rain fell during this 5-day period in a relatively narrow swath over the area. Most of this rain fell during the night of 07/27/2022 and continued into the day of 07/28/2022. The rainfall values observed between this 5-day period were over 600% of normal rainfall for the given area and given time of year. Additionally, these rates of rainfall occurring in such a short time period are incredibly rare, with less than a 0.001% chance for such levels of rainfall in a given year (NOAA, 2022). The average VWC at various depths over the study area from the LIS Landslide Period (01/01/2021 – 10/01/2022) run are shown in Figure 3-13.

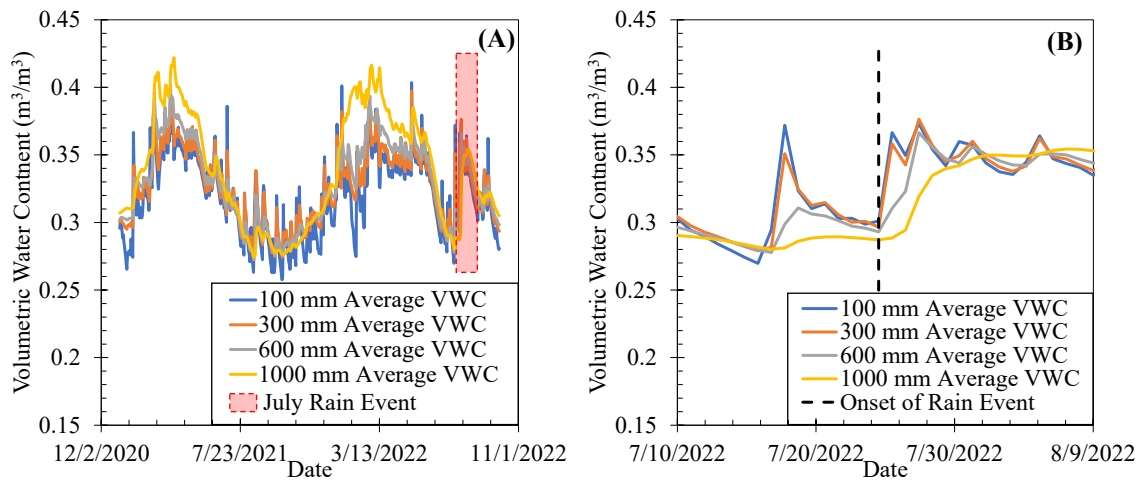


Figure 3-13: Average of LIS Estimates of Volumetric Water Content from Landslide Area over: (A) Full LIS Landslide Period Run (01/01/21 - 10/01/22) and (B) Snapshot of LIS Landslide Period Run Over Period Including Rain Event (07/10/22 - 08/9/22)

Figure 3-13A shows the VWC estimated via LIS NOAA 3.6 over the entire “Landslide Period”, defined as 01/01/2021 to 10/01/2022 herein. The time frame corresponding to the associated rain event and landslide occurrences is indicated by the red dashed box. Within this period, enclosed by the boxed region, the LIS estimates perform well at capturing the intense rain event. The VWC can be seen to sharply increase, likely in response to the intense rainfall over the area. Figure 3-13B shows the VWC over the time period enclosed within the dashed box in Figure 3-13A. Additionally, the black dashed line in Figure 3-13B represents the date of 07/25/2022, which is the onset of the intense rainfall and subsequent landslide occurrences. As seen in 3-14B, there was a sharp increase in VWC immediately after the onset of the intense rain. It can also be seen that a sharp increase occurred approximately a week before the rainfall event. However, it can be observed that only the upper layers experienced a sharp increase at that time. At the time of the

associated rainfall event, it was observed that all analyzed layers experience an increase in their respective VWCs. From these observations, the LIS estimates, based solely on spatiotemporal parameters, were able to successfully capture increases in VWC at the onset of a known rainfall event.

### 3.6.3 INFINITE SLOPE STABILITY

The associated slope stability analyses conducted throughout this work made use of a modified version of the Lu and Godt (2008) infinite slope stability equation. As most slides investigated during herein were observed to be that of shallow colluvial slides, the usage of this equation was assumed to be valid. The investigated equation is as follows:

$$FS = \frac{\tan(\phi')}{\tan(\beta)} + \frac{2c'}{\gamma H_{ss} \sin(2\beta)} + r_u [\tan(\beta) + \cot(\beta)] \tan(\phi') \quad (3-4)$$

where,  $FS$  = factor of safety;  $\phi'$  = soil friction angle (estimated using ANN model);  $c'$  = effective soil cohesion (assumed negligible due to normally consolidated soil);  $\beta$  = slope angle;  $\gamma$  = soil unit weight ( $\text{kN/m}^3$ );  $H_{ss}$  = depth to bedrock;  $r_u$  = pore pressure ratio =  $\sigma^s / \gamma H_{ss}$ ;  $\sigma^s$  = suction stress.

It is necessary to note that the matric suction variable of Equation 3-4,

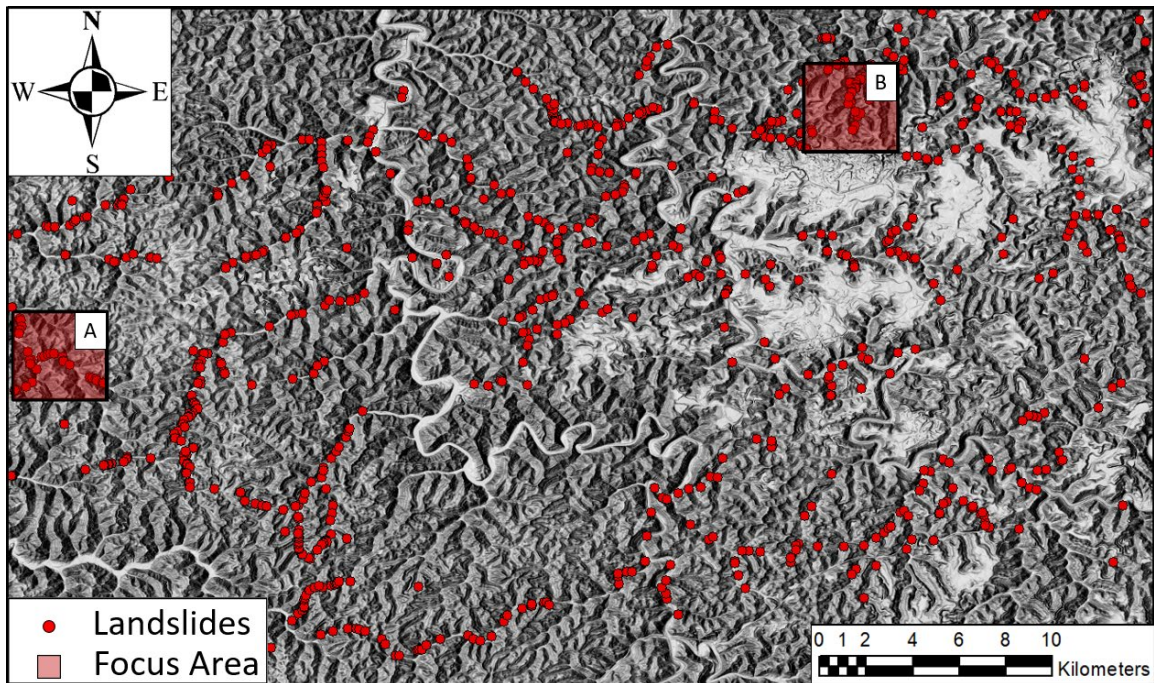
$\{r_u [\tan(\beta) + \cot(\beta)] \tan(\phi')\}$ , was considered to be positive. This differs from the equation defined by Lu and Godt (2008), in that this variable was assumed negative. Lu and Godt (2008) assumed the suction stress to be negative. Thus, a negative multiplied by a negative yields a positive. The overall assumption from this is that matric suction within the soil increases the factor of safety. As matric suction decreases (i.e., as the soil becomes saturated), the additional strength from matric suction decreases and thus the factor of safety is reduced. These reductions in strength become indicative of landslide occurrences through the associated infinite slope analyses.

### 3.6.4 LIS NOAH 3.6 BASED STABILITY ANALYSES

As discussed previously, the scope herein was the investigation into the detection of landslide occurrence utilizing spatiotemporal parameters. The investigation was conducted over the entire landslide study area. However, for demonstration purposes, two roughly 3 km by 3 km areas observed to encompass a high number of landslides were selected. These focus areas can be seen in Figure 3-14. It is worth noting that, as mentioned previously, the observed landslide



occurrences (indicated as points) are not the initiation point of the slide. Instead, the noted locations are that of the observed runout of the associated slides. This can be seen in Figure 3-14 whereas most landslide locations appear to be at the foot of a hill rather than upon the face of the hill. So again, due to the nature of these landslides occurring during an extreme rain event, the exact time of occurrence and exact point of initiation are not known. Instead, the runout of the slide is known, and the point of initiation is inferred from this.



*Figure 3-14: Hill Shade map of Landslide Area with 3 x 3 km Landslide Focus Areas A and B*

#### **3.6.4.1 FINE RESOLUTION INFINITE SLOPE STABILITY ANALYSES**

As the efforts herein made use of very fine resolution (~100 m) WSS data, the associated landslide analyses were also conducted at this same resolution. This was able to be achieved due to the fact that suction stress (Equation 3-2), which is a key variable of the infinite slope equation (Equation 3-4), was able to be estimated as a function of LIS estimates and WSS data. The LIS estimates were available at ~1 km resolution while WSS data is readily available at a very fine resolution of ~100 m. As suction stress is a function of both datasets, estimation was able to be conducted at the same resolution as that of WSS data. This is due to the fact that, across a ~1 km soil moisture cell, the fine resolution WSS data would vary. This variance in soil data led to a variance in suction stress across the estimated soil moisture cell. Therefore, with suction stress estimated at these very fine resolutions, the slope stability analyses were in turn able to be conducted at the same very fine resolution.

As the heaviest rainfall over the area was reported to have occurred on the night of July 27<sup>th</sup> into the day of July 28<sup>th</sup>, the stability analyses from the dates after this heavy rainfall (07/28/2022-07/30/2022) are shown. The stability analyses prior to the rainfall event, on 07/22/2022, are also shown. Additionally, it was observed that the lowest factors of safety (closest to or below that of 1.0) were observed through use of 1000 mm soil moisture estimates. Therefore, the following analyses show data obtained at 1000 mm depth. The stability analyses over these dates and at 1000 mm depth from Focus Area A are shown in Figure 3-15.

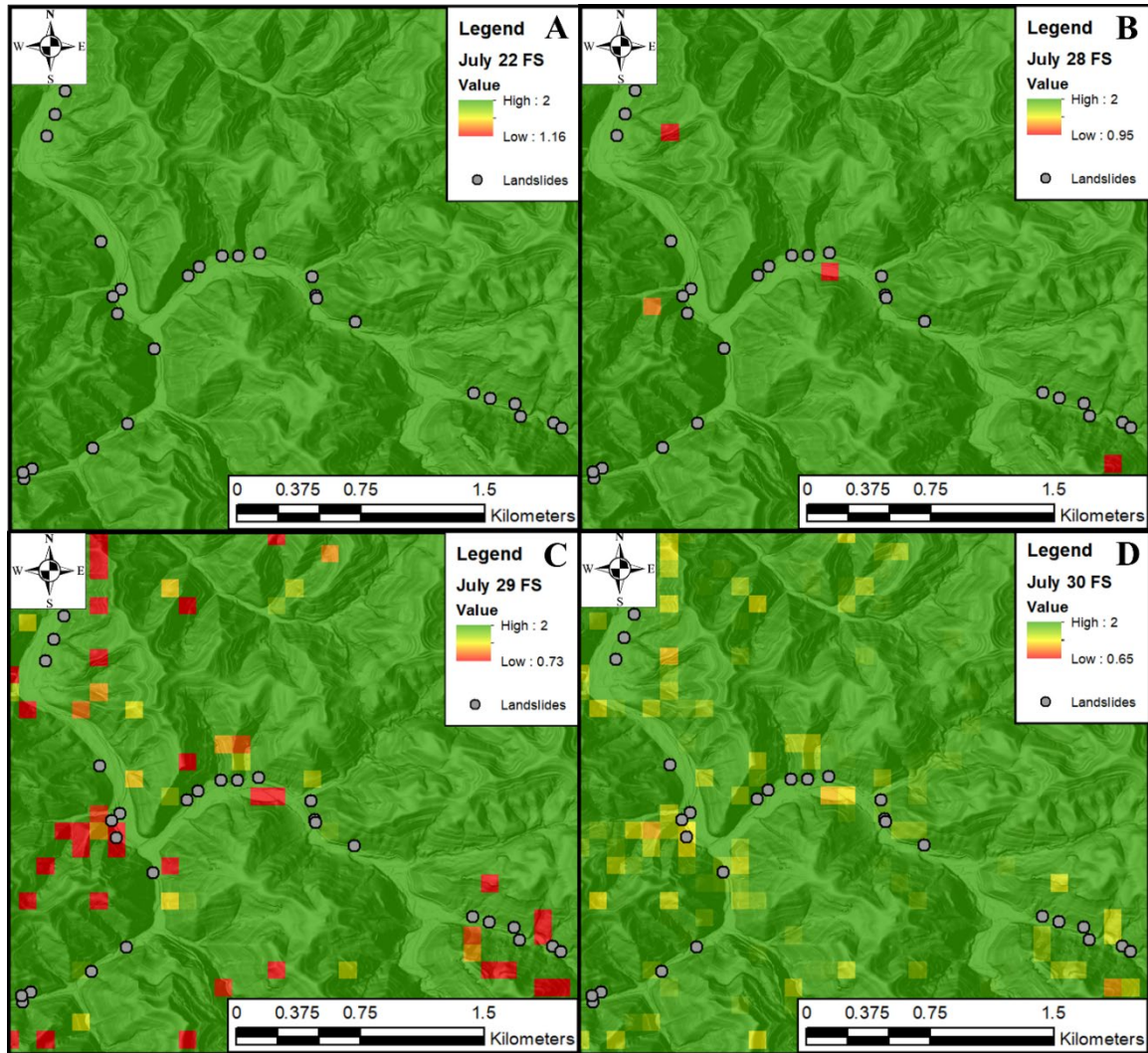


Figure 3-15: Stability Analyses from Focus Area A at 1000 mm Depth on: (A) 07/22/2022, (B) 07/28/2022, (C) 07/29/2022, and (D) 07/30/2022

Figure 3-15 shows the stability analyses over Focus Area A. In general, it can be observed that reported landslides occurred either in or very close (within 100 m) to areas estimated to have low factors of safety. This discrepancy of landslide location is again due to the fact that the runout of



the slides, rather than the initiation point, are known. This carries the implication that, even though a slide may not appear to lie within an area of low factor of safety, the slide may have initiated in a nearby low factor of safety and ran out to its observed point. In addition to success in correlating landslide occurrence to areas of low factor of safety, it was also observed that low factors of safety were present on 07/29/2022. This implies that the soils became most saturated and soil moisture drove the initiation of these slides after the heavy rain event on 07/27/2022 and 07/28/2022. This same success in detection of landslide occurrence was observed over Focus Area B, shown in Figure 3-16.

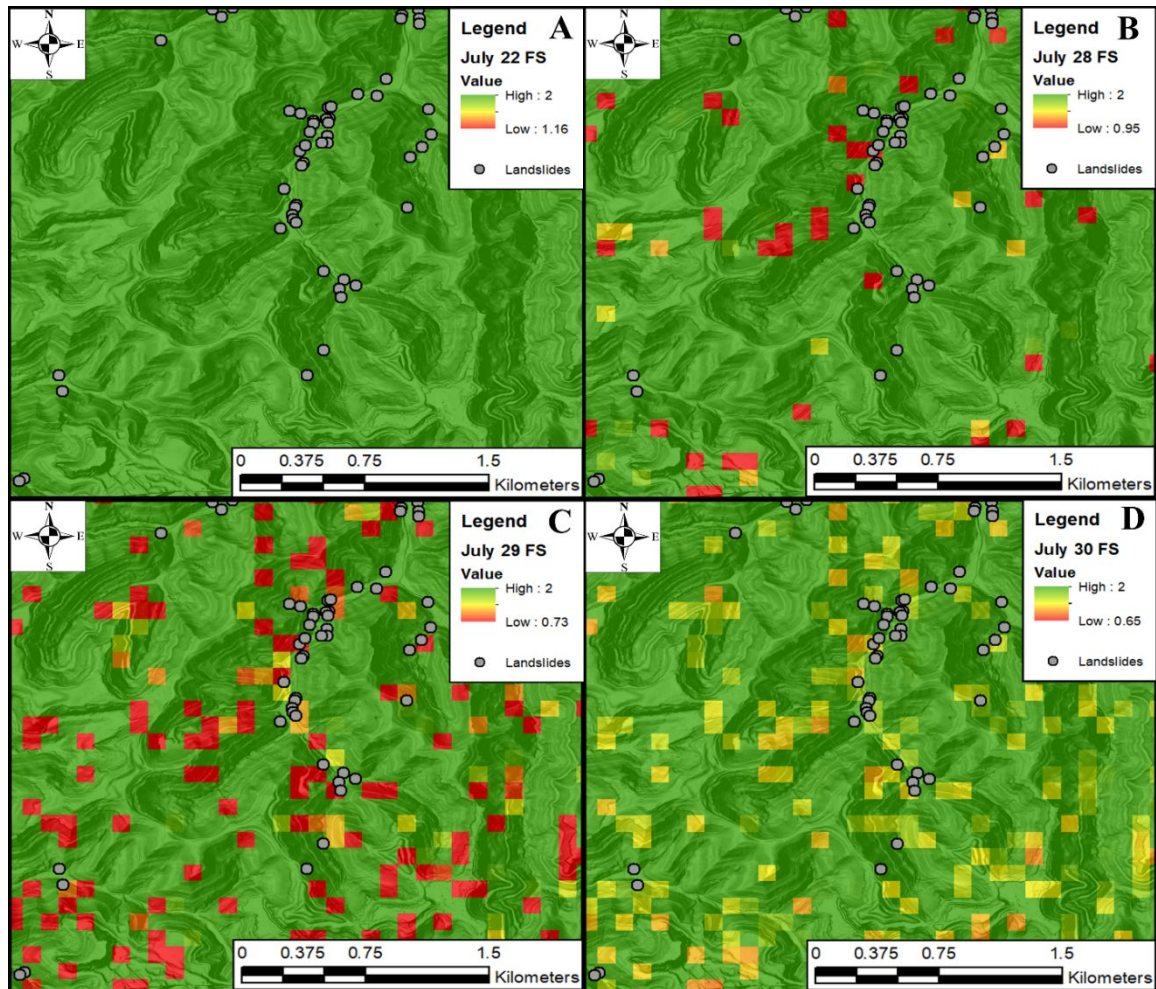


Figure 3-16: Stability Analyses from Focus Area B at 1000 mm Depth on: (A) 07/22/2022, (B) 07/28/2022, (C) 07/29/2022, and (D) 07/30/2022

As with Figure 3-15, Figure 3-16 shows successful stability analyses over the given area. It can be observed that the lowest factors of safety were present on 07/29/2022 after saturation due to heavy rain on the previous days. This again implies that, as it was the primary flux in this scenario, soil moisture was the driving force behind the occurrence of these slides.

The analyses over both Focus Areas A and B perform well at indicating landslide occurrences after the heavy rainfall event. However, it can also be observed that low factors of safety exist in areas where no landslide occurrence was reported. It is speculated that one potential cause of these false positives is that the associated stability equation does not account for the presence of roots within a soil system. The infinite slope stability equation only accounts for weights of the soil layer and strength added through friction and suction stress. It has been observed that stability conditions of slopes have an additional dependence on the mechanical reinforcement of plant roots that are present within the soil (Chirico et al., 2013). As the investigated area has been observed to be heavily vegetated, there is likely to be a strong influence from the roots of this vegetation upon the stability of the respective slope. It is due to this lack of representation of strength via root systems that the false positives are speculated to exist. Future work will conduct investigations into the additional strength added by these roots, and it is expected that the occurrence of these false positives will be reduced.

#### **3.6.4.2 VARIATIONS OF PARAMETERS WITHIN STABILITY ANALYSES**

Three implicit assumptions were made in respect to the input parameters for the associated stability analyses. These assumptions are as follows slope angles less than that of 30 degrees are considered to not be at risk of failure, estimated friction angles fall within a range of values rather than a singular point, and saturation of soil in a rain event differs from saturated water contents that may be observed over an extended soil moisture analysis.

As this effort investigated slope stability upon hillslopes, a wide range of slope angles were investigated. These angles ranged from 0 degrees (i.e., flat land) to more extreme angles of ~45 degrees. However, through the associated stability analyses, it was observed that lower slope angles yielded very high factors of safety (upwards of  $FS = 20$ ). These high factors of safety did not need to be accounted for as this study focused primarily on factors of safety indicative of landslide occurrence. Therefore, slope angles less than that of 30 degrees were not considered within the factor of safety analyses. As these slides were those of shallow slides induced by influx of moisture through a rainfall event, the shallow saturated zone plays a role in the triggering of the slide. Results from previous studies indicate that this influential zone easily forms on slopes of 30 degrees~40 degrees. It is because of these findings that a threshold of 30 degrees was established during this work (He et al., 2021).

The second assumption that was made pertained to estimated friction angles. The empirical relationship between PI and friction angle, by Mesri et al. (1996), was adopted herein. As this relationship saw a line of best fit (Equation 3-1) between points of PI vs friction angle, outliers

were observed to be present. This implies that estimated friction angles could in theory lie above or below this line of best fit. It was assumed that Mesri et al. (1996) estimated friction angles could lie in a range of  $\pm 5$  degrees from the line of best fit. This assumption was utilized to yield lower factors of safety at known occurrence locations. When friction angles via the original Mesri et al relationship were used, factors of safety higher than 1.0 were observed over sites where landslides occurred. However, when estimated friction angles were reduced by 5 degrees (following the implication of outliers within the Mesri et al relationship), the respective factors of safety were observed to move towards the expected values (i.e.,  $FS = 1.0$ ). Therefore, as outliers are observed within Mesri et al, and due to the nature of fitting a curve through data, this variation of the friction angle parameter was deemed to be a sound approach.

The final parameter varied was that of the saturated water content ( $\theta_s$ ). Initially, this parameter was estimated by taking the maximum water content at each depth over the entire Landslide Period LIS run (01/01/2021 – 10/01/2022). However, as seen in Figure 3-13A, the maximum water content for each layer, over this entire period was visibly higher than that of the maximum water content observed over the investigated period of intense rainfall. This discrepancy between water contents became problematic when calculations of matric suction were conducted. As the historical saturated water content was higher than the water content over the analyzed period, the implication became that matric suction existed and was adding additional strength to the soil. This added strength led to higher factors of safety being observed when it was instead expected to see values nearer to that of 1.0. Therefore, rather than taking the maximum water content over the entire period, a maximum was taken over a finer temporal period of 1 month before and after that of the rainfall event (06/01/2022-08/31/2022). This shorter period led to a saturated water content more indicative of the soil becoming saturated through the rain event and having little to no strength added via matric suction. This assumption is based upon that of saturation changes between antecedent saturations. Antecedent saturation in this case pertains to the saturation in a soil before the onset of a rain event. Antecedent saturation change focuses upon the saturation change from antecedent saturation to that of saturation after the onset of the rain event. As seen in Figure 3-13A, all layers experience a drastic increase of saturation, with the increase starting at antecedent values (i.e., prior to the rain event). This increase of saturation can cause a critical increase in pore water pressure and loss of matric suction, which becomes indicative of saturated conditions. This again leads to an increased probability of landslide occurrence (Wicki et al., 2020). A rapid increase of saturation from that of antecedent conditions is observed over the investigated rainfall period. From Wicki et al (2020) this increase likely led to a drop in matric

suction, which is indicative of saturated conditions. Therefore, based on Wicki et al (2020), the maximum water content over this shorter period was considered to be the saturated water content for the associated analyses.

### **3.7 CONCLUSIONS**

This study saw the development of a means to conduct slope stability analyses which made use of spatially retrieved data across Kentucky. These analyses were investigated as functions of both static geomorphologic data of the soil and the dynamic soil moisture measurements at a given site. The geomorphologic data for these analyses was easily obtainable at given sites of interest through implementation of the Web Soil Survey tool and machine learning. However, as landslides are observed to occur at various depths within a soil layer, soil moisture at various depths was also desired. Yet, measurements of soil moisture at various levels within the soil were not observed to be as readily available at these sites of interest. To remedy this lack of soil moisture availability with depth, the Land Information System (LIS) framework was utilized. Through implementation of LIS, satellite-based surface moisture estimates were assimilated with estimates from the NOAA 3.6 land surface model (LSM). The NOAA 3.6 LSM and satellite-based assimilation yields fine resolution estimates at depths of 100 mm, 300 mm, 600 mm, and 1000 mm. Therefore, the desired soil moisture estimates at various depths were now obtainable.

As the efforts herein implicitly assumed soil moisture to be the controlling factor in landslide occurrence, the soil moisture estimates via LIS needed validation. This validation was conducted by comparing LIS-based estimates to that of in-situ data, at equivalent depths, at 20 in-situ stations across Kentucky. From these comparisons, it was observed that LIS-based estimates of soil moisture performed extremely well as compared to that of in-situ data. Therefore, the LIS-based soil moisture was validated for use in the subsequent stability analyses conducted during this work.

Slope stability analyses were conducted to investigate the role of soil moisture and suction as precursors indicative of incipient conditions for rainfall-induced landslide occurrences. These analyses relied on the availability of the shear strength parameter, friction angle. However, it was observed that spatial availability of friction angle over the study area was lacking. To remedy this, an ANN machine learning workflow was developed and validated such that friction angle became spatially available across the study area. Herein, it was shown that ANN machine learning can serve as a viable means to obtain spatial estimates of data over areas observed to be lacking data availability.

With all requisite spatial data available, the work herein shifted to landslide hazard analyses as a function of limit equilibrium stability analyses. These analyses focused upon the hydro-mechanical relationship of slope stability as a function of soil moisture fluxes. This relationship was investigated through detection of incipient failure conditions at known landslide sites. These incipient conditions are indicated by low factors of safety via the associated infinite slope stability analyses. These analyses were conducted over a region within Eastern Kentucky that experienced an extreme rainfall event. This rainfall event led to the occurrence of over 500 small to medium shallow landslides. Through the associated slope stability analyses, a majority of these slides were able to be detected in areas observed to experience increases in soil moisture. These slides were considered to be at incipient failure conditions by the observation of low factors of safety at or near the observed landslide locations. However, it was also observed that low factors of safety were present in areas where no landslide occurrence was reported. These false positives are attributed to the lack of representation of strength gain via plant root systems in the soil system. Future work will be conducted to investigate strength gain via these roots and is expected to yield stability analyses with less occurrence of these false positives. Nonetheless, the investigated stability analyses performed well at detection of known landslide occurrences and showing that soil moisture can serve as a hydro-mechanical precursor for landslide occurrence. Therefore, the efforts herein show that spatial geomorphologic data and spatiotemporal estimates of soil moisture can be effectively used to detect and better understand hydro-mechanical conditions leading to incipient failure conditions in the occurrence of rainfall-induced landslides.

## CHAPTER 4: Coupled Landslide Analyses Through Dynamic Susceptibility and Forecastable Hazard Analysis

### 4.1 INTRODUCTION

Landslides are geohazards typically observed to occur during, or soon after, periods of intense rainfall that lead to increased levels of soil moisture. This increase in soil moisture reduces overall strength of the soil while increasing overall stresses within the soil (Hong et al., 2007; Ray and Jacobs, 2007). As global climatological influences (global warming, climate variability, etc.) continue to increase, so do observed occurrences of extreme rainfall events. This subsequent increase of rainfall extremes leads to the overall increase of rainfall induced landslide likelihood (Sidle and Bogaard 2016, Swain et al., 2020). Therefore, a means to accurately monitor potential landslide occurrences in susceptible areas is desired. Two approaches for the monitoring of these occurrences are readily available. The first approach is that of landslide susceptibility analyses. These analyses make use of static factors over a landslide prone region to describe the spatial distribution of the probability of landslide occurrences based upon the geographic settings (Stanley and Kirschbaum 2017; Crawford et al. 2021; Dashbold et al. 2023). The key part of these susceptibility analyses is the knowledge that such analyses typically rely solely upon static data. Static data, including geomorphologic parameters such as slope angle, soil textures, and soil depths, are those that are not observed to vary with time. As a function of static parameters, traditional susceptibility analyses provide insight into *where* a landslide occurrence is possible but provides no insight into *when* or *why* the landslide will occur. Landslide hazard analyses, the second approach, provide insights into not only the *where*, but also *when* and *why* a landslide occurs. Hazard analyses consider both static data, such as geomorphic data, *and* dynamic data that is seen to vary with time, such as precipitation or soil moisture levels. These hazard analyses are typically coupled with landslide Early Warning Systems (EWS) or nowcasts of hazard predictions. EWS's and nowcasts provide warnings of potential landslide hazards through analyses of precipitation levels indicative of landslide occurrences. If these analyses observe precipitation that is likely to lead to landslide occurrences, then warnings are provided for potential landslide hazards in near real-time (i.e., over the next few coming days) (Kirschbaum and Stanley, 2018; Hidayat et al., 2019). A coupling of these analyses, susceptibility and hazard, provides planners with a better understanding of what regions are prone to landslide occurrence as well as the ability to directly analyze, in real time, when and where a landslide occurrence is likely.

This paper presents a unique approach to yield dynamic susceptibility analyses as well as forecastable, rather than nowcasted, landslide hazard analyses for landslide prone regions in Eastern Kentucky. The implementation of both dynamic susceptibility and forecastable landslide hazard analyses will undoubtedly provide area planners with more warning of landslide occurrences than with existing static susceptibility and near real-time nowcasting systems. This extended period of warning will likely provide more time for preparation and mitigation for areas or infrastructure within affected areas. Through this study, static susceptibility analyses were shifted to dynamic through the inclusion of measures of vegetation at a given site. Vegetation plays a role in landslide occurrence through potential losses of soil strength due to root reduction or through potential added surcharges due to moisture retainment within vegetation (Guo et al., 2020; Phillips et al., 2021). Temporally varying estimates of vegetation were retrieved via the Moderate Resolution Imaging Spectroradiometer instrument aboard the Aqua and Terra Earth Observing Satellite (EOS) Earth missions. Inclusion of these temporally varying data within the developed susceptibility analyses yielded susceptibility observed to vary with time rather than remaining static. Landslide hazard forecasting was developed through implementation of forecasted soil moisture levels within an infinite slope stability equation. Historic retrievals of spatial soil moisture were estimated through the Land Information System (LIS) framework, available through NASA. These soil moisture retrievals were then used, in conjunction with climatologic data, in the development of a Long Short-Term Memory (LSTM) machine-learning based forecasting regime. Once developed, the LSTM model was then used to forecast soil moisture levels at a site of interest. These forecasted values were assimilated with the aforementioned infinite slope stability equation, effectively yielding forecasted, rather than nowcasted, landslide hazard analyses.

For the development of both dynamic susceptibility and forecastable hazard analyses, a requisite geomorphologic variable was that of depth to bedrock (DTB). This is due to the observation that Eastern Kentucky landslides typically occur in shallow (i.e., shallow DTB) colluvial soils. Therefore, spatial estimates of DTB were desired for the creation of spatial susceptibility and hazard analyses. However, it was observed that these desired spatial data were seen to be lacking. To remedy this lack of data, a novel DTB map was developed. This map provided spatial estimates of DTB for use herein over areas previously seen to be lacking this data. This innovative DTB map proved to be an effective means for acquisition of requisite data for the development of dynamic susceptibility and forecastable hazards herein.

This study validated the developed novel approaches for landslide hazards through investigations into known landslide occurrences. The dynamic susceptibility analyses were validated in conjunction with the proposed infinite slope stability over known landslides within the investigated Eastern Kentucky domain. Once validated, the efforts of this study moved to conducting a 7-day pseudo-forecast of landslide hazard analyses. These forecasted hazards were investigated over *known* landslides that occurred outside the period used in the development of the LSTM forecasting model. Doing so allowed for: (1) Forecasted hazards to be validated against known occurrences and (2) The forecasted soil moisture to be truly forecasted as this forecast period was one that the LSTM model had not been exposed to. Additionally, no landslides are known beyond the date of this study. Therefore, a true forecast beyond this study is possible, but validations of landslide occurrence are not. The results of these forecasted hazard analyses indicated failure conditions to be present at or near the time of failure of the investigated known landslides. The study presented herein provides a novel workflow for both dynamic susceptibility analyses as well as forecastable landslide hazard analyses. The potential increased warning provided by these systems will undoubtedly impact geohazard predictions for planners and designers in mountainous regions around the world.

#### **4.2 LIS-BASED SPATIAL SOIL MOISTURE RETRIEVALS**

Landslide occurrence has been observed to be dependent upon soil moisture levels within a soil layer (Hong et al., 2007; Ray and Jacobs, 2007). Therefore, spatial measurements of soil moisture are desired for implementation within spatial landslide hazard analyses. Spatial estimates of soil moisture are globally available through low Earth orbit satellite-based retrievals. However, these satellite-based retrievals are typically at too coarse a resolution (9 km) for adequate use in landslide hazard analyses. To remedy this, the Land Information System (LIS) framework was utilized to obtain desired spatial estimates of soil moisture. LIS is a flexible land surface modeling and data assimilation system that is readily available for use by researchers. The LIS framework is structured such that researchers are able to utilize modeling, data, and assimilation resources to recreate existing models, or create novel models, over a given area of interest (Kumar et al., 2006; Peters-Lidard et al., 2007). The LIS framework was used to obtain soil moisture estimates at a fine resolution of approximately 1 km for use in subsequent landslide hazard analyses.

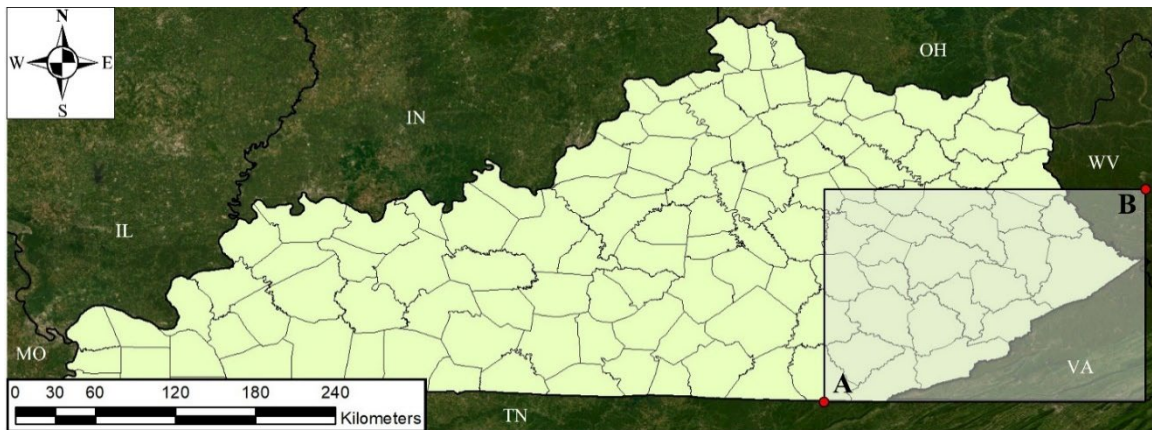


## 4.2.1 LIS EXECUTABLES AND WORKFLOW

The LIS framework is composed of a subset of three executables: (1) The Land Data Toolkit (LDT), (2) The Land Information System (LIS) core, and (3) The Land surface Verification Toolkit (LVT). These executables, as with the overarching LIS framework, are designed for development and implementation within a Linux-based operating system. Instructions and resources for development of these executables, and overall implementation of the LIS framework, are readily available to the public via the NASA-LIS GitHub repository (<https://github.com/NASA-LIS/LISF>).

### 4.2.1.1 LAND DATA TOOLKIT (LDT) EXECUTABLE

The Land Data Toolkit provides preprocessed data for subsequent Land Surface Model (LSM) runs within the LIS core. This preprocessed data is acquired within a user defined domain over which associated soil moisture estimates will be later obtained. Through this study, a domain, with a spatial resolution of 0.01-degree ( $\sim 1$  km), was established over an area of Eastern Kentucky. Eastern Kentucky was identified as an area of interest herein due to frequent landslide occurrence in the region (Crawford et al., 2021). Figure 4-1 shows the established domain over Eastern Kentucky.



*Figure 4-1: LIS Domain Over Eastern Kentucky*

The Eastern Kentucky domain shown in Figure 4-1 extends from Point A (36.58, -84.11) to Point B (-81.96, 38.00). The preprocessed data needed for LSM modeling includes the following physical parameters: (1) Land cover/vegetation, (2) Land/water masks, (3) soil texture information, and (4) topography (Arsenault et al., 2018). LDT preprocesses these physical parameters for use in subsequent LSM modeling within the LIS core. The sources and resolutions of the LDT preprocessed physical parameters are shown in Table 4-1.

Table 4-1: LDT Preprocessed Physical Parameter Information

Geomorphic Parameter	Source	Available Resolution
Land Cover/Vegetation	USGS	30 arcsecond (~0.01 degree)
Land/Water Masks	MODIS	
Soil Texture	STATSGO_FAO	
Topography	SRTM	

Land cover and vegetation classifications were obtained through the USGS Earth Resources Observation and Science (EROS) Center (<https://www.usgs.gov/centers/eros>). Land and water mask information was obtained via the Moderate Resolution Imaging Spectroradiometer instrument aboard the Aqua and Terra Earth Observing Satellite (EOS) Earth missions. The soil texture map is a global hybrid soil map derived from the NRCS State Soil Geographic (STATSGO) and the Food and Agriculture Organization (FAO) of the United Nations Educational, Scientific, and Cultural Organization (UNESCO) soil maps. The STATSGO\_FAO map is produced by the National Center for Atmospheric Research (NCAR) and is developed as a 16-texture two-horizon soil system (Horizon 1: 0 – 30 cm, Horizon 2: 30 – 100 cm). The final physical parameter, topography, was obtained via the NASA Shuttle Radar Topography Mission (SRTM) (SRTMGL30 V021, 2023). All LDT preprocessed geomorphologic parameters shown in Table 4-1 are available at a fine resolution of 30 arcseconds (~0.01 degree).

#### 4.2.1.2 LAND INFORMATION SYSTEM (LIS) CORE EXECUTABLE

The Land Information System is the core executable within the LIS framework. User selected LSM model runs are conducted within this core executable. These LSMs require both geomorphologic data (extracted and preprocessed by LDT) *and* climatologic forcing data. Forcing data via the North American Land Data Assimilation System (NLDAS-2) was utilized herein. Once provided with both geomorphic and climatologic data, the LIS core effectively becomes an LSM. Additionally, the LIS core allows for assimilation of satellite-based and LSM-based soil moisture retrievals. The LIS core has the following options for assimilation techniques: Direct Insertion, Ensemble Kalman Filter (EnKF), and the Ensemble Kalman Smoother (EnKS). The EnKF was selected as the assimilation technique based upon various past research that has shown successful assimilation of soil moisture products through EnKF assimilation (Fertig et al., 2009; Baldwin et al., 2017).

#### **4.2.1.3 LAND VERIFICATION TOOLKIT (LVT) EXECUTABLE**

The Land surface Verification Toolkit is the final executable within the LIS framework. The LVT serves as a post processor for that of the LIS core. LVT was designed as a means to provide an integrated workflow whereas verifications and validations of LIS-based estimates can be conducted (Kumar et al., 2012). LVT can be utilized to directly compare LIS estimates to that of either in-situ or satellite-based retrievals of similar data. LVT was primarily utilized as a means for data retrieval for use in validations of LIS estimates as compared to in-situ data and for later use within subsequent landslide hazard analyses.

#### **4.2.2 LIS LAND SURFACE MODEL AND ASSIMILATED PRODUCT**

This study made use of the LIS framework to retrieve spatiotemporal estimates of soil moisture across the domain established in Figure 4-1. For these retrievals the chosen LSM was that of the NOAH 3.6 LSM. For data assimilation, estimates of soil moisture via NOAH 3.6 were assimilated with that of SMAP soil moisture retrievals. The assimilated moisture retrievals were then validated using direct comparisons to in-situ measured moisture at monitoring sites within the investigated domain.

##### **4.2.2.1 NOAH 3.6 LSM**

The NOAH 3.6 LSM, available to researchers within the LIS-core, was used to obtain desired spatiotemporal estimates of soil moisture. As the NOAH 3.6 LSM was utilized within the LIS framework, this implies that the LSM soil moisture estimates were at the same resolution as that of the established LIS domain (i.e., 0.01 degree or ~ 1 km). Additionally, the NOAH 3.6 LSM has a four-layer configuration with default increasing depths of 100 mm, 300 mm, 600 mm, and 1000 mm. However, these four depths were altered through this study to be 100 mm, 200 mm, 500 mm, and 1000 mm. This alteration was conducted to provide better alignment of LSM soil moisture estimates with that of in-situ measurements during the LIS-based soil moisture validation process of this study. NOAH 3.6 provides estimates of soil moisture at these established layers through implementation of the Richard's equation. The NOAH 3.6 model also accounts for both evaporation, through a diurnally dependent Penman potential evaporation model, and vegetation, through a primitive canopy model (Shrestha et al., 2020).

##### **4.2.2.2 SATELLITE-BASED ASSIMILATION PRODUCT**

Soil moisture retrievals utilized throughout this study became a function of both NOAH 3.6 estimates and globally available satellite-based retrievals of soil moisture. This was conducted through EnKF assimilation within the LIS-core of the LIS framework. Herein, NASA 9 km Soil

Moisture Active Passive (SMAP) L3SMP\_E surface moisture retrievals were utilized for assimilation purposes. The L3SMP\_E product is an enhanced Level 3 radiometer passive soil moisture product, providing daily composite estimates of surface soil moisture (0-5 cm). However, before assimilation of L3SMP\_E product could be conducted, alterations to the LIS source code were required. It was observed that quality control flags were restricting assimilation of L3SMP\_E due to the product having uncertain quality over heavily vegetated areas. These flags restricted assimilation of L3SMP\_E over these areas which in turn limited the availability of the desired assimilated soil moisture product. These alterations to the LIS source code, which allowed for assimilation of L3SMP\_E with uncertain quality, are fully outlined in Francis and Bryson (2023). It is understood through this study that this source code alteration implies that retrievals of uncertain quality are now allowed for assimilation. However, this implication is noted as acceptable due to the fact that the alteration now allows for assimilated soil moisture to be readily available across the entire domain.

#### 4.2.2.3 GROUND BASED MOISTURE VALIDATION SITES

Validation of LIS-based soil moisture estimates was conducted by directly comparing said estimates to that of soil moisture measured at three existing soil moisture monitoring stations within the investigated domain. These in-situ stations are a subset of the Kentucky Mesonet, maintained by Western Kentucky University (WKU). These stations provide soil moisture measurements at depths of 102 mm, 203 mm, 508 mm, and 1016 mm. These depths are comparable to the adjusted layers of the NOAA 3.6 LSM (100, 200, 500, and 1000 mm). The locations of the three validation sites are shown in Figure 4-2.

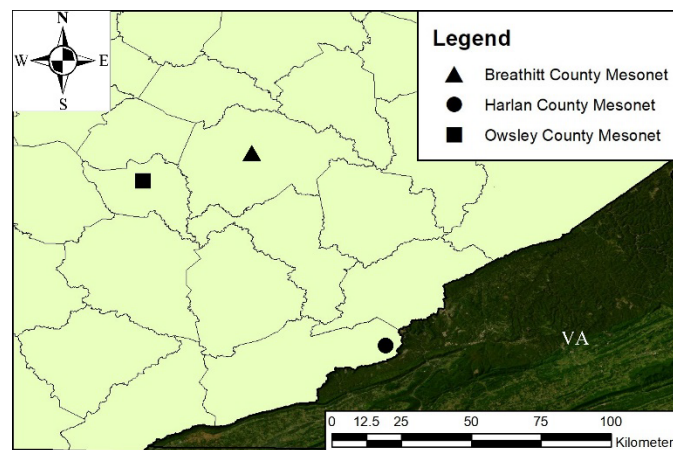


Figure 4-2: Ground Based Validation Sites within LIS Domain

#### **4.2.3 LIS NOAH 3.6 ASSIMILATION RUNS**

Spatiotemporal estimates of soil moisture across the defined Eastern Kentucky domain were retrieved through LIS-based assimilation of NOAH 3.6 and SMAP L3SMP\_E estimates. However, assimilation first requires a rescaling of L3SMP\_E to that of NOAH 3.6 modeled data. This rescaling is performed through cumulative density function (CDF) matching within the LIS-core. CDF matching allows for the LIS-core to bias correct and reduce differences in observed and modeled data (Arsenault et al., 2018). This implies that a CDF is required for both L3SMP\_E and NOAH 3.6 modeled moisture. As historic L3SMP\_E data was used for assimilation, a CDF was constructed based on the historic retrievals. To obtain a CDF for NOAH 3.6 modeled moisture, an initial run of LIS, whereas only NOAH 3.6 was used to provide estimates, was conducted. This initial run, which ran from 01/01/2013 – 05/01/2023, was used to obtain a cumulative density function (CDF) for NOAH 3.6 estimates.

With CDFs for both products (NOAH 3.6 and L3SMP\_E retrievals) now available, the study then moved to conducting assimilation of NOAH and SMAP moisture products. Data assimilation was conducted over the date range of 01/01/2017 – 05/01/2023. This range was selected to provide a temporally large dataset of soil moisture for use in subsequent landslide hazard analyses. Figure 4-3 shows assimilated SMAP and NOAH 3.6 soil moisture estimates as compared to in-situ moisture retrieved from the Breathitt County Mesonet monitoring station.

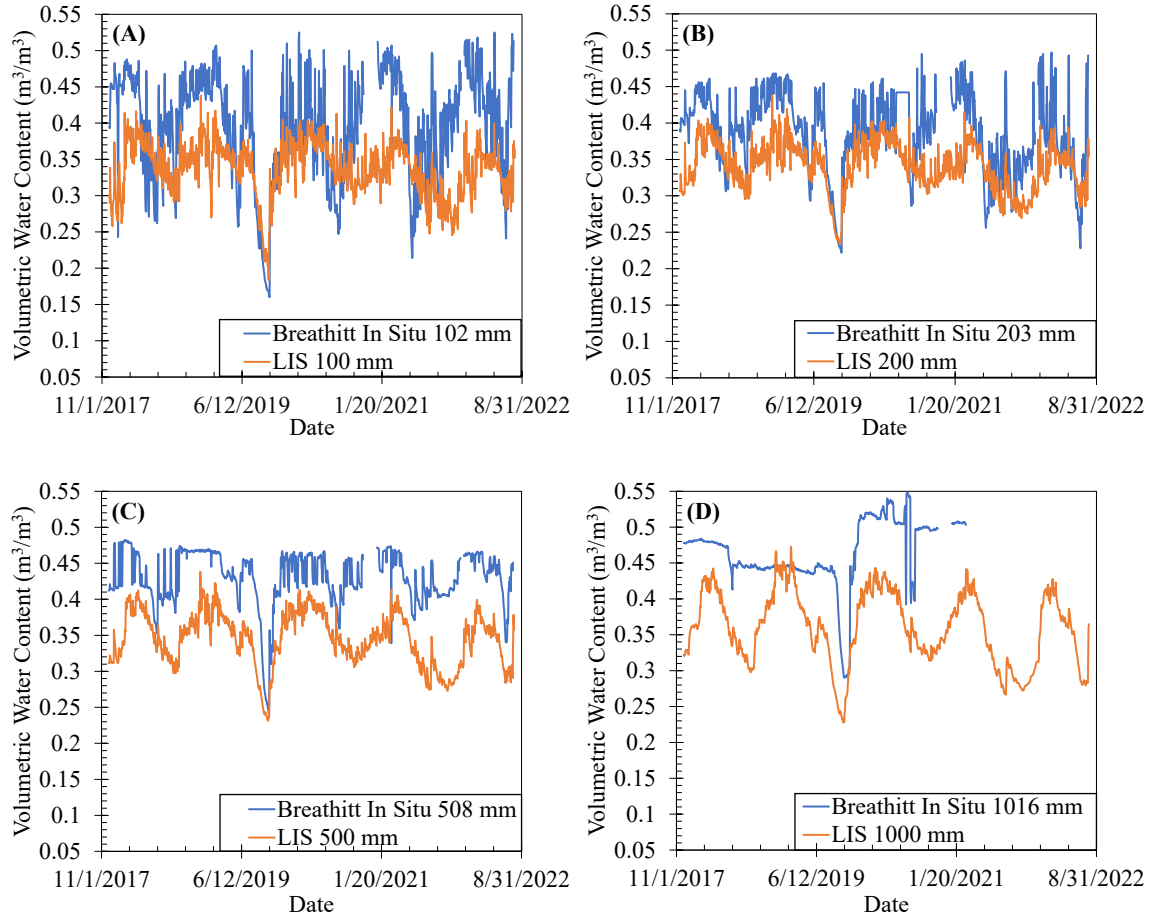


Figure 4-3: Comparison of Assimilated SMAP and NOAA 3.6 Estimated Moisture to Mesonet Breathitt County In-Situ Moisture at: (A) 100 mm, (B) 200 mm, (C) 500 mm, and (D) 1000 mm

As seen in Figure 4-3, LIS-based moisture estimates performed well in capturing the overall seasonal trends (wetting and drying) as compared to in-situ measured data. Additionally, LIS-based moisture performed exceptionally well at capturing the large decrease in soil moisture visible at all depths approximately four months after 06/12/2019. However, it can be observed that discrepancies exist between LIS moisture and that of in-situ measurements. The most observable discrepancy is that LIS appears to underpredict moisture as compared to in-situ conditions. However, after further investigation, it appears that in-situ moisture levels tend to return to, or stay very near to, local maxima. It can be observed that the moisture values rapidly increase, or remain near, 0.5, 0.47, and 0.48  $\text{m}^3/\text{m}^3$  at depths of 102, 203, and 508 mm, respectively. It can also be observed that moisture levels stay relatively constant for extended periods of time at 1016 mm in depth. These characteristics of the data lead to speculations that the moisture sensors at these respective depths are experiencing pooling of water, which causes the moisture levels to remain at or rapidly climb towards observable maxima. This phenomena was

observable at all monitoring sites shown in Figure 4-2. Considering this speculation, and the observation of LIS capturing overall wetting and drying trends, the overall performance of LIS estimates is thought to be quite suitable for implementation of LIS-based spatiotemporal estimates of soil moisture in subsequent landslide hazard analyses.

#### 4.3 SPATIAL RETRIEVALS OF GEOMORPHOLOGIC DATA

Throughout this study, heightened soil moisture levels, typically those seen soon after a rainfall event, were assumed capable of serving as precursors for incipient conditions indicative of landslide occurrence. To support this assumption, landslide hazard analyses were conducted that made use of both soil moisture and area specific geomorphologic retrievals. A majority of the required geomorphic data required by these analyses were retrieved via Web Soil Survey (WSS) (<https://websoilsurvey.nrcs.usda.gov/app/>). WSS is an online soil parameter database operated by the USDA National Resources Conservation Service (NRCS). WSS provides digitized soil survey maps across the contiguous United States that are available at very fine resolutions (100 – 300 m). Herein, WSS data was retrieved through implementation of the ArcMap addon, SoilViewer. SoilViewer is a tool which allows the user to import digitized soil survey maps into a spatial environment. Once soil data is imported, SoilViewer can be used to obtain estimates of geomorphic data at a given desired depth (1000 mm throughout this study). In addition to soil data, slope angles, which have significant influence upon landslide susceptibility (Çellek, 2020), were also obtained. These slopes were retrieved through analysis of Digital Elevation Maps (DEMs) that are readily available through ArcMap. Table 4-2 shows the geomorphologic data throughout this study as well as the source from which the data was obtained. The data shown here, as well as data for subsequent landslide hazard analyses, can be found at the UKnowledge Civil Engineering data repository, <https://doi.org/10.13023/ce7.2023.dat>.

*Table 4-2: Soil Geomorphologic and Slope Characteristic Data Sources*

Characteristic Data	Obtained Via
Soil Texture (% Sand, % Silt, % Clay)	WSS  ( <a href="https://websoilsurvey.nrcs.usda.gov/app/">https://websoilsurvey.nrcs.usda.gov/app/</a> )
Bulk Density of Soil (g/cm <sup>3</sup> )	
Plasticity Index (%)	
Slope Angle (degree)	ArcMap DEM Esri. "Terrain: Slope in Degrees" [basemap]. No Scale



#### 4.3.1 SPATIAL ESTIMATES OF DEPTH TO BEDROCK

Throughout this work, it was assumed that the bedrock underlying a soil layer served as the interface along which a landslide would translate down a hillslope. This has been observed by past research that has shown a majority of landslides within Eastern Kentucky to occur in shallow ( $< 3$  m) colluvial soils (Crawford et al., 2021). Therefore, Depth to Bedrock (DTB), became an important variable within the conducted landslide hazard analyses. Estimates of DTB are available through the WSS tool. However, it was observed that WSS only provides DTB at up to 2.01 m. Anything beyond 2.01 m is simply classified as “ $> 2.01$  m” within the WSS database. To remedy this lack of availability with depth, a novel DTB map, which allowed for spatial retrievals of DTB across the defined LIS domain, was developed. The Kentucky Transportation Cabinet (KYTC) provides an online database of borehole information for projects beginning in 2007 (<https://kgs.uky.edu/kgsmap/kytclinks.asp>). This database was utilized to obtain DTB data from more than 3,000 boreholes across the investigated domain. To transform DTB from these boreholes into a spatial map, Kriging interpolation within ArcMap was implemented. Kriging is a geostatistical interpolation approach that is commonly implemented for use in estimations of spatial distribution of a given variable (Belkhirri et al., 2020). The distribution of KYTC boreholes across the domain, as well as the developed DTB map, can be seen in Figure 4-4.

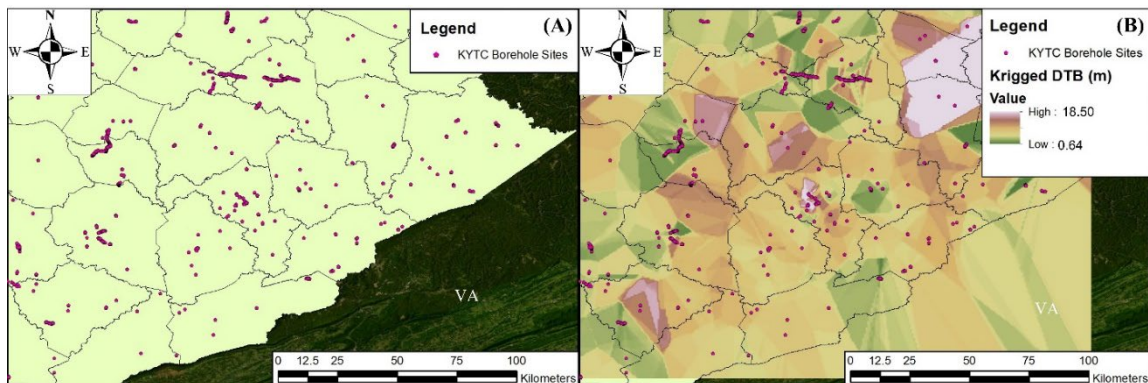


Figure 4-4: (A) Locations of KYTC Boreholes with DTB Data Across Domain and (B) Novel DTB Map Developed Through Kriging of DTB Data from 3000 Boreholes

As seen in Figure 4-4A, the borehole sites used for DTB data retrieval are very well spatially distributed across the investigated domain. Areas that show clustering of borehole sites correlate to roadway projects which required borehole drilling to be conducted along a given stretch of developed/proposed roadway. Figure 4-4B shows the developed DTB map, which shows spatial estimates of DTB to now range from approximately 0.64 m to 18.50 m. Therefore, the developed DTB map provides spatial estimates of DTB at greater depths than were readily available via the



WSS tool. Spatial DTB estimates were retrieved from this developed map for use in subsequent landslide hazard analyses.

#### **4.3.2 SPATIAL ESTIMATES OF FRICTION ANGLE**

A key physical parameter of the conducted landslide hazard analyses was that of friction angle ( $\phi'$ ). However, it was observed that spatial estimates of friction angle in a soil were lacking across the investigated domain. Past research developed an approach which made use of Artificial Neural Network (ANN) machine learning to relate soil textures (% Sand, % Silt, and % Clay) to that of the friction angle within a soil (Francis and Bryson, 2023). As WSS provides spatial estimates of soil texture, this machine learning workflow enabled the acquisition of spatial estimates of friction angle throughout the domain. Therefore, the Francis and Bryson (2023) ANN machine learning workflow was incorporated into this study to acquire estimates of friction angle at an investigate location.

#### **4.4 MACHINE LEARNING-BASED LANDSLIDE SUSCEPTIBILITY**

This study focused heavily upon landslide hazard analyses as a function of soil moisture acting as a precursor indicative of incipient failure conditions. However, this study also investigated landslide susceptibility analyses. The key difference between these two analyses is as follows: Susceptibility analyses are traditionally performed using static factors to distinguish between areas where landslide occurrence is possible and areas where landslide occurrence is unlikely. Hazard analyses make use of dynamic factors, such as soil moisture and related variables, to show when and where landslide occurrence is likely. Therefore, the landslide analyses conducted throughout this study became a coupled approach between susceptibility and hazard analyses. The coupled susceptibility analyses were conducted through implementation of a machine learning classification approach. Additionally, the landslide susceptibility analyses herein were developed as a function of normalized difference vegetation index (NDVI). Addition of NDVI added a temporal element to an otherwise static analysis.

##### **4.4.1 NDVI ACQUISITION FOR LANDSLIDE SUSCEPTIBILITY**

Vegetation on a hillslope has been shown to have a role in the occurrence of landslides. When vegetation decays, or the rooting system weakens, a weakening of the soil previously reinforced by healthy roots can occur. Additionally, vegetation on a hill slope can apply an additional surcharge on the slope due to moisture retainment within said vegetation. (Guo et al., 2020; Phillips et al., 2021). Due to these potential roles in landslide occurrence, vegetation levels were incorporated into the landslide susceptibility analyses conducted during this work. Spatial

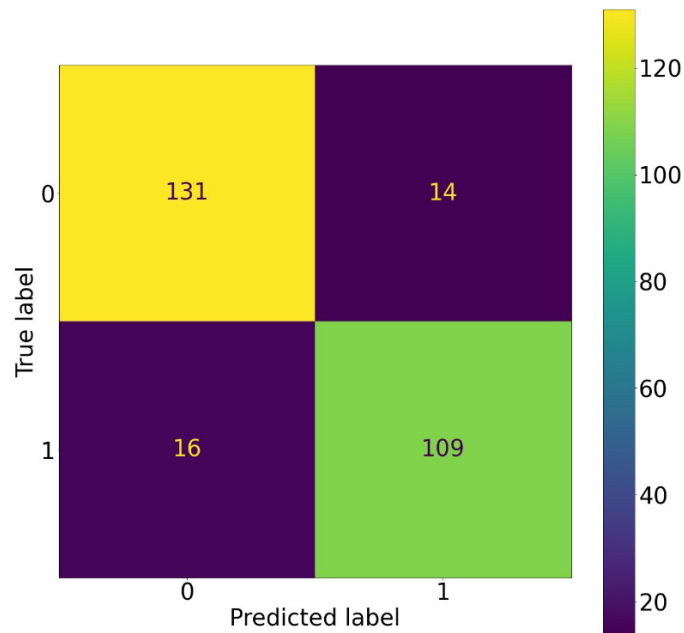
estimates of vegetation across the investigated domain were obtained via the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the EOS Aqua and Terra satellites. These estimates are in the form of Normalized Difference Vegetation Indices (NDVI) and are readily available in a 16-day (i.e., consistent across the 16-day period) product at a resolution of 250 m. The retrieved products from the Aqua and Terra satellite were that of the MYD13Q1.006 and MOD13Q1.006, respectively. Acquisition of NDVI estimates over investigated landslides was conducted through utilization of an Application Programming Interface (API). This API was run within a Python environment whereas latitude, longitude, and date of occurrence for each landslide were utilized to obtain respective NDVI estimates over all included landslides.

#### **4.4.2 LOGISTIC REGRESSION SUSCEPTIBILITY MODEL**

Regional landslide susceptibility maps serve as a response to potential threats posed by landslide occurrence. These maps are constructed through determination of spatial association between various influencing factors (location, geomorphology, vegetation, etc.) and existing landslides. Determination of these spatial associations is typically done through multivariate statistical approaches, such as logistic regression classification. (Chen et al., 2018; Sujatha and Sridhar, 2021). Logistic regression is a statistical modeling approach which is widely used in the parameterization of non-linear relationships between dependent (i.e., landslide occurrences) and independent variables (i.e., geomorphic data). Additionally, logistic regression is widely used when the dependent variable is binary in nature (Sujatha and Sridhar, 2021). Logistic regression classification of Kentucky landslide susceptibility has been previously investigated (Crawford et al., 2021; Dashbold et al., 2023). However, the logistic regression susceptibility approach developed through this study differs from past studies through incorporation of NDVI as an independent factor of landslide occurrence. This incorporation of a temporally variable factor allows for susceptibility through this study to vary with time rather than remain a static tool.

The logistic regression susceptibility model was developed within a Python coding environment during this work. The training dataset for this model consisted of geomorphic and NDVI data from 675 known landslide locations across the Eastern Kentucky domain. These landslides, retrieved via the KGS landslide inventory have dates of occurrence ranging from 01/01/2004 to the most recent recorded occurrence on 07/28/2022. Equivalent site data over 675 randomly selected locations assumed to have no landslide occurrences were also obtained for training. This ratio of occurrence to non-occurrence was selected due to observed sensitivity of the model to unbalanced ratios (i.e., anything other than 1:1). Herein, occurrences were labelled as 1, while non-occurrences were labeled as 0. Data from both landslide and non-occurrence sites were

shuffled together to prevent potential model learning from trends within the dataset. To train the model for use in susceptibility classifications, an 80-20 test/train split was utilized. During training, it was observed that implementation of the liblinear solver and the L1 penalty resulted in slightly higher accuracy as compared to the default parameters. Through training and development of the susceptibility model, an 89% success rate in susceptibility classifications was achieved. In addition to accuracy, model performance was gauged through inspection of the confusion matrix. Confusion matrices are powerful tools in investigation of classification ability as the matrix shows the amount of correct and incorrect classifications. The confusion matrix for the developed logistic regression model is shown in Figure 4-5.



*Figure 4-5: Confusion Matrix for Trained Logistic Regression Susceptibility Classifier*

As can be seen through the confusion matrix in Figure 4-5, the logistic regression model performed adequately in classification of susceptibility. As 20% of data was retained for testing of the trained model, this implies that 270 data points, consisting of both occurrences (1) and non-occurrences (0), were used for testing. The trained model successfully classified non-occurrences 131 times, occurrences 109 times, and misclassified only 30 test points. Therefore, the logistic regression classification model has been shown to perform well in susceptibility classifications as investigated through this work.

#### 4.4.3 TEMPORAL VARIATION OF SUSCEPTIBILITY

As discussed, NDVI retrievals were included within the developed landslide susceptibility model in order to allow for static susceptibility to become temporally variable. The intent behind this addition was to add a temporal element to an otherwise static analysis. An example of 1 km spatial resolution susceptibility over Harlan County, Kentucky can be seen in Figure 4-6.

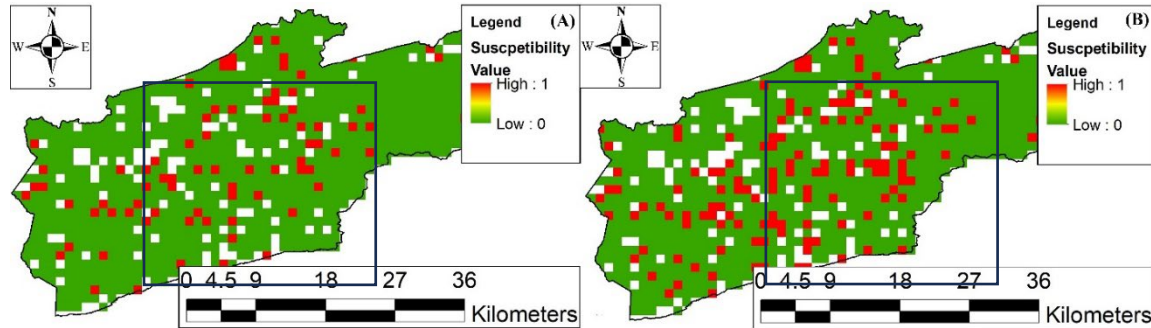


Figure 4-6: Susceptibility Over Harlan County, Kentucky Using NDVI from: (A) 02/10/2023 to 02/18/2023 and (B) 03/30/2023 to 04/07/2023

As can be seen through Figure 4-6, the inclusion of temporally varying NDVI data allows for susceptibility to change with time. Figure 4-6A shows susceptibility using NDVI data retrieved from February, 2023, while Figure 4-6B shows susceptibility using NDVI retrieved from April, 2023. Areas in these figures that are shown as neither susceptible nor non-susceptible indicate areas where WSS data was unavailable for retrieval. It can be observed that areas not shown as susceptible (i.e., as a value of 1) in February shift to become susceptible in April, 2023. The greatest observable changes occur within the bounded regions within the Figure. This again shows that the developed susceptibility model is now a temporally varied model rather than a static model, yielding a more powerful susceptibility tool.

#### 4.4.4 VALIDATION OF SUSCEPTIBILITY MODEL

In an effort to validate performance of the developed logistic regression model, susceptibility classifications were conducted over four known landslides. These landslides were selected based upon the confidence recorded within the KGS inventory. The inventory makes use of a confidence ranking system which ranges from 1-8 for each slide within the inventory. A confidence of 1 indicates “possible landslide occurred in the area” while a confidence of 8 indicates a “high confidence that the nature and extent of the landslide is well characterized”. Based on this ranking system, landslides 6456, 8596, 9677, and 10757 were selected for investigation. To validate the developed susceptibility classifier, data over these four landslides was obtained at a resolution of 100 m. This resolution was selected primarily to enable the

accurate capture of slope angles at, or near, the validation sites. The susceptibility via the developed classifier over these four landslides is shown in Figure 4-7.

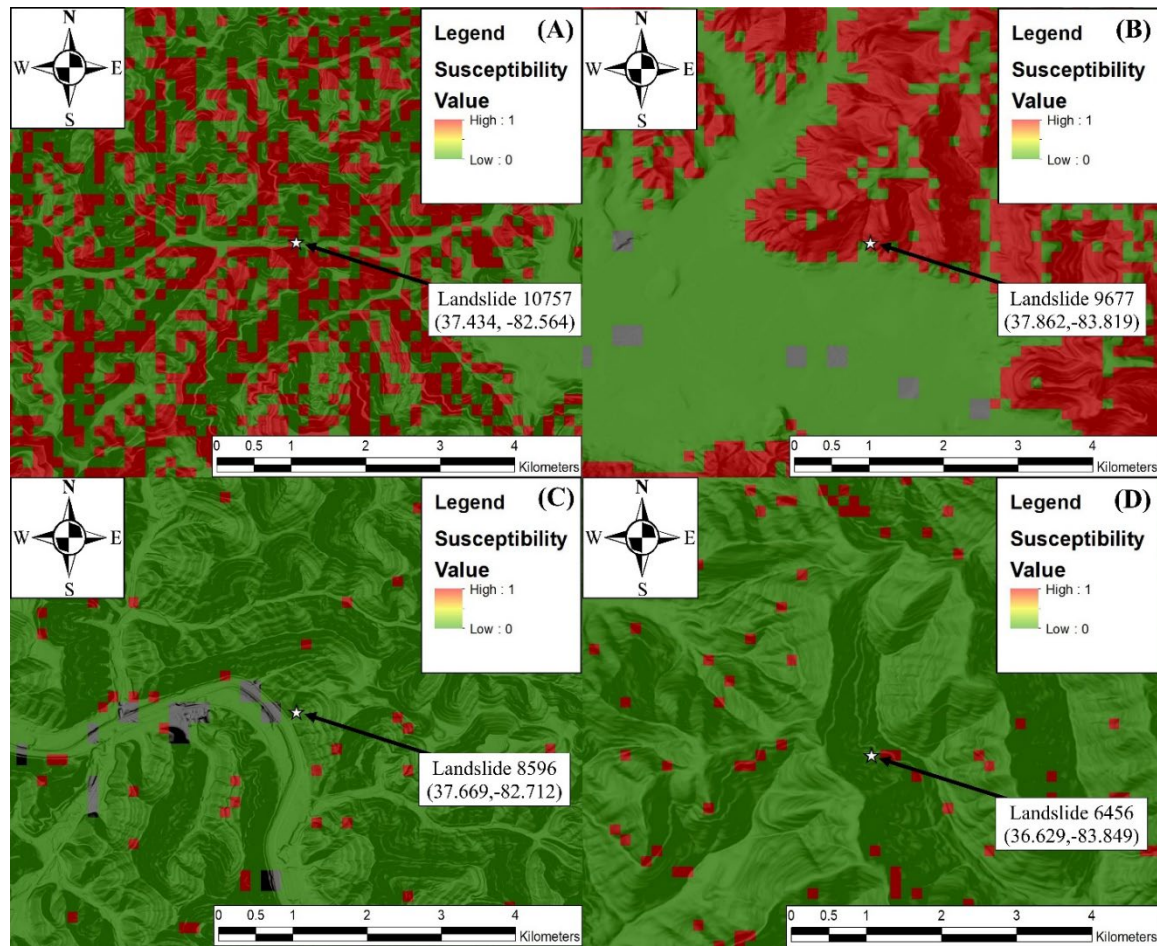


Figure 4-7: Susceptibility Over: (A) Landslide 10757, Occurred 03/17/2021, (B) Landslide 9677, Occurred 06/17/2019, (C) Landslide 8596, Reported 01/04/2019, and (D) Landslide 6456, Occurred 02/10/2018

As seen in Figure 4-7, the susceptibility classifications perform well at aligning with known landslide occurrences. Landslides 10757 and 9677 (Figure 4-7A and 4-7B) are indicated to have occurred in an area deemed susceptible for landslide occurrence. Landslide 6456 (Figure 4-7D) occurred very near to an area that is deemed susceptible. However, it has been observed that runout locations of landslides are sometimes recorded within the KGS inventory, rather than the initiation point of the slide. Therefore, it is speculated that Landslide 6456 initiated in the nearby susceptible cell and then ran out to its geographically recorded location. Finally, landslide 8596 (Figure 4-7C) can be observed to have no susceptible areas nearby. The speculation as to the cause of this lies in the date associated with the landslide occurrence. The KGS inventory maintains both reported and failure dates, whereas reported dates correspond to the date the landslide was observed, and failure dates correspond to actual times of failure. Landslide 8596

was only recorded as having a reported date, and not a known failure date. Therefore, it is speculated that this slide occurred earlier than its reported date and susceptibility at this earlier time would align better with the occurrence. This discrepancy will be investigated further in the following landslide hazard analyses. Regardless, the performance of the susceptibility classifier at indicating areas that align well with known slides leads to the satisfactory validation of the model.

#### 4.5 HISTORIC LANDSLIDE HAZARD ANALYSES

It can be observed that the landslide susceptibility analyses shown in Section 4 of this paper identify large areas to be susceptible whereas a landslide may occur in only a subset of this area. This is again due to the nature of susceptibility analyses, which generally only relate site specific characteristics to the likelihood of landslide occurrence. However, landslide hazard analyses make use of dynamic factors, such as soil moisture and related variables, to further investigate when and where landslide occurrence is likely. Therefore, susceptibility analyses indicate areas where landslides are possible, while hazard analyses indicate when landslides are possible. The hazard analyses conducted throughout this work make use of soil moisture retrieved over historic landslide locations in conjunction with an infinite slope stability analysis to determine when a landslide is likely to occur.

##### 4.5.1 HYDROLOGIC VARIABLES VIA LIS-BASED MOISTURE RETRIEVALS

The infinite slope stability equation used to conduct landslide hazard analyses, through stability analysis, was that of the Lu and Godt infinite slope equation. An important variable in these analyses was that of suction stress. Suction stress is a function of matric suction. The decrease of matric suction is considered one of the most important factors in initiation, through a loss in shear strength, of rainfall-induced landslides (Bittelli et al., 2012; Mauri et al., 2022). Therefore, as a function of matric suction, suction stress also becomes a controlling factor in the initiation of a rainfall-induced landslide. Suction stress is given as follows:

$$\sigma^s = S_e s \quad (4-1)$$

where,  $S_e$  = effective degree of saturation =  $(\theta - \theta_r)/(\theta_s - \theta_r)$ ;  $\theta$  = volumetric water content (VWC) (1000 mm VWC estimated through LIS-based NOAH 3.6 LSM and SMAP assimilations);  $\theta_r$  = residual VWC;  $\theta_s$  = saturated VWC;  $s$  = matric suction =  $(u_a - u_w)$ ;  $u_a$  = pore air pressure;  $u_w$  = pore water pressure; and  $\sigma^s$  = suction stress (kPa).

As seen in Equation 4-1, suction stress is calculated as a function of both effective degree of saturation ( $S_e$ ) and matric suction ( $s$ ).  $S_e$  was able to be estimated through soil moisture estimates retrieved via LIS. However, measurements of matric suction were not observed to be readily available at sites where infinite slope stability analyses were conducted. To remedy this lack of data availability, the soil water characteristic curve (SWCC) model, established by van Genuchten (1980), was manipulated to obtain estimates of matric suction. The manipulated van Genuchten (1980) SWCC equation for acquisition of matric suction is as follows:

$$s = \frac{1}{\alpha} \left[ \left( \frac{1}{S_e} \right)^{1/m} - 1 \right]^{1/n} \quad (4-2)$$

where,  $s$  = matric suction (kPa);  $S_e$  = effective degree of saturation; and  $\alpha, n, m$  = SWCC fitting parameters that reflect the air entry value, the slope at the inflection point of the SWCC, and the curvature of the SWCC near the residual point, respectively.

The fitting parameters,  $\alpha, n$ , and  $m$ , were obtained through use of the online tool, Handbook60 (<https://www.handbook60.org/rosetta/>). Handbook60 makes use of the Schaap et al. (2001) pedotransfer functions (PTFs) to estimate  $\alpha, n$ , and  $\theta_r$ . The inputs for these PTFs are the soil texture information (% Sand, % Silt, and % Clay) and bulk density from a given landslide location. Spatial estimates of these data, shown in Table 4-2 of this paper, were again retrieved through implementation of the WSS tool. Handbook60 provides estimates of  $\theta_r$ ,  $\theta_s$ ,  $\alpha$ , and  $n$ . However, the Handbook60 estimate of  $\theta_s$  was not used within these slope stability analyses. Instead, the maximum VWC retrieved from LIS-based moisture estimates were considered to be reflective of  $\theta_s$  at a given analysis site. Additionally, estimates of the fitting parameter,  $m$ , were not readily available through Handbook60. To obtain this parameter, the Mualem (1976) approximation,  $m = 1 - 1/n$ , was used.

As these parameters were estimated through WSS-based retrievals, with resolutions of 100 – 300 m, this in turn implies that these parameters were also at these very fine resolutions. As matric suction, and in turn suction stress, are calculated through these fitting parameters, this also implies that matric suction and suction stress are also obtainable at these fine resolutions. Therefore, as the Lu and Godt infinite slope equation is a function of suction stress, the slope stability analyses can also be conducted at these fine resolutions.

#### 4.5.2 INFINITE SLOPE STABILITY

The slope stability analyses conducted through this work made use of the Lu and Godt infinite slope stability equation. This equation is as follows:

$$FS = \frac{\tan(\phi')}{\tan(\beta)} + \frac{2c'}{\gamma H_{ss} \sin(2\beta)} + r_u [\tan(\beta) + \cot(\beta)] \tan(\phi') \quad (4-3)$$

where,  $FS$  = factor of safety;  $\phi'$  = friction angle (estimated using Francis and Bryson (2023) ANN model);  $c'$  = effective soil cohesion (assumed negligible herein due to soil assumed normally consolidated);  $\beta$  = slope angle (degree) (estimated using a 50 ft buffer around each slide);  $\gamma$  = soil unit weight (assumed to be 17.28 kN/m<sup>3</sup>);  $H_{ss}$  = depth to bedrock (estimated via DTB map);  $r_u$  = pore pressure ratio =  $\sigma^s / \gamma H_{ss}$ ;  $\sigma^s$  = suction stress (kPa).

It is necessary to note that two deviations from the original Lu and Godt equation were conducted through this study. The first is simply that Lu and Godt defined  $H_{ss}$  as the height to the slip surface within the soil layer. As the investigated landslides herein were observed to be shallow colluvial slides,  $H_{ss}$  was instead defined to be the height of the soil layer upon the bedrock. The second deviation pertains to the matric suction variable of Equation 4-3,

$\{r_u [\tan(\beta) + \cot(\beta)] \tan(\phi')\}$ , which was assumed positive herein. Lu and Godt defined this term to be negative and also assumed suction stress to be negative. Therefore, a negative multiplied by a negative yields a positive value. The assumption is that matric suction within the soil leads to an increased factor of safety. This again aligns with the belief that decreases of matric suction are indicative of reduced strength within the soil, leading to initiation of rainfall-induced landslides.

#### 4.5.3 LIS-MOISTURE BASED LANDSLIDE HAZARD ANALYSES

As discussed, suction stress ( $\sigma^s$ ) is an important variable within the Lu and Godt infinite slope stability equation. Herein, suction stress was estimated using LIS-based moisture estimates and SWCC fitting parameters at a depth of 1000 mm. This depth was selected for investigation as it was the most comparable to  $H_{ss}$  of the available LIS-based moisture depths. However, it is worth noting that if very large DTB (i.e., >20 m) were retrieved via the developed DTB map, a value of 3 m was substituted based on observations of shallow soil by Crawford et al., (2021). Herein, suction stress, and subsequent infinite slope analyses, became a function of LIS-based moisture.



Inclusion of temporally varying soil moisture estimates allowed for the infinite slope stability equation to yield temporally varying factors of safety. This allowed for the investigation into *when* a landslide would. The Lu and Godt infinite slope stability equation was used to conduct landslide hazard analyses at the same landslides investigated in the previously discussed susceptibility analyses. The results of these analyses as a function of LIS-based moisture and spatial retrievals of geomorphologic data, are shown in Figure 4-8.

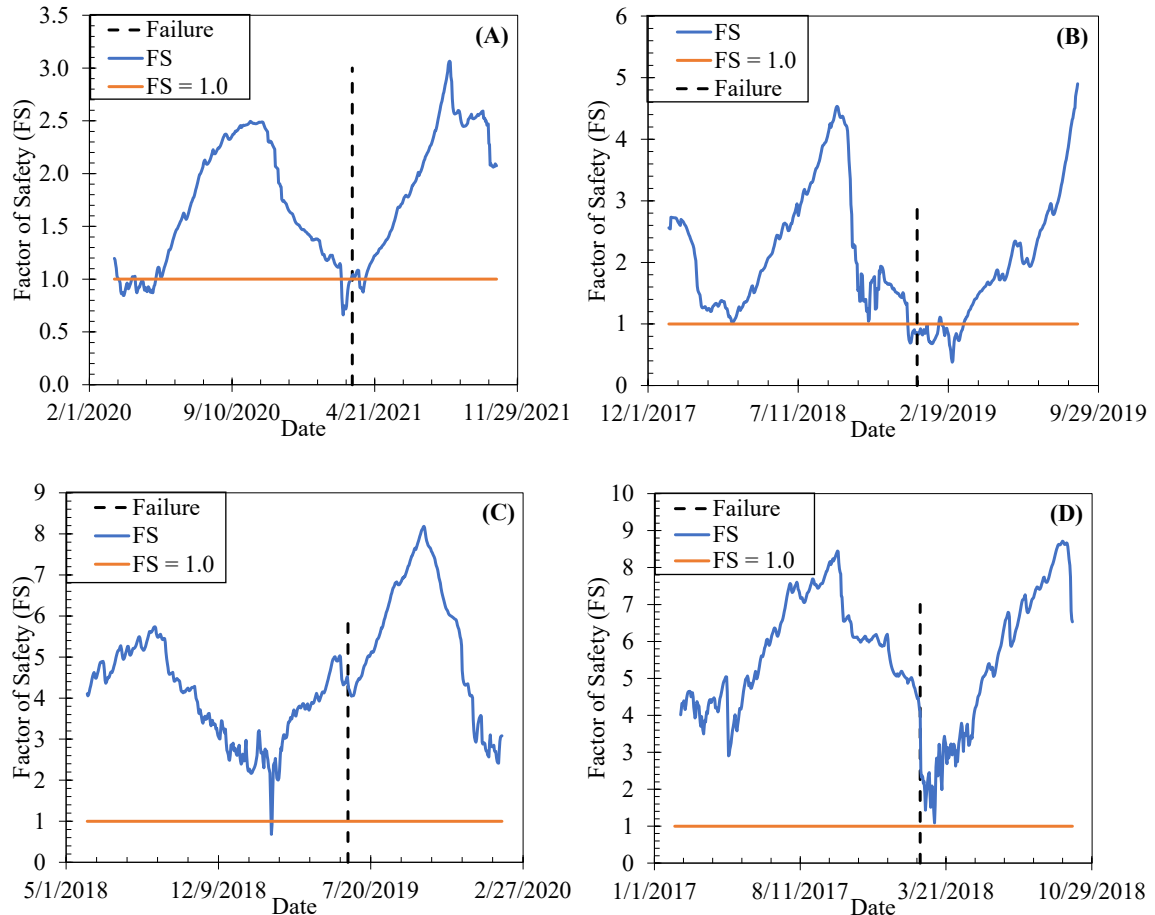


Figure 4-8: Infinite Slope Stability Analysis at: (A) Landslide 10757, Occurred 03/17/2021, (B) Landslide 8596, Reported 01/04/2019, (C) Landslide 9677, Occurred 06/17/2019, and (D) Landslide 6456, Occurred 02/10/2018

As can be seen in Figure 4-8, sharp decreases in the Factor of Safety can be observed to occur at or near the associated dates of failure for each investigated landslide. This can be seen in Figures 4-8A and 4-8D, which show the factor of safety for landslides 10757 and 6456 to be at 1.0 at or near the time of failure. However, the factor of safety in Figure 4-8C, for landslide 9677, can be observed to be far higher than a factor of safety of 1.0 (indicative of failure) at the associated date. It can also be observed that the FS decreased rapidly to below 1.0 approximately four

months before failure. It is speculated that a catastrophic failure did not occur at the time of the rapid decrease, but instead the slope was severely weakened. It appears that a restrengthening occurred after this point, with failure occurring after a second rapid drop in the factor of safety. Therefore, it is speculated that this was a compound failure which saw a severe reduction in strength and then failure with the next rapid drop in factor of safety. This same trend can be observed in the factor of safety shown in Figure 4-8B, for landslide 8596. A rapid drop in the factor of safety can be observed well before reported failure with a restrengthening occurring prior to reported failure. This similar phenomena adds confidence in the speculation discussed for landslide 9677. Additionally, it can be observed that the factor of safety in Figure 4-8B, landslide 8596, dropped before the indicated failure date. As discussed in Section 4-4, landslide 8596 was only recorded to have a reported date (a date where it was observed). A known date of occurrence for this slide is not known. Therefore, it is speculated that landslide 8596 occurred at the time of the observable drop in factor of safety and was simply not observed until a later date. Regardless, the infinite slope stability analyses within the investigated hazard analyses indicate sharp drops in factors of safety at or near recorded dates for each investigated slide. This satisfactory performance leads to confidence in further use of the Lu and Godt infinite slope stability equation for further landslide hazard analyses across the established Eastern Kentucky domain.

#### **4.6 FORECASTING OF LANDSLIDE HAZARD ANALYSES**

The previously discussed landslide hazard analyses made use of historic landslide occurrences and historic soil moisture values. However, this same workflow is utilized for near real-time analyses to develop landslide early warning systems (EWS). EWS's are used in landslide prone regions in an effort to reduce losses and damage caused by rainfall-induced landslides (Hidayat et al., 2019). Typically, EWS's make use of nowcasts as a means to provide warnings of potential landslide hazards. Nowcasts typically consider levels of precipitation indicative of landslide occurrence and indicate areas over which landslides are probable in near real-time (i.e., over the next few coming days) (Kirschbaum and Stanley, 2018). An observable shortcoming of EWS's and nowcasts is that these hazard warnings are applicable only in near real-time. A more powerful hazard analysis tool would provide a potential forecast (i.e., 3-day, 7-day, 10-day, etc.) of landslide hazards. This forecast would allow for additional preparation and/or mitigation time of those in affected areas. Herein, a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) machine learning workflow was developed as a means of hazard forecasting. It has been shown through this study that soil moisture is a factor of landslide occurrence. Therefore, the forecasting of soil moisture in turn allows for the forecasting of landslide hazards.

#### 4.6.1 LSTM OVERARCHING FRAMEWORK

Past research has seen success in the forecasting of landslide occurrences through implementation of Long Short-Term Memory (LSTM) models (Orland et al., 2020; Utomo et al, 2020; Mubashar et al., 2021; Zhang et al., 2022). However, this past research focused primarily upon landslide forecasting through either: (1) Pure forecasting of climatic conditions (rainfall, temperature, wind, etc.) that did not account for geomorphologic information, (2) Forecasting of landslide hazards as a function of rainfall, rather than the driving factor of soil moisture through matric suction, or (3) Forecasting of landslide susceptibility through climatic variability forecasting. Therefore, this study developed a novel workflow for hazard analysis through LSTM-based forecasting of soil moisture for implementation into soil moisture dependent stability analyses. The general framework of this LSTM model is shown in Figure 4-9.

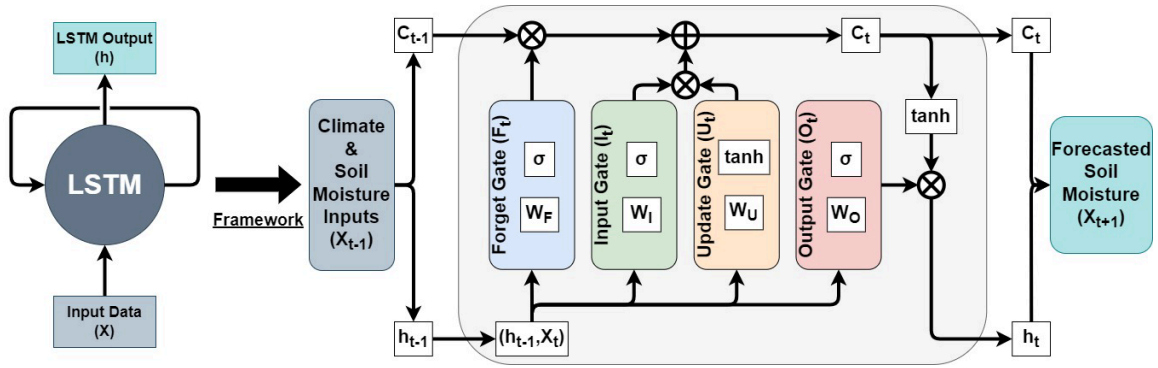
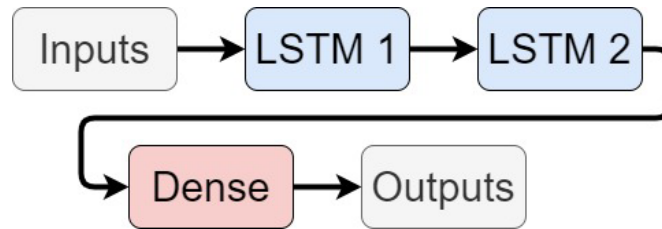


Figure 4-9: Generalized Framework of Long Short Term Memory Model

As can be seen in Figure 4-9, the LSTM framework consists of multiple “gates” (Forget, Input, Update, and Output). These gates, which are all individual neural networks, are used to update both the cell ( $C_t$ ) and hidden ( $h_t$ ) states. The cell state corresponds to the current long-term memory of the network, while the hidden state corresponds to the model outputs, or the short-term memory, at the previous point in time ( $t-1$ ). In general, data flows through these gates in the following order: Forget gate, Input and Update gates in unison, and then through the Output gate. The Forget gate, a sigmoid activated neural network, looks at new input data (i.e., climate and soil moisture data) at the current time step ( $X_t$ ) and data from the hidden state at the previous time step ( $h_{t-1}$ ). These data are used to determine which *old* components of the cell state (i.e., the current memory) should be forgotten (i.e., have less weight moving forward). Results from the Forget gate are then pointwise multiplied (indicated as  $\otimes$ ) with the previous cell state ( $C_{t-1}$ ), effectively “forgetting” less important parts of the memory. The Input and Update gates,

sigmoid and tanh (hyperbolic tangent) activated networks, respectively, then determine which *new* data to be added to the cell state. These gates again consider data from  $X_t$  and  $h_{t-1}$ . However, as opposed to the Forget gate, these gates add data deemed important to the memory cell, rather than forgetting data deemed unimportant. Results from the Input and Update gates are first pointwise multiplied with each other, and the product is then pointwise added ( $\oplus$ ) to the current memory cell ( $C_t$ ). The final step is to push the updated cell state, the previous hidden state and new input data through the Output gate. The Output gate, a sigmoid activated network, is pointwise multiplied with a tanh activated version of the current memory cell. Doing so allows for weights of irrelevant data, potentially added in the current time step, of the memory cell to be reduced. These reduced weights are then saved as the new hidden state and then used to repeat this process at the next time step. Herein, sigmoid activations yield a vector ranging from  $[0,1]$  while tanh activations yield a vector from  $[-1,1]$ . Sigmoid activations are used to update weights of features while tanh activations are used to update feature importance. Pointwise multiplication of tanh and sigmoid activated gates is done so that a vector of  $[0,1]$  is multiplied with a vector of  $[-1,1]$ . Doing so effectively enables feature importance to be converted to feature weights for an update of the current cell memory. (Dolphin, 2021).

The LSTM framework shown in Figure 4-9 represents a “vanilla” LSTM, or a singular LSTM model. However, stacked LSTMS, which were introduced by Graves et al (2013) and have been shown to outperform vanilla LSTMS with sequence prediction problems. The stacking of LSTMs yields an overall LSTM model composed of subsequent LSTM layers. Additionally, when LSTMs are stacked, this increases the depth of the model, effectively creating a deep learning model. Therefore, a stacked LSTM approach was investigated through this study. Figure 4-10 shows the layers of the stacked LSTM used herein.



*Figure 4-10: Stacked LSTM Configuration*

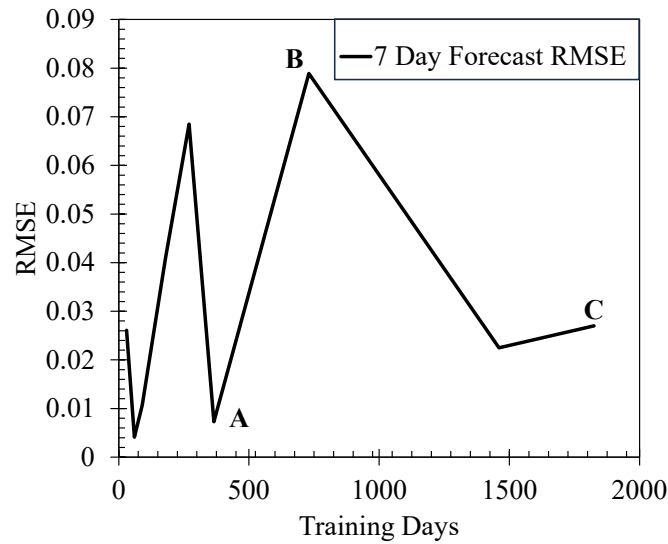
The stacked LSTM, shown in Figure 4-10, functions by providing a sequence of outputs (one output per time step) to each of the subsequent layers. In the case of the stacked LSTM in Figure 4-10, LSTM 1 outputs a sequence of timesteps (with a length equivalent to the number of

timesteps of the input data). LSTM 2 takes this output, learns from it, and then outputs its own sequence of outputs (Brownlee, 2017). This sequence is then pushed through the “Dense” layer, which effectively translates the outputs from the stacked LSTM into useful outputs for the prediction problem. This stacked framework effectively allows the LSTM layers to learn in unison, yielding a better performing model overall. A stacked LSTM consisting of 2 layers (LSTM 1 with 3 neurons and LSTM 2 with 1 neuron) was selected based on the observation that 2 layers generally perform well at detection of complex features within a prediction/forecasting problem (Eckhardt, 2018). Additionally, the optimizer “Adam” with an optimized learning rate of 0.022 was used within this stacked LSTM workflow.

#### **4.6.2 LSTM MODEL TRAINING**

This study made use of a stacked LSTM to forecast soil moisture for subsequent landslide hazard forecasting. This LSTM was trained with LIS-based soil moisture at 1000 mm as the dependent variable and following independent forcing climatological data: (1) Precipitation and (2) Evapotranspiration. Measurements of precipitation were retrieved via NLDAS (NLDAS\_FORA0125\_H\_v2.0) and are available at a resolution of 0.125 degrees (~ 9 km). Measurements of 8-day evapotranspiration were retrieved from MODIS aboard the Aqua (MYD16A2GF\_061\_ET\_500m) and Terra (MOD16A2GF\_061\_ET\_500m) EOS satellites at a resolution of 500 meters. A 3-day sliding window was used in the training of the LSTM model during this study. This implies that for each time step of soil moisture, 3-days of prior climatologic data were investigated. This window was selected due to the observation that soil moisture at 1000 mm did not immediately respond to climatologic variations at the surface. Therefore, this window accounts for potential lag-periods between climatologic events and deep-seated soil moisture response. As LSTMs are sensitive to scaling of data, all climatologic data was normalized prior to training to ensure scales aligned well. Once scaled, an investigation into how much training data was required to effectively train the LSTM as well as how many days the LSTM could accurately forecast soil moisture was needed.

To conduct this investigation, varying amounts of time series soil moisture and respective forcing data were used to train the LSTM. At each iteration of training, a 7-day forecast of soil moisture was compared to historic LIS moisture over the same period. Root Mean Squared Error (RMSE) was used as the primary measure of performance throughout this study. The achieved RMSE values across these iterations of amounts of training data for 7-day forecasts can be seen in Figure 4-11.



*Figure 4-11: RMSE of 7 Day Forecasted LSTM Data Compared to Amount of Training Data*

The RMSE shown in Figure 4-11 can be seen to vary as the amount of training data increases. The achieved RMSE using 365 days of training data (Point A) is very low. However, when the amount of training data increases to 730 days (Point B), the RMSE increases as well, rather than further decreasing as expected. Once the maximum available amount of training data, 1825 days, is implemented, the RMSE can be seen to return to a low value (Point C). Traditionally in machine learning, as training data increases, the model performance is observed to increase as well. That trend is not necessarily observable here. However, this is likely due to the behavior of soil moisture measurements over these respective amounts of training days. These behaviors, as well as a 7-day forecast using the selected 1825 days of training date, are shown in Figure 4-12.

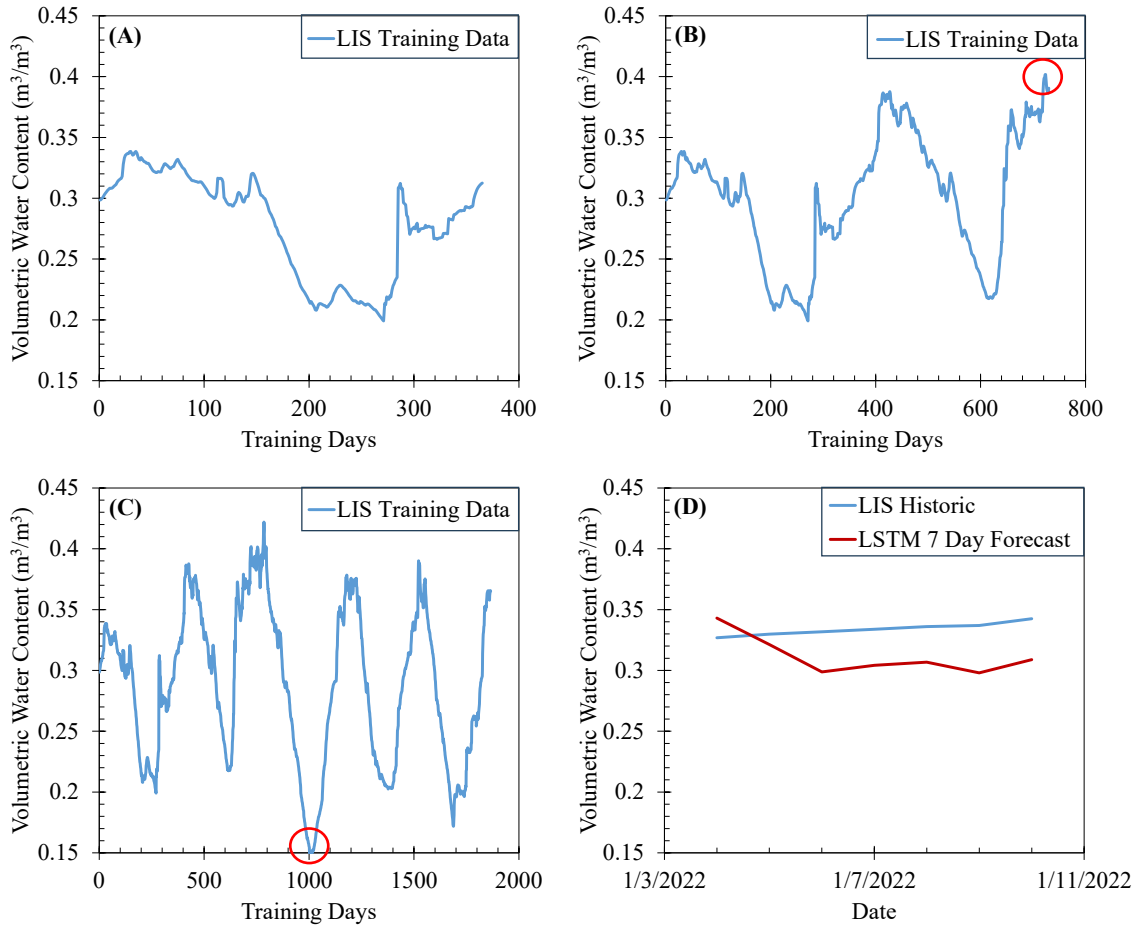


Figure 4-12: Behavior of Soil Moisture Over: (A) 365 Days, (B) 730 Days, and (C) 1825 Days of Training Data and (D) Example of LSTM-based 7 Day Forecast Using 1825 Days of Training Data

As seen in Figures 4-12A-12C, the observable seasonal patterns within the soil moisture training data become more apparent with additional training data. However, occurrences of minima and maxima within these training datasets can likely provide an explanation for the variations of RMSE observed in Figure 4-12. Figure 4-12A shows no large observable local minima or maxima. Figure 4-12B shows a local maxima to occur at the tail end of the training data (circled within the figure). In comparison to this, data in Figure 4-12C shows that the model would be exposed to far more occurrences of local maxima and minima (e.g., peaks and valleys of seasonal training data). However, a local minima that is significantly less than other measurements can also be seen in Figure 4-12C at 1000 days (circled within the figure). It is likely that the observable lack of large maxima and minima in the 365-day training period is what led to success in predictions over that period. As training data increases, so do potential occurrences of these maxima and minima. However, it can be observed that a singular local maxima/minima can lead to a decrease in performance. This can be seen in model performance using 2 years of training

data with one observable maxima (Figure 4-12B). The implication of this is that with extensive amounts of training data come more appearances of these maxima and minima, as seen in Figure 4-12C. These appearances would undoubtedly allow the model to better generalize the conditions leading to these outliers as well as generalizing conditions of soil moisture seasonality. However, this extensive data (e.g., decadal data) was not readily available during this study. Therefore, a training period of 1825 days (5 years), with various occurrences of maxima and minima, was selected with a forecasting period of 7 days. Figure 4-12D shows an example of this 7-day forecast as compared to historical LIS-based moisture retrieved over Landslide 10757. The soil moisture data over this 7-day forecast is observed to align satisfactorily with that of historic data. This leads to confidence in the implementation of LSTM-based soil moisture in the forecasting of landslide hazard analyses.

#### **4.6.3 LSTM-BASED FORECASTED HAZARDS**

The intention behind developing an LSTM workflow capable of forecasting soil moisture was to enable the subsequent assimilation of forecasted moisture within hazard analyses. Doing so effectively allows for hazard analyses to become functions of forecasted soil moisture and therefore become forecastable. To investigate the efficacy of forecasted hazard analyses, pseudo-forecasting was investigated over two known landslides within the Eastern Kentucky domain. The term “pseudo-forecasting” is used because this study does not make an effort to forecast beyond the current date of this study. Such forecasting is possible through inclusion of forecasted climatological data. However, it was desired to have known landslide occurrences, which are not available beyond the date of this study, to be able to effectively gauge the efficacy of forecasted hazard analyses. Therefore, landslides 10793 and 10794, from the KGS landslide inventory, were chosen for pseudo-forecasting. It is necessary to note that these pseudo-forecasted landslides were *not* included in the development of the landslide susceptibility model. Additionally, these landslides occur directly after the 5-year training period (01/01/2017 – 12/31/2021) used in the LSTM development. Therefore, forecasts of soil moisture over these landslides are true forecasts from the developed LSTM model. Herein, these landslides are pseudo-forecasted using the 5-year trained LSTM in conjunction with climatological data retrieved over the slides. Forecasted soil moisture is then used in conjunction with the Lu and Godt infinite slope equation to develop forecasted hazard analyses. The susceptibility and pseudo-forecasted hazard analyses over these two landslides can be seen in Figure 4-13.



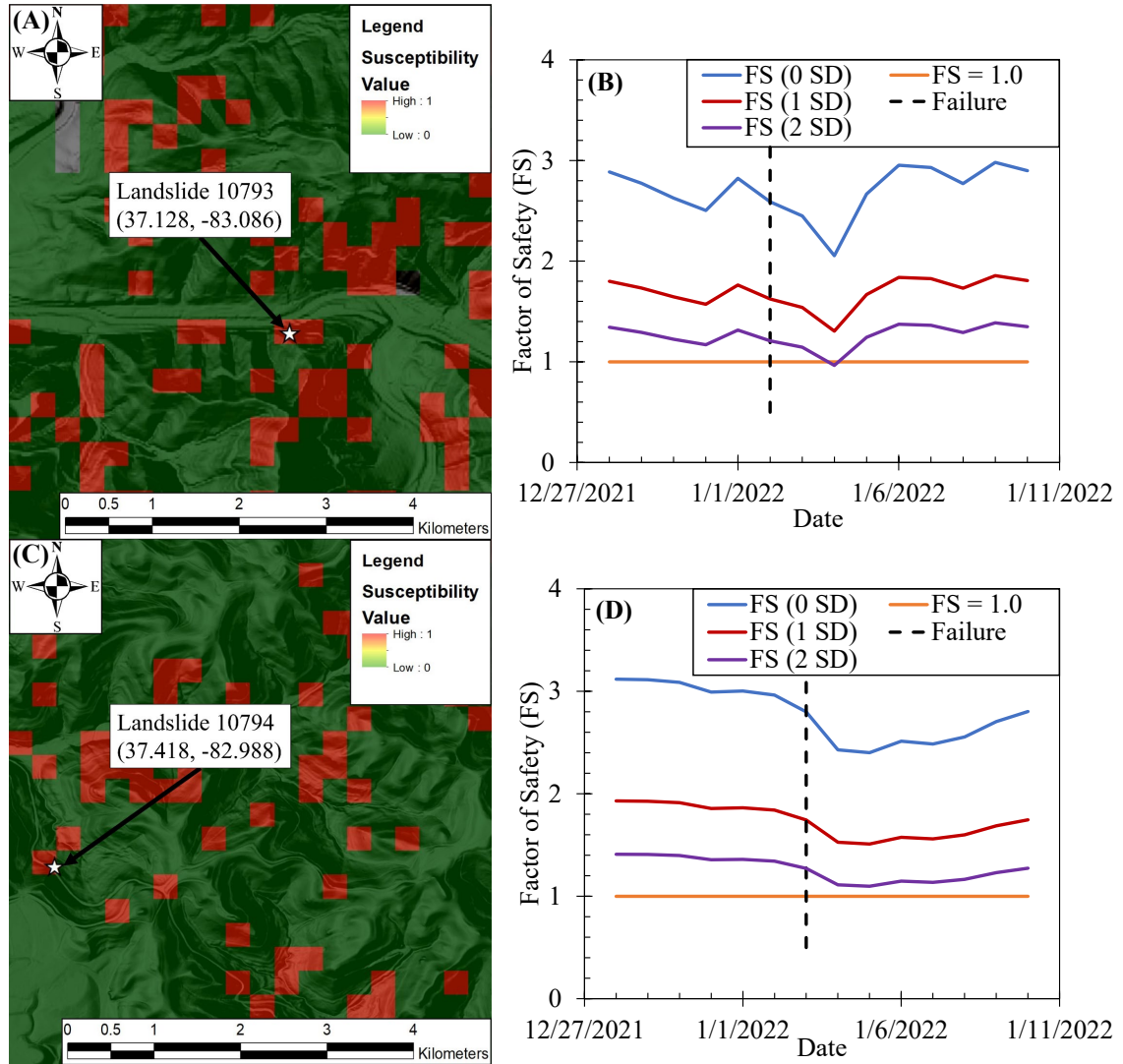


Figure 4-13: (A) 100 m Susceptibility Analysis for Landslide 10793 (Occurred 01/02/2022), (B) Pseudo-Forecasted Hazard Analysis for Landslide 10793, (C) 100 m Susceptibility Analysis for Landslide 10794 (Occurred 01/03/2022), and (D) Pseudo-Forecasted Hazard Analysis for Landslide 10794

As can be seen in Figures 4-13A and 4-13C, the susceptibility analysis developed through this work successfully identifies areas (indicated as areas of 1.0) where landslides 10793 and 10794 occurred as susceptible. Figures 4-13B and 4-13D show pseudo-forecasted hazard analyses over the respective landslides. These forecasts show a 14-day forecast of factor of safety. This forecast was developed by combining two 7-day forecasts to allow for a longer span of temporal analyses. These 7-day forecasts were developed as follows: (1) First 7-day forecast was created using 1818 days of training data, and forecasting 7-days beyond the 1818<sup>th</sup> day, (2) Second 7-day forecast was created using 1825 days of training data to forecast beyond the 1825<sup>th</sup> day. Effectively, the first forecast covers 12/28/2021 – 1/3/2022 with the second covering 1/4/2022 – 1/10/2022. Combining the two forecasts yields the 14-day pseudo-forecast used herein. However, when

investigating the forecasted FS (i.e., landslide hazards), it is apparent that the FS does not reach 1.0 at the time of failure for either slide, whereas 1.0 is indicative of catastrophic failure. To improve the efficacy of these forecasted results, variations of the geomorphologic data used to conduct these analyses were investigated. These variations were investigated through alterations of the data through respective standard deviations (SD). It was found that spatially retrieved estimates of  $\phi'$ ,  $H_{ss}$ , and  $\beta$  experienced a SD of 1.14 degrees, 2.6 m, and 8.4 degrees, respectively. These alterations were investigated by reducing  $\phi'$  while increasing  $H_{ss}$  and  $\beta$  by a respective SD. From Figure 4-13B and 4-13D, it can be observed that factors of safety achieved by adjusting these data by 2 SD far better represented failure conditions. The geomorphologic data after these adjustments for landslide 10793 were  $\phi' = 29.98$  degrees,  $H_{ss} = 9.34$  m, and  $\beta = 58.8$  degrees and for landslide 10794 were  $\phi' = 30.27$  degrees,  $H_{ss} = 9.64$  m, and  $\beta = 54.8$  degrees. The implication of this is that forecasted landslide hazards, as a function of forecasted soil moisture, can successfully detect reductions in strength. This can be seen through the FS with no reductions via SD showing decreases in strength (i.e., reductions of FS) at or near failure. However, it is imperative that geomorphologic data that is representative of site conditions also be included in these analyses. As this study makes use of geomorphologic data retrieved at resolutions of approximately 300 m, it is understood that discrepancies are possible between these retrievals and site-based characteristics. Regardless, the pseudo-forecasted hazard analyses successfully indicate reductions in strength, as well as showing failure (at or near 1.0) to occur with adjustments to spatial retrievals of geomorphologic data. These results lead to confidence in the developed framework for the forecasting of landslide hazards through inclusion of forecasted soil moisture.

#### 4.7 CONCLUSIONS

This study saw the development of a workflow for dynamic landslide susceptibility classifications as well as the forecasting of landslide hazard analyses across a domain within Eastern Kentucky. Susceptibility classifications were developed as a function of static geomorphologic data and dynamic Normalized Difference Vegetation Index (NDVI) retrievals. Hazard analyses were developed as a function of geomorphologic data as well but became dynamic through inclusion of soil moisture retrievals over investigated sites. Dynamic NDVI retrievals were easily obtainable over areas of interest via the MODIS platform aboard the Aqua and Terra EOS satellite. However, spatial retrievals of soil moisture were observed to not be as readily available. To remedy this lack of spatial data, the NASA Land Information System (LIS) framework was

utilized. Through utilization of LIS, NASA SMAP satellite-based retrievals were assimilated with NOAA 3.6 LSM estimates of soil moisture. This assimilation allowed for retrieval of soil moisture at a resolution of 0.01 degree (~1 km) at depths of 100-, 200-, 500-, and 1000-mm. To ensure accuracy of LIS-based soil moisture retrievals, direct comparisons were conducted at various sites throughout the investigated domain with in-situ soil moisture monitoring capabilities. These comparisons yielded satisfactory results. Therefore, spatial retrievals of soil moisture were now readily available for use by this study.

A majority of landslide occurrences throughout Eastern Kentucky have been observed to occur on hillslopes that have fairly shallow depths of soil overlaying bedrock. Therefore, spatial retrievals of depth to bedrock (DTB) became necessary for both the developed susceptibility and hazard analyses herein. As with soil moisture, it was observed that spatial measurements of DTB were lacking. Therefore, a novel DTB map was created through this study to allow for spatial retrievals of DTB at a given site of interest. To create this map, DTB data from 3,000 Kentucky Transportation Cabinet (KYTC) boreholes across the investigated domain were obtained. Kriging interpolation was then used to interpolate a spatial map between these 3,000 boreholes. The resulting DTB map became a tool which yielded spatial estimates of DTB over areas previously seen to be lacking this requisite data.

Landslide susceptibility analyses were developed herein through the development of a logistic regression machine learning classification model. This model was developed upon a dataset pertaining to 675 known landslides across the investigated domain. Geomorphologic data was retrieved at each of these 675 known landslides. Additionally, a dynamic aspect was added to these susceptibility analyses through inclusion of NDVI retrievals representative of each landslide's site and date of occurrence. Therefore, the developed susceptibility dataset contained site specific geomorphic data as well as vegetation conditions at the time of failure. Once developed, this classifier saw an accuracy of 89% in classifications and was capable of yielding susceptibilities observed to vary spatially and temporally. Therefore, it was shown that the developed susceptibility model is able to satisfactorily classify susceptible areas as well as yielding susceptibilities observed to vary dynamically due to inclusion of dynamic vegetation levels.

With successful development of a dynamic susceptibility classification tool, the efforts of this study moved towards the development of a workflow for the forecasting of landslide hazards. These hazard analyses were conducted through implementation of the Lu and Godt infinite slope stability equation. As this equation is a function of suction stress, this allowed for these hazard

analyses to become a function of soil moisture at a given site. Therefore, as soil moisture levels varied with time, so would the results from these hazard analyses. However, it was desired to investigate the accuracy of these hazard analyses prior to development of a forecastable workflow. Accuracy was investigated through a historic analysis of four previous landslides, with dates and locations known, within the domain. LIS-based soil and geomorphic data were retrieved at each landslide and implemented within the Lu and Godt equation. The results from this historic analysis yielded results indicative of failure at or near the time of these four known landslides. This success in hazard analyses as a function of the Lu and Godt infinite slope stability provided confidence in implementation of said equation in a hazard analysis forecasting workflow.

With confidence in both spatial retrievals of historic soil moisture and the efficacy of Lu and Godt based hazard analyses, the efforts of this study moved to the forecasting of hazard analyses. Soil moisture is assumed herein to be the controlling factor in initiation of landslide occurrence. Therefore, the first step of this forecasting workflow was the development of a means to obtain forecasts of soil moisture levels at a given site. To do so, a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) was developed. This LSTM forecast model was trained upon 1825 days (5 years) of LIS-based soil moisture, NLDAS precipitation, and MODIS evapotranspiration at each site of interest. After training and optimization of the LSTM, forecasted soil moisture levels were obtained that were observed to align well with historic LIS-based moisture over the forecasted period. Based on these successes in forecasting of soil moisture levels, the developed LSTM became the framework for subsequent hazard analysis forecasting.

As no known landslides were readily available beyond the date of this study, a pseudo-forecast was instead investigated. This pseudo-forecast made use of forecasted soil moisture over *known* landslides. However, these known landslides occurred shortly after the 5-year training period of the LSTM model. This allowed the trained LSTM to effectively forecast soil moisture over these slides as this period was outside what the model had been trained upon. Additionally, these known slides were *not* included in the development of the dynamic susceptibility classifier. Therefore, a coupled approach, of both susceptibility and hazard analyses, was investigated over these known landslides. The dynamic susceptibility model was able to accurately identify areas in which these slides occurred as susceptible. The pseudo-forecasted hazard analyses indicated a reduction in strength occurred at or near time of failure but did not indicate failure conditions. However, statistic-based alterations to spatial retrievals of geomorphic data within these analyses

were investigated. After these alterations, the hazard analyses indicated failure conditions at or near the time of failure of these investigated slides. Therefore, it became apparent that forecasted soil moisture was capable of showing reductions in strength through forecasted hazard analyses. However, geomorphologic data must be ensured to be representative of site conditions for effective use of these forecasted hazard analyses. Regardless, the efforts herein yield a novel workflow for the creation of an effective dynamic landslide susceptibility approach as well as forecastable landslide hazard analyses as a function of forecasted soil moisture and geomorphic data.

## CHAPTER 5: Conclusions

The first goal of this study was the investigation into the increasing the efficacy of SMAP satellite-based soil moisture retrievals for use in site-based analyses. Many existing methodologies are readily available for this process. These include downscaling (i.e., the improvement of resolution of satellite-based retrievals) as well as bias-reducing techniques through assimilation and modeling approaches. However, these approaches were observed to require extensive knowledge of the associated approach as well as being tedious to conduct. Therefore, a novel approach for improvement of satellite-based retrievals at a site-based scale was developed. This approach related satellite-based soil moisture correction factors with readily available soil texture information (available via web databases) through multivariate regression. This novel approach yielded a far less intensive workflow observed to be as capable as existing workflows in the obtainment of satellite-based soil moisture for site-based analyses.

With the use of satellite-based retrievals within site-based analyses came the observation that these retrievals are seen to be lacking moisture at varying depths. Currently, SMAP satellite-based moisture retrievals are available at the surface layer (0 – 50 mm) and at the root zone layer (0 – 1000 mm). However, the investigated landslides throughout this study have the likelihood of occurring within various depths of the soil layer. Therefore, a method to retrieve spatial retrievals of soil moisture at varying depth was desired. The Land Information System (LIS) framework was modified herein to provide these spatial estimates across Kentucky at a fine resolution of 1 km at varying depths of 100 mm, 300 mm, 600 mm, and 1000 mm. The LIS framework yielded these soil moisture measurements through assimilation of modeled (NOAH 3.6 LSM) and satellite-based (SMAP) soil moisture measurements. These measurements at varying depths were then used in spatiotemporal landslide hazard analyses over a known severe landslide event. Incorporation of LIS-based moisture estimates in these analyses led to identification of incipient conditions indicative of occurrence at or near the times of failure of the associated landslide event.

The final aim of this study was the improvement of existing landslide susceptibility and hazard analyses systems. Current susceptibility systems rely primarily upon static geomorphologic information to yield susceptibility classifications over an area. This implies that these susceptibility classifications remain static as does the static geomorphic data. Landslide hazard analyses make use of geomorphic and dynamic data but are only capable of providing warnings of potential landslide occurrences in near real-time. Therefore, this study developed a dynamic

susceptibility approach built upon the inclusion of static geomorphologic data *and* dynamic measurements of vegetation. Inclusion of vegetation allows for susceptibility classifications to now vary with time rather than remain static. Regarding hazard analyses, the ability to provide a forecast (i.e., over a 7-day window) of hazards, rather than in real time, would undoubtedly provide planners and residents with more preparation time for potential landslide occurrences. Therefore, a predictive analytics Long Short-Term Memory (LSTM) forecasting machine learning workflow was developed to forecast soil moisture over sites of interest. As the landslide analyses conducted throughout this study assumed soil moisture to be the controlling factor of these analyses, the forecasting of soil moisture implies these analyses are now forecastable as well. These dynamic and forecastable analyses developed herein provide more warning of potential landslide occurrences for those in mountainous regions. Additionally, these analyses, as well as all analyses herein, provide more insight into the role of soil moisture as a hydromechanical precursor indicative of incipient conditions prior to landslide occurrence.

## APPENDIX



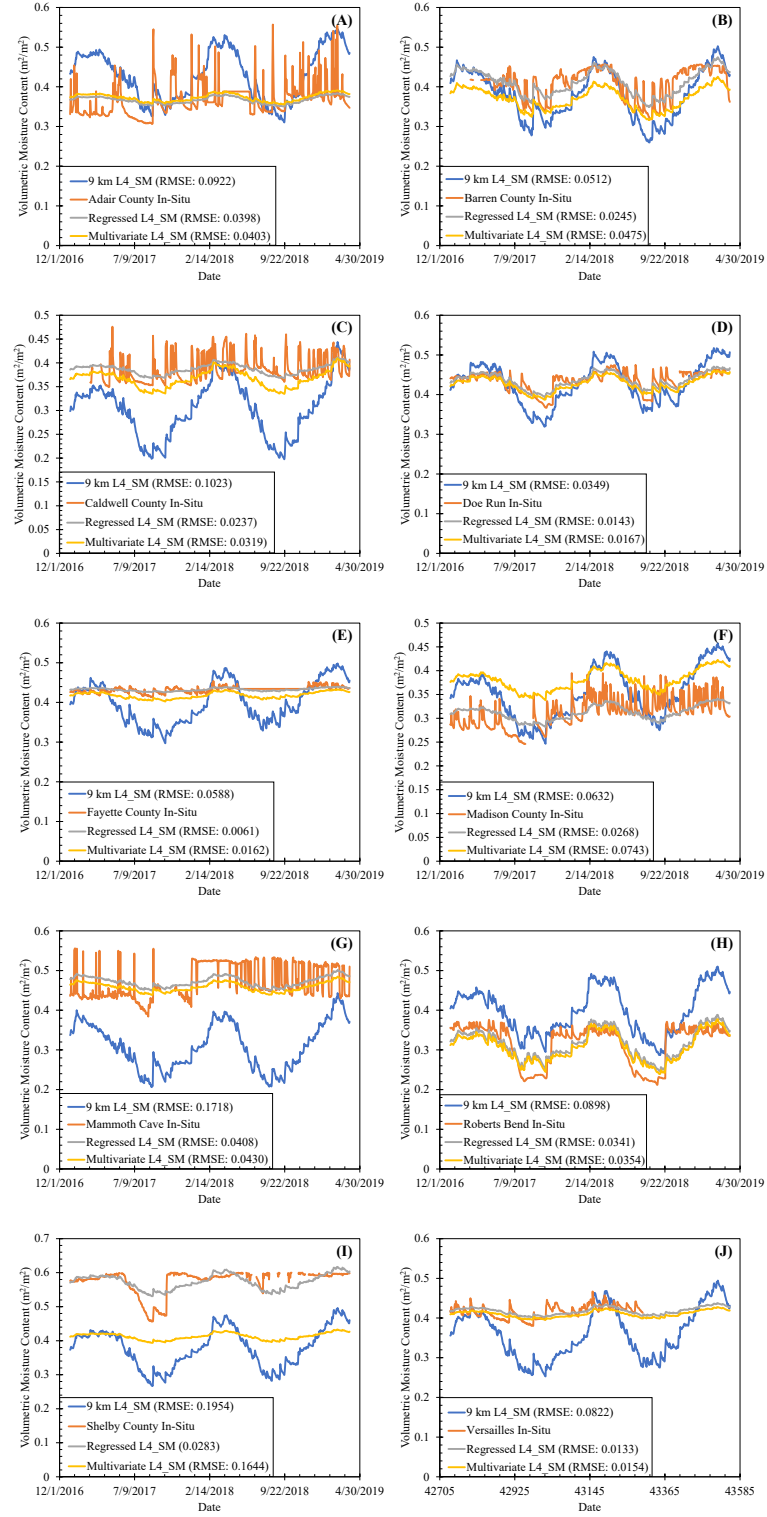


Figure A1: (A) 9 km L4\_SM corrections at Adair County site, (B) 9 km L4\_SM corrections at Barren County site, (C) 9 km L4\_SM corrections at Caldwell County site, (D) 9 km L4\_SM corrections at Doe Run site, (E) 9 km L4\_SM corrections at Fayette County site, (F) 9 km L4\_SM corrections at Madison County site, (G) 9 km L4\_SM corrections at Mammoth Cave site, (H) 9 km L4\_SM corrections at Roberts Bend site, (I) 9 km L4\_SM corrections at Shelby County site, and (J) 9 km L4\_SM corrections at Versailles site

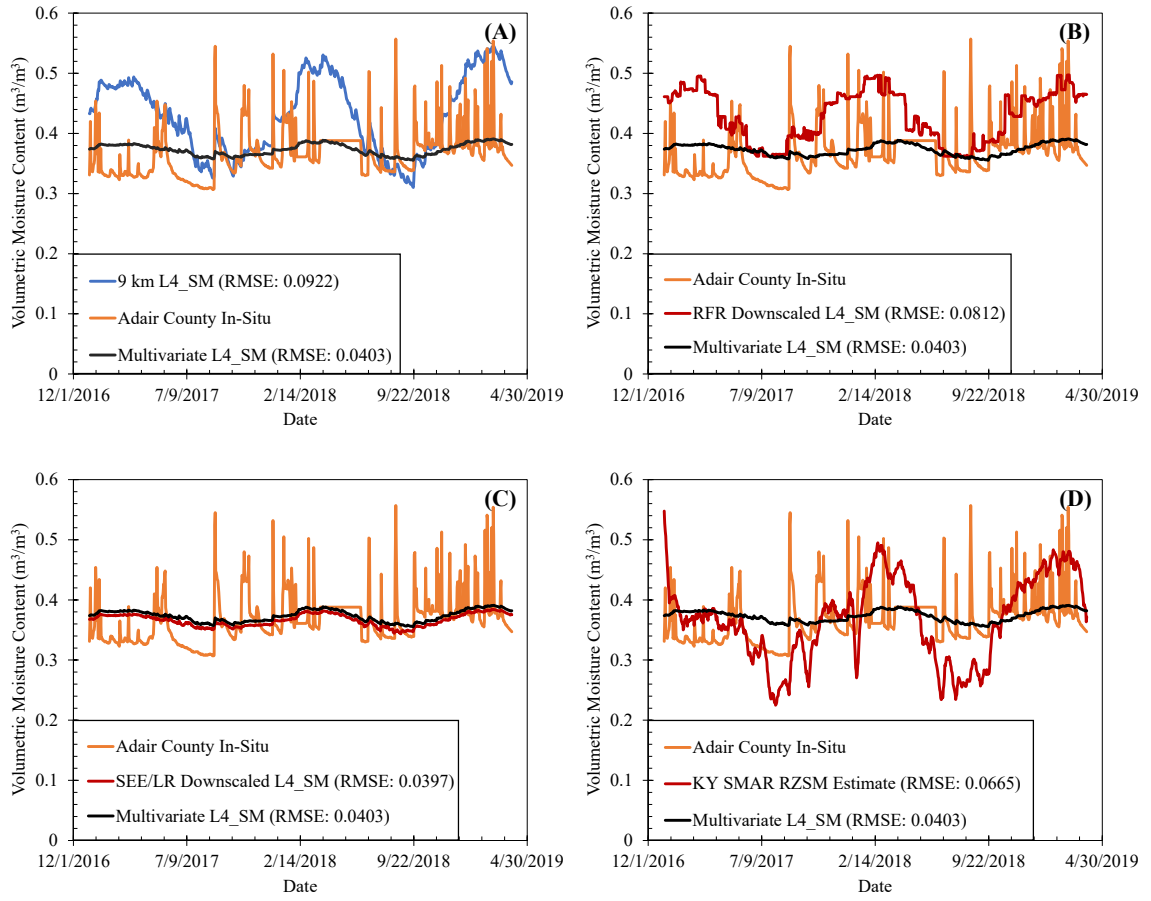


Figure A2: Comparison of Adair County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

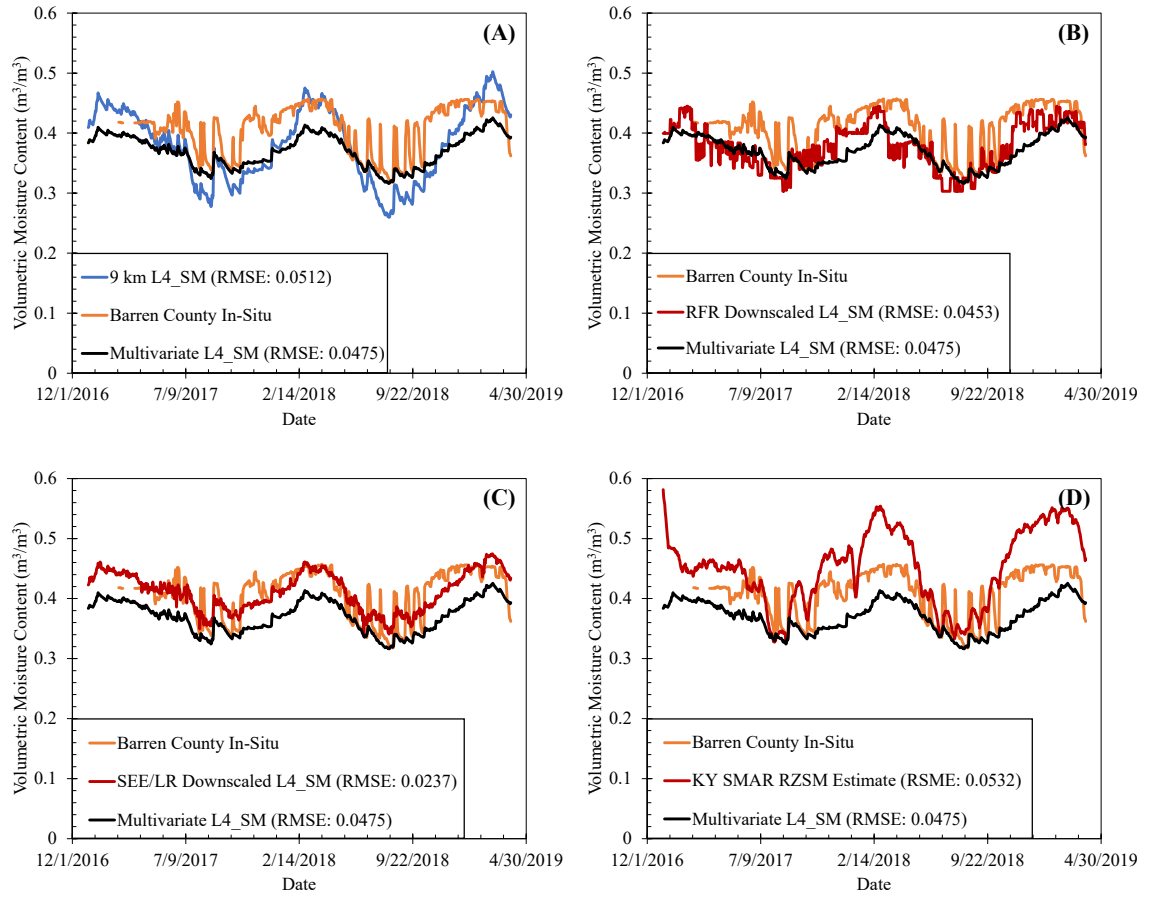


Figure A3: Comparison of Barren County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

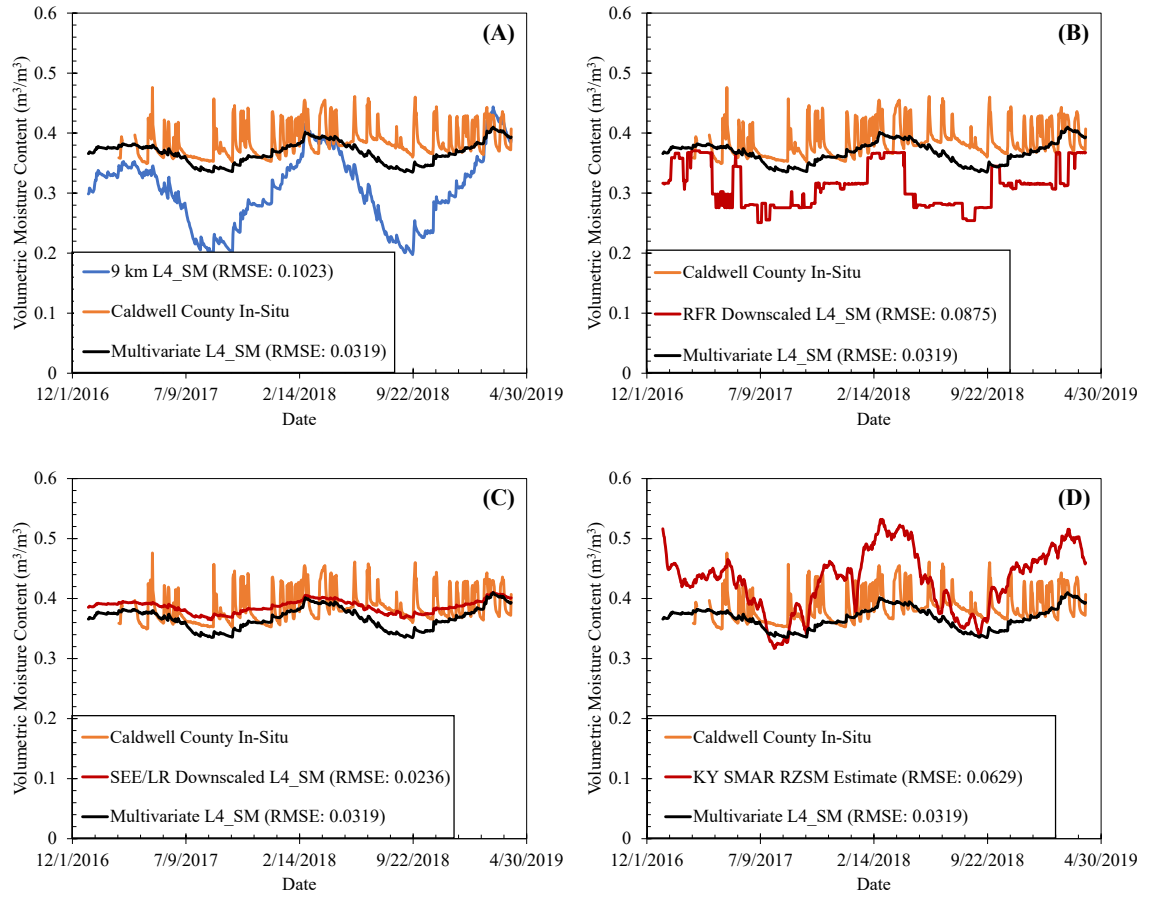


Figure A4: Comparison of Caldwell County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

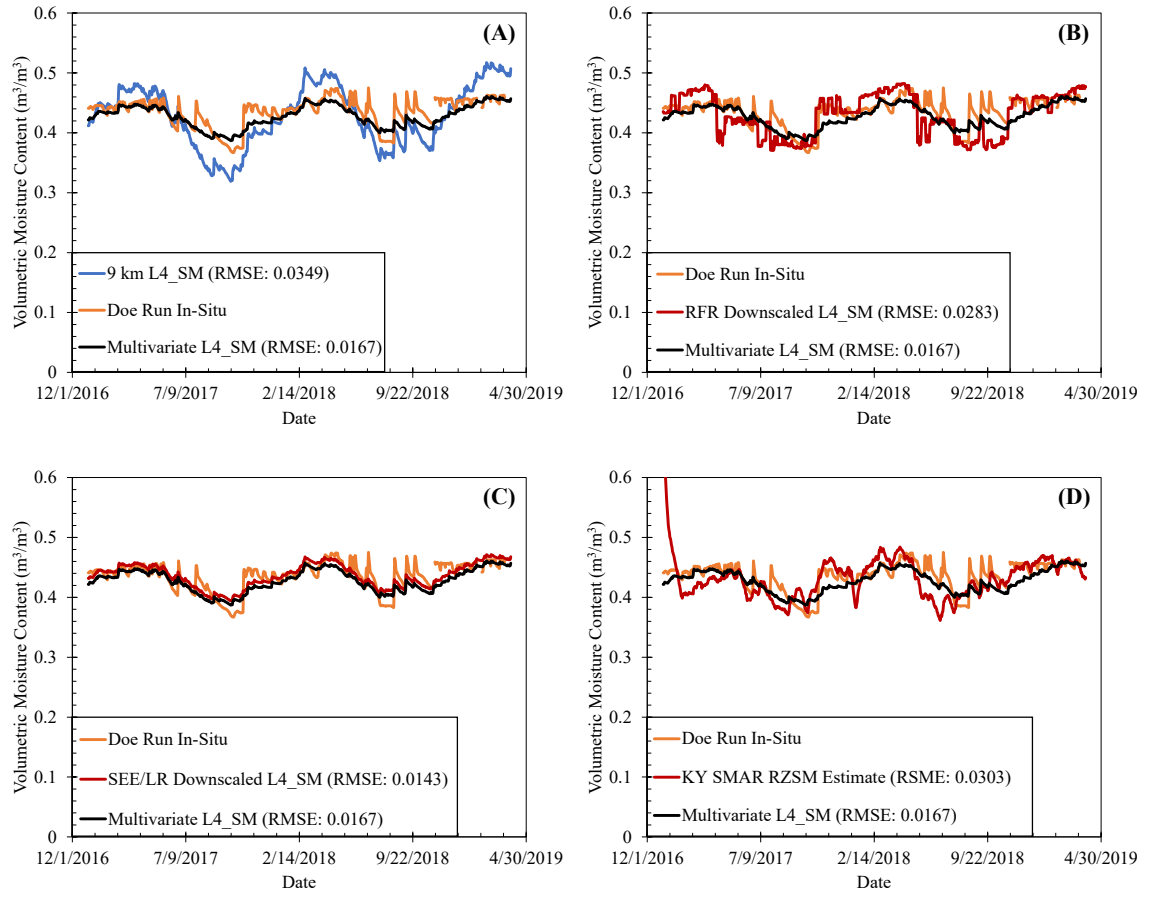


Figure A5: Comparison of Doe Run In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

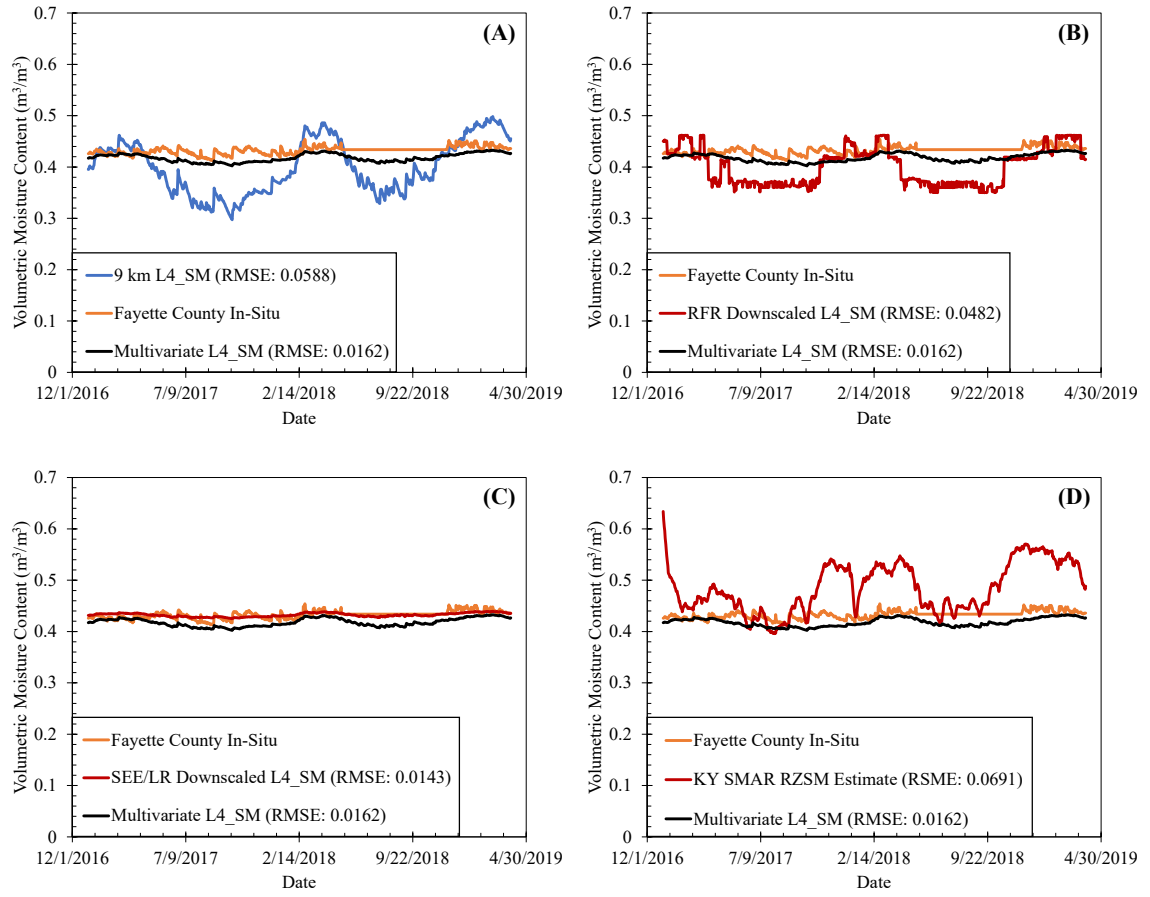


Figure A6: Comparison of Fayette County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

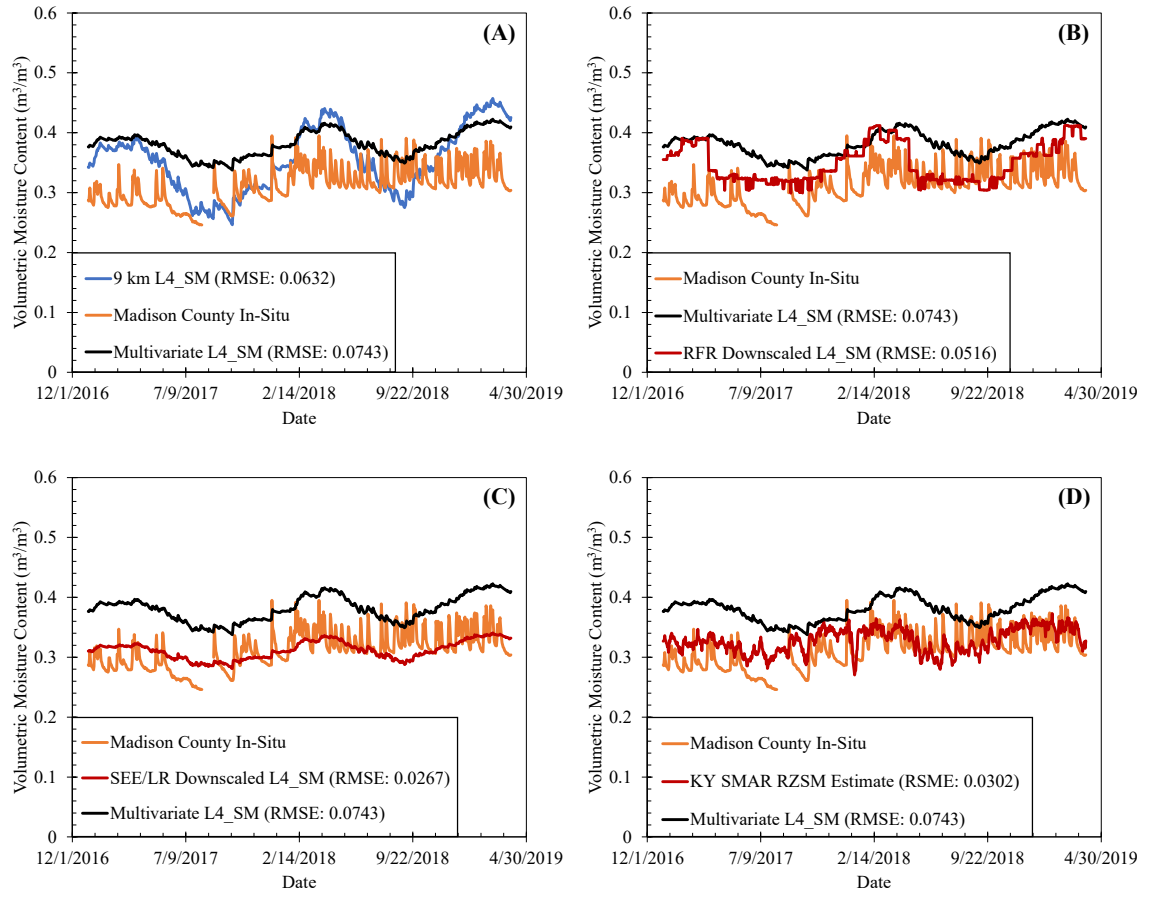


Figure A7: Comparison of Madison County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

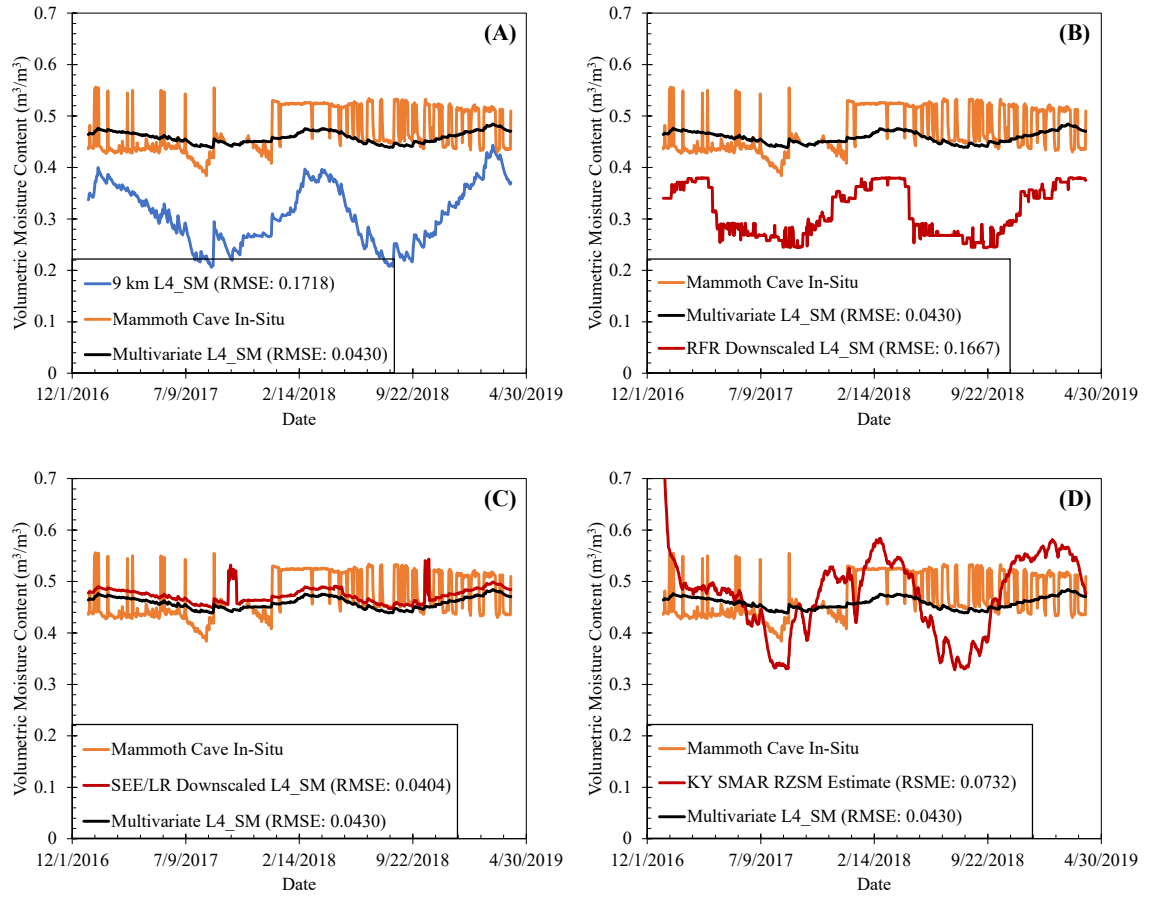


Figure A8: Comparison of Mammoth Cave In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR



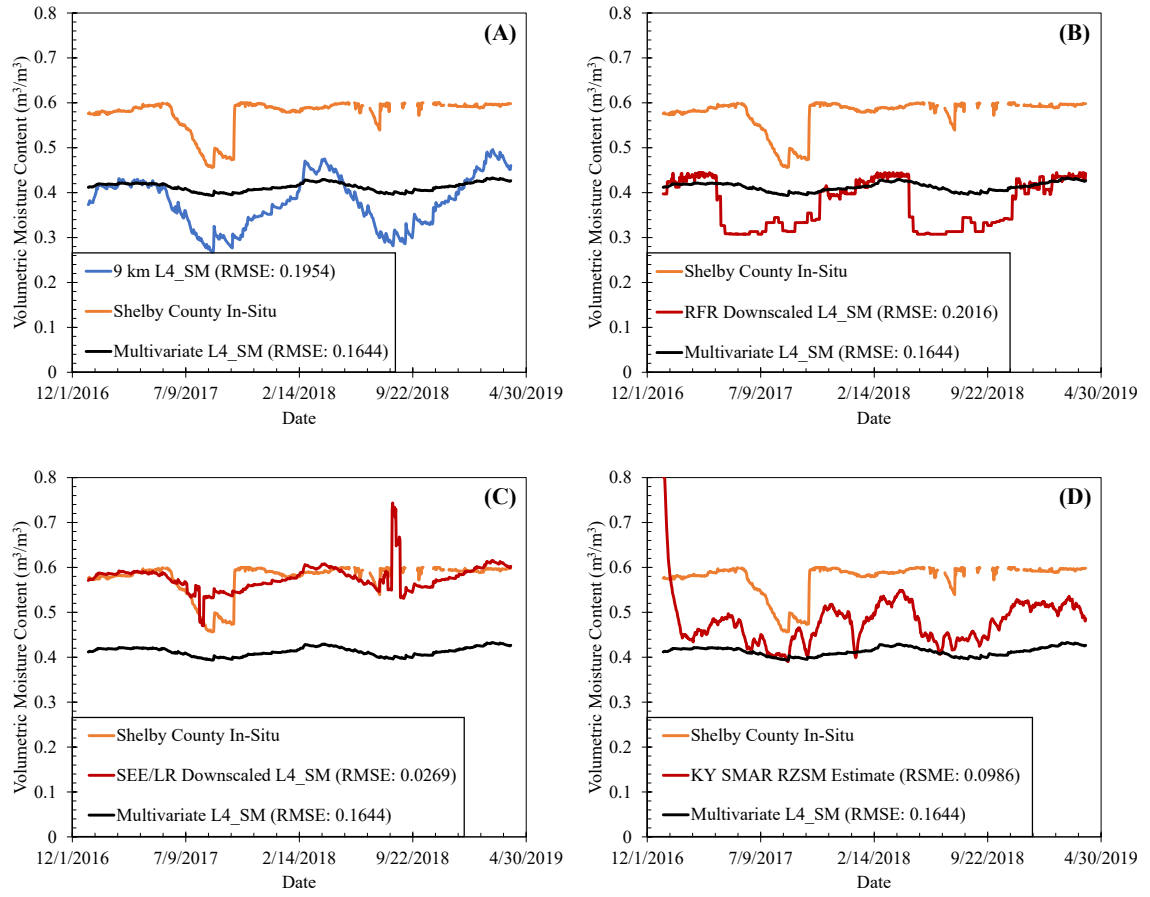


Figure A9: Comparison of Shelby County In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

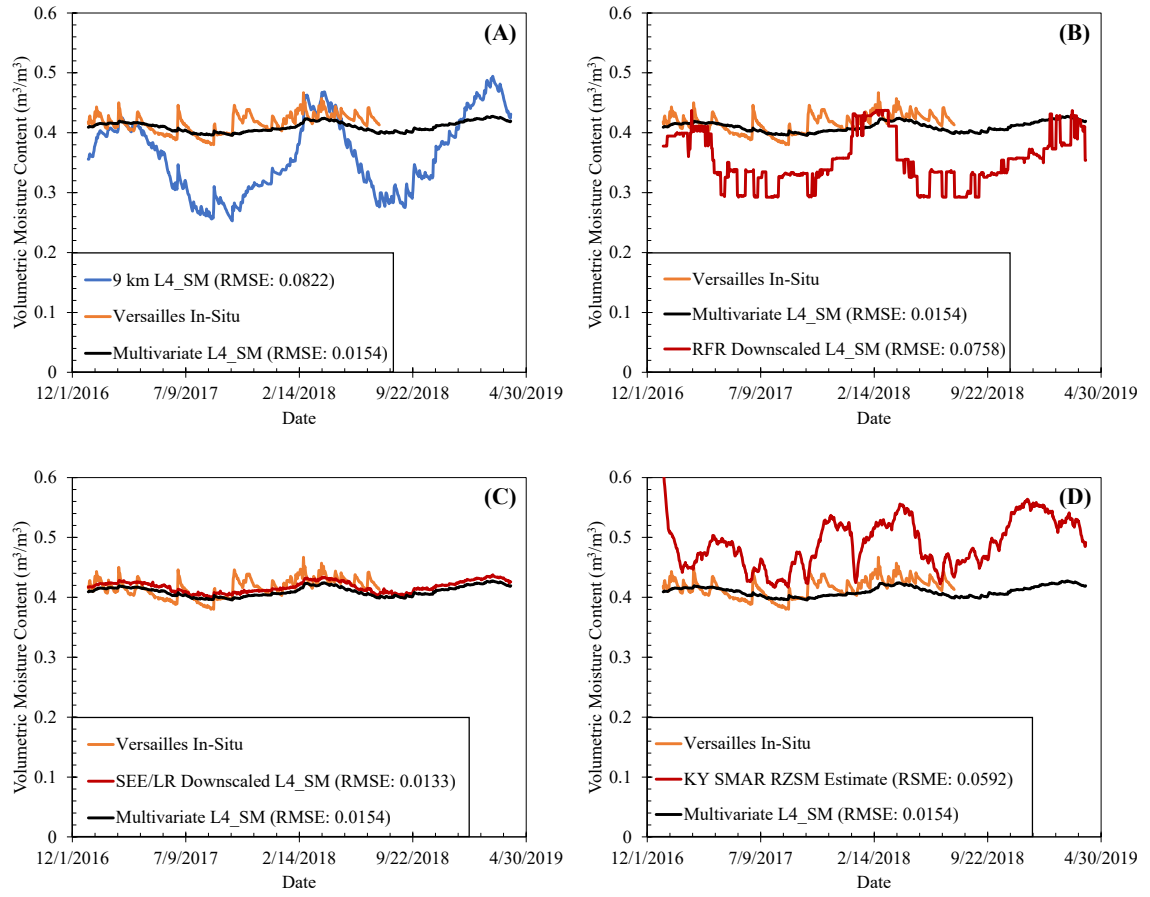


Figure A10: Comparison of Versailles In-Situ RZSM to site-specific RZSM through: (A) Multivariate Regression vs 9 km L4\_SM, (B) Multivariate Regression vs RFR, (C) Multivariate Regression vs SEE, and (D) Multivariate Regression vs EnKF/SMAR

## REFERENCES

- Akbar, R., Short Gianotti, D., McColl, K. A., Haghighi, E., Salvucci, G. D., & Entekhabi, D. (2018). Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation. *Water Resources Research*, 54(3), 1476-1492.
- Arsenault, K. R., Kumar, S. V., Geiger, J. V., Wang, S., Kemp, E., Mocko, D. M., ... & Peters-Lidard, C. D. (2018). The Land surface Data Toolkit (LDT v7. 2)—a data fusion environment for land data assimilation systems. *Geoscientific Model Development*, 11(9), 3605-3621.
- Baldwin, D., Manfreda, S., Keller, K., & Smithwick, E. A. H. (2017). Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *Journal of Hydrology*, 546, 393-404.
- Belkhiri, L., Tiri, A., & Mouni, L. (2020). Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. *Groundwater for Sustainable Development*, 11, 100473.
- Bittelli, M., Valentino, R., Salvatorelli, F., & Pisa, P. R. (2012). Monitoring soil-water and displacement conditions leading to landslide occurrence in partially saturated clays. *Geomorphology*, 173, 161-173.
- Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets. *ISPRS International Journal of Geo-Information*, 1(1), 32-45.
- Brownlee, J. (2017). *Long short-term memory networks with Python: Develop Sequence Prediction models with Deep Learning*, Machine Learning Mastery, pp 78–79.
- Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., ... & Kerr, Y. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4994-5007.
- Chen, W., Shahabi, H., Zhang, S., Khosravi, K., Shirzadi, A., Chapi, K., ... & Ahmad, B. B. (2018). Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression. *Applied Sciences*, 8(12), 2540.
- Chirico, G. B., Borga, M., Tarolli, P., Rigon, R., & Preti, F. (2013). Role of vegetation on slope stability under transient unsaturated conditions. *Procedia Environmental Sciences*, 19, 932-941.

- Colliander, A., Fisher, J. B., Halverson, G., Merlin, O., Misra, S., Bindlish, R., ... & 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., & Bryson, L. S. (2018). Assessment of active landslides using field electrical measurements. *Engineering Geology*, 233, 146-159.
- Crawford, M. M., Bryson, L. S., Woolery, E. W., & Wang, Z. (2019). Long-term landslide monitoring using soil-water relationships and electrical data to estimate suction stress. *Engineering Geology*, 251, 146-157.
- Crawford, M. M., Dortch, J. M., Koch, H. J., Killen, A. A., Zhu, J., Zhu, Y., ... & Haneberg, W. C. (2021). Using landslide-inventory mapping for a combined bagged-trees and logistic-regression approach to determining landslide susceptibility in eastern Kentucky, USA. *Quarterly Journal of Engineering Geology and Hydrogeology*, 54(4), qjagh2020-177.
- da Silva, I. N., Alves, S. F. dos R., Liboni, L. H. B., Flauzino, R. A., and Spatti, D. H. (2018). *Artificial Neural Networks: A practical course*. Springer, Cham, pp 117–138.
- Das, S. K., & Basudhar, P. K. (2008). Prediction of residual friction angle of clays using artificial neural network. *Engineering Geology*, 100(3-4), 142-145.
- Dashbold, B., Bryson, L. S., & Crawford, M. M. (2023). Landslide hazard and susceptibility maps derived from satellite and remote sensing data using limit equilibrium analysis and machine learning model. *Natural Hazards*, 116(1), 235-265.
- Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European radiology experimental*, 4(1), 1-8.
- Dolphin, R. (2021). “LSTM networks: A detailed explanation.” *Medium*, Towards Data Science, <https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9> (Jun. 27, 2023).
- Eckhardt, K. (2018). “Choosing the right hyperparameters for a simple LSTM using keras.” *Medium*, Towards Data Science, <https://towardsdatascience.com/choosing-the-right-hyperparameters-for-a-simple-lstm-using-keras-f8e9ed76f046> (Jun. 28, 2023).

- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53, 343-367.
- Fertig, E., Baek, S. J., Hunt, B., Ott, E., Szunyogh, I., Aravéquia, J., ... & Liu, J. (2009). Observation bias correction with an ensemble Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, 61(2), 210-226.
- Francis, D. M., and Bryson, L. S. (2023). "Rainfall-Induced Landslide Hazard Analyses using Spatiotemporal Retrievals of Soil Moisture and Geomorphologic Data." *Water Resources Research*. SUBMITTED.
- Francis, D. M., and Bryson, L. S. (2022). "Random Forest based downscaling of SMAP L4 soil moisture and subsequent use in landslide slope stability analysis." Geo-Congress 2022: State of the Art and Practice in Geotechnical Engineering, Charlotte, North Carolina, USA, 20-23 March 2022.
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15), 2627-2636.
- Graves, A., Mohamed, A., and Hinton, G. (2013). "Speech recognition with deep recurrent neural networks." *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. (pp. 6645-6649). Vancouver, Canada, 26-31 May 2013.
- Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, 63(4), 308-319.
- Guo, W. Z., Chen, Z. X., Wang, W. L., Gao, W. W., Guo, M. M., Kang, H. L., ... & Zhao, M. (2020). Telling a different story: The promote role of vegetation in the initiation of shallow landslides during rainfall on the Chinese Loess Plateau. *Geomorphology*, 350, 106879.
- He, J., Wang, S., Liu, H., Nguyen, V., & Han, W. (2021). The critical curve for shallow saturated zone in soil slope under rainfall and its prediction for landslide characteristics. *Bulletin of Engineering Geology and the Environment*, 80, 1927-1945.
- Hidayat, R., Sutanto, S. J., Hidayah, A., Ridwan, B., & Mulyana, A. (2019). Development of a landslide early warning system in Indonesia. *Geosciences*, 9(10), 451.
- Hong, Y., Adler, R., & Huffman, G. (2007). Use of satellite remote sensing data in the mapping of global landslide susceptibility. *Natural hazards*, 43, 245-256.

- Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated machine learning: methods, systems, challenges* (p. 219). Springer Nature.
- Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, 75, 1-19.
- Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. *Applied Artificial Intelligence*, 33(10), 913-933.
- Jiang, Y., Chen, W., Wang, G., Sun, G., and Zhang, F. (2017). “Influence of initial dry density and water content on the soil–water characteristic curve and suction stress of a reconstituted loess soil.” *Bulletin of Engineering Geology and the Environment*, 76, 1085–1095.
- Kirschbaum, D., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, 6(3), 505-523.
- Kumar, S. V., Peters-Lidard, C. D., Santanello, J., Harrison, K., Liu, Y., & Shaw, M. (2012). Land surface Verification Toolkit (LVT)—a generalized framework for land surface model evaluation. *Geoscientific Model Development*, 5(3), 869-886.
- Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., ... & Houser, P. R. (2008). A land surface data assimilation framework using the land information system: Description and applications. *Advances in Water Resources*, 31(11), 1419-1432.
- Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., ... & Sheffield, J. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental modelling & software*, 21(10), 1402-1415.
- Lee, J. H., Zhao, C., & Kerr, Y. (2017). Stochastic bias correction and uncertainty estimation of satellite-retrieved soil moisture products. *Remote Sensing*, 9(8), 847.
- Lu, N., & Godt, J. (2008). Infinite slope stability under steady unsaturated seepage conditions. *Water resources research*, 44(11).
- Lu, N., and Likos, W. J. (2006). “Suction stress characteristic curve for unsaturated soil.” *Journal of Geotechnical and Geoenvironmental Engineering*, 132(2), 131–142.

- Likos, W. J., & Lu, N. (2004). Unsaturated soil mechanics. *ed: John Wiley and Sons Inc., New Jersey.*
- Manfreda, S., Brocca, L., Moramarco, T., Melone, F., & Sheffield, J. (2014). A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrology and Earth System Sciences, 18*(3), 1199-1212.
- Merlin, O., Walker, J. P., Chehbouni, A., & Kerr, Y. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment, 112*(10), 3935-3946.
- Mesri, G., Terzaghi, K., and Peck, R. B. (1996). *Soil Mechanics in Engineering Practice*. Wiley, New York.
- Miralles, D. G., Crow, W. T., & Cosh, M. H. (2010). Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology, 11*(6), 1423-1429.
- Monsieurs, E., Dewitte, O., & Demoulin, A. (2019). A susceptibility-based rainfall threshold approach for landslide occurrence. *Natural Hazards and Earth System Sciences, 19*(4), 775-789.
- Mauri, L., Cucchiaro, S., Grigolato, S., Dalla Fontana, G., & Tarolli, P. (2022). Evaluating the interaction between snowmelt runoff and road in the occurrence of hillslope instabilities affecting a landslide-prone mountain basin: A multi-modeling approach. *Journal of Hydrology, 612*, 128200.
- Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water resources research, 12*(3), 513-522.
- Mubashar, M., Khan, G. M., & Khan, R. (2021, April). Landslide prediction using long short term memory (LSTM) neural network on time series data in Pakistan. In *2021 International conference on artificial intelligence (ICAI)* (pp. 175-181). Montreal, Canada, 19-26 August 2021.
- Nguyen, T. A., Ly, H. B., & Pham, B. T. (2020). Backpropagation neural network-based machine learning model for prediction of soil friction angle. *Mathematical Problems in Engineering, 2020*, 1-11.
- Orland, E., Roering, J. J., Thomas, M. A., & Mirus, B. B. (2020). Deep learning as a tool to forecast hydrologic response for landslide-prone hillslopes. *Geophysical Research Letters, 47*(16), e2020GL088731.

- Panchal, G., Ganatra, A., Kosta, Y. P., & Panchal, D. (2011). Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering*, 3(2), 332-337.
- Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., ... & Sheffield, J. (2007). High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, 3, 157-165.
- Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, 40(10), 927-930.
- Phillips, C., Hales, T., Smith, H., & Basher, L. (2021). Shallow landslides and vegetation at the catchment scale: A perspective. *Ecological Engineering*, 173, 106436.
- Piciullo, L., Calvello, M., & Cepeda, J. M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, 179, 228-247.
- Priscilla, C. V., & Prabha, D. P. (2020, August). Influence of optimizing XGBoost to handle class imbalance in credit card fraud detection. *2020 third international conference on smart systems and inventive technology (ICSSIT)* (pp. 1309-1315). Tirunelveli, India, 20 – 22 August, 2020.
- Quiring, S. M., Ford, T. W., Wang, J. K., Khong, A., Harris, E., Lindgren, T., ... & Li, Z. (2016). The North American soil moisture database: Development and applications. *Bulletin of the American Meteorological Society*, 97(8), 1441-1459.
- Rahman, M. M., & Lu, M. (2015). Model spin-up behavior for wet and dry basins: a case study using the Xinanjiang model. *Water*, 7(8), 4256-4273.
- Ray, R. L., & Jacobs, J. M. (2007). Relationships among remotely sensed soil moisture, precipitation and landslide events. *Natural Hazards*, 43, 211-222.
- Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, 31(19).
- Reichle, R., Koster, R., De Lannoy, G., Crow, W., and Kimball, J. (2014). "SMAP Level 4 Surface and Root Zone Soil Moisture (L4\_SM) Data Product Algorithm Theoretical Basis Document."



- Reichle, R., De Lannoy, G., Liu, Q., Ardizzone, J., Kimball, J., & Koster, R. (2016, July). SMAP Level 4 surface and root zone soil moisture. *2016 IEEE international Geoscience and Remote sensing Symposium (Igarss)* (pp. 136-138). Beijing, China, 10 – 15 July, 2016.
- Ridler, M. E., Madsen, H., Stisen, S., Bircher, S., & Fensholt, R. (2014). Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. *Water Resources Research*, 50(11), 8962-8981.
- Schaap, M. G., Leij, F. J., & 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.
- Schmugge, T., O'Neill, P. E., & Wang, J. R. (1986). Passive microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, (1), 12-22.
- Segoni, S., Tofani, V., Rosi, A., Catani, F., & Casagli, N. (2018). Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Frontiers in Earth Science*, 6, 85.
- Sharifi, E., Saghafian, B., & Steinacker, R. J. J. O. G. R. A. (2019). Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *Journal of Geophysical Research: Atmospheres*, 124(2), 789-805.
- Shrestha, A., Nair, A. S., & Indu, J. (2020). Role of precipitation forcing on the uncertainty of land surface model simulated soil moisture estimates. *Journal of Hydrology*, 580, 124264.
- Sidle, R. C., & Bogaard, T. A. (2016). Dynamic earth system and ecological controls of rainfall-initiated landslides. *Earth-science reviews*, 159, 275-291.
- Šimůnek, J., Van Genuchten, M. T., & Šejna, M. (2016). Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone Journal*, 15(7).
- “SRTMGL30 V021.” (n.d.). *LP DAAC - SRTMGL30*, <https://lpdaac.usgs.gov/products/srtmgl30v021/> (Jan. 6, 2023).
- Stanley, T., & Kirschbaum, D. B. (2017). A heuristic approach to global landslide susceptibility mapping. *Natural hazards*, 87, 145-164.
- Sujatha, E. R., & Sridhar, V. (2021). Landslide susceptibility analysis: A logistic regression model case study in Coonoor, India. *Hydrology*, 8(1), 41.

Swain, D. L., Wing, O. E., Bates, P. D., Done, J. M., Johnson, K. A., & Cameron, D. R. (2020). Increased flood exposure due to climate change and population growth in the United States. *Earth's Future*, 8(11), e2020EF001778.

Tiwari, P., & Singh, V. (2021). Diabetes disease prediction using significant attribute selection and classification approach. In *Journal of Physics: Conference Series* (Vol. 1714, No. 1, p. 012013). IOP Publishing.

US Department of Commerce, N. O. A. A. (2022). “Historic July 26th-July 30th, 2022 Eastern Kentucky flooding.” *National Weather Service*, NOAA's National Weather Service, <https://www.weather.gov/jkl/July2022Flooding> (Dec. 28, 2022).

Utomo, D., Hu, L. C., & Hsiung, P. A. (2020, September). Deep neural network-based data reconstruction for landslide detection. In *IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium* (pp. 3119-3122). Waikola, Hawaii, USA, 26 – 2 September/October, 2020.

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.

Vinnikov, K. Y., Robock, A., Qiu, S., & Entin, J. K. (1999). Optimal design of surface networks for observation of soil moisture. *Journal of Geophysical Research: Atmospheres*, 104(D16), 19743-19749.

Wan, Z. (1999). “MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD) Version 3.3.”

Wicki, A., Lehmann, P., Hauck, C., Seneviratne, S. I., Waldner, P., & Stähli, M. (2020). Assessing the potential of soil moisture measurements for regional landslide early warning. *Landslides*, 17, 1881-1896.

Xu, X. (2020). Evaluation of SMAP level 2, 3, and 4 soil moisture datasets over the Great Lakes region. *Remote Sensing*, 12(22), 3785.

Yuan, Q., Xu, H., Li, T., Shen, H., & Zhang, L. (2020). Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental US. *Journal of Hydrology*, 580, 124351.

Zhang, D., Yang, J., Li, F., Han, S., Qin, L., & Li, Q. (2022). Landslide risk prediction model using an attention-based temporal convolutional network connected to a recurrent neural network. *IEEE Access*, *10*, 37635-37645.

Çellek, S. (2020). Effect of the slope angle and its classification on landslide. *Natural Hazards and Earth System Sciences Discussions*, 1-23.

## CURRICULUM VITAE

### Education

Ph.D. Candidate:	Spring 2019 – Spring 2023, Civil Engineering, University of Kentucky
Master of Science:	2018, Civil Engineering, University of Kentucky
Bachelor of Science:	2017, Civil Engineering, University of Kentucky

### Recent Appointments

Jul. 2023 – Present	Faculty <i>University of Kentucky – Lexington, UKY First Year Engineering</i>
Aug. 2022 – May 2023	Instructor of Record <i>University of Kentucky – Lexington, UKY First Year Engineering</i>
Aug. 2019 – May 2023	Teaching Assistant <i>University of Kentucky – Lexington, UKY First Year Engineering</i>
Aug. 2017 – Aug. 2019	Teaching Assistant <i>University of Kentucky – Lexington, KY UKY Civil Engineering</i>

### Scholarships and Awards

2012 – 2015	Provost Scholarship
2021	Tau Beta Pi Outstanding PhD Student in Civil Engineering
2021	College of Engineering Outstanding Teaching Assistant

### Recent Publications and Presentations

Francis, D. M., and Bryson, L. S. (2023). “Proposed Methodology for Site Specific Soil Moisture Obtainment Utilizing Coarse Satellite-Based Data.” *Environmental Earth Sciences*, <https://doi.org/10.1007/s12665-023-11057-0>.

Presentation: “Use of Downscaled SMAP L4 Soil Moisture Data in Landslide Stability Analysis”. (2022). 20th International Conference on Soil Mechanics and Geotechnical Engineering. Sydney, Australia, April 2022.

Francis, D., Bryson, S. L., and Imarah, M. (2020). “Use of Downscaled SMAP L4 Soil Moisture Data in Landslide Stability Analysis”. 20th International Conference on Soil Mechanics and Geotechnical Engineering. Sydney, Australia, April 2022.

Presentation: “Random Forest Based Downscaling of SMAP L4 Soil Moisture and Subsequent use in Landslide Slope Stability Analysis”. (2022). Geo-Congress, Charlotte, North Carolina, March, 2022.

Francis, D. M., and Bryson, S. L. (2022). “Geo-Congress 2022.” Random Forest Based Downscaling of SMAP L4 Soil Moisture and Subsequent use in Landslide Slope Stability Analysis, Charlotte.

Presentation: “Investigation and Comparison of Various SMAP L4 SM Downscaling Routines”. (2021). American Geophysicist Union Fall Meeting, New Orleans, Louisiana, December, 2021.

Francis, D.M., Bryson, S. L., and Dashbold, B. (2019). “Landslide Analysis Using Hydrologic Parameters Derived from Calibrated Satellite Data”. 32nd Central Pennsylvania Geotechnical Conference. Hershey, PA, September 2021.

Presentation: “Calibration of Satellite Based Hydrologic Parameters Using Ground Based Sensor Data”. (2019). American Geophysicist Union Fall Meeting, San Francisco, California, December 2019.

Presentation: “Landslide Analysis Using Hydrologic Parameters from Calibrated Satellite Data”. (2020). 31st Pennsylvania Geotechnical Conference, Hershey, Pennsylvania, April 2020.

Francis, D.M., Erhardt, G.D., and Tsang, F. (2019). “Estimating the Cost and Utility of Statewide Travel Models Using Scenario-Based Interviews”. Transportation Research Board. Washington, D.C., January 2019. (MSCE Independent Study)

Presentation: “NCHRP Project 8-36C, Task 137. Assessing the Utility and Costs of Statewide Travel Demand Models”. (2018). University of Kentucky MSCE Exit Exam, Lexington, Kentucky, November 2018.

Francis, D.M., Erhardt, G.D., and Tsang, F. (2018). “Assessing the Utility and Costs of Statewide Travel Demand Models.” NCHRP Project 8-36C, Task 137. Transportation Research Board. Washington, D.C., April 2018.

Presentation: “NCHRP Project 8-36C, Task 137. Assessing the Utility and Costs of Statewide Travel Demand Models”. (2018). University of Kentucky Transportation Joint Lab Meeting, Lexington, Kentucky, February 2018.