VARIATION OF SOIL SUCTION AND APPLICATION OF REMOTE SENSING IN EVALUATING UNSATURATED SOIL BEHAVIOR WITHIN VADOSE ZONE

By

AMIR HOSSEIN JAVID

Bachelor of Science in Civil Engineering
Amirkabir University of Technology
Tehran, Iran
2011

Master of Science in Civil Engineering
Texas Tech University
Lubbock, Texas
2017

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
MAY, 2023

VARIATION OF SOIL SUCTION AND

APPLICATION OF REMOTE SENSING IN

EVALUATING UNSATURATED SOIL BEHAVIOR

WITHIN VADOSE ZONE

Dissertation Approved:

Dr. Rifat Bulut

Dissertation Adviser

Dr. Mohamed Soliman

Dr. Gregory Wilber

Dr. Omer San

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my advisor Dr. Rifat Bulut for his wonderful guidance, endless patience, and continuous support throughout my Ph.D. pursuit. His invaluable assistance and insights led me to the writing of this dissertation. Additionally, under Dr. Bulut's fabulous mentorship and support, I was honored to receive several department-wide and campus-wide scholarships and awards.

My special gratitude is extended to Dr. Mohamed Soliman, Dr. Omer San, and Dr. Gregory Wilber, who agreed to serve on my Ph.D. advisory committee. Their invaluable feedback, support, and kindness are greatly appreciated.

I gratefully acknowledge the support by grants from (a) the National Science Foundation (OIA-1946093), (b) Oklahoma Department of Transportation (ODOT) Award SP&R ITEM NUMBER 2272, and (c) the Transportation Consortium of South-Central States (Tran-Set) Award 18ITSOKS01.

I would like to thank my family and friends for their help and continues support. I thank my loving parents, whose unconditional love, support, and constant reassurance have made me the person that I am today. I am unable to express in words my respect and love for them. I would like to also thank my sister Arghavan and my brother Ardavan for always being there and providing their help and support. I would like to present my sincere gratitude to my dear grandfather, who passed away last December, for his great role in my life. Last but not least, I want to sincerely thank my beautiful and loving wife, Homa, for her love, support, patience, and understanding. She always believed in me and inspired me to aim higher.

Name: AMIR HOSSEIN JAVID

Date of Degree: MAY, 2023

Title of Study: Variation of Soil Suction and Application of Remote Sensing in Evaluating Unsaturated Soil Behavior within Vadose Zone

Major Field: CIVIL ENGINEERING

Abstract: Moisture movement in pavements and road embankments is receiving more attention in pavement and geotechnical engineering. Water from rainfall is the primary source of moisture in soils. Following a rainstorm event, large quantities of moisture can be absorbed by the soil when the water migrates into the soil mass. The passage of moisture has an impact on the mechanical performance and functionality of the pavement infrastructure. When the pavement infrastructure is built on expansive soils, water flow can cause damage to pavements due to the swelling and shrinking of expansive soils through adsorption and desorption of moisture. Such damage can result in severe financial loss; in fact, the estimated yearly cost of damage from expanding soil problems is $2.3 billion in the United States. Oklahoma contains large expanses of medium to highly expansive clays. The state's largest cities are located in these areas, and significant and costly highway systems have been built to support the population density. These areas have relatively high average annual precipitation, which worsens the expansive clay problems. The swelling or shrinking of expansive clays causes distortion and cracking in pavements, reducing pavement service life. Thus, it is critical to understand how water moves in road embankments of expansive soils subjected to seasonal rainfall and to predict the vertical movement of pavements built on expansive soils. This study used Oklahoma Mesonet measurements to develop a data-driven statistical model for estimating soil diffusivity and soil suction in order to predict the movement of expansive soils over time. The first component of the study used unsupervised learning and a nonlinear least squares model to estimate soil diffusivity. The second component of the study presents a numerical model for predicting equilibrium suction that considers the diffusion coefficient's effects and uses surface field suction measurements. The final component of the study utilized Interferometric Synthetic Aperture Radar (InSAR) technique for effective displacement monitoring using time series of SAR data. The study investigated the performance of moisture barriers on two state highways in Oklahoma, where expansive soils are a major problem.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION, RESEARCH NOVELTY, AND OBJECTIVES 1
1.1 Introduction 1
1.2 Research Objectives 3
1.3 Technical Contributions of The Research 3

II. ESTIMATING VAN GENUCHTEN DIFFUSIVITY PARAMETER OF UNDISTURBED SOILS IN OKLAHOMA 5

2.1 Abstract 5
2.2 Introduction 6
2.3 Method and Data Analysis 12
2.3.1. Sample Size and Data Processing 13
2.3.2. Develop a Diffusivity Prediction Model 15
2.3.3. Cluster Analysis 16
2.3.4. Non-Linear Model 20
2.3.5 Accuracy Assessment 21
2.3.6 Cross-Validation 21
2.4 Results 22
2.5 Conclusion 28

III. EQUILIBRIUM SUCTION PREDICTION MODEL FOR SUBGRADE SOILS IN OKLAHOMA… 29

3.1 Abstract 29
3.2 Introduction 30
3.3 Development of A GIS-Based Contour Map of TMI 33
3.4 A Numerical Model to Determine Equilibrium Suction 38
3.4.1 Calculation of the Diffusion Coefficient 40
3.4.2 Calculation of the Matric Suction in Subgrade Soil 41
3.4.3 Calculation of the Equilibrium Suction 42
3.5 Development of Prediction Models for Equilibrium Suction 46
3.5.1 Estimating the Accuracy of the Prediction Model 49
3.5.2 Ridge Regression 50
3.6 Conclusions 52

Chapter Page

IV. MONITORING THE PERFORMANCE OF HORIZONTAL MOISTURE BARRIERS AT TWO PAVEMENT SECTIONS IN OKLAHOMA, USING SAR INTERFEROMETRY....54

4.1 Abstract....54
4.2 Introduction....55
4.3 Method and Data Analysis....60
4.3.1 Remote Sensing Data....61
4.3.2 PSI Processing....63
4.3.3 Time Series Analysis....66
4.3.4 Time Series Properties....66
4.3.5 Fitting an ARIMA Model....68
4.3.6 Anomaly Detection in Time Series....69
4.3.7 Time Series Correlation....70
4.4 Results....71
4.5 Field Validation....79
4.6 Conclusions....79

V. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK....81

Conclusions....81
Recommendations For Future Work....84

REFERENCES....85

LIST OF TABLES

Table Page

Table 2.1. First 15 rows of test data....14
Table 2.2. Descriptive statistics of clay content and Ks by cluster....24
Table 2.3. nonlinear least-squares models fit summary....25
Table 2.4. Characteristics of soil (Parker et al. 1985)....27
Table 2.5. Soil properties used on numerical simulation (Ma et al. 2009)....28
Table 2.6. Comparison of estimated diffusivity and improved method....28
Table 3.1. Analysis of Variance table....49
Table 3.2. AIC value and test MSE for two GLM models....49
Table 3.3. The comparison of estimates for coefficient standard errors....50
Table 3.4. Comparison between regression models....52
Table 4.1. Research data from Sentinel-1A....62
Table 4.2. Main characteristics of the selected Sentinel-1A master scene....62
Table 4.3. Augmented Dickey-Fuller test results for two monitoring locations....76

LIST OF FIGURES

Figure Page

Figure 2.1. Mean monthly matric suction at 5 cm depth in FAIR station, Major County, Oklahoma (2019) 13
Figure 2.2. Measured diffusivity versus measured suction in training data 15
Figure 2.3. Ks versus clay content in training data....15
Figure 2.4. The total within-cluster variation for k clusters 23
Figure 2.5. The AIC for k clusters 23
Figure 2.6. Optimal k clusters selection using 10-fold cross validation....24
Figure 2.7. Predicted nonlinear models by cluster....25
Figure 2.8. Comparison of measured diffusivity (Parker et al. 1985) and predicted diffusivity using the nonlinear model 27
Figure 3.1. GIS map of the average annual precipitation in 2019 36
Figure 3.2. GIS map of the temperature for 2019....36
Figure 3.3. GIS map of the average annual potential evapotranspiration (2019) 37
Figure 3.4. TMI contour map of Oklahoma (2019)....37
Figure 3.5. Histogram of the standard deviation of annual TMI in Oklahoma 38
Figure 3.6. Mean monthly matric suction at 5 cm depth in FAIR station, Major County, Oklahoma (2019) 42
Figure 3.7. Suction distribution profile for measured suction at 5 cm depth at FAIR station, Major County, Oklahoma (2019)....43
Figure 3.8. GIS-based contour map of equilibrium suction in Oklahoma 2016....44
Figure 3.9. GIS-based contour map of equilibrium suction in Oklahoma 2017....44
Figure 3.10. GIS-based contour map of equilibrium suction in Oklahoma 2018....45
Figure 3.11. GIS-based contour map of equilibrium suction in Oklahoma 2019....45
Figure 3.12. 10-fold cross-validation for polynomial relative humidity 48
Figure 4.1. DInSAR deformation measurement scheme (Crosetto et al. 2016) 58
Figure 4.2. Study locations coordinates and area of interest for Sentinel-1 ascending track....62
Figure 4.3. Workflows (i) and (ii) for Sentinel-1 data (Cian et al. 2019)....64
Figure 4.4. Workflow (iii): analyze deformation maps and create plots (Cian et al. 2019) 66
Figure 4.5. Standard deviation of deformation velocity map over the monitoring period (a) Madill (b) Hugo....72
Figure 4.6. LOS deformation velocity map over the monitoring period (a) Madill (b) Huge....73
Figure 4.7. Mean vertical displacement velocity at Madill Jan. 2018–Dec. 2019....73
Figure 4.8. Mean vertical displacement velocity at Hugo Apr. 2018–Apr. 2020....74

Figure Page

Figure 4.9. Sentinel-1 displacement time series at Madill for Jan. 2018–Dec. 2019 (A) control section; (B) moisture barrier left section; (C) moisture barrier right section ..74
Figure 4.10. Sentinel-1 displacement time series at Hugo for Apr. 2018–Apr. 2020 (A) control section; (B) moisture barrier left section; (C) moisture barrier right section ..75
Figure 4.11. Best subset ARIMA models selection based on BIC with the maximum AR and MA orders of 4 for Madill time series 77
Figure 4.12. Best subset ARIMA models selection based on BIC with the maximum AR and MA orders of 4 for Hugo time series 77
Figure 4.13. Sample CCF of pre-whitened time series at Madill (A) control section vs. moist. section (left); (B) control section vs. moist. section (right); (C) moist. section (left) vs. moist. section (right).....78
Figure 4.14. Sample CCF of pre-whitened time series at Hugo (A) control section vs. moist. section (left); (B) control section vs. moist. section (right); (C) moist. section (left) vs. moist. section (right).....79

CHAPTER I

INTRODUCTION, RESEARCH NOVELTY, AND OBJECTIVES

1.1 INTRODUCTION

This research focuses on three major topics: estimating van Genuchten soil-water diffusivity of undisturbed soils in Oklahoma, predicting equilibrium soil suction within the vadose zone, and monitoring and assessing vertical ground and pavement surface movements using SAR Interferometry. These three topics encompass the behavior of unsaturated expansive soils within the vadose zone with respect to moisture flow, moisture equilibrium, and vertical swelling and shrinkage.

The objective of the first topic is to establish a well-documented foundation for assessing soil suction variations within the so-called moisture active zone or vadose zone. Soil diffusivity controls the rate of suction (moisture) changes in unsaturated soils. For establishing suction envelopes and predicting swelling and shrinkage in expansive soils, it is vital that a realistic estimation of unsaturated diffusivity be made. The topic utilizes a numerical solution for estimating the soil diffusivity using the Oklahoma Mesonet matric suction measurements five centimeters below the ground surface. This study's statistical model is based on the van Genuchten (1980) approach to estimating transient changes in suction beneath an uncovered surface. The predicted diffusivity parameters are validated against laboratory measurements.

Equilibrium soil suction establishes the depth of the moisture active zone. Currently utilized computational approaches frequently fail to appropriately account for the effects of climate and soil properties for establishing equilibrium suction values. The objective of the second topic is to develop a prediction model of equilibrium suction that takes into account various climatic parameters and soil properties. Under this topic, a statistical model is presented to predict equilibrium suction by considering the effects of the diffusion coefficient and utilizing the Oklahoma Mesonet field suction measurements at five centimeters below the ground surface. The predicted equilibrium suction values are compared with the measured equilibrium suctions in the field.

Interferometric Synthetic Aperture Radar (InSAR) is a particularly intriguing method for analyzing ground surface deformations. By exploiting existing SAR data, it is possible to provide regional-scale monitoring as well as a historical assessment of deformations. The third topic in the dissertation aims to use InSAR to monitor and assess vertical ground and pavement surface movements for swelling and shrinking soils. The study evaluated the effectiveness of horizontal moisture barriers in reducing subgrade swelling and shrinkage by processing SAR satellite data. The observed vertical soil surface deformations are further assessed using time series analysis and visual investigation.

This dissertation discusses available literature, provides solutions, and presents results related to the identified knowledge gaps in the following chapters:

Chapter 2. Estimating Van Genuchten Diffusivity Parameter of Undisturbed Soils in Oklahoma

Chapter 3. Equilibrium Suction Prediction Model for Subgrade Soils in Oklahoma

Chapter 4. Monitoring The Performance of Horizontal Moisture Barriers at Two Pavement Sections in Oklahoma, Using SAR Interferometry

1.2 RESEARCH OBJECTIVES

The primary purpose of this dissertation is to develop data-driven statistics for estimating soil diffusivity, soil equilibrium suction, and the performance of moisture barriers under gradual and abrupt changes in moisture effects. In particular, the research aims at solving the following research problems:

- Developing an efficient approach for establishing soil diffusivity using a supervised learning method.
- Incorporating climate change effects and soil parameters in equilibrium suction estimation.
- Evaluating equilibrium suction for statewide Mesonet stations in Oklahoma.
- Utilizing a non-destructive displacement monitoring for highways and infrastructure using the InSAR technique.
- Quantifying the reliability of horizontal moisture barriers using a time series analysis.

Chapter II to Chapter IV of this dissertation aims to provide solutions for these objectives.

1.3 TECHNICAL CONTRIBUTIONS OF THE RESEARCH

- Developing the statistical model for assessing soil diffusivity within the so-called moisture active zone. Current methods to estimate soil diffusivity are complex, time-consuming, and require quite expensive instruments. For this reason, the determination of diffusivity has seldom been carried out. The present approach estimates diffusivity for various Oklahoma soils using nonlinear least square models based on the clustering of soil parameters. Thus, the statistical model could provide a reasonable estimate of diffusivity in situations where limited resources are available to collect test samples from the field. Employing a numerical model based on mechanics principles is presented to improve the estimate of the equilibrium suction in subgrade soils. The model in this analysis approach will allow for a more reliable

prediction of soil parameters dependent on suction within the unsaturated zone, such as resilient modulus, soil movement, unsaturated permeability coefficient (of diffusivity), shear strength, and shear wave velocity. Compared with the existing methods, the advantage of the new approach lies in the effect of the diffusion coefficient considered in the development of the model.

- Developing equilibrium suction prediction models from readily available climatic parameters (i.e., relative humidity and TMI) and soil parameters. The framework uses the Oklahoma Mesonet data and soil physical property database for the Oklahoma Mesonet to evaluate the uncertainties associated with different parameters that may affect the equilibrium suction. The presented framework uses ridge regression to reduce the effects of sampling variation.
- Employing satellite remote sensing technology, specifically SAR for ground monitoring, to assess the effectiveness of horizontal moisture barriers in controlling subgrade deformation. The presented framework combines the persistent scatterers interferometry (PSI) method and time series analysis for effective displacement monitoring of time series of SAR data. The framework employs a processing chain for monitoring ground displacement using the European Space Agency's SentiNel Application Platform and the Stanford Method for Persistent Scatterers (StaMPS).

CHAPTER II

ESTIMATING VAN GENUCHTEN DIFFUSIVITY PARAMETER OF UNDISTURBED SOILS IN OKLAHOMA

2.1 ABSTRACT

Accurate estimations of soil water diffusivity are required for defining and forecasting water movement in unsaturated soils. Moisture movement through unsaturated soils is controlled by the total suction gradient within the soil profile, with moisture migrating from regions of low total suctions to regions of high total suctions. Because soil water diffusivity is a fundamental soil hydraulic factor, it is critical to accurately measure and parameterize it in order to understand soil moisture dynamics and hydrology.

The study's objective is to establish a well-documented foundation for assessing soil suction variations within the so-called moisture active zone or vadose zone. The study utilized a supervised method for estimating soil diffusivity five centimeters below the ground surface using Oklahoma Mesonet matric suction measurements. The study presents a set of equations for estimating van Genuchten diffusivity in Oklahoma soils. The study's findings show that the prediction model produced results comparable to those obtained by the one-step pressure outflow procedure.

2.2 INTRODUCTION

Reliable estimates of soil water diffusivity are critical for describing and predicting water movement in unsaturated soils. Moisture diffusion through unsaturated soils is governed by the total suction gradient within the soil profile, with moisture moving from low to high suctions (Mitchell 1980). Therefore, soil water diffusivity is regarded as one of the most fundamental soil hydraulic parameters (Assouline et al., 1998), and accurate acquisition and parameterization of diffusivity are essential for comprehending soil moisture dynamics and soil hydrology (Jarvis et al. 1991).

The Richards Equation (Richards 1931) is widely used to describe fluid flow in unsaturated porous media (Hertaeg et al. 2020). By simplifying the Richards Equation in terms of the volumetric moisture content (θ) for spontaneous capillary flow become

$$\frac{\partial \theta}{\partial t} = \nabla.(D(\theta)\nabla\theta) \tag{2.1}$$

The moisture diffusivity equation and the moisture diffusivity function D(θ) are critical for the success of this equation as a viable model for soil moisture flow (Bear and Cheng 2010).

There are two main approaches to obtaining soil water diffusivity: the first is experimental determination, and the second is a derivation from an analytical approach. The pressure plate method proposed by (Gardner 1956) and the horizontal infiltration method proposed by Bruce and Klute (1963) is the two methods most commonly used in laboratories to determine soil water diffusivity properties. These methods rely on measurements of water content. Gardner's approach is based on measuring the outflow of water over time from a soil specimen subjected to matriculation changes. In the Bruce-Klute method, water is introduced at one end of a horizontal soil column, and measurements of water content distribution along a

horizontal soil column are taken. These methods are beset by costly, difficult, and/or time-consuming laboratory procedures and calculations.

Gardner (1962)

Gardner (1962) approximate solution to Richards' (1931) equation is the foundation for the one-step outflow concept. Assuming that (i) plate and boundary impedance are negligible and (ii) water content does not vary significantly with sample depth at any given time during the outflow process. Gardner (1962) showed that diffusivity could be calculated directly from the instantaneous flow rate, average water content, and dimensions of the sample with the expression.

$$D = -\frac{4L^2}{\pi^2(\bar{\theta}-\theta_f)}\frac{d\bar{\theta}}{dt} \tag{2.2}$$

where D is soil water diffusivity; L is sample length; $\bar{\theta}$ is average volumetric water content, which is obtained by gravimetric means; and $\frac{d\bar{\theta}}{dt}$ instantaneous outflow rate plotted and evaluated.

Passioura (1977)

Passioura (1977) proposed a new method for calculating diffusivity from one-step outflow data that can be used in the laboratory on a regular basis. The method of Passioura is based on the assumption that the water content change rate is uniform throughout the draining column of soil at any given time. Passioura's procedure determines soil water diffusivity as a function of soil water content at position $z = L$, which is the top end of the soil column.

$$D(\theta_L) \cong \frac{L^2}{2}\frac{dF}{dW} \tag{2.3}$$

Where $D(\theta_L)$ is soil water diffusivity at a water content θ_L (i.e., θ at $z = L$) and L is sample length; F is the rate of outflow, and W is the amount of water remaining in the soil at any time.

Although the experimental approach is time-consuming and expensive, it is unquestionably more exact and dependable for the diffusivity of specific soils. An assortment of empirical or semi-empirical models have been developed and revised to fit the discrete measured data (Assouline et al. 1998; Hallema et al. 2015). Among them, several models (for example, van Genuchten, Brooks and Corey) have clear physical importance and demonstrated their practicality for a wide range of soils (Ghanbarian-Alavijeh et al. 2010; Nasta et al. 2013; Sommer and Stöckle 2010).

van Genuchten (1980)

The van Genuchten (1980) equation presented the unique water retention curve for each site and depth.

$$\Theta = \left[\frac{1}{1+(\alpha\psi)^n}\right]^m \tag{2.4}$$

The parameters include θ_r ($cm^3\ cm^{-3}$), which is the residual volumetric water content (at matric potentials $\ll 0$); θ_s ($cm^3\ cm^{-3}$), which is the saturated volumetric water content; and α, n, and m, which are fitting parameters. α is the inverse of the air-entry value (or bubbling pressure), and n is a pore-size distribution index. A typical simplification followed in this study is setting $m = 1 - \frac{1}{n}$ (Schaap et al. 1998). Estimates of four van Genuchten parameters (θ_s, θ_r, α, and n) are thus required to calculate Θ given matric potential, ψ, measured by the heat dissipation sensors.

van Genuchten describes the diffusivity expression to estimate diffusivity as shown below.

$$D(\Theta) = \frac{K_s(1-m)}{\alpha m(\theta_s-\theta_r)}\Theta^{\left(\frac{1}{2}-\frac{1}{m}\right)}\left[\left(1-\Theta^{\frac{1}{m}}\right)^{-m}+\left(1-\Theta^{\frac{1}{m}}\right)^{m}-2\right] \tag{2.5}$$

van Genuchten's expression for diffusivity contains six parameters α, m, n, θ_s, θ_r, and K_s. where Θ is the normalized water content obtained from the van Genuchten equation, and K_s is the saturated hydraulic conductivity.

Parker et al. (1985)

Parker et al. (1985) studied the numerical inversion of lab pressure outflow data to learn more about the hydraulic behavior during monotonic drying. To more accurately predict transient hydraulic behavior, it will be necessary to take wetting phenomena caused by hysteresis in soil hydraulic properties. Parker et al. (1985) have compared three sums of squared deviations objective functions for parameter estimation using observed data. These observed data include (Method I) cumulative outflow with time from initial saturation following a one-step pneumatic head increment of 10 m, (Method II) cumulative outflow data supplemented by the measured water content at h = -150 m, and (Method III) equilibrium $\theta(h)$ data only without transient flow measurements. The method I produces satisfactory results in the range of water contents seen in the transient experiment. Still, it cannot be reliably extrapolated to lower water contents, especially for fine-textured soils. With only minor effects on the predictions at high θ, Method II increases the range of validity of predicted properties to low θ. Compared to Methods I and II, Method III produces a more accurate description of $\theta(h)$, but occasionally at the expense of accuracy in *K(h)* and predictions of transient flow, reflecting the biases present in the various objective functions used.

Wang et al. (2004)

Wang et al. (2004) developed a diffusivity expression based on hydraulic expressions provided by Brooks and Corey (1964) and the assumption of constant water flux proposed by Parlange (1971). Wang et al. 2004 used the expression: to estimate the soil water diffusivities

using observed data of cumulative infiltration versus time, changes in infiltration rate, and wetting front distance with time.

$$D = \frac{k_s h_d}{n(\theta_s - \theta_i)} \left(\frac{\theta - \theta_i}{\theta_s - \theta_r}\right)^{\frac{m-n-1}{n}} \quad (2.6)$$

where D is soil water diffusivity; k_s is the saturated coefficient of permeability; h_d is air-entry suction; θ_s is saturated water content; θ_r is residual soil water content; θ_i is initial water content; m and n are fitting parameters.

Aubeny and Lytton (2004)

The soil moisture diffusivity can be calculated indirectly by measuring suction changes in a soil column at various locations. The relationship can be used to test the method's accuracy (Aubeny and Lytton 2004).

$$D = \frac{k_0 h_0 \gamma_w}{c \gamma_d} \quad (2.7)$$

where k_0 is the saturated permeability of the soil, h_0 is the suction at which the soil saturates (roughly given by the air-entry value), γ_w is the unit weight of water, γ_d is the dry unit weight of soil, and c is the slope of the suction (in pF) versus the gravimetric water content curve.

Quan-Jiu et al. (2006)

Quan-Jiu et al. (2006) devised a simple approach for calculating sorptivity using hydraulic diffusivity. Water absorption experiments were performed in a one-dimensional horizontal soil column to confirm the relationship between diffusivity and sorptivity. A regression of cumulative infiltration against time from water absorption tests was performed on three different soils to test the hypothesized function between sorptivity and diffusivity. The theoretical relationship for estimating diffusivity and sorptivity (S) is shown in Equation 2.8.

$$S = ln\frac{\sqrt{2}}{\sqrt{\alpha+2}} + \frac{1}{2}lnD_s + \frac{\alpha+2}{2}ln\theta_s \quad (2.8)$$

where D_s is diffusivity constant, and α is an empirical parameter.

Ma et al. (2009)

Ma et al. (2009) presented an analytical technique based on the assumption of exponential flux distribution. They used numerical soils in a wider range of textures to improve Wang et al. (2004)'s method. The horizontal absorption equation was determined from the shape coefficient of the normalized water content profile using the assumption of exponential flux distribution. So Brooks–Corey model parameters were computed from the shape coefficient of the normalized water content profile. Soil water diffusivity can be obtained from the soil water retention curve and the unsaturated hydraulic conductivity curve.

$$D(S) = D_s S^L \quad (2.9)$$

where *S* is effective water saturation distribution using Equation 2.10:

$$S = \frac{\theta-\theta_r}{\theta_s-\theta_r} \quad (2.10)$$

D_s is saturated soil water diffusivity. Details about this model can be found in the original work cited here.

In this work, a statistical model was proposed to predict $D(\theta)$ for representing soils in Oklahoma. First, careful data mining exploration was conducted. The primary goal of this level of research is to identify any non-random patterns or structures that need to be investigated further. Clustering analysis was undertaken to organize the dataset with respect to two significantly correlated parameters in the van Genuchten diffusivity equation. Then, a nonlinear least-squares analysis was used to fit to train datasets. The accuracy of models was assessed graphically and by calculating the root mean square error of test datasets. Additionally, experimental data are

presented and compared with the $D(\theta)$ function obtained for each model, providing more information about the suitability of the methods for the whole range of soil and the correctness of the assumptions made.

2.3 METHOD AND DATA ANALYSIS

This study collected and evaluated the data obtained from Oklahoma Mesonet from 2000 to 2019. The dataset was a massive collection of climate data to assess and analyze. The Oklahoma Mesonet system is preferred in this study because the climate data is rigorously monitored and controlled. More importantly, the 121 weather stations across the state provide uniformly distributed data over short distances. These features allow the Mesonet data to be used, making it an ideal platform for developing a predictive model for moisture variations in subgrade soils.

The Oklahoma Mesonet weather stations are equipped with CSI 229-L heat dissipation sensors at depths of 5 cm. The sensors can indirectly measure the matric suctions via heat transfer, characterized by measuring the temperature difference. The temperature difference is converted into an estimate of the ceramic and soil matric suction using an empirical calibration Equation 2.11. The matric suction calibration equation is given by the following equation (Zhang et al. 2019).

$$|\Psi| = \frac{2083}{1+e^{-3.35(\Delta T_{ref}-3.17)}} \tag{2.11}$$

Where Ψ is soil matric suction (kPa) and ΔT_{ref} is reference temperature differential (°C).

Equation 2.11 was used in this study to determine surface matric suction at 5 cm depth for temperature references obtained from Mesonet Stations. Figure 2.1 shows the mean monthly matric suctions for the FAIR station, Major County, Oklahoma, during 2019.

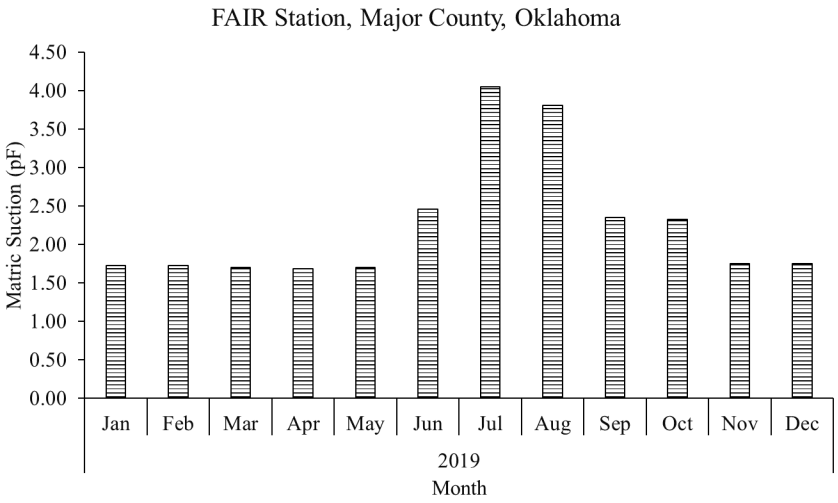

Figure 2.1. Mean monthly matric suction at 5 cm depth in FAIR station, Major County, Oklahoma (2019)

The Oklahoma Mesonet soil property database (MesoSoil) includes 13 soil properties determined using replicated samples from 545 sites and depth combinations representing 121 Oklahoma Mesonet stations (Scott et al. 2013). The database contains the percent of sand, silt, and clay; the bulk density; the volumetric water content at −33 and −1500 kPa; the van Genuchten parameters of residual volumetric water content θ_r; saturated volumetric water content θ_s ($cm^3\ cm^{-3}$), alpha α (kPa^{-1}), and n (unitless); the saturated hydraulic conductivity K_s ($cm\ day^{-1}$); as well as the matching point conductivity K_0 ($cm\ day^{-1}$) and the empirical parameter L (unitless). This study used the MesoSoil database in the development process of a prediction model to estimate the diffusion coefficient (Scott et al. 2013).

2.3.1. *Sample Size and Data Processing*

Oklahoma Mesonet started to collect heat dissipation data at the beginning of 2000, and during the past 20 years, more stations have been equipped with these sensors. This study utilized 20 years of the Mesonet dataset from 2000 to 2019 to calculate soil matric suction. Diffusivity at each station was estimated from van Genuchten's expression using van Genuchten parameters

from the Mesosoil dataset and matric suction from an earlier calculation. MATLAB was used to calculate the matric suction and diffusivity in this process. Finally, data pooling was conducted to have all the results in a single dataset. The number of rows of the pooled dataset was 697,230. Table 2.1 summaries of the first 15 rows of the test data. Figure 2.2 demonstrates the measured diffusivity versus measured suctions in training data. Figure 2.3 shows the Ks versus clay content in the training data.

Table 2.1. First 15 rows of test data

Year	Theta (predictor)	Clay (predictor)	Ks (predictor)	Diffusivity (response)
2016	0.001797123	15.3	53	1.5093
2016	0.001762528	15.3	53	1.5145
2018	0.001669934	15.3	53	1.5288
2018	0.001602878	15.3	53	1.5396
2018	0.001601559	15.3	53	1.5398
2018	0.001593419	15.3	53	1.5412
2018	0.001552607	15.3	53	1.5480
2018	0.00154707	15.3	53	1.5490
2018	0.001533344	15.3	53	1.5513
2018	0.000231341	24.1	26.8	1.5523
2018	0.000229094	24.1	26.8	1.5550
2018	0.00022881	24.1	26.8	1.5553
2018	0.000226882	24.1	26.8	1.5577
2018	0.000225594	24.1	26.8	1.5593
2018	0.000224462	24.1	26.8	1.5607

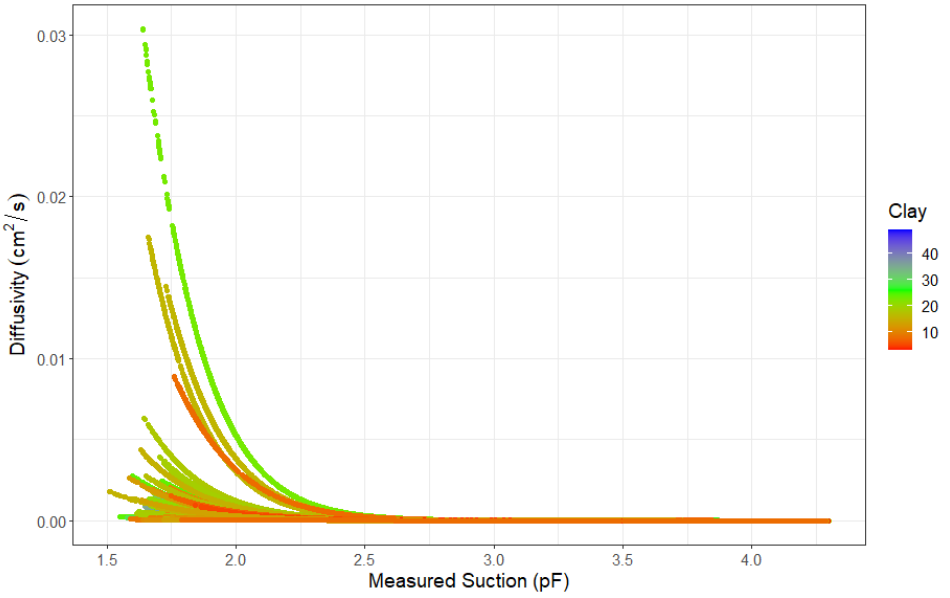

Figure 2.2. Measured diffusivity versus measured suction in training data

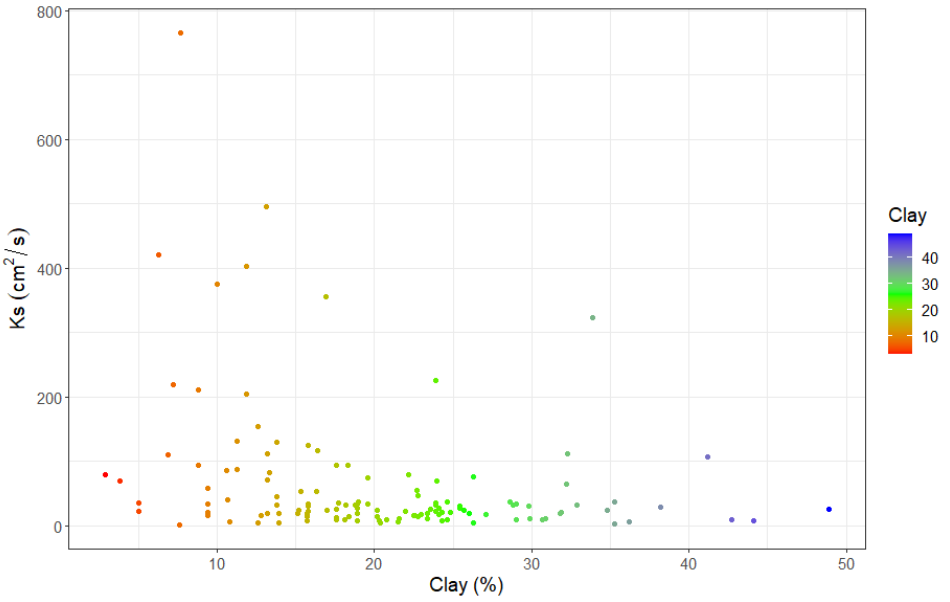

Figure 2.3. K_s versus clay content in training data

2.3.2. *Develop a Diffusivity Prediction Model*

The prediction model was developed to estimate the diffusivity from measured surface suction and soil properties. The following predictors were considered in this analysis:

- Mean daily measured surface suction (pF)
- Clay content (%)
- K_s (cm/s)

The analysis consists of three steps. First, a clustering analysis was conducted to define a group based on clay content and K_s. Clustering of clay and K_s simplify statistical model complexity. The second step was to develop a nonlinear model using the clustered training dataset to estimate diffusivity. Finally, the prediction error was estimated using the test dataset.

2.3.3. *Cluster Analysis*

Clustering is the most fundamental unsupervised learning problem. It seeks to uncover a structure (intrinsic grouping) in a collection of unlabeled data, as do all other issues of this type. Therefore, a cluster is a collection of objects that are 'similar' to one another but 'dissimilar' to objects in other clusters (Madhulatha 2012). Four representatives of the clustering algorithms are the centroid-based cluster analysis (k-means, k-medoid, etc.), connectivity-based clustering (hierarchical clustering), distribution-based clustering (Gaussian mixture models), and density-based clustering (DBSCAN, OPTICS). K-Centroids clustering provides a significant advantage over hierarchical cluster clustering. In contrast to hierarchical clustering, centroid-based clustering arranges data into non-hierarchical groupings. In K-Centroids clustering, we can constrain the algorithm to provide different centroid (similarity) measures (Euclidean, Manhattan, Jaccard, and Angle). In return, these different centroid measures allow control of clusters' shape; hence, sub-populations can be detected in data with distinctive characteristics.

The Euclidean measure is rather straightforward as the distance is calculated from the cartesian coordinates of the points using the Pythagorean theorem. Equation 2.12 shows Euclidean distance.

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.12}$$

Manhattan distance refers to the distance between two vectors if they can only move at right angles. There is no diagonal movement involved in calculating the distance. Equation 2.13 demonstrates the Manhattan distance.

$$D(x,y) = \sum_{i=1}^{k}|x_i - y_i| \tag{2.13}$$

Pearson correlation is a metric used to calculate the similarity and diversity of sample sets. The metric measures how highly correlated two variables are and is measured from -1 to +1. Similar to the modified Euclidean Distance, a Pearson Correlation Coefficient of 1 indicates that the data objects are perfectly correlated. In this case, a score of -1 means that the data objects are not correlated. Equation 2.14 shows Pearson correlation distance.

$$D(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{2.14}$$

The k-medoids algorithm (Kaufman, L., and P. Rousseeuw, 1987) is a clustering technique similar to the k-means and medoid shift algorithms. The k-means and k-medoids methods are both partitional, attempting to reduce the distance between points labeled as belonging to a cluster and a point designated as the cluster's center. Unlike the k-means algorithm, k-medoids select data points as centers (medoids) and work with any distance metrics between data points. A silhouette width is a convenient tool for calculating k. K-medoids are more resistant to noise and outliers than k-means because they minimize a sum of pairwise dissimilarities rather than the sum of squared Euclidean distances. A medoid is an object in a cluster whose average dissimilarity to all the objects in the cluster is minimized, i.e., it is the cluster's most centrally located point.

The Partitioning Around Medoids (PAM) technique is the most frequent implementation of k-medoid clustering. PAM is divided into two stages: BUILD and SWAP. The algorithm seeks a good set of initial medoids during the BUILD phase, and during the SWAP phase, all possible swaps between the BUILD-medoids and the observations take place to ensure that the objective is not reduced further (Struyf et al. 1997).

The fundamental concept of PAM includes:

1. Find a set of *k* Medoids (*k refers to the number of clusters, and M is a collection of medoids*) from the data points of *size n* (n being the number of records).

2. Using any distance metric (e.g., Euclidean, Manhattan, etc.), try and locate Medoids that minimize the overall distance of data points to their closest Medoid.

3. Finally, swap Medoid and non-Medoid pairs that reduce the loss function L among all possible *k(n-k)* pairs. The loss function is defined as:

$$L(M) = \sum_{i=1}^{n} min_{m \in M} d(m, x_i) \tag{2.15}$$

Gaussian Mixture Models are probabilistic models used to represent normally distributed subpopulations within a larger population. Two types of values are used to parameterize a Gaussian mixture model: the mixture component weights and the component means and covariances (for the multivariate case). If the number of components is known, expectation maximization (EM) is the most widely used strategy to estimate the parameters of the mixture model.

The EM algorithm consists of two steps, an E-step or Expectation step and an M-step or Maximization step. Let's say we have some latent variables $\boldsymbol{\gamma}$ (which are unobserved and denoted by the vector Z below) and our data points **X**. Our goal is to maximize the marginal likelihood of X given our parameters (denoted by the vector **θ**).

The E-Step: Gaussian Mixture distribution as a combination of Gaussians with weights equal to π as below. Where K is the number of Gaussians we want to model.

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{2.16}$$

We can calculate the posterior distribution using the formula below. Equation 2.17 is the Bayes rule, where π is the prior weights, and the likelihood is normal.

$$\Upsilon(z_{nk}) = \frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j,\Sigma_j)} \tag{2.17}$$

The M-Step: After obtaining our posterior, we estimate the parameters of each Gaussian given by the equations below and then evaluate the log-likelihood. This process is then repeated until convergence is reached. Equations 2.8 through 2.9 are the mean of the Gaussians, the covariance of the Gaussians, weights, the sum of responsibilities in each Gaussian k, and marginal likelihood.

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \Upsilon(z_{nk}) x_n \tag{2.18}$$

$$\sum_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \Upsilon(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T \tag{2.19}$$

$$\pi_k^{new} = \frac{N_k}{N} \tag{2.20}$$

$$N_k = \sum_{n=1}^{N} \Upsilon(z_{nk}) \tag{2.21}$$

$$\ln p(X|\mu, \Sigma, \pi) = \sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k\right\} \tag{2.22}$$

The DBSCAN algorithm views clusters as areas of high density separated by low-density areas. DBSCAN identifies clusters of any shape instead of k-means, which assumes clusters are convex in shape. The concept of core samples, which are samples that are in high-density areas, is essential to the DBSCAN. A cluster is thus a collection of core samples that are close to one

another (as measured by some distance measure) and a set of non-core samples that are close to a core sample (but are not themselves core samples).

2.3.4. *Non-Linear Model*

A nonlinear model is any model of the basic form.

$$Y = f(X; \beta) + \epsilon \tag{2.23}$$

where ϵ is residual with zero mean error and $f(X; p) = E(y|x)$ is non-linear with regard to the unknown parameters, $\beta = (\beta_0, \beta_1, ..., \beta_p)$.

For each cluster *G*, the following nonlinear model was fitted to the current dataset:

$$f(x) = e^{\beta_0 + \beta_1 x} \tag{2.24}$$

with β_0 and β_1 being parameters, X is theta, and y is diffusivity.

Non-linear least squares problems are generally solved by the Gauss-Newton algorithm under the assumption that the solution is unique, equivalent to minimizing the sum of squared.

Defined

$$SSE(\beta_0, \beta_1) = \sum_i (y_i - e^{\beta_0 + \beta_1 x_i})^2$$

However, a simpler approach is to focus on the fact that.

$$E(\log(y|x)) = \beta_0 + \beta_1 x$$

and estimate β_0 and β_1 by minimizing.

$$\sum_i (\log y_i - (\beta_0 + \beta_1 x_i))^2$$

If $Var(\epsilon)$ is small, both approaches will yield similar estimates. Further, least squares estimates are simpler and avoid an iterative procedure. Because our methodology is already complex and results indicate the variance is indeed small, we utilize these simpler estimates in this study.

2.3.5 *Accuracy Assessment*

The reliability of the fitting technique was assessed graphically and using the Root Mean Square Error (RMSE). RMSE is computed as follows:

$$RMSE = \sqrt{\frac{(y_i - \hat{y}_i)^2}{n_{tot}}} \tag{2.26}$$

Where n_{tot} is the number of observations, $\hat{y}_i$ and y_i are the estimated and measured diffusivity, respectively.

2.3.6 *Cross-Validation*

First, we divide the data set into training data and validation data. Here 70% of the data is training, and 30% is validation. 10-fold cross-validation is used to identify the optimal number of clusters using the training data. That is, we divide the training data into 10 datasets.

For $k = 1, 2, 3, \ldots, K$ clusters, β_0 and β_1 are estimated for each cluster sing 9 of the folds. Y is then predicted for the 10^{th} fold, and the test means the squared error is computed. This process is repeated 10 times, once for each fold, and the MSEs are averaged. Finally, for a given clustering method, *k* can be chosen to minimize this cross-validation estimate of the prediction error. A model ($\hat{k}$) is then validated by predicting the validation set and computing the RMSE.

2.4 RESULTS

First, A cluster partitioning method's primary goal is to construct clusters in such a way that the total intra-cluster variance (also known as a total within-cluster variation or total within-cluster sum of squares) is minimized. By reducing the total within-cluster variance, observations are partitioned into *k* clusters. The kth cluster's within-cluster variation is the sum of all pairwise squared Euclidean distances between observations in the kth cluster divided by the total number of observations in the kth cluster (James et al. 2013). Figure 2.4 depicts the change in total between-cluster and total within-cluster ratios for various *k* clusters. Figure 2.5 shows AIC statistics for different *k* clusters. The result of variance is explained, and AIC suggests that *k* between 5 and 8 is a reasonable candidate. To further verify the optimal *k* and the clustering method for this study, a cross-validation approach was used to find the *k*.

Recall that six clustering techniques were considered the cross-validation RMSEs for the nonlinear model, with *k* ranging from 3 to 16 clusters, which was computed. Figure 2.6 summarizes the results. Observe that DBSCAN and Kmean clustering at seven clusters provide the least test RMSEs when compared to the other four clustering methods. It should be noted that Kangle and Kmedoid had comparable test RMSEs when *k* was 16; however, since the goal was to select a reasonably parsimonies model, these models were not considered further. This study used Kmean with seven clusters for developing a prediction model. Table 2.2 shows the statistics for clay content and K_s by cluster.

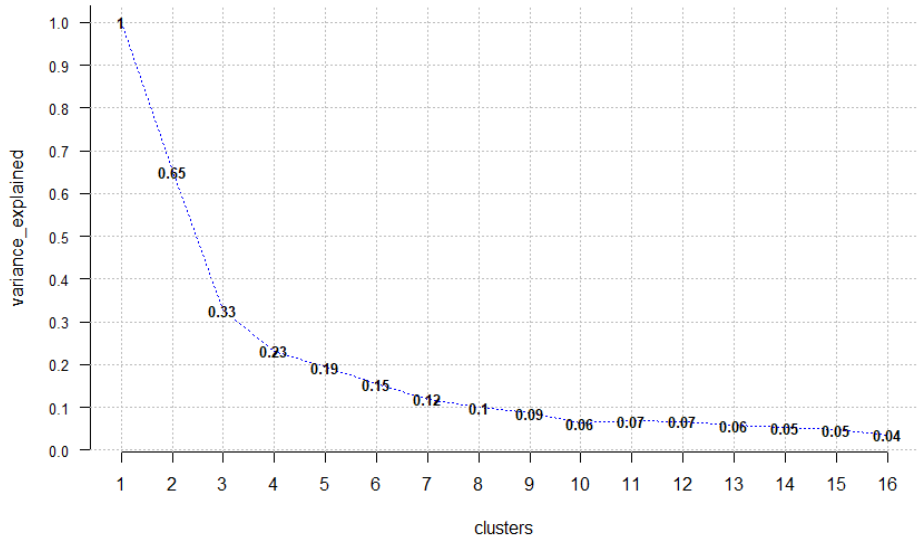

Figure 2.4. The total within-cluster variation for k clusters

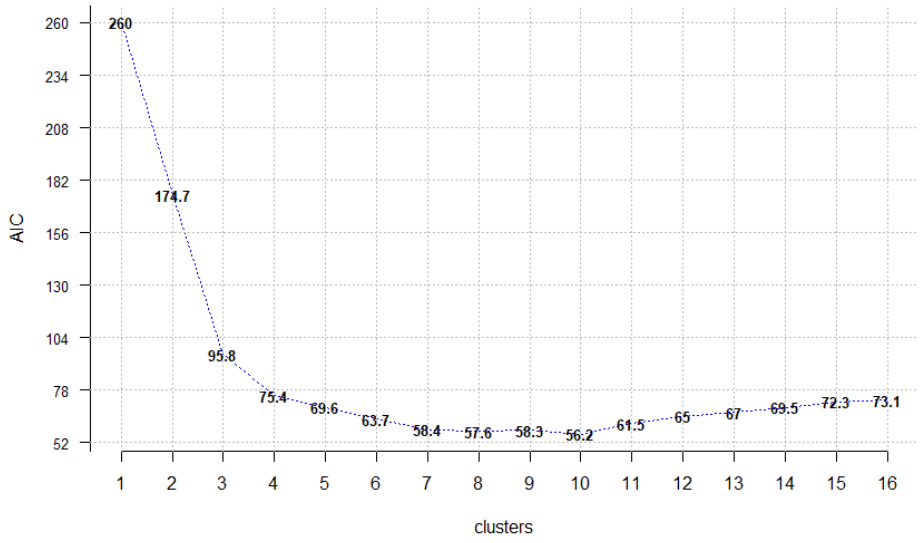

Figure 2.5. The AIC for *k* clusters

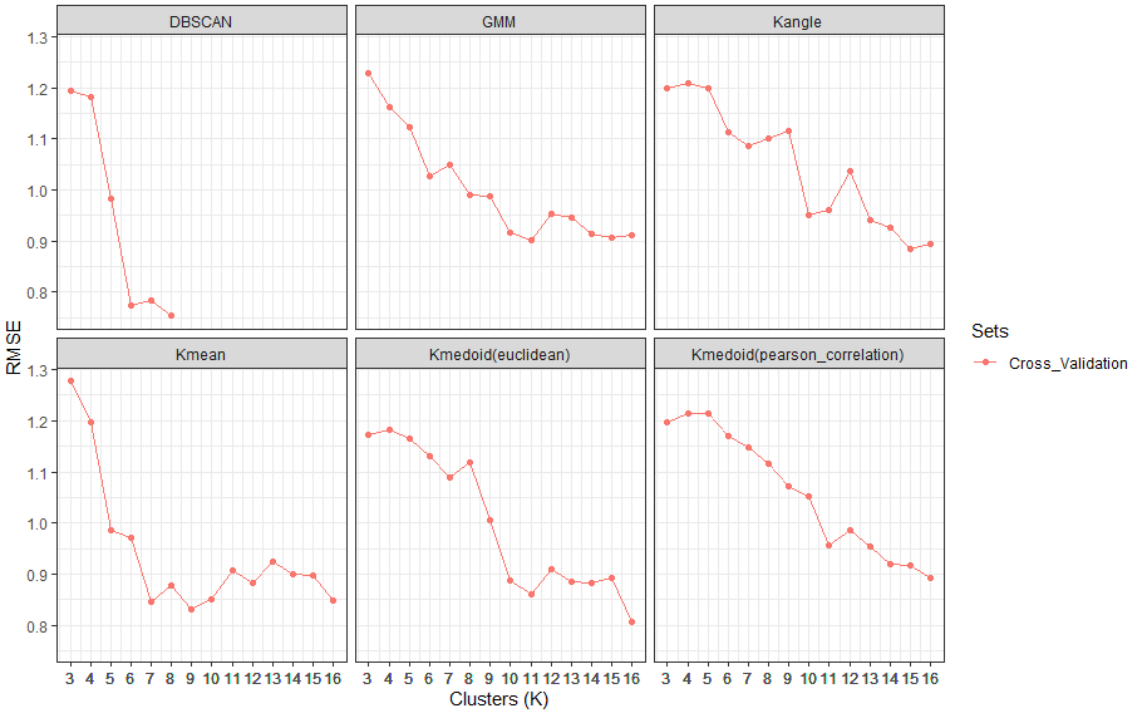

Figure 2.6. Optimal *k* clusters selection using 10-fold cross-validation

Table 2.2. Descriptive statistics of clay content and K_s by cluster

Class	N	clay content (%)				K_s (cm day^{-1})			
		Min	Max	Mean	SD	Min	Max	Mean	SD
G1	79,061	31	49	37	5	3	111	35	37
G2	8,049	24	34	25	3	225	323	238	33
G3	24,526	6	17	13	3	356	495	408	54
G4	234,229	20	30	24	3	4	78	23	16
G5	233,506	8	20	16	3	2	94	27	21
G6	113,833	3	16	11	3	22	219	117	51
G7	4,026	8	8	8	0	765	765	765	0

A nonlinear model was fitted to each class in the training set. The training set and test set, and predictive residual mean squared error were calculated. The cross-validation and test RMSEs established how well a given statistical learning procedure could be predicted to perform on

independent data. Table 2.3 summarizes the fit statistics, cross-validation RMSE, and test RMSE for different classes. Figure 2.7 visualizes the predicted lines for different clusters. The fit summary statistics show that the nonlinear model fits optimally on the train and test dataset. The low test RMSEs demonstrate that the nonlinear model established a reasonable relationship between predictor and response variables.

Table 2.3. nonlinear least-squares models fit summary

Class	Coefficients		CV RMSE	Test RMSE
	a	b		
G1	0.063936	0.0264150	1.06633	1.06343
G2	0.052633	0.0263530	0.11528	0.11750
G3	0.063840	0.0119470	1.70692	1.69251
G4	0.224140	0.0202910	0.88986	0.89628
G5	0.848200	0.0147300	1.12876	1.12861
G6	0.063808	0.0233430	1.00101	1.00223
G7	0.069184	0.0083189	0.00687	0.00686

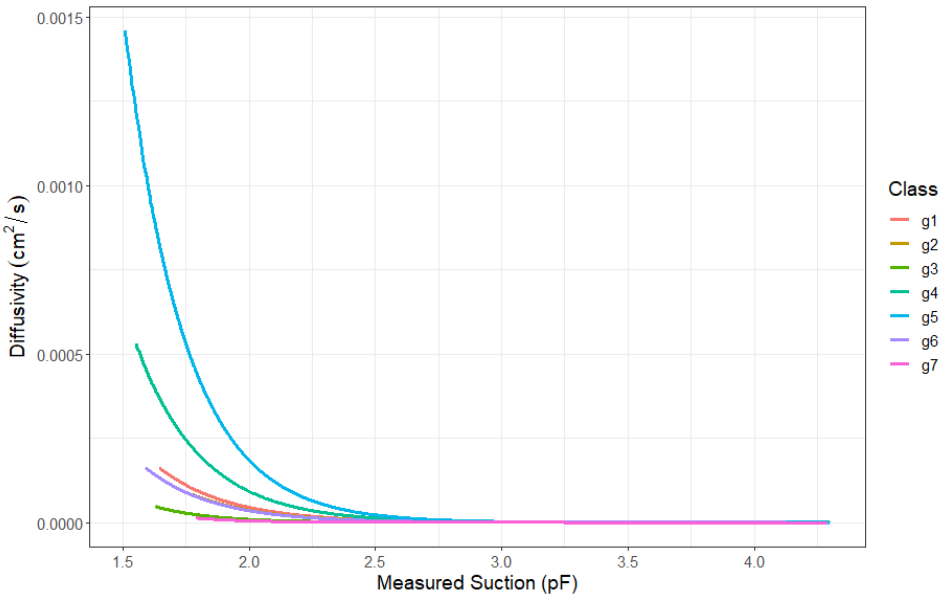

Figure 2.7. Predicted nonlinear models by cluster

Recall that a nonlinear model was fitted to each cluster in the training sets. These models were used to predict the holdout data and get each cluster's predictive residual mean squared error. Table 2.3 summarizes fit statistics, training RMSE, and holdout RMSE for different classes. Figure 2.7 visualizes the predicted lines for different clusters. The fit summary statistics show that the holdout RMSEs is reasonable. Especially for some clusters. It is not more than 1.7 and is as small as 0.006. We also see that the training and test RMSE are nearly identical. That is, overfitting is not occurring.

A comparison was made between an independent dataset from the experimental study (Parker et al. 1985). Table 2.4 summarizes the soil properties of the four samples. Comparisons of diffusivity measured by one-step pressure outflow tests with those derived from estimated parameters using the nonlinear least-squares model are shown in Figure 2.8. There is good agreement between calculated and measured diffusivity values for Clay and Silt loam soils. A slight difference was observed between predicted and measured diffusivity for Sandy clay loam at low matric suction (< 2.5 pF). The 1:1 scale plot also showed that calculated and measured diffusivity followed similar trends with no noticeable bias Figure 2.8.

The proposed statistical model was compared with the analytical model developed by Ma et al. (2009) based on the numerical simulation using HYDRUS-1D. Table 2.5 shows soil properties used in the numerical simulation study (Ma et al. 2009). Table 2.6 summarizes the comparison between the current study, HYDRUS, and the Ma et al. (2009) model. The results indicated that the analytical method predicted greater diffusivity than HYDRUS, and the prediction model has better agreement with HYDRUS.

Table 2.4. Characteristics of soil (Parker et al. 1985)

Soil	Particle size distribution			Fitting parameters*				
	Sand	Silt	Clay	Ks	α	n	θr	θs
Sandy loam	61	24	15	7.0×10^{-5}	1.3	1.51	0.18	0.36
Silt loam	28	56	15	1.5×10^{-3}	3.4	1.39	0.16	0.39
Sandy clay loam	56	18	26	1.1×10^{-4}	2.2	1.22	0.2	0.4
Clay	21	31	48	2.2×10^{-7}	0.2	1.49	0.4	0.59

* Units of K_s in cm s^{-1}, α in m^{-1}, θ_s and θ_r in m^3 m^{-3}, n dimensionless.

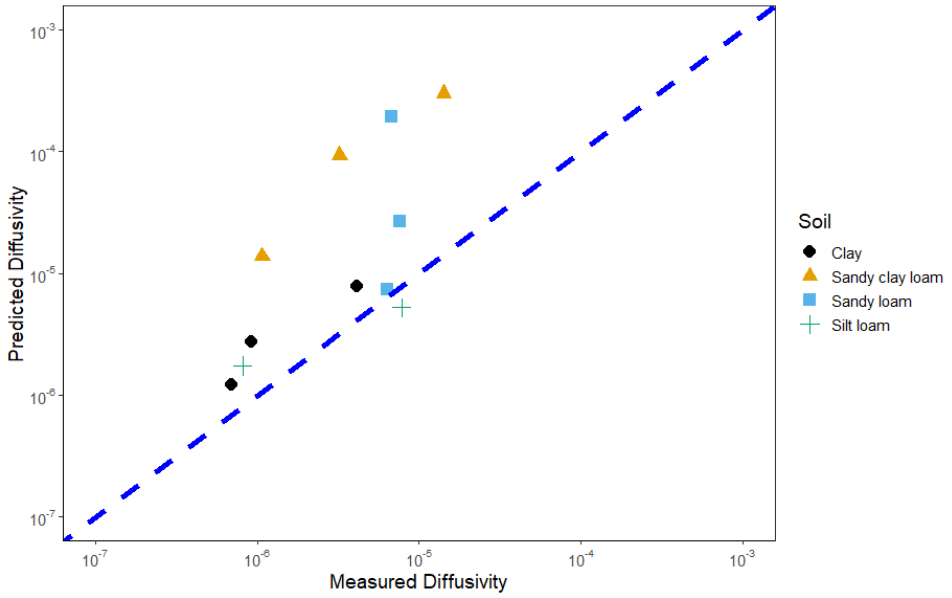

Figure 2.8. Comparison of measured diffusivity (Parker et al. 1985) and predicted diffusivity using the nonlinear model

Table 2.5. Soil properties used in the numerical simulation (Ma et al. 2009)

Soil	Fitting parameters*				
#	K_s	L	n	θ_r	θ_s
2	1.69×10^{-3}	4.11	0.474	0.035	0.401
4	3.67×10^{-4}	6.55	0.22	0.027	0.434
9	4.67×10^{-5}	8.62	0.151	0.04	0.432

* Units of K_s in cm s^{-1}, θ_s and θ_r in m^3 m^{-3}, n dimensionless.

Table 2.6. Comparison of estimated diffusivity and improved method

Soil	HYDRUS-1D		Estimated parameters	Estimated parameters
	Ma et al. (2009)		Ma et al. (2009)	(Current study)
	Matric Potential	Diffusivity	Diffusivity	Diffusivity
	pF	cm^2 s^{-1}	cm^2 s^{-1}	cm^2 s^{-1}
2	1.669	3.21×10^{-3}	9.71×10^{-6}	3.78×10^{-3}
4	2.736	1.68×10^{-4}	3.73×10^{-3}	2.43×10^{-4}
9	4.090	1.01×10^{-5}	4.48×10^{-4}	8.56×10^{-6}

2.5 CONCLUSION

The most extensively used soil-water diffusivity model is the van Genuchten model. This study proposes a set of equations for estimating the van Genuchten diffusivity for representing soils in Oklahoma. Clustering is a flexible, continuous, and automatic way to group key parameters, clay content, and saturated permeability coefficient. The nonlinear model technique produced findings comparable to those obtained by the one-step pressure outflow procedure. Determining the diffusivity by the nonlinear least-squares method is simple and relatively quick and can be carried out much easier for practitioners. It is theoretically sound and provides reasonable results with an acceptable degree of certainty and robustness.

CHAPTER III

EQUILIBRIUM SUCTION PREDICTION MODEL FOR SUBGRADE SOILS IN OKLAHOMA

3.1 ABSTRACT

A rational method for estimating soil suction variations within the so-called moisture active or vadose zone is required. Within the unsaturated zone, a soil suction envelope is defined by the maximum and minimum suctions at the ground surface (i.e., the suction amplitude), equilibrium suction, and depth to the equilibrium (or constant) suction. This paper presents a numerical model for predicting equilibrium suction that considers the diffusion coefficient's effects and uses surface field suction measurements. The rational attenuation function in this study represents the rate of suction change within the unsaturated zone. The model can predict the equilibrium suction in response to changes in the subgrade moisture regime. The geographic information system (GIS) platform was used to create contour maps of Oklahoma's equilibrium suction values. In addition, the paper presents a reliable statistical-based equilibrium suction prediction model based on readily available parameters such as Thornthwaite moisture index (TMI), clay content, and relative humidity.

3.2 INTRODUCTION

Seasonal variations in environmental factors, such as moisture content, can significantly impact subgrade performance. More than any other environmental factor, Moisture directly affects subgrade hydromechanical properties (Bulut et al. 2013; Perera et al. 2004; Puppala et al. 2009; Zapata et al. 2007). For instance, variations in moisture content have a significant impact on the state of stress and, as a result, the modulus of the subgrade (Gu et al. 2014; Khoury and Zaman 2004; Liang et al. 2008; Luo et al. 2017; Oloo 1998; Wolfe and Butalia 2004; Zhang et al. 2019), permeability (Huang et al. 1998; Mabirizi and Bulut 2011), volume change (Adem and Vanapalli 2015; Fredlund and Houston 2013), shear strength (Vanapalli et al. 1996), and shear wave velocity (Dong and Lu 2016; Ngoc et al. 2020). Compacted soils' hydraulic and mechanical behavior should also be interpreted in unsaturated soil mechanics, with suction serving as the primary stress-state variable (Banerjee et al. 2020; Han et al. 2017). As a result, determining the long-term hydromechanical properties of subgrade soils requires a realistic estimate of suction distribution beneath lightweight structures such as pavements.

The Enhanced Integrated Climatic Model (EICM), a component of the current AASHTOWare Pavement ME, is critical in defining the short and long-term pavement material properties used in the design guide. Water and heat flow through pavement layers as a result of climatic and moisture boundary conditions above and below the ground surface in pavement structures are analyzed by the EICM. Hourly climate data is one of the most critical inputs to the EICM. These are used to forecast temperature and moisture distribution in the pavement over time and depth. However, in most cases, these climate data are limited in number and region within each state and thus cannot represent site-specific climate information (Breakah et al. 2011; Bulut and Javid 2019).

Furthermore, EICM relies extensively on the ambient climatic condition (i.e., groundwater table depth) to determine the subgrade's moisture content and related suction. If the water table data is unavailable or the depth of the water table exceeds 7 m, this model will provide erroneous suction estimates (Saha et al. 2019). Previous studies have indicated that, in addition to the groundwater table, other critical components (e.g., precipitation, evapotranspiration, field capacity, and so on) influence the suction profile (Coleman 1965; Guymon et al. 1993; Russam and Coleman 1961).

Suction is composed of two components: matric and osmotic pressures. Matric or capillary suction is caused by the capillary phenomenon, which is caused by the surface tension of water. Osmotic suction occurs when there are dissolved salts in water. The matric suction in the subgrade soil reaches an equilibrium condition at a specific depth after a certain period of time following construction (i.e., active moisture zone) (Aitchison 1965; Basma and Al-Suleiman 1991).

Despite the universal acceptance of the importance of past development history on future wetting-induced soil movements, computational approaches usually fail to fully account for the effects of development on the soil's equilibrium (final) stress states (change in stress state). Incorrect assessment of the impacts of boundary conditions on equilibrium suction leads to erroneous predictions of soil heave or collapse, as well as substantial variability in heave or collapse estimates among engineering firms/geotechnical professionals (Vann 2019).

A climatic index is determined by rainfall and evapotranspiration frequency analysis and is related to the soil environment. Thornthwaite Moisture Index (TMI) (Thornthwaite 1948) and Penman Index (Penman 1963) are two climatic indices created to correlate with different yearly moisture balance indicators. These indices were discovered to be connected to the depth of the active moisture zone as well as the equilibrium suction. Gay (1995) used meteorological data

from 12 Texas sites to create a predictive model between TMI values and mean moisture depth. The Post Tensioning Institute (PTI 2004) and the Australian standard AS2870 (2011) developed a correlation between subgrade equilibrium suction and TMI (AustralianStandard 2011).

There are three approaches to estimating soil suction based on the results of conventional and widely used testing procedures. The first method is associated with soil suction "sign-posts," which assess soil suction based on varying water contents and their variance from conventional index tests. The second approach is directly related to SWCC fitting curve correlations. Much effort has been expended on SWCC fitting parameters to arrive at relevant relationships that can be utilized to estimate soil suction (Pasha et al. 2016). Recently, Pasha et al. (2017) presented a model that captures the WRC volume change dependency without incorporating new soil parameters. While "fitting parameter" methods have great analytical merits, they are highly complicated and, as a result, are not used in practice. The third approach is known as statistical relationships. Statistical relationships include, but are not limited to, relations between suction, TMI, commonly completed index property tests, gradation to a certain extent, and any combination of such values (Vann 2019).

Witczak et al. (2006) proposed a statistical model based on P200 and wPI to predict the equilibrium suction of subgrade soil, where the first predictor is the material passing through no. 200, and the second term is the interaction between P200 and the plasticity index (PI). The model was developed using a sample size of highly variable parameters, yet, not large enough to predict the equilibrium suction accurately. Vann and Houston (2021) used measured data from over 40 geotechnical studies to develop an improved second-order polynomial regression model between the 30-year TMI and equilibrium soil suction. However, the high standard error and the relative flatness of the curve relating equilibrium suction to TMI indicate a relatively weak relationship (Vann and Houston 2021). These studies found an association between equilibrium soil suction and Thornthwaite's moisture index (TMI); however, these suction correlations are highly variable.

The purpose of this study is to build a prediction model of equilibrium suction that takes into account a range of influence parameters such as TMI, clay content, and relative humidity. The primary goal of this paper is to present a statistical model for estimating equilibrium suction based on TMI and climatic variables. To generate equilibrium suction contour maps for Oklahoma, a numerical approach is used. The sections that follow explains how the TMI contour map is created using a GIS platform. This contour map provides the necessary input data for calculating the equilibrium suction of subgrade soil. A generalized linear model (GLM) and ridge regression were created and tested using a validation-based method. The bootstrap method was also used to validate the model's prediction accuracy. Finally, a ridge regression model was developed to further shrinks all regression coefficients of the GLM model. In terms of fit and test MSE, the results show that the ridge regression model outperforms the generalized linear model.

3.3 DEVELOPMENT OF A GIS-BASED CONTOUR MAP OF TMI

The TMI is a relative measure of the moisture level of a region that Thornthwaite initially introduced in 1948. The entire procedure is computationally demanding and necessitates soil and moisture storage information that may not be readily available in many locations. Many additional approaches for calculating TMI values have been proposed over the years. Thornthwaite and Mather (1955) modified the original TMI method and used a reasonably simple model to calculate adjusted potential evapotranspiration. Due to its simplicity, the technique has been widely used (Bhagat 2014; Fityus et al. 1998; Karunarathne et al. 2016; Legates and Junghenn 2018; McKeen and Johnson 1990; Sun et al. 2017; Yue and Bulut 2014).

Witczak et al. (2006) proposed a revised TMI equation based on the simplified Thornthwaite and Mather (1955) technique for use in the Mechanistic-Empirical Pavement Design Guide (MEPDG). Olaiz et al. (2018) recently conducted a comparison study between the four different TMI calculation processes developed from 1948 to 2006, Thornthwaite and Mather

(1955); Thornthwaite (1948); Willmott and Feddema (1992); and Witczak et al. (2006) and discovered that the model developed by Witczak et al. (2006) is comparable to the original TMI calculation process. As a result, this study used Witczak's model to calculate the TMI for Oklahoma. The following section provides a complete overview of the method for creating the TMI map on the GIS platform.

The GIS-based TMI contour map used in this study was created using the Oklahoma Mesonet dataset's precipitation and monthly average temperature. The Oklahoma Mesonet is a network of around 121 automated meteorological stations that send current weather and soil moisture conditions to a centralized computer facility every 5 minutes (Oklahoma Climatological Survey, climate.ok.gov). In the United States, no comparable network collects climatic data across small distances (Illston et al. 2008). One of the critical goals of the Mesonet network was to guarantee that each station location represented as much of the county as feasible. As a result, Mesonet station site sites must meet several common requirements for meteorological and environmental purposes. Figures 3.1 and 3.2 depict the annual average precipitation (*P*) and temperature contour map of Oklahoma in the GIS platform for 2019.

Using Equation 3.1, the adjusted potential evapotranspiration PET_i for the month, *i* is calculated.

$$PET_i = e_i\left(\frac{D_i N_i}{30}\right) \tag{3.1}$$

where D_i is McKeen and Johnson's (1990) day length correction factor for a month *i*; N_i is the number of days in the month *i* and e_i is the uncorrected potential evapotranspiration (cm) for the month *i* calculated as

$$e_i = 1.6\left(\frac{10t_i}{H_y}\right)^a \tag{3.2}$$

where t_i denotes the average monthly temperature in degrees Celsius, and H_y denotes the annual heat index, calculated by adding the 12 monthly heat index values. Each month's heat index is calculated as follows.

$$h_i = (0.2t_i)^{1.514} \tag{3.3}$$

where a is a coefficient calculated as follows

$$a = 6.75 \times 10^{-7} H_y^3 - 7.71 \times 10^{-5} H_y^2 + 0.017921 H_y + 0.49239 \tag{3.4}$$

The average annual potential evapotranspiration (*PE*) was calculated by PET_i summing over 12 months, as shown below. The unit of PE is centimeters.

PET_i summing over 12 months yielded the average yearly potential evapotranspiration (*PE*), as shown in Equation 3.5.

$$PE = \sum_{i=1}^{12} PET_i \tag{3.5}$$

Figure 3.3 depicts the calculated *PE* map of Oklahoma on the GIS platform for 2019.

Finally, the TMI value is calculated using Witczak et al. (2006) equation given as

$$TMI = 75\left(\frac{P}{PE} - 1\right) + 10 \tag{3.6}$$

The contour map created using the Witczak et al. (2006) method is shown in Figure 3.4.

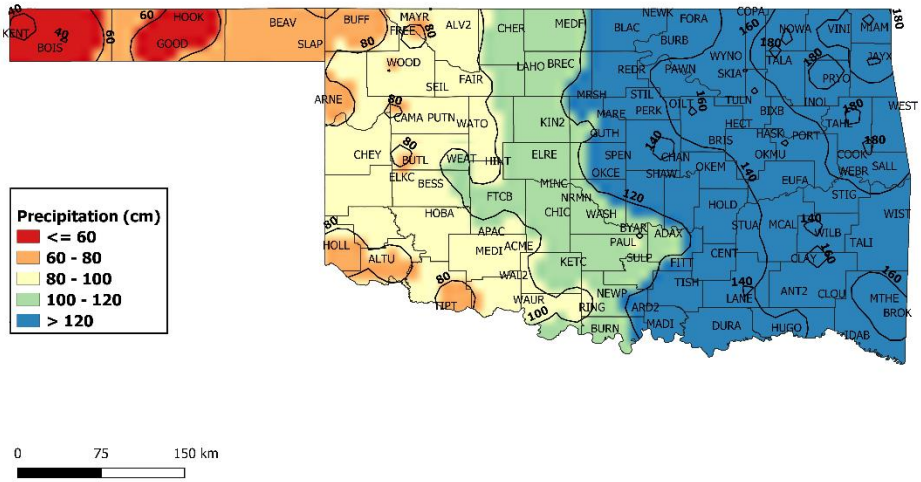

Figure 3.1. GIS map of the average annual precipitation in 2019

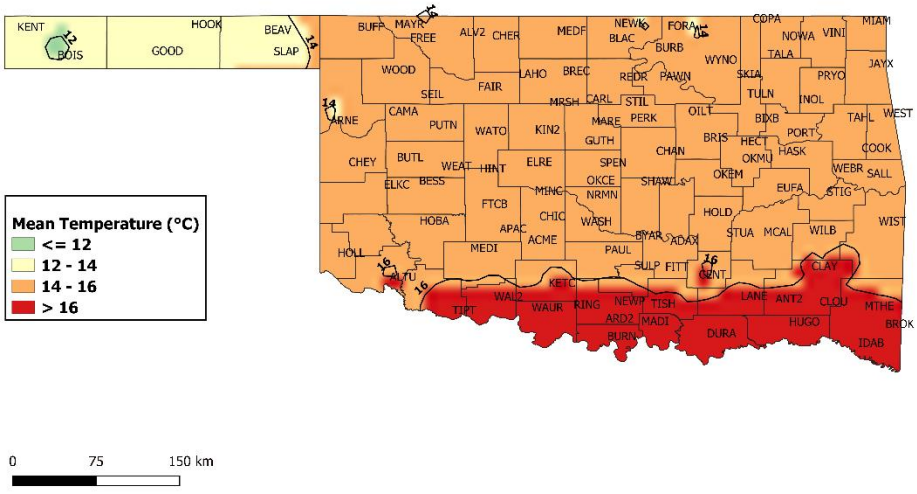

Figure 3.2. GIS map of the temperature for 2019

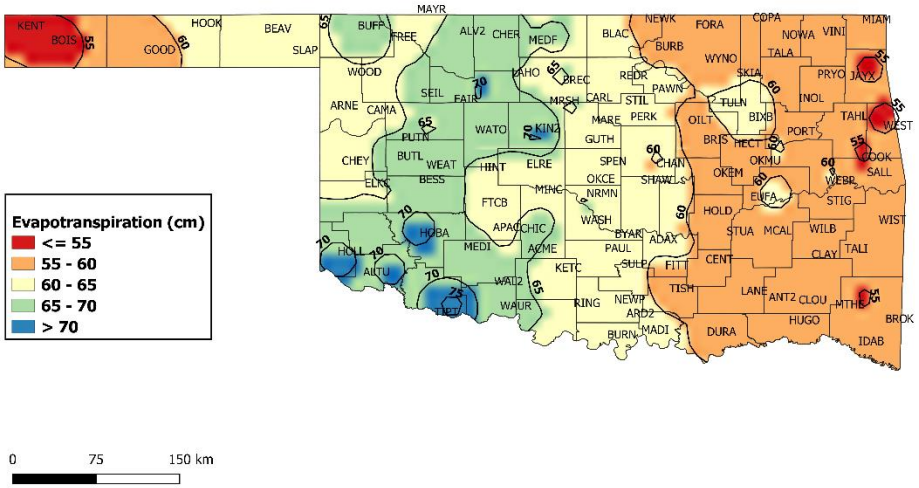

Figure 3.3. GIS map of the average annual potential evapotranspiration (2019)

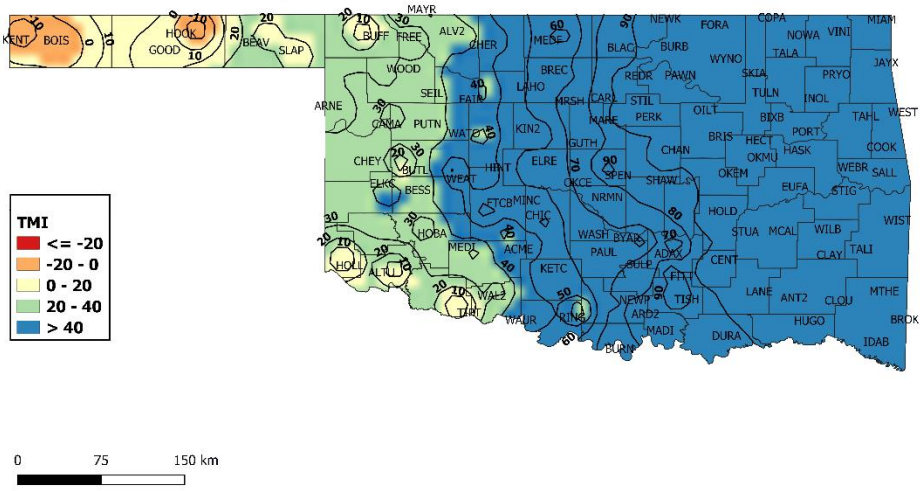

Figure 3.4. TMI contour map of Oklahoma (2019)

It is common practice to report the mean TMI based on 20 years or more extended periods. During this study project, the annual TMI was computed for 116 stations between 1997 and 2019, and univariate analysis was carried out on a sample size of 2784 observations. The histogram for the standard deviation of annual TMI for each station is presented in Figure 3.5. The histogram shows the high variability of annual TMI from year to year. In this dataset, the mean, standard deviation was 31, which indicates the average distance between the annual TMI and mean annual TMI for each station. Therefore, this study adopted the annual TMI for the correlation analysis between the TMI and equilibrium suction. The following section discusses a numerical approach to determining the soil's equilibrium suction.

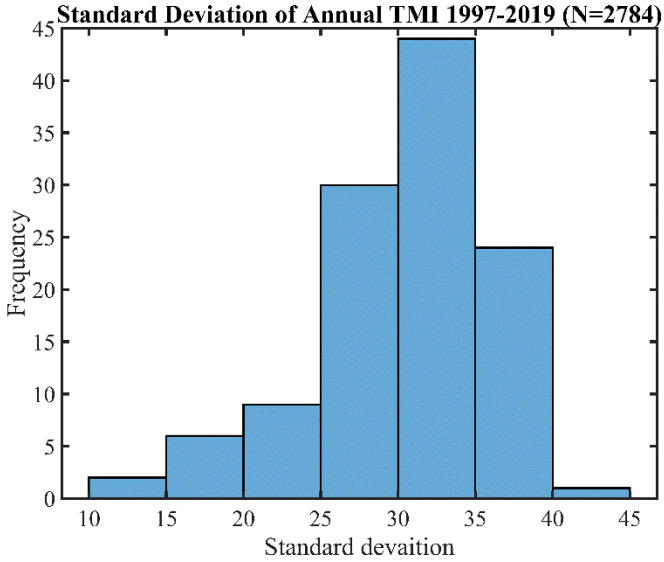

Figure 3.5. Histogram of the standard deviation of annual TMI in Oklahoma

3.4 A Numerical Model To Determine Equilibrium Suction

A Numerical Model was built to predict equilibrium suction using Mitchell's (1980) diffusion equation and a numerical solution presented by Javid and Bulut to estimate the diffusion coefficient. According to Mitchell, suction change related to climate, drainage, and site cover is a

periodic function of time. It can be calculated by solving the diffusion equation for this boundary condition. At any depth y, the suction can be expressed as

$$u(y,t) = U_e + U_o e^{-\sqrt{\frac{n\pi}{\alpha}}y} \cos\left(2\pi nt - \sqrt{\frac{n\pi}{\alpha}}y\right) \tag{3.7}$$

Where U_e represents equilibrium suction below the moisture active zone depth [pF (log kPa)], U_o represents suction variation amplitude, and *n* represents the number of suction cycles per second (1 year = 365×24×60×60 seconds). Equation 3.7 is a function of the diffusion coefficient; α and the suction reduce exponentially as depth increases.

The influence of climate change can be depicted by imposing a random suction state. A Fourier series can be computed for any arbitrary state of suction as a function of time $u(o,t)$ and for any periodic function with period $2p = \frac{1}{n}$. As a result, the total effect of $u(y,t)$ equals the sum of all partial wave effects.

$$u(y,t) = \frac{U_o}{2} + U_1 \, \mathrm{e}^{-y\sqrt{n\pi/\alpha}} \cos\left(n2\pi t - y\sqrt{n\pi/\alpha}\right) + U_2 \, \mathrm{e}^{-y\sqrt{2n\pi/\alpha}} \cos\left(2n2\pi t - y\sqrt{2n\pi/\alpha}\right) + \cdots etc. \tag{3.8}$$

Where *y* is the measured suction depth and $U_0, U_1, U_2 \ldots . etc.$ are the Fourier coefficients calculated using Equation 3.9. In this study, an eight terms Fourier series was fitted to time series data.

$$U_m = \frac{2}{P}\int_0^p u(o,t) \cos\left(\frac{2m\pi t}{p}\right) dt, \; m = 0, 1, 2 \ldots, 8 \tag{3.9}$$

Fourier coefficients can therefore be represented over a period of 12 months as

$$U_m = \frac{2}{12}\left[\int_0^1 S_1 \, cos\frac{m\pi t}{6} dt + \int_1^2 S_2 \, cos\frac{m\pi t}{6} dt + \cdots + \int_{11}^{12} S_{12} \, cos\frac{m\pi t}{6} dt\right] \tag{3.10}$$

Where monthly averages of the surface suction are $S_1, S_2, \ldots, S_{12}$.

The Mitchell (1980) formulation for estimating suction profiles was utilized in this research to estimate soil suction. Mitchell made several assumptions in simplifying Richards' equation for unsaturated flow. Mitchell, in particular, assumed that the soil's hydraulic conductivity decreased linearly with the log of suction (pF) and that the soil-water characteristic curve (SWCC) varied linearly with the log of suction. These two simplifying assumptions imply that volume change dependency of suction is not taken into account, and the diffusion coefficient (a function of both the slope of the SWCC and the soil vs. log suction) remains constant. According to research studies, Mitchell's solutions appear comparable with the gathered field data (Lytton 1994; Vann 2019).

3.4.1 *Calculation of the Diffusion Coefficient*

The diffusion coefficient (α) in unsaturated soils governs the depth of the active moisture zone (i.e., the depth of substantial seasonal moisture changes) (Aubeny et al. 2003; Mabirizi and Bulut 2010); thus, it is an essential factor in determining. In this study, a grid search method was deployed to estimate the diffusion coefficient that represented the actual field behavior of the soils by using the in-situ surface suction measurement. In the back-calculating process, a grid of diffusion coefficients ranging from 2×10^{-3} to 10^{-5} cm^2/s was created, representing the minimum and maximum nominal values for the diffusion coefficient from previous studies (Lambe and Whitman 1991; Lytton 1994; McKeen and Johnson 1990). The maximum iteration number k was 19901. The precision criteria were the absolute difference between a fitted surface suction from Equation 3.8 and a measured suction at the same depth was less than 0.001pF.

The procedure is repeated for the next value of the diffusion coefficient. This iterative procedure was repeated for all possible diffusion coefficient values. Finally, the suction profiles for each site and month could be calculated using monthly surface suction measurements and

diffusion coefficient values. Because diffusion is a time-dependent phenomenon and diffusion coefficient is a site-dependent parameter (McKeen and Johnson 1990), the numerical modeling was found to be a reasonable representative estimate of the diffusion coefficient at any specified time for which the surface suction was measured (Javid and Bulut 2019).

3.4.2 *Calculation of the Matric Suction in Subgrade Soil*

The Oklahoma Mesonet meteorological stations have CSI 229-L heat dissipation sensors at depths of 5 cm, 25 cm, and 60 cm. The sensors can indirectly measure matric suctions by detecting temperature variations, described as heat transfer. Only 70 Oklahoma Mesonet stations are equipped with CSI 229-L at depths of 25 cm and 60 cm, and there are several missing measurements at depths of 25 cm and 60 cm throughout the year. As a result, this study did not include measures of suctions at these depths. The temperature difference is used to calculate the ceramic and soil matric suction using an empirical calibration equation. The following equation (Zhang et al. 2019) gives the matric suction calibration Equation 3.11

$$|\psi| = \frac{2083}{1+e^{-3.35(\Delta T_{ref}-3.17)}} \tag{3.11}$$

Where ΔT_{ref} is the reference temperature differential (°C), and ψ is the soil matric suction (kPa).

For temperature references acquired from Mesonet Stations, Equation 3.11 was utilized to calculate surface matric suction at 5 cm depth. Figure 3.6 shows the mean monthly matric suction for the FAIR station in Major County, Oklahoma, in 2019.

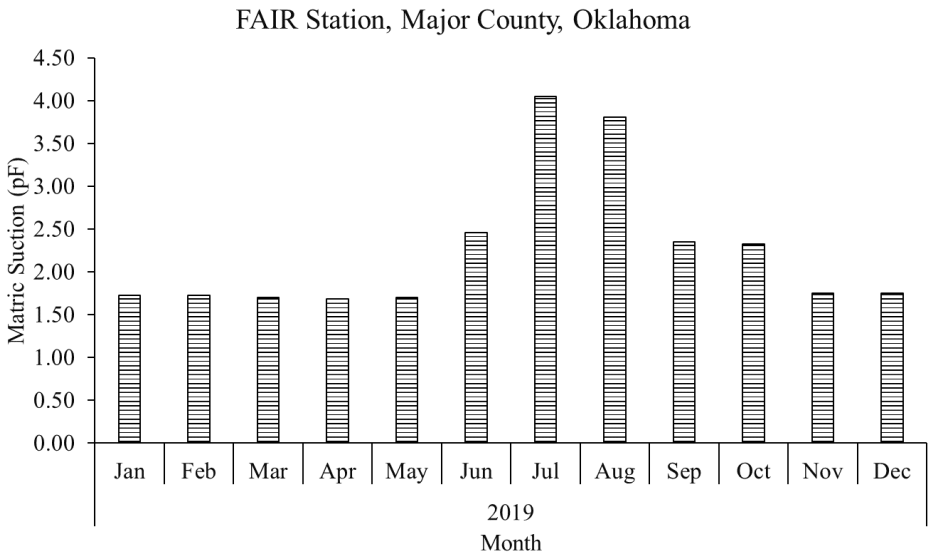

Figure 3.6. Mean monthly matric suction at 5 cm depth in FAIR station, Major County, Oklahoma (2019)

3.4.3 *Calculation of the Equilibrium Suction*

The equilibrium suction at any Mesonet station can be determined by inspecting the suction profile. The suction profile was calculated using a Mitchell model. The suction at a convergence depth was used to determine the equilibrium suction for a specific site. The convergence, defined as the absolute difference between wet and dry envelopes, was less than 0.01 pF. Figure 3.7 depicts suction profiles at the FAIR station for surface suction measurement at 5 cm from March to October 2019. The equilibrium suction at the FAIR station in 2019 was found to be 2.25 pF.

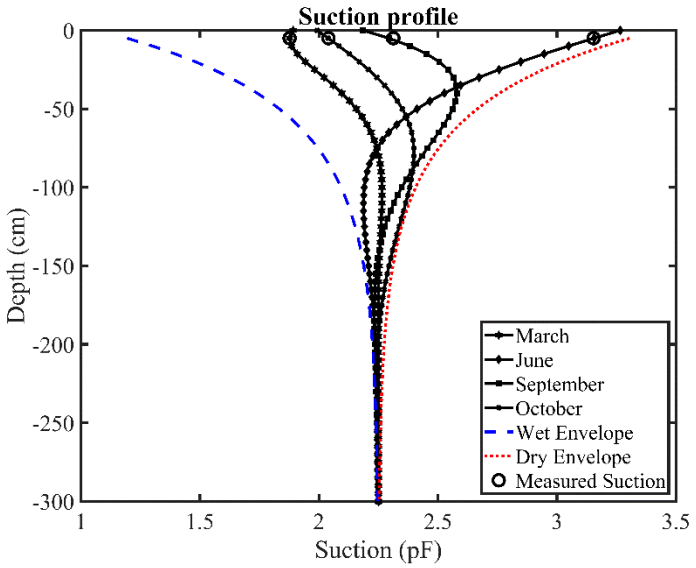

Figure 3.7. Suction distribution profile for measured suction at 5 cm depth at FAIR station, Major County, Oklahoma (2019)

The numerical model facilitates the calculation of the equilibrium suction for all counties in Oklahoma. The geographic information system (GIS) platform was used to create contour maps of the equilibrium suction values for Oklahoma from 2011 to 2019. Because of space limitations, only contour maps from 2016 to 2019 are shown in Figures 3.8 to 3.11. Two statistical models were developed to estimate the equilibrium suction from meteorological factors and soil types and simplify the calculation process. The following section discusses this relationship and its evolution.

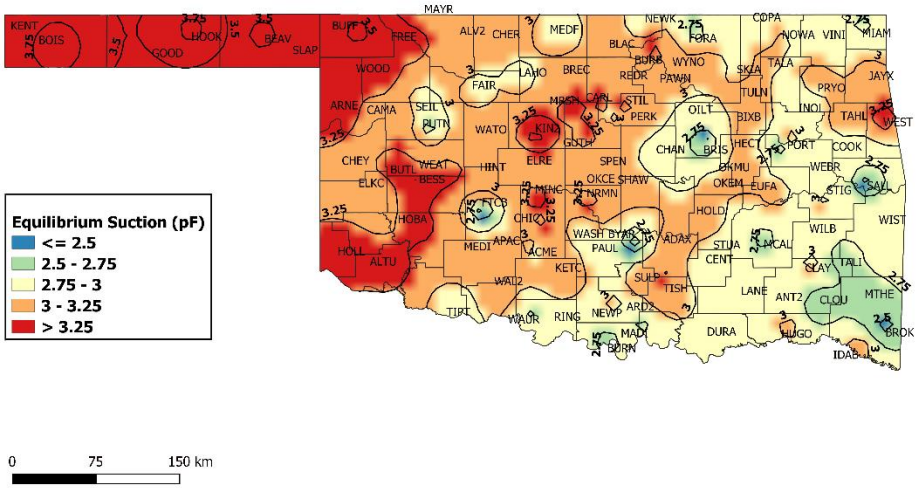

Figure 3.8. GIS-based contour map of equilibrium suction in Oklahoma 2016

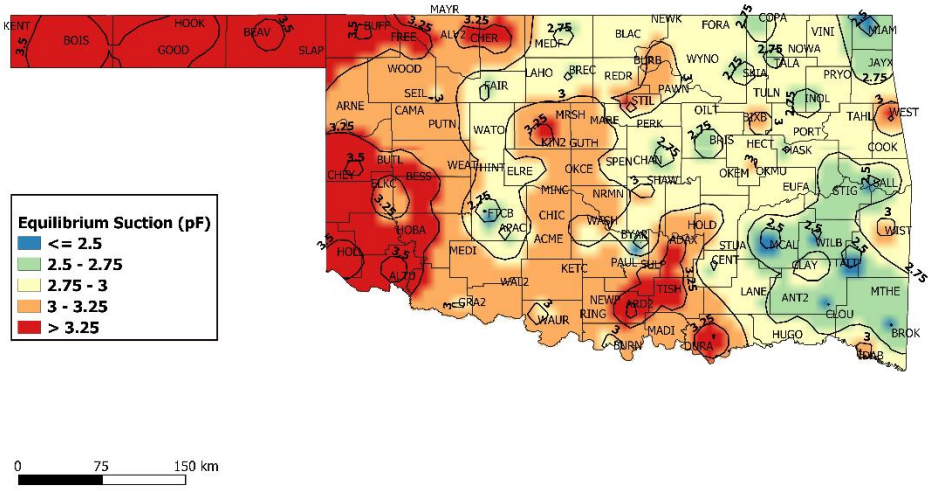

Figure 3.9. GIS-based contour map of equilibrium suction in Oklahoma 2017

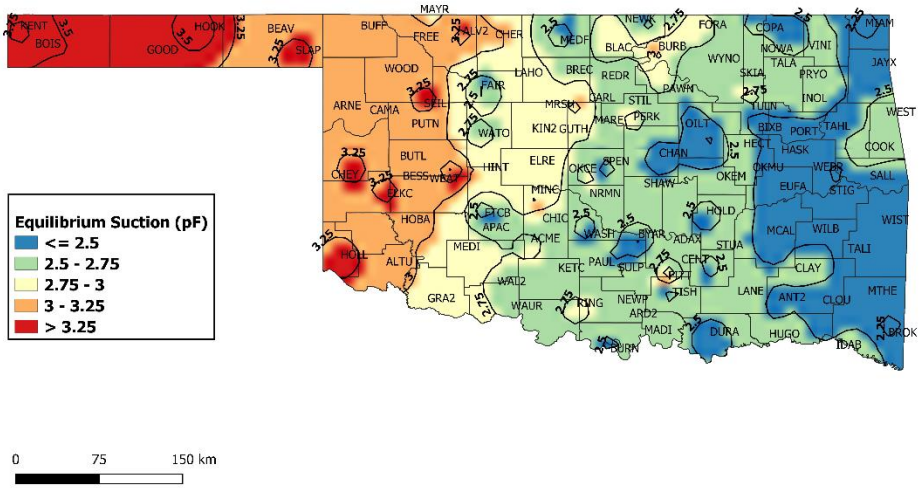

Figure 3.10. GIS-based contour map of equilibrium suction in Oklahoma 2018

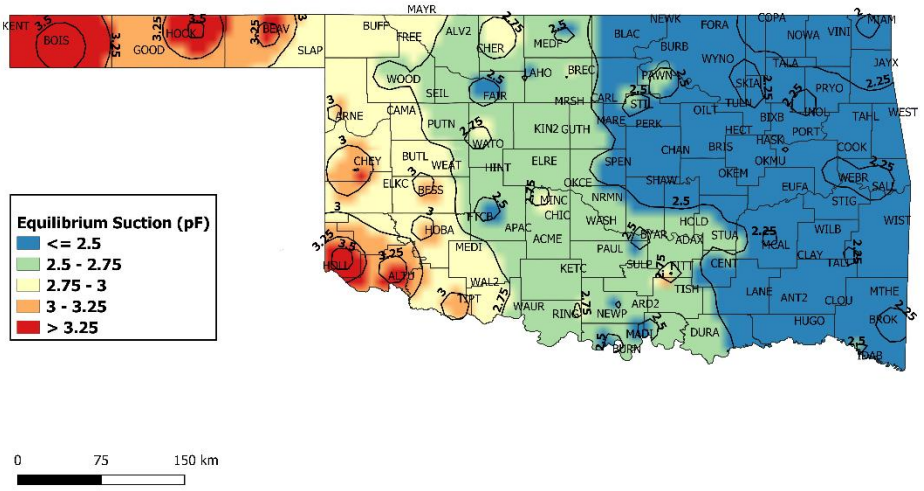

Figure 3.11. GIS-based contour map of equilibrium suction in Oklahoma 2019

3.5 DEVELOPMENT OF PREDICTION MODELS FOR EQUILIBRIUM SUCTION

A GLM and ridge regression models were built to estimate the equilibrium suction in the subgrade soil using climate factors and soil properties. In this analysis, the following predictors were taken into account.

- Mean annual relative humidity,
- Mean annual temperature,
- Annual rainfall,
- Mean annual percent sunshine,
- Clay content (%) at 5cm depth,
- Annual TMI

The climatic data for this study were gathered from the Oklahoma Mesonet stations. The clay contents were obtained using the published data (Scott et al. 2013). As stated in the previous section, the equilibrium suction dataset utilized in the correlations was obtained from the mechanical-numerical model. The annual TMIs from 1994 to 2019 were calculated, as explained in the previous section. This study has a sample size of n = 2096 after removing missing values and the data cleaning process.

In the correlation process, a partial correlation coefficient was calculated for each variable to prevent the effects of confounding variables. The matric suction, for example, decreased when the annual mean relative humidity increased. Suction is directly proportional to clay content (CC), and as clay content rises, matric suction also increases. This study evaluated the correlation between the equilibrium suction and clay content at 5 cm depth. The current research reveals that relative humidity (RH), clay content at 5 cm, and TMI are statistically significant variables that influence equilibrium suction (pF).

The data in this study were fitted with a generalized linear regression model (GLM). The GLM relies on large-sample approximations since it utilizes maximum likelihood estimation (MLE) instead of ordinary least squares (OLS) to estimate the parameters. Errors in the GLM must be independent, but they do not have to be normally distributed. Finally, the independent (explanatory) variables could be the original independent variables' power terms or other nonlinear transformations. *K*-fold cross-validation was carried out to get an accurate estimate of the MSE test, which was done to determine how well a particular statistical learning method is likely to perform when applied to independent results. This method randomly divides the set of observations into *k* groups, or folds, of roughly equal size (James et al. 2013). The first fold is treated as a validation set, and the method fits on the remaining $k - 1$ folds. The mean squared error, MSE1, is determined by the observations in the held-out fold. This technique is repeated *k* times, with a different group of observations serving as a validation set each time. This procedure yields k estimates of the test error, denoted by MSE_1, MSE_2, ..., MSE_k. Averaging those values yields the *k*-fold CV estimate (James et al. 2013).

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k} MSE_i \tag{3.12}$$

Several generalized linear models with a different polynomial order were developed for each predictor to find the degree of correlation between predictors and the response variable. After calculating the test MSE for each model, the model with the smallest test MSE was selected for each predictor. This iterative procedure was repeated for increasingly complex polynomial fits with a 10-fold cross-validation estimate for different GLMs. In this study, the associated cross-validation error was computed for the iteratively fits of the polynomial regressions of order $i = 1$ to $i = 5$. The procedure revealed that TMI and clay contents had linear relations with the response variable. However, the quadratic relative humidity fit has a lower MSE than the linear fit. The estimated test MSE is sharply reduced between the linear and quadratic fits, as shown in Figure 3.12; there is no significant improvement when higher-order polynomials are used.

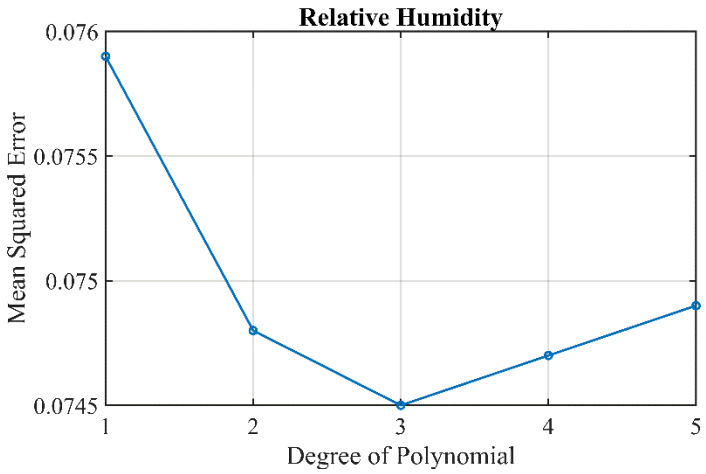

Figure 3.12. 10-fold cross-validation for polynomial relative humidity

Two different models were fitted to the data and compared against each other to select a model with better prediction accuracy and interpretability. Model 1 has three predictors, clay content, relative humidity, and TMI.

$$Equilibrium\ Suction = \beta_0 + \beta_1\ (CC) + \beta_2\ (RH) + \beta_3\ (TMI) \tag{3.13}$$

Model 2 consists of four predictors, including three predictors similar to model 1 and a second-degree relative humidity term.

$$Equilibrium\ Suction = \beta_0 + \beta_1\ (CC) + \beta_2\ (RH) + \beta_3\ (TMI) + \beta_4\ (RH^2) \tag{3.14}$$

An analysis of variance (ANOVA, using an F-test) was used in this study to compare the null hypothesis that Model 1 is sufficient to explain the data to the alternative hypothesis that a more complex Model 2 is required. Model 1 and Model 2 must be nestled to use the ANOVA. That is, predictors in Model 1 must subset of predictors in Model 2. In this case, two different models were fitted, and the simpler model was sequentially compared to the more complex model (Table 3.1).

Table 3.1. Analysis of Variance table

	Residual Df	RSS	Df	Sum of Square	F	P-Value
Model 1	2092	143.85				
Model 2	2091	142.91	1	0.9421	13.784	2.104×10^{-4}

The p-value, which compares linear Model 1 to quadratic Model 2 (2.104×10^{-4}), indicates that a linear fit is insufficient. The overall fit of models was also ranked according to their Akaike Information Criterion (AIC) value and test MSE to select the best model further. Table 3.2 summarizes the AIC values for the two GLM models. Model 2, with four predictors, has the lowest AIC and test MSE and is a better predictive model.

Table 3.2. AIC value and test MSE for two GLM models

	AIC	Test MSE
Model 1	169.39	0.0753
Model 2	161.55	0.0737

3.5.1 *Estimating the Accuracy of the Prediction Model*

Bootstrap is a versatile and effective statistical tool for quantifying the uncertainty associated with a particular estimator or statistical learning procedure. Efron (1992) was the first to present the bootstrap methodology in can be obtained by Efron and Tibshirani (1994). The nonparametric and parametric bootstrap approaches are the most common. The nonparametric bootstrap makes no assumptions about the population distribution and generates a sample of size *n* replacement from the original sample. The bootstrap sample is then used to evaluate the estimates' variability.

This procedure is then performed several times, yielding *B* bootstrap estimates. The bootstrap estimates are then used to calculate the estimated bias and standard error.

This study utilized the nonparametric bootstrap to determine the uncertainty of the coefficient estimates and predictions from a statistical learning method. In this analysis, the bootstrap sample was used to assess the variability of the estimates for β_0, β_1, β_2, β_3, and β_4, the intercept and coefficient terms for the quadratic model (James et al. 2013). The simulation process for 1,000 paired observations was repeated 100,000 times to estimate the standard error of coefficients, i.e., the intercept and slope terms. Table 3.3 shows estimates of the standard intercept error and the bootstrap coefficient terms relative to those obtained from the GLM model. The results show that the quadratic GLM fit coefficients standard errors and the bootstrap estimates are very similar.

Table 3.3. The comparison of estimates for coefficient standard errors

	Coef. Estimate	GLM Std. Error	Bootstrap Std. Error
Intercept	4.8863	0.13540	0.15309
Clay content (CC)	0.0082	0.00087	0.00067
Relative Humidity	-0.0278	0.00209	0.00177
TMI	0.0051	0.00234	0.00183
Relative Humidity^2	-0.0001	0.00003	0.00002

3.5.2 *Ridge Regression*

Ridge regression is a shrinkage method that limits or regularizes coefficient estimates. Shrinking the estimates of the coefficient could substantially reduce their variance and improve the fit. Ridge regression, like least-squares, seeks coefficient estimates that fit the data well by making

the RSS small. Especially the estimates of the ridge regression coefficient are the values that minimize the penalized sum of squares (James et al. 2013).

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}{\beta_j}^2 = RSS + \lambda\sum_{j=1}^{p}{\beta_j}^2 \quad (3.15)$$

Where $\lambda \geq 0$ is a tuning parameter. The second term, $\lambda\sum_{j=1}^{p}{\beta_j}^2$, which is called a shrinkage penalty, is small if $\beta_1, \ldots, \beta_p$ are close to zero and therefore have the effect of shrinking the estimates of β_j towards zero. It is important to note that the shrinkage penalty does not apply to intercept, which is simply a measure of the mean response value when all predictors are set to zero. The tuning parameter λ controls the relative impact of these two terms on the regression coefficient estimations (James et al. 2013).

The ridge regression was used for this analysis to improve the quadratic fit to the dataset. The validation set and cross-validation approaches were used to provide reliable estimates of the ridge regression test error. The observations are divided into a training set and a test set in the first step. The training sample was chosen randomly from 70 percent of the total data set in this analysis, and the test observations were selected from the remaining 30 percent of the dataset.

The best value for the tuning parameter λ was selected using the cross-validation on the training set. We have chosen to implement a grid of λ values ranging from 10^{-2} to 10^{10}, covering mainly the full range of scenarios from the null model containing only the intercept to the full model and calculating the cross-validation error for each value of λ. The value of the tuning parameter with the minimum cross-validation error was chosen. Finally, the model was re-fitted using all of the available observations, and the tuning parameter value was determined via cross-validation.

Table 3.4 summarizes the results of the ridge regression. In ridge regression, each least squares coefficient estimate is shrunk by the same proportion. The performance of the developed

regression model in this study was compared with those proposed by Cuzme (2018), Saha et al. (2019), and Vann and Houston (2021). The test MSE and AIC for all models were calculated and summarized in Table 3. The results showed that the ridge regression model with an optimum choice of λ had outperformed the regression model developed by Cuzme (2018), Saha et al. (2019), and Vann and Houston (2021) with lower AIC and test MSE.

Table 3.4. Comparison between regression models

Coef. Estimate	Ridge regression model (current study)	Regression model (Cuzme 2018)	Regression model (Saha et al. 2019)	Regression model (Vann and Houston 2021)
Test MSE	0.0726	0.1142	0.0792	0.0864
AIC	93.36	572.96	265.39	339.94

The derived ridge regression equation for equilibrium suction is written as follows.

$$U_e = 4.6499 + 0.0071\ (CC) - 0.0236\ (RH) - 0.0014\ (TMI) - 0.00002\ (RH \times TMI) \quad (3.16)$$

3.6 CONCLUSIONS

An improved prediction model based on mechanics principles is proposed to improve the estimate of the equilibrium suction in subgrade soils. The model used in this analysis approach will allow for a more accurate prediction of suction-dependent soil parameters within the unsaturated zone, such as resilient modulus, soil movement, unsaturated permeability coefficient (of diffusivity), shear strength, and shear wave velocity.

Compared with the existing methods, the advantage of the new method lies in three categories. First, the effect of the diffusion coefficient is considered in the development of the Mitchell model to calculate equilibrium suction. Furthermore, a large cluster of raw climate and soil moisture data was obtained from the Oklahoma Mesonet for evaluation. Finally, the new

method can be conveniently programmed into computer codes using the formulation presented. The following is a list of the study's important contributions.

- Oklahoma's GSI-based TMI contour maps were created using precipitation and temperature data from Oklahoma Mesonet.
- A numerical model was used to estimate equilibrium suction based on Mitchell's steady-state diffusion equation and the diffusion coefficient calculated from recorded soil moisture data.
- A GSI-based equilibrium suction contour map of Oklahoma was created by estimating values for 116 representative Mesonet stations in Oklahoma. As a result, the contour map can be used to determine the equilibrium suction value.
- A ridge regression model was developed to predict equilibrium suction based on readily available climatic variables (i.e., relative humidity and TMI) and clay content.

It should also be emphasized that the equilibrium suction prediction model is based on Mesonet meteorological data collected at specific locations in Oklahoma under specified weather and drainage conditions. Changes in drainage (i.e., ponding) or soil fabric (i.e., deeper fissures) are predicted to result in different predictions. More field observations and research are required to expand the database to include additional climates, drainage conditions, and soil fabrics.

CHAPTER IV

MONITORING THE PERFORMANCE OF HORIZONTAL MOISTURE BARRIERS AT TWO PAVEMENT SECTIONS IN OKLAHOMA USING SAR INTERFEROMETRY

4.1 ABSTRACT

This study will look into the capability of remote sensing satellite data, including Synthetic Aperture Radar (SAR) satellite data, to monitor the performance of horizontal moisture barriers. SAR is a tangible advancement in remote sensing technology that enables millimetric accuracy in assessing pavement deformations at single points. Recent advances in remote sensing satellite systems and the availability of high-resolution SAR systems have created an opportunity for SAR-based pavement and infrastructure management monitoring. The Sentinel-1 satellite constellation has globally and freely provided frequent and abundant SAR data and enabled nationwide deformation monitoring through InSAR time series analysis.

The study was carried out on two state highways in Oklahoma, where the amount of deformation was determined and calculated using the Persistent Scatterer InSAR method. Displacement values were estimated by reducing error sources related to temporal and geometrical decorrelation and atmospheric phase delay. Based on vertical deformation data, a detailed time series analysis was performed. The time series analysis results revealed that the moisture barriers section is not different from the controlling section at two research sites during the monitoring period. A field investigation was conducted and further validated the findings of this study.

4.2 INTRODUCTION

Persistent Scatterer Interferometry (PSI) is a powerful remote sensing technique that can measure surface displacements on the Ground (Crosetto et al. 2016). The terms displacement and deformation are used interchangeably in this work. As it is later described in greater detail, PSI is a subclass of Differential Interferometric Synthetic Aperture Radar (DInSAR) techniques. These methods make use of the information contained in the radar phase of at least two complexes of SAR images acquired at different times over the same area and used to form an interferometric pair.

The DInSAR principle is summarized below. Taking into account a single pixel footprint on the ground P, the sensor obtains the first SAR image from a satellite position M, measuring a phase φ_M

$$\varphi_M = \varphi_{geom-M} + \varphi_{scatt-M} = \frac{4.\pi.MP}{\lambda} + \varphi_{scatt-M} \tag{4.1}$$

where MP is the sensor to target distance, φ_{scatt} is the phase shift generated during the interaction between the microwaves and the target P, λ is the radar wavelength, and factor 4π is related to the two-way path, radar-target-radar. Assuming that the sensor acquires a second image from a satellite position S, measuring the phase φ_S over the same pixel footprint P, then

$$\varphi_S = \varphi_{geom-S} + \varphi_{scatt-S} = \frac{4.\pi.SP}{\lambda} + \varphi_{scatt-S} \tag{4.2}$$

The Interferometric SAR (InSAR) technique exploits the phase difference $\varphi_S - \varphi_M$

$$\Delta\varphi_{Int} = \varphi_S - \varphi_M = \frac{SP-MP}{\frac{\lambda}{4.\pi}} + \varphi_{scatt-S} - \varphi_{scatt-M} \tag{4.3}$$

This phase is known as the interferometric phase. It is related to the distance difference $SP - MP$, which is essential for the generation of Digital Elevation Models (DEMs), i.e.,

estimating the topography of the observed scene see (Bamler and Hartl 1998) and (Rosen et al. 2000). The satellite baseline SM determines the sensitivity of InSAR to topography. More specifically, the projection of *SM* in the direction perpendicular to the SAR Line-Of-Sight (LOS) is known as a perpendicular baseline. In the case of DInSAR deformation measurement, a phase φ_M is measured using a single-pixel footprint P and a first acquisition from the satellite position *M* (see Figure 4.1). Then, assuming that the target moves from P to P' and that the sensor then acquires a second image from the satellite position *S*, φ_S is calculated as follows.

$$\varphi_S = \varphi_{geom-S} + \varphi_{scatt-S} = \frac{4.\pi.SP'}{\lambda} + \varphi_{scatt-S} \tag{4.4}$$

The interferometric phase $\Delta\varphi_{Int}$ in this case, is given by

$$\Delta\varphi_{Int} = \varphi_S - \varphi_M = \frac{SP'-MP}{\frac{\lambda}{4.\pi}} + \varphi_{scatt-S} - \varphi_{scatt-M} \tag{4.5}$$

The following equation is generated by adding and subtracting the expression $\frac{SP}{\frac{\lambda}{4.\pi}}$

$$\Delta\varphi_{Int} = \varphi_S - \varphi_M = \frac{SP-MP}{\frac{\lambda}{4.\pi}} + \frac{SP'-SP}{\frac{\lambda}{4.\pi}} + \varphi_{scatt-S} - \varphi_{scatt-M} \tag{4.6}$$

where the first term is the topographic phase component φ_{Topo}, which includes the reference ellipsoidal phase component, and the second term is the displacement phase component φ_{Displ}, which is connected to the LOS displacement d illustrated in Figure 4.1. If a DEM of the imaged scene is available, φ_{Topo} can be simulated and subtracted from $\Delta\varphi_{Int}$ (this is the inverse procedure performed in InSAR DEM production), yielding the so-called DInSAR phase $\Delta\varphi_{D-Int}$

$$\Delta\varphi_{D-Int} = \Delta\varphi_{Int} - \varphi_{Topo-simu} = \varphi_{Displ} \tag{4.7}$$

where $\varphi_{Topo-simu}$ is the simulated topographic component, which includes the flat-earth phase component implicitly. It should be noted that orbital errors have an effect on this simulated topography component even if the flattening operation is not explicitly performed.

This principle has been widely applied over the last 25 years, yielding significant results in the fields of seismology (Dalla Via et al. 2012; Massonnet et al. 1993; Peltzer and Rosen 1995), vulcanology (Antonielli et al. 2014; Massonnet et al. 1995; Massonnet and Sigmundsson 2000), glaciology (Goldstein et al. 1993; Rignot et al. 1997), landslides (Carnec et al. 1996; García-Davalillo et al. 2014), ground subsidence and uplift (Amelung et al. 1999; Galloway et al. 1998). Massonnet and Feigl (1998) and Hanssen (2001) provide comprehensive studies of several DInSAR applications. The DInSAR working concept is summarized in the following equation, which enables the displacements of the imaged scene to be determined from two complex SAR images.

$$\Delta\varphi_{D-Int} = \Delta\varphi_{Int} - \varphi_{Topo-simu} = \varphi_{Displ} + \varphi_{Topo_{res}} + \varphi_{Atm_S} - \varphi_{Atm_M} + \varphi_{Orb\_S} - \varphi_{Orb\_M} + \varphi_{Noise} + 2.K.\pi \quad (4.8)$$

where φ_{Topo_res} is the residual topographic error (RTE) component, φ_{Atm} is the atmospheric phase component at the time of image acquisition, φ_{Orb} is the orbital error component of each image (errors that impact the position of *M* and *S* in Figure 4.1), and φ_{Noise} is the phase noise. The final component, $2.K.\pi$, where *k* is an integer number known as phase ambiguity, arises from the wrapped nature of $\Delta\varphi_{D-Int}$, i.e., the fact that the DInSAR phases are bounded in the range $(-\pi, \pi]$.

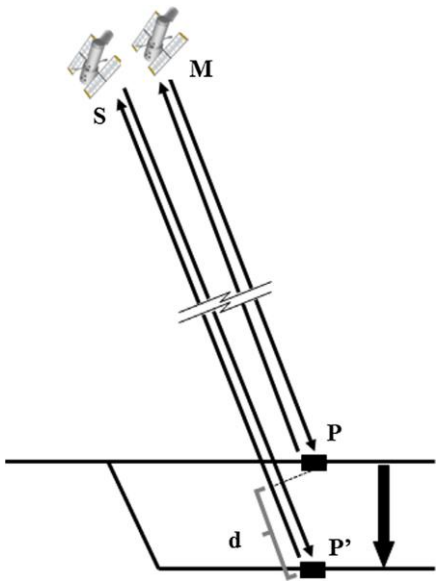

Figure 4.1. DInSAR deformation measurement scheme (Crosetto et al. 2016)

The purpose of any DInSAR approach is to obtain φ_{Displ} from $\Delta\varphi_{D-Int}$. This necessitates the separation of φ_{Displ} from the other phase components of the last equation. An essential condition for achieving this separation is to examine pixels with low φ_{Noise}, which are typically associated with one of two types of reflectors: those where the response to the radar is dominated by a strong reflecting object and is constant over time (Permanent Scatterer, PS) and those where the response is constant over time but due to different small scattering objects (Distributed Scatterers, DS). The main constraints of DInSAR are: (i) temporal and geometric decorrelations that affect the φ_{Noise} component (Hanssen 2001); (ii) phase unwrapping that affects the estimation of k (Ghiglia and Pritt 1998); and (iii) the atmospheric component (Zebker et al. 1997). DInSAR stacking strategies (Sandwell and Price 1998; Wright et al. 2001; Zebker et al. 1997) employ various methods for reducing atmospheric influences by averaging several interferograms.

PSI is a DInSAR technique that uses several SAR images obtained over the same area, as well as proper data processing and analysis procedures, to extract φ_{Displ} from the other phase

components described in the last equation. The deformation time series and deformation velocity estimated over the examined PSs or DSs are the major outcomes of a PSI study. Hereafter, the term PSs will refer to both PSs and DSs. Another outcome of a PSI analysis is the RTE, which is the difference between the true height of the scattering phase center of a specific PS and the height of the DEM at this point. The RTE is a critical component in achieving accurate PS geocoding.

Hooper et al. (2004) made a significant PSI contribution. They suggested a novel PS selection method based on phase characteristics suited for identifying low-amplitude natural targets with phase stability that are not discovered by amplitude-based methods. One significant advantage is that it does not necessitate using a prior deformation model. StaMPS, one of the most frequently used PSI software packages, arose from this study (Hooper 2008; Hooper and Zebker 2007).

Differential SAR Interferometry (DInSAR) and advanced methods such as permanent scatterers interferometry (PSInSAR) can measure even minor deformations and determine deformation velocities across a set of images (Delgado Blasco et al. 2019; Hooper et al. 2012; Mikhailov et al. 2014). Furthermore, high-resolution SAR images may improve end-product quality and level of detail. As a result, this newly emerging SAR-based monitoring has proven useful for monitoring the nation's deteriorating roadway infrastructure elements such as bridge settlements and displacements, roadway surface deformations, geohazard, sinkhole detection, etc.

This study used Copernicus Sentinel-1 SAR images and open-source tools to study the characteristics of ground deformation of two state highways in Marshall and Choctaw counties in Oklahoma. These study areas were part of the reconstruction two bridges in State Highway 32 in Kingstone and State Highway 109 in Grant, Oklahoma. A previous investigation by the Oklahoma department of transportation revealed that the soil subgrade has suffered from high

volume changes due to expansive soil in the area. Thus, these projects utilized the horizontal moisture barrier to control the subgrade volume change as part of rehabilitation.

This research used a processing chain for ground deformation monitoring using the European Space Agency (ESA) SentiNel Application Platform (SNAP) (Veci et al. 2014) and Stanford Method for Persistent Scatterers (StaMPS) (Hooper 2008). The processing chain was applied to more than 100 Sentinel-1A images taken over orbits to depict the Line-Of-Sight ground deformation rates primarily. The geometry result was then adjusted using a corrected incidence angle to calculate the vertical motion component, resulting in over 2 million points targets over their common area. The study assessed the effectiveness of horizontal moisture barriers in controlling subgrade deformation using satellite remote sensing technology, specifically SAR, for ground monitoring. It was found that remote sensing techniques, specifically InSAR with Sentinel-1 data, show great potential within land deformation and surface pavement monitoring. Significant surface pavement deformation can easily be exposed for further analysis or risk reduction measures.

4.3 METHOD AND DATA ANALYSIS

The study is focused on State Highway 32 in Kingstone County and State Highway 109 in Grant County in Oklahoma. This study names these sites after their cities, Madill and Hugo. The principal criteria for selecting these sites were the history of expansive soil problems and the wide range of climatic variations. The Madill project, which included the replacement of an existing bridge crossing Buncombe Creek on State Highway 32 with a new one, was finished in December 2017. The approach pavements were rebuilt across 256 meters (840 feet) in this project. A 160-meter (528 ft) stretch of road on both sides of the bridge was selected to install a moisture barrier. The moisture barrier was constructed using a heavy prime coat with separator fabric per section 408 of the ODOT specification book. The prime coat was applied at 0.5 gal/sy minimum across

the grading section over 18 m (60 ft) and covered by separator fabric on the fresh prime coat. It was ensured that the separator was uniformly saturated with the prime coat. After hardening, the second prime coat was applied on top of the fabric. The prime coat was allowed to harden before the aggregate base was laid over it. These moisture barriers have very low hydraulic conductivity, usually less than 10^{-10} cm/s if constructed properly (Asphalt Institute).

Similarly, an existing bridge across Horse Creek was replaced as part of the Hugo project. The moisture barrier was constructed on a 160-meter (528-foot) portion of state highway 109 in March 2018. In this project, the pavements were replaced across 256 meters (840 feet). The Intertape Polymer Group (IPG) geomembrane product "NovaLiner 24" with a covered length of 24 m (80 ft), was used as a moisture barrier. The hydraulic conductivity of the NovaLiner 24 is 1.77×10^{-12} cm/s (Geomembrane Solutions).

4.3.1 *Remote Sensing Data*

Sentinel-1A, launched by the European Space Agency (ESA) in April 2014, is an earth observation satellite for the Copernicus Initiative. Copernicus, formerly known as global monitoring for environment and security (GMES), is a European program for developing environmental and security-related information services. This satellite's equatorial revisit time is 12 days. The Sentinel-1A satellite operates in the C-band (central frequency = 5.404 GHz), with VH and VV polarizations and a range and azimuth spatial resolution of 5 m by 20 m.

Sentinel-1A data was collected in this investigation using the interferometric wide swath (IW) collecting mode with a sweep width of 250 km. Furthermore, data utilized level-1 single look complex (SLC) format (https://sentinel.esa.int/). Sentinel-1A IW level-1 SLC data with VV polarizations and ascending orbit modes was obtained from https://search.asf.alaska.edu/#/. In addition, these acquired Sentinel-1A data are freely available on the internet. Table 4.1 shows each dataset's first and last image, orbit pass, track, and the number of acquisitions. Table 4.2

displays the complete acquisition dates and characteristics of Sentinel-1A research data. Figure 4.2 illustrates study locations coordinates and areas of interest (AOI) on Sentinel-1 ascending (A34) orbit.

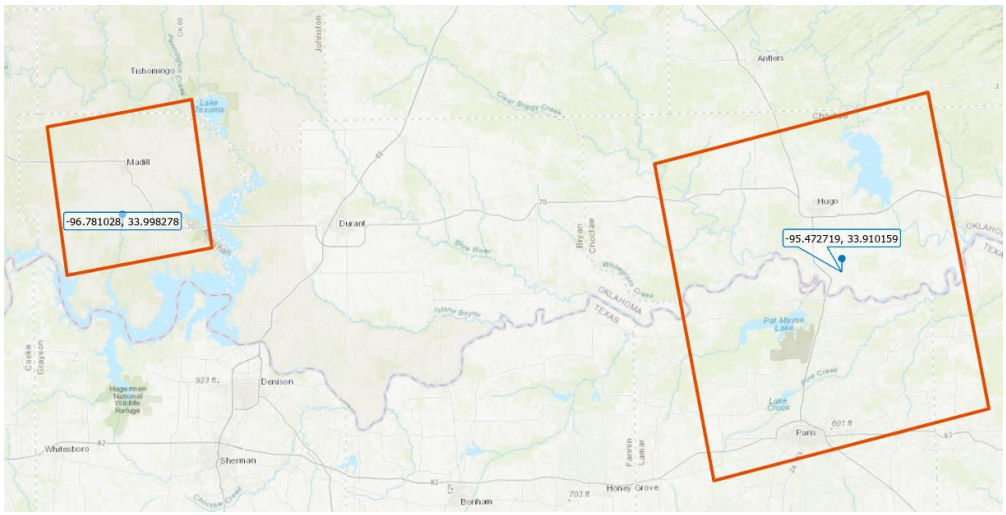

Figure 4.2. Study locations coordinates and area of interest for Sentinel-1 ascending track

Table 4.1. Research data from Sentinel-1A

Site	Satellite	First Image	Last Image	Orbit Pass	Track	N Acquisitions
Madill	S1A	2018/01/08	2019/12/29	Ascending	34	51
Hugo	S1A	2018/04/21	2020/04/22	Ascending	34	51

Table 4.2. Main characteristics of the selected Sentinel-1A master scene

Site	Acquisition Date	Mean Inc. Angle (rad/degrees)	Sub-Swath	Polarization	Initial Burst	Last Burst
Madill	2018/01/08	0.63/36.22	IW2	VV	7	9
Hugo	2018/04/21	0.72/41.54	IW3	VV	5	6

4.3.2 *PSI Processing*

This section explains the main foundations for SAR interferometry and the workflow adopted to process the S1 SAR images. PSI chooses one of the available (n+1) images as the master image. Interferograms are created between the master and the n available slave images acquired at various times. Each one is distinguished by a distinct perpendicular baseline or the perpendicular distance between the two satellite sites. Knowing the satellite's precise position on the two acquisition dates allows us to eliminate the components linked to the perpendicular baseline (flat earth and topographic components) from the interferogram phase. Man-made structures show persistent scatterers (PSs) in urban environments because they reflect the majority of the energy transmitted by the SAR directly backward or in a double-bounce process (i.e., from the ground to a perpendicular structure) back to the sensor).

PSI separates the deformation phase from the ambient phase and noise through time and spatial filtering. In reality, deformation is time-dependent, whereas atmospheric impacts are spatially but not temporally associated (i.e., they are visible in a single interferogram but not in other interferograms because the atmosphere varies randomly from date to date) (Cian et al. 2019). Finally, noise is spatially and temporally uncorrelated.

PSI can connect deformation with a specific scatterer, enabling very high-resolution infrastructure monitoring. As illustrated in Figure 4.3, the first step was to create a subset of the n available S1 images. The S1 pictures were subdivided into three sub-swaths. The city of interest was usually housed inside a single sub-swath. Each sub-swath is made up of numerous bursts (a longitudinal portion of the image). In order to construct the subset, we selected the sub-swath containing the area of interest and the bursts encompassing the area of interest exactly (Yagüe-Martnez et al., 2016). This enabled us to decrease the data size and accelerate processing.

The second step was to generate individual interferograms by merging the master image with the remaining n slave images. The open-source software SNAP was used to compute the subset and interferogram. The stack of n unique interferograms was used as input for the PSI analysis (third step), which identifies and analyzes PSs. This step was carried out using the StaMPS open-source toolbox (Hooper et al. 2018; Hooper and Zebker 2007) for MATLAB. The toolbox generates a map displaying the mean velocity of deformation for all PSs in the study area, their deformation on each observation date, and the mean standard deviation of deformation velocity over the entire time series. In addition, for atmospheric phase reduction, we used the Toolbox for Reducing Atmospheric InSAR Noise (TRAIN) (Bekaert et al. 2015) and the linear technique (phase versus elevation).

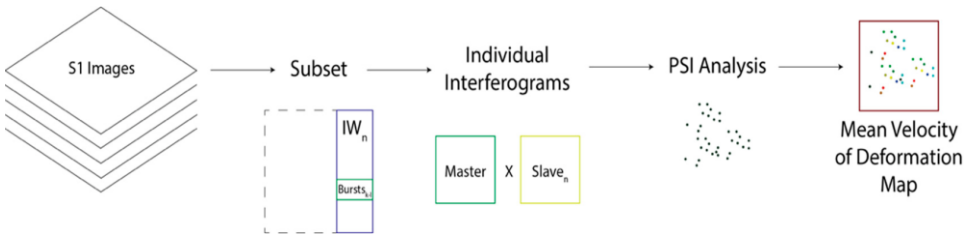

Figure 4.3. Workflows (i) and (ii) for Sentinel-1 data (Cian et al. 2019)

A set of scripts called "snap2stamps" was used to automate the process of interferogram formation. These scripts automatically compute interferograms based on the setting of a few parameters (sub-swath of S1 images to be processed, the bounding box of the area of interest, path to the data folder, path and name of the master image, parameters regarding the computational resources to employ, etc.). These programs are Python wrappers that use SNAP as the InSAR processor and produce PSI analysis output compatible with StaMPS. Open-source scripts of this type are available (Delgado Blasco and Foumelis). The overall analysis consisted of three main workflows: (i) single master interferogram processing using SNAP, (ii) PSI processing using StaMPS, and (iii) results from analysis in a geographic information system (GSI).

Following that, the area of interest (AOI) within the master image must be identified and used to split the master image using the SNAP graphical user interface (GUI). In this context, a single image swath is chosen, along with the bursts required to cover the AOI. This operation reduces data volume while optimizing processing time and resources. At this point, we can use the snap2stamps scripts to prepare the slave images and compute the interferograms. The scripts sort the slave images by acquisition date, divide them into subsets based on the master extent, co-register each of them with the master image, and compute the interferograms, which are then exported for analysis in StaMPS.

The first step in workflow (ii) is to use the mt prep snap script to prepare the exported data for analysis in StaMPS. Then, within MATLAB, StaMPS can be run from steps 1 to 7 as described in the StaMPS user manual (Hooper et al. 2018), using TRAIN for APS mitigation and the *aps_linear* approach to perform the PS analysis. The result is a map representing land deformation values throughout the time series.

The land deformation maps are loaded into a GIS in workflow (iii) (Figure 4.4), and points with deformation greater than or equal to 2 mm/year are selected and analyzed. For each area of interest (AOI), three plots are generated: (i) a plot showing the distribution of the velocity of deformation values for all analyzed points, (ii) a plot showing the distribution of the velocity of deformation standard deviation values for all analyzed points, and (iii) a plot showing the deformation trend of each point measured at each time step of the time series. The plots enable us to comprehend the average behavior of the points within the AOI. The first plot depicts the moderate deformation (subsidence, uplift, or stability), the second illustrates the stability of velocity of deformation values, and the third shows the deformation trend over time.

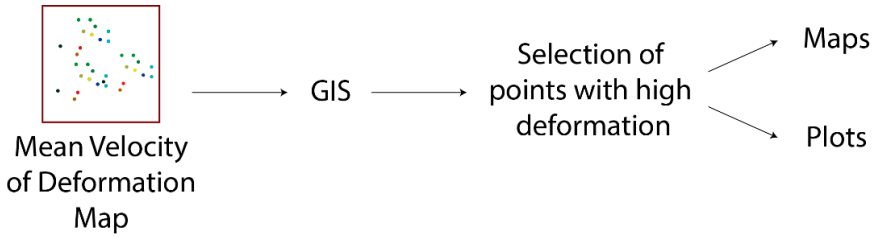

Figure 4.4. Workflow (iii): analyze deformation maps and create plots (Cian et al. 2019)

4.3.3 *Time series Analysis*

Time series analysis examines the attributes of the independent variable, the response variable, with regard to time. This study uses time series analysis to understand the underlying causes of trends or systemic patterns over time (Cryer and Chan 2008). The following steps have been taken to conduct the time series analysis.

In the first step, an autoregressive integrated moving average (ARIMA) model was fitted to the dataset. Outliers analysis was performed in the second step to detect errors or abrupt short-term changes. The correlation was done to relate a time series with other covariate time series in the final step. The outcome of the time series analysis answers a critical question, whether fitted time series are related. The findings are expanded to discuss whether moisture barriers successfully control ground settlement/heave during the monitoring period at two research sites.

4.3.4 *Time series Properties*

A time series is a chronologically ordered sequence of data points at equally spaced intervals in time. Non-stationarity, autocorrelation, and seasonality are common characteristics of time series. The time series must be stationary in order to be used in ARIMA modeling. A stationary series has three characteristics: a constant mean, a constant variance, and a constant covariance dependent only on the time interval between values. Because it can be modeled with fewer parameters, a stationary series (also known as a "white noise process") is easier to analyze. While

it fluctuates, it always returns to a constant mean and is thus easier to predict. The first is changing variance over time (heteroscedasticity), which can often be addressed by applying a log transformation; the second is an increasing or decreasing trend, which can often be eliminated by taking the first difference (i.e., $Y_t - Y_{t-1}$). A second differencing may be required to achieve stationary on rare occasions, but third-order differencing and above is uncommon (Hyndman and Athanasopoulos 2018). The Augmented Dickey-Fuller (ADF) (Cao et al. 2013) unit-root test can be used to determine whether a time series is stationary.

Time series observations are frequently correlated with observations from previous time points and thus are not distributed independently. This type of correlation is known as autocorrelation or serial correlation. Autocorrelation in time series does not satisfy standard regression analysis assumptions. Due to the fact that autocorrelated data are rarely stationary, differencing the data is often sufficient to eliminate autocorrelation; therefore, any necessary data transformations should be performed prior to testing for autocorrelation.

Stationery and autocorrelation can be checked using autocorrelation functions (ACFs). An ACF depicts the correlation between each observation and previous values at different lags, where a lag is the number of time points between two observations. The partial ACF (PACF), which is the correlation between an observation and historical values that are not explained by correlations at lower order lags, complements the ACF. For example, after adjusting for the correlation between Y_t and Y_{t-3}, Y_{t-2}, and Y_{t-1}, the PACF value at lag 4 is the correlation between an observation (Y_t) and the prior observation at lag 4 (Y_{t-4}).

ARIMA models have a single dependent variable (Y_t), a function of previous Y values and the error term (ϵ_t). ARIMA models may accommodate any continuous result (such as rates or means) as well as high counts that are not bounded by zero because they assume errors are normally distributed. ARIMA components are:

1. Autoregressive (AR) model: Y_t is predicted by one or more lagged values of Y_t. This is expressed by the equation below, where c is a constant, is the magnitude of the autocorrelation, p is the number of delays, and t is the error.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

2. Moving average (MA) model: Y_t is predicted by one or more lagged error values (t). This is not to be confused with moving average smoothing. The equation below is the value of the autocorrelation of the mistakes, and q is the number of delays.

$$Y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

3. Differencing (Integration): To generate meaningful predictions from an ARIMA model, the time series being modeled must be stationary. Stationary is achieved through differencing, which is the process of determining the difference between adjacent data points.

$$Y_t' = Y_t - Y_{t-1}$$

An ARIMA model is a mix of AR, MA, and differencing models (Integration). The notation for describing an ARIMA model is (p, d, q), where p, d, and q are positive integers:

- p = the order of the AR part of the model;
- d = the degree of non-seasonal differencing; and
- q = the order of the MA part of the model.

4.3.5 *Fitting an ARIMA Model*

The ARIMA model's parameters are determined using the Box-Jenkins technique. The Box-Jenkins technique is a popular strategy that involves model identification and selection, parameter

estimates, and model validation (Box et al. 2015). The method for determining the best-fitting ARIMA model is based on reducing the information criteria (AIC, BIC). ARIMA model fitting processes are:

1. Plot data to understand patterns
2. Transform data to stabilize variance (if necessary)
3. Determine differencing order to induce stationarity
4. Plot the ACF, PACF, and extended autocorrelation function (EACF) of stationarity data to determine potential AR/MA orders
5. Estimate the model and use information criteria to find the best model

4.3.6 *Anomaly Detection in Time Series*

Outliers are abnormal observations that can occur due to measurement and/or copying mistakes or as a result of sudden, short-term changes in the underlying process. Outliers in time series may be classified into additive outliers and innovative outliers. (Fox 1972) proposed the innovative outlier (IO), in which the value of the 'innovation' or noise is excessive. This impacts not just the initial observation but also subsequent observations. In contrast, the additive outlier (AO) behaves like an error of observation that occurs just at that point in time and has no effect on following observations since it does not enter the structure of the series.

Fox (1972) recommended using likelihood ratio test statistics to test for outliers in autoregressive models. Tsay (1986) suggested an iterative approach for detecting outliers, removing their impacts, and determining a preliminary model for the underlying process. (Chang et al. 1988) provided an iterative approach for estimating the time series parameters in ARIMA models and derived likelihood ratio criteria for assessing the existence of both types of outliers as

well as criteria for discriminating between them. Chang et al. (1988) developed a method for detecting outliers based on computing the likelihood ratio test statistics for all of the series' data under the null hypothesis that there are no outliers.

4.3.7 *Time Series Correlation*

The last step was to conduct a spurious correlation test between pairs of ARIMA models. The spurious association is considered to have occurred when statistical summaries show that two variables are correlated when there is no theoretical relationship between the two variables. In this case, the theoretical CCF is nonzero only at lag $-d$, reflecting that X is "leading" Y by d units of time. The CCF can be estimated by the sample cross-correlation function (sample CCF) defined by

$$r_k(X,Y) = \frac{\sum(X_t-\bar{X})(Y_{t-k}-\bar{Y})}{\sqrt{\sum(X_t-\bar{X})^2}\sqrt{\sum(Y_t-\bar{Y})^2}} \tag{4.9}$$

One difficulty in calculating cross-correlation is that the CCF is affected by the time series structure of the x-variable and any "in common" trends the x and y series may have over time. One strategy for dealing with this difficulty is called "pre-whitening." Pre-whitening is the technique of removing or decreasing short-term stochastic persistence to detect deterministic change. Pre-whitening steps are followed:

1. Determine a time series model for the *x*-variable and store the residuals from this model.

2. Filter the *y*-variable series using the *x*-variable model (using the estimated coefficients from step one). This step finds differences between observed *y*-values and "estimated" *y*-values based on the *x*-variable model.

3. Examine the CCF between the residuals from Step one and the filtered *y*-values from Step two. This CCF can be used to identify the possible terms for a lagged regression.

4.4 RESULTS

The results from Sentinel-1A data (12-day repeat cycle) are presented for both case studies (Table 4.1). In addition, the GACOS atmospheric correction data (Yu et al. 2018) was used to mitigate topography-correlated atmospheric effects, which was sufficient given the magnitude of ground displacements expected and the availability of a large number of acquisitions over the area of interest.

The results show maps of land deformation velocity in mm/year for the PSs found in each time series across the observation period. Deformation velocity can be negative, indicating land subsidence, or positive, indicating land uplift. The deformation values pertain to the satellite's line of sight (LOS) and can have a vertical as well as a horizontal component. Given that the incidence angle for all of the acquisitions used in the study is less than 45 degrees, the vertical component is more likely to be the dominant component, which was considered noise to highlight substantial deformations. The vertical component for each Persistent Scatter (PS) point is calculated to discriminate the two components. Using the following equation, we can calculate $d_{vertical}$, which is the projection of vertical displacement in ascending azimuth-look direction.

$$d_{vertical} = d_{LOS}.\cos\left(\theta_{incident}\right) \tag{4.10}$$

For each case study, three plots were created: (i) the distribution of deformation velocity values, (ii) the distribution of deformation velocity standard deviation values, and (iii) the trend of deformation for each point analyzed over the entire time series, which also shows the overall mean velocity of deformation, that is, the mean of the mean velocity of deformation of each PS within the AOI, and its standard deviation, that is, the standard deviation of the mean velocity of deformation. The plots help us comprehend the amount (velocity) of the deformation, as well as its consistency (standard deviation) and evolution over time (trend). Figure 4.5 shows the velocity

of deformation standard deviation. Areas with larger standard deviation indicate more significant errors (due perhaps to atmosphere or unwrapping errors). Figure 4.6 depicts the deformation velocity obtained for ascending tracks in Madill and Hugo. The majority of these PSs exhibit subsidence greater than 2 mm/year, and thousands exhibit subsidence greater than 5 mm/year. These figures appear credible given the low standard deviation (the majority of which is less than 1 mm/year).

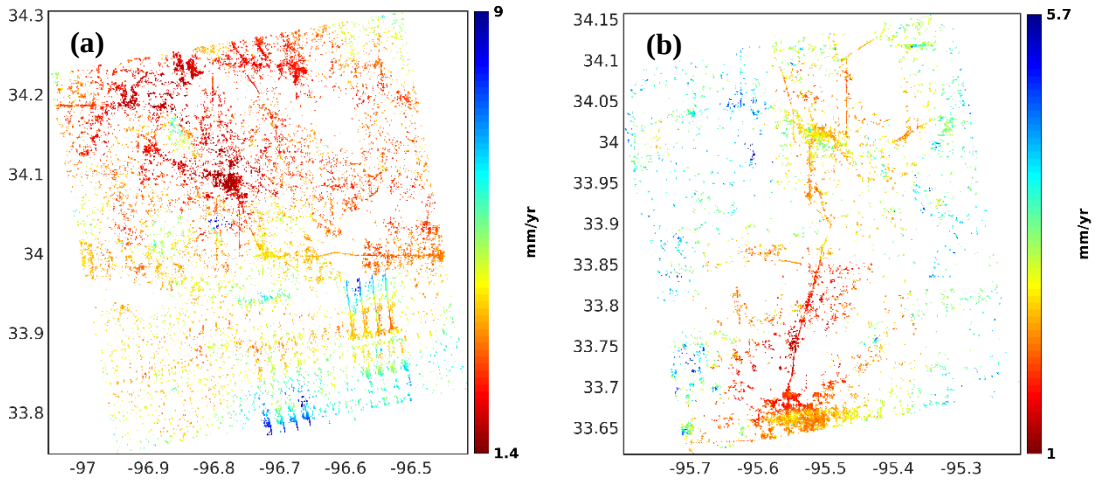

Figure 4.5. Standard deviation of deformation velocity map over the monitoring period

(a) Madill (b) Hugo

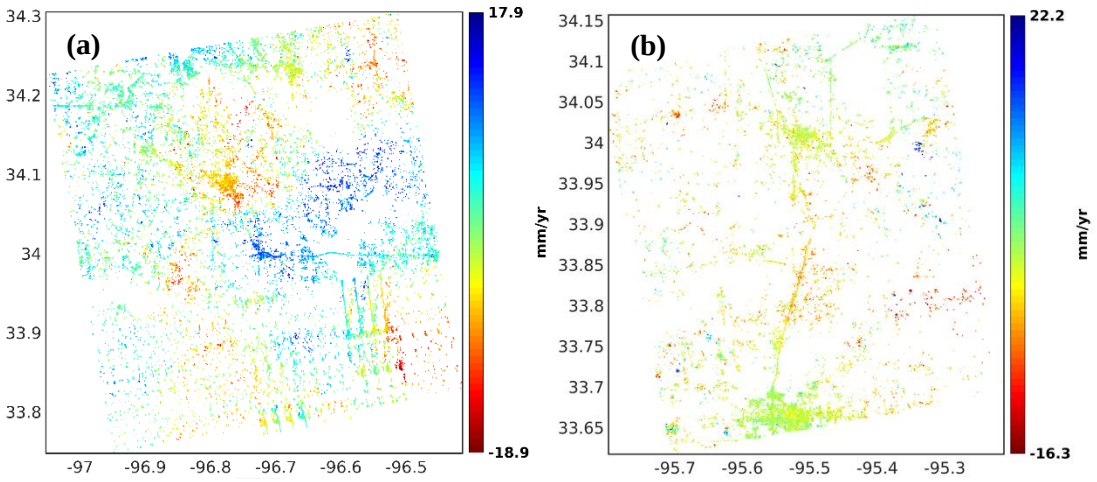

Figure 4.6. LOS deformation velocity map over the monitoring period (a) Madill (b) Hugo

The average vertical displacement velocity was computed for two study areas. The results at Madill revealed that the average displacement velocity on the moisture barrier section on the left side was 12.71 mm/year and on the right side of the bridge was 9.1 mm/year. In contrast, the control section has a -6.1 mm/year displacement velocity. Figure 4.7 shows the average vertical displacement velocity for the Madill.

Figure 4.7. Mean vertical displacement velocity at Madill Jan. 2018–Dec. 2019

The results at Hugo revealed that the average displacement velocity on the moisture barrier section was -7.07 mm/year on the left side, and the right side of the bridge was -3.4 mm/year. In comparison, the control section has a displacement velocity of -4.47 mm/year.

Figure 4.7 shows the average displacement velocity for the Hugo study area. Figures 4.9 and 4.10 illustrate The vertical displacement time series for each section of the Madill and Hugo study areas during the monitoring period.

Figure 4.8. Mean vertical displacement velocity at Hugo Apr. 2018–Apr. 2020

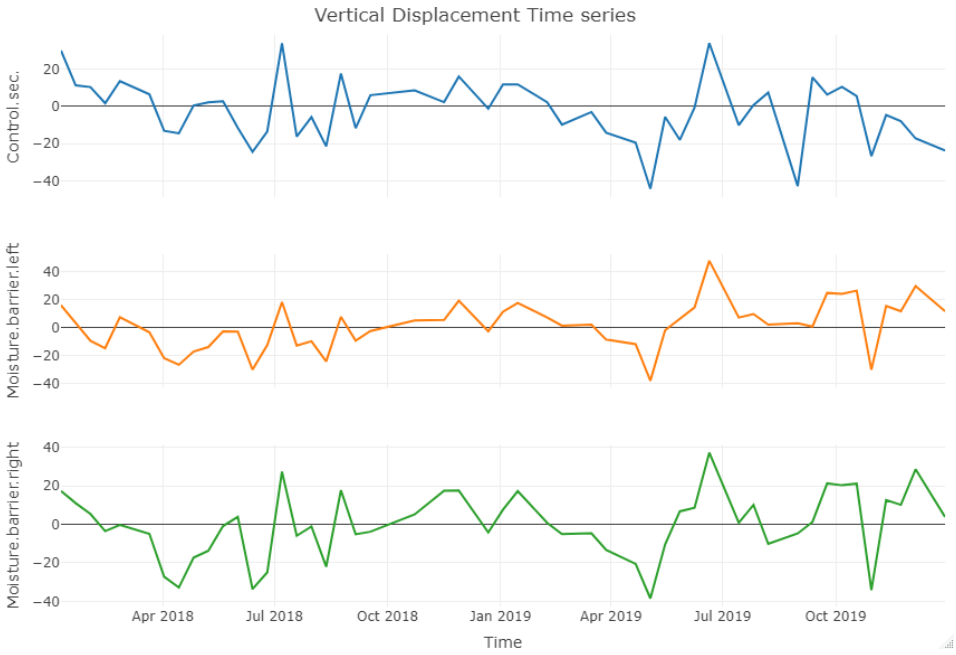

Figure 4.9. Sentinel-1 displacement time series at Madill for Jan. 2018–Dec. 2019 (A) control section; (B) moisture barrier left section; (C) moisture barrier right section

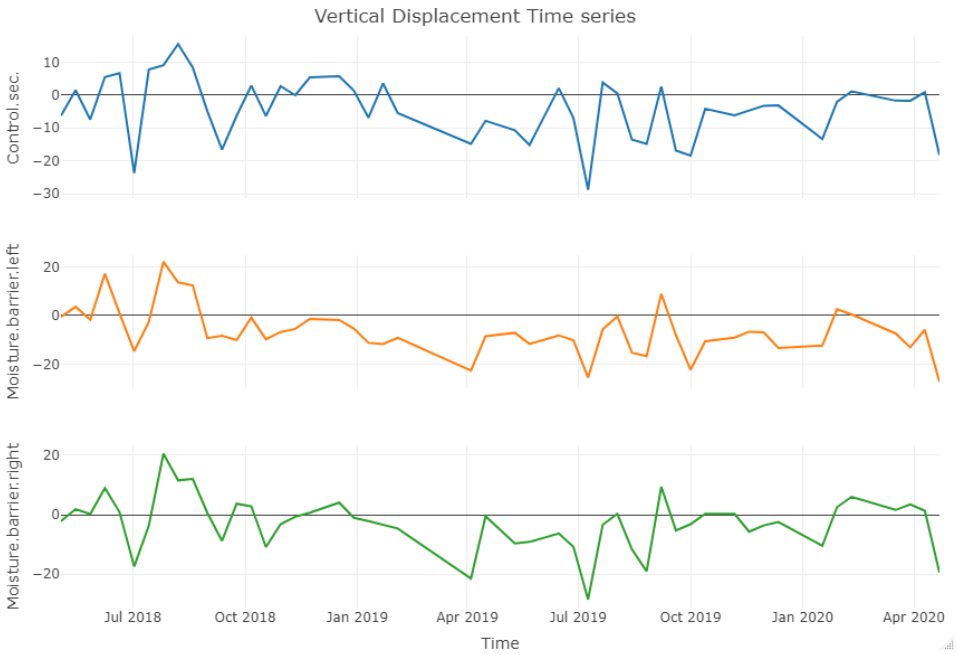

Figure 4.10. Sentinel-1 displacement time series at Hugo for Apr. 2018–Apr. 2020 (A) control section; (B) moisture barrier left section; (C) moisture barrier right section

The time series analysis was carried out in detail based on vertical deformation data. In the first step, the ADF test was conducted to determine whether time series were nonstationary. Table 4.3 summarizes the results of the ADF test. The ADF test results for the Madill time series have relatively small p-values and are close to a significant alpha of 0.05 and indicate that data are, in fact, nonstationary. The ADF test results for the Hugo time series have p-values greater than a significant alpha of 0.05 and show that data are, in fact, stationary.

Table 4.3. Augmented Dickey-Fuller test results for two monitoring locations

Site	Location	Statistics	Lag order	p-value
Madill	control section	-3.4923	3	0.05182
	moisture barrier section (left)	-3.6278	3	0.03975
	moisture barrier section (right)	-3.4038	3	0.06557
Hugo	control section	-3.2892	3	0.08385
	moisture barrier section (left)	-3.1191	3	0.1263
	moisture barrier section (right)	-2.7663	3	0.2675

The ACF and PACF of the differenced Madill dataset were plotted. The ACF and PACF plots indicate that ARIMA (1,1,1) is a good candidate. In the model selection process, the EACF chart was computed, and the best subset analysis based on BIC was conducted. The results revealed that ARIMA (1,1,1) with the lowest BIC is a good prediction model for the Madill time series dataset. Figure 4.11 shows the computed various subset ARIMA models for the Madill time series.

ACF and PACF were plotted for the Hugo dataset, which revealed ARMA (1,1) is a good candidate. Further assessment was made to select the best model. The EACF chart and best subset analysis revealed that ARMA (1,1) has the lowest BIC and is a good prediction model for the Hugo time series. Figure 4.12 demonstrates the computed various subset ARMA models for the Hugo time series.

Outlier detection was done as part of the time series analysis to identify anomalies. This study utilized test statistics for outlier detection proposed by Chang et al. (1988). The test statistics, $\lambda_{2,t}$ were not significantly large for both case studies' time series, and there is no evidence of AO and AI.

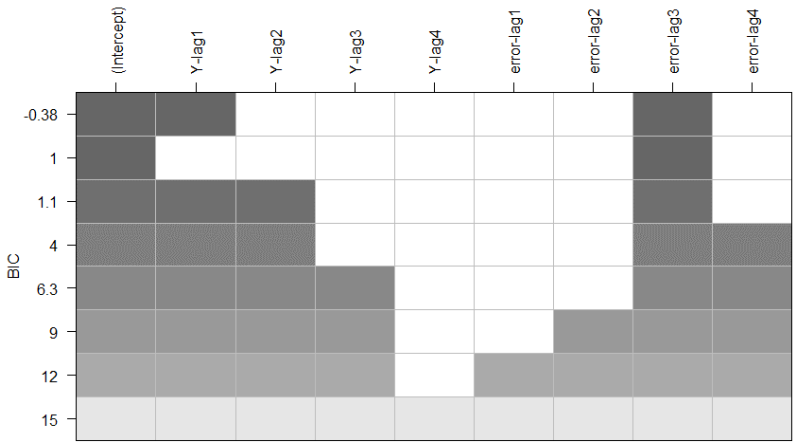

Figure 4.11. Best subset ARIMA models selection based on BIC with the maximum AR and MA orders of 4 for Madill time series

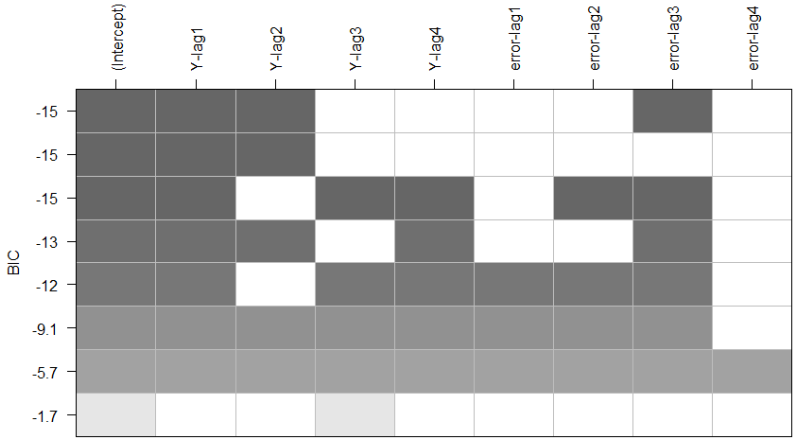

Figure 4.12. Best subset ARIMA models' selection based on BIC with the maximum AR and MA orders of 4 for Hugo time series

For displacement time series in Madill, sample CCF for all paired displacement time series combinations were examined. The CCF of the pre-whitened bivariate series is then computed and plotted. Figure 4.13 shows plotted CCF for all possible Madill time series ARIMA model combinations. The sample CCF is significant at lag 0, suggesting a positive contemporaneous correlation between control and moisture barrier sections. Figure 4.14 shows

the plotted CCF for all plausible combinations of Hugo time series ARIMA models. At lag 0, the sample CCF is significant, indicating a positive cross-correlation between the control and moisture barrier sections.

By looking at these results, we can conclude that Sentinel-1 data allows deformation detection with good coherence throughout the whole time series. The big advantage brought by Sentinel-1 is the frequency of acquisition, which increases the ability to detect PSs and the precision of measuring their velocity of deformation. The time series analysis reveals insufficient evidence to suggest a significant difference between the moisture barrier and the control section time series.

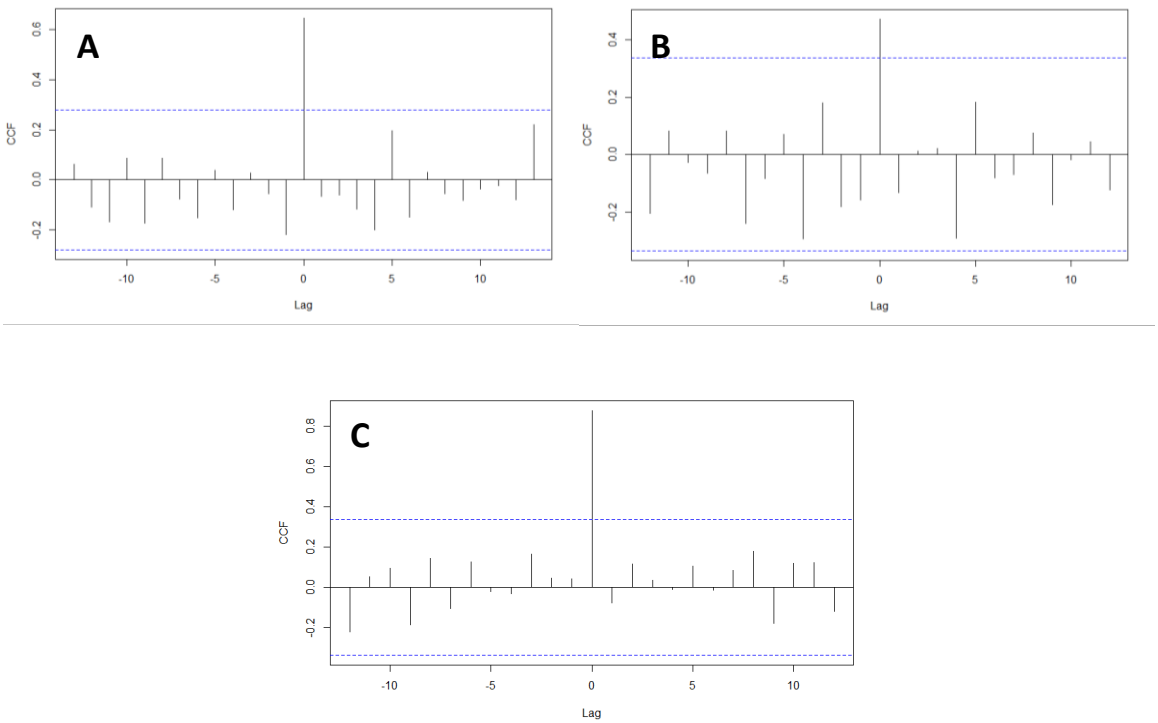

Figure 4.13. Sample CCF of pre-whitened time series at Madill (A) control section vs. moist. section (left); (B) control section vs. moist. section (right); (C) moist. section (left) vs. moist. section (right)

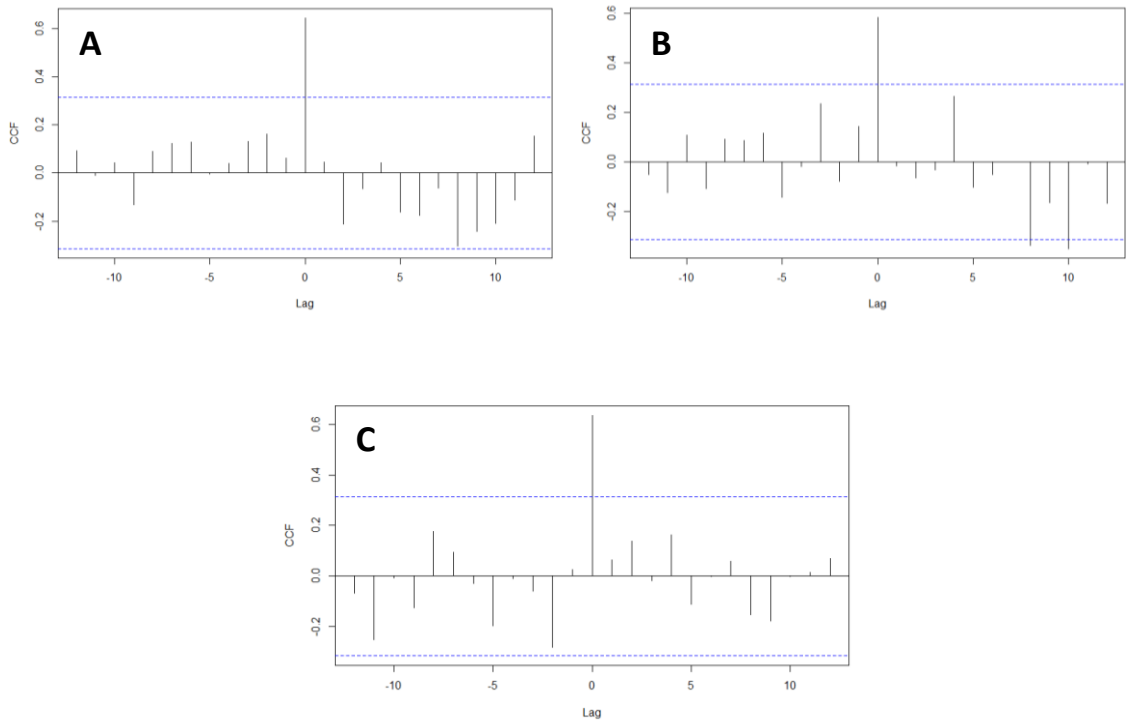

Figure 4.14. Sample CCF of pre-whitened time series at Hugo (A) control section vs. moist. section (left); (B) control section vs. moist. section (right); (C) moist. section (left) vs. moist. section (right)

4.5 FIELD VALIDATION

Oklahoma Department of Transportation (ODOT) is responsible for collecting and processing data from more than 300 highways across the state in urban and suburban areas. Oklahoma Department of Transportation Materials Division / Geotechnical Branch assessed the pavement for roughness and distress on two monitoring sites in November 2022. ODOT concluded that is virtually no distress in either section with and without membrane on both sites.

4.6 CONCLUSIONS

Maintaining a high level of service (LOS) on roadways is one of the primary advantages of satellite remote sensing technology. This is possible because early detection of problems can lead

to significant rehabilitation/maintenance issues, resulting in time, energy, and financial loss. In light of current pavement monitoring practices, this study aims to detect various types of ground deformation by processing freely available Sentinel-1 spaceborne radar data. Sentinel-1 data appear to be a perfect compromise given the level of detail they can provide and the reliability of the measurements, and they are publicly available. Assessing the surface deformation of pavement was accomplished through a remote sensing method based on the combined application of SNAP and StaMPS. The methodology has proven simple, thorough, and dependable in producing InSAR time series analysis from large datasets, such as the long Sentinel-1 time series. The performance of horizontal moisture barriers on two Oklahoma state highways was evaluated using time series analysis.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

5.1 CONCLUSIONS

This dissertation presented statistical and numerical approaches that employ machine learning and statistical modeling techniques in unsaturated soil diffusivity and suction estimate and performance assessment of civil infrastructure. The presented research can help predict the diffusivity and suction under the effect of a rainstorm event or large quantities of moisture while considering future climate variability. The developed approaches were formulated to consider the various sources of uncertainty associated with soil mechanical parameters. Statistical modeling and numerical simulation techniques are used to accelerate mechanistic simulations and the model's prediction accuracy. The dissertation discusses an interferometric synthetic aperture radar (InSAR) method to monitor and assess vertical ground and pavement surface movements for swelling and shrinking soils. InSAR (synthetic aperture radar interferometry) is a particularly intriguing method for analyzing ground surface deformations. Using existing SAR data, it is possible to provide regional-scale monitoring as well as a historical assessment of displacement. Based on the presented research, the following conclusions are drawn:

- The goal is to find hidden patterns or subgroups in data that have not been explicitly labeled. The analysis is simplified by clustering continuous predictors into discrete ones. Three well-known clustering algorithms and four distance metrics were examined and compared for this study. The results of the study show that DBSCAN and Kmean clustering are both effective methods for dividing soils into groups based on their physical characteristics.
- Nonlinear least squares regression extends linear least squares regression to a much larger and more general class of functions. A nonlinear regression model can incorporate almost any function written in closed form. In contrast to linear regression, there are few constraints on how parameters can be used in the functional part of a nonlinear regression model. The greatest advantage of nonlinear least squares regression over many other techniques is the wide range of functions that can be fit.
- The proposed nonlinear model yielded results comparable to the one-step pressure outflow procedure. The nonlinear least-squares model for determining diffusivity is simple and quick, making it much simpler for practitioners to perform. It is theoretically sound and yields reasonable results with a reasonable level of certainty and robustness.
- An improved statistical model based on mechanics principles is proposed to improve the estimate of the equilibrium suction in subgrade soils. The model used in this analysis approach allows for a more accurate prediction of suction-dependent soil parameters within the unsaturated zone, such as resilient modulus, soil movement, unsaturated permeability coefficient (of diffusivity), shear strength, and shear wave velocity.
- A ridge regression model was developed to predict equilibrium suction based on readily available climatic variables (i.e., relative humidity and TMI) and clay content. The ridge regression model surpassed the regression models developed by previous studies. The main advantage of ridge regression is that it can be used on datasets with a large number

of correlated features. Another benefit of ridge regression is that it can be used when there are more features than observations. The L2 penalty introduced for ridge regression models will continue to be effective at minimizing overfitting. This is due to the fact that the penalty reduces the number of close-to-zero coefficients and, consequently, the model's complexity.

- Sentinel-1 data appears to be a particularly good compromise given the level of detail and measurement reliability, and they are publicly available. Pavement surface deformation was evaluated using a method based on the integration of SNAP and StaMPS. In producing InSAR time series analysis from large datasets, such as the long Sentinel-1 time series, the methodology has proven simple, thorough, and dependable. Based on the validation procedure, the workflow enables the generation of decomposed velocity maps with an accuracy of 2 mm/yr and expected uncertainty levels of less than 2 mm/yr.
- The incorporation of time series analysis in infrastructure monitoring has the potential to significantly improve the current state of the art in risk and reliability assessment, as well as damage identification. The main advantage of time series analysis is that it aids in understanding the underlying causes of trends or systemic patterns over time. Non-stationary time series or outliers, on the other hand, can have an impact on the analysis, and appropriate action must be taken to address the concerns.

5.2 RECOMMENDATIONS FOR FUTURE WORK

- The presented research in Chapter II focused on estimating van Genuchten diffusivity parameters for Oklahoma Mesonet soil under various moisture conditions. However, the statistical model is based on Mesonet meteorological data collected at specific locations in Oklahoma under specified weather. Accordingly, future efforts should aim at more field observations to expand the database to include additional climates and soil fabrics.
- The presented research in Chapter III focused on developing a statistical model to estimate equilibrium suction using Oklahoma Mesonet climate data and the Mesosoil database. However, changes in drainage (i.e., ponding) or soil fabric (i.e., deeper fissures) are predicted to result in different predictions. Consequently, future efforts should focus on increasing field observations, and research is required to expand the database to include more climates, drainage conditions, and soil types.
- Sentinel-1 SAR C-band images with high temporal resolution (6–12 days) and moderate spatial resolution (5 m × 20 m) can provide ground deformation maps. High-resolution SAR, such as CSK satellites, use X-band with a revisit interval of 16 days and are well suited for pavement monitoring.
- The PSI has been demonstrated to be a dependable tool for S1 data processing. However, new approaches have been introduced in recent years. When compared to PSI, these techniques improved interferogram coherence. Future efforts are required to implement the new method to assess pavement deformation and infrastructure monitoring better.

REFERENCES

Adem, H. H., and Vanapalli, S. K. (2015). "Review of methods for predicting in situ volume change movement of expansive soil over time." *Journal of Rock Mechanics and Geotechnical Engineering*, 7(1), 73-86.

Aitchison, G. (1965). “Moisture equilibria and moisture changes in soils beneath covered areas.”

Amelung, F., Galloway, D. L., Bell, J. W., Zebker, H. A., and Laczniak, R. J. (1999). "Sensing the ups and downs of Las Vegas: InSAR reveals structural control of land subsidence and aquifer-system deformation." *Geology*, 27(6), 483-486.

Antonielli, B., Monserrat, O., Bonini, M., Righini, G., Sani, F., Luzi, G., Feyzullayev, A. A., and Aliyev, C. S. (2014). "Pre-eruptive ground deformation of Azerbaijan mud volcanoes detected through satellite radar interferometry (DInSAR)." *Tectonophysics*, 637, 163-177.

Assouline, S., Tessier, D., and Bruand, A. (1998). "A conceptual model of the soil water retention curve." *Water Resources Research*, 34(2), 223-231.

Aubeny, C., Lytton, R., and Tang, D. (2003). “Simplified analysis of unsteady moisture flow through unsaturated soil.” *Transp*. *Res*. *Rec*. 1821(1), 75-82.

Aubeny, C., and Lytton, R. (2004). "Shallow slides in compacted high plasticity clay slopes." *Journal of geotechnical and geoenvironmental engineering*, 130(7), 717-727.

AustralianStandard (2011). “Residential slabs and footings.” *Sydney NSW: Standard*.

Banerjee, A., Puppala, A. J., Congress, S. S., Chakraborty, S., Likos, W. J., and Hoyos, L. R. (2020). "Variation of Resilient Modulus of Subgrade Soils over a Wide Range of Suction States." *Journal of Geotechnical and Geoenvironmental Engineering*, 146(9), 04020096.

Bamler, R., and Hartl, P. (1998). "Synthetic aperture radar interferometry." *Inverse problems*, 14(4), R1.

Basma, A. A., and Al-Suleiman, T. I. (1991). “Climatic considerations in new AASHTO flexible pavement design.” *J. Transp. Eng*. 117(2), 210-223.

Bear, J., and Cheng, A. H.-D. (2010). *Modeling groundwater flow and contaminant transport*, Springer.

Bhagat, V. (2014). "Agriculture water balance of micro-watershed using GIS techniques." *Journal of Earth Science Research*, 2(1), 1.

Breakah, T. M., Williams, R. C., Herzmann, D. E., and Takle, E. S. (2011). “Effects of using accurate climatic conditions for mechanistic-empirical pavement design.” *J. Transp. Eng*. 137(1), 84-90.

Bekaert, D., Walters, R., Wright, T., Hooper, A., and Parker, D. (2015). "Statistical comparison of InSAR tropospheric correction techniques." *Remote Sensing of Environment*, 170, 40-47.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.

Brooks, R. H., and Corey, A. T. (1964). "Hydraulic properties of porous media and their relation to drainage design." *T. ASAE*, 7(1), 26-0028.

Bruce, R., and Klute, A. (1963). "Measurements of soil moisture diffusivity from tension plate outflow data." *Soil Science Society of America Journal*, 27(1), 18-21.

Bulut, R., and Javid, A. H. (2019). “Developing Implementable Climatic Input Data and Moisture Boundary Conditions for Pavement Analysis and Design.”

Bulut, R., Zaman, M., Amer, O., Mantri, S., Chen, L., Tian, Y., and Taghichian, A. (2013). “Drying shrinkage problems in high-plastic clay soils in Oklahoma.” Oklahoma Transportation Center.

Cao, S., Wang, F., Tam, W., Tse, L. A., Kim, J. H., Liu, J., and Lu, Z. (2013). "A hybrid seasonal prediction model for tuberculosis incidence in China." *BMC medical informatics and decision making*, 13(1), 1-7.

Carnec, C., Massonnet, D., and King, C. (1996). "Two examples of the use of SAR interferometry on displacement fields of small spatial extent." *Geophysical research letters*, 23(24), 3579-3582.

Chang, I., Tiao, G. C., and Chen, C. (1988). "Estimation of time series parameters in the presence of outliers." *Technometrics*, 30(2), 193-204.

Cian, F., Blasco, J. M. D., and Carrera, L. (2019). "Sentinel-1 for monitoring land subsidence of coastal cities in Africa using PSInSAR: a methodology based on the integration of SNAP and StaMPS." *Geosciences*, 9(3), 124.

Coleman, J. (1965). "Geology, climate and vegetation as factors affecting soil moisture." *Moisture Equilibria and Moisture Changes in Soils Beneath Covered Areas. Butterworth, Sydney, Australia*, 93-99.

Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N., and Crippa, B. (2016). "Persistent scatterer interferometry: A review." *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 78-89.

Cryer, J. D., and Chan, K.-S. (2008). *Time series analysis: with applications in R*, Springer.

Cuzme, A. J. (2018). "Estimating Expansive Soil Field Suction Profiles Using a Soil Suction Surrogate." Arizona State University.

Dalla Via, G., Crosetto, M., and Crippa, B. (2012). "Resolving vertical and east-west horizontal motion from differential interferometric synthetic aperture radar: The L'Aquila earthquake." *Journal of geophysical research: solid earth*, 117(B2).

Delgado Blasco, J. M., and Foumelis, M. "Automated SNAP Sentinel-1 DInSAR Processing for StaMPS PSI with Open Source Tools (Version 1.0.1)."

Delgado Blasco, J. M., Foumelis, M., Stewart, C., and Hooper, A. (2019). "Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry." *Remote Sensing*, 11(2), 129.

Doering, E. (1965). "Soil-water diffusivity by the one-step method." *Soil Science*, 99(5), 322-326.

Dong, Y., and Lu, N. (2016). "Dependencies of shear wave velocity and shear modulus of soil on saturation." *Journal of Engineering Mechanics*, 142(11), 04016083.

Efron, B. (1992). “Bootstrap methods: another look at the jackknife.” *Breakthroughs in statistics*. Springer, 569-593.

Efron, B., and Tibshirani, R. J. (1994). *An introduction to the bootstrap*, CRC press.

Fagrhi, A., and Ozden, A. (2015). "Satellite Assessment and Monitoring for Pavement Management." *Center for Advanced Infrastructure and Transportation, State University of New Jersey: Piscataway, NJ, USA*.

Fityus, S., Walsh, P., and Kleeman, P. "The influence of climate as expressed by the Thornthwaite index on the design depth of moisture change of clay soils in the Hunter Valley." *Proc., Conference on geotechnical engineering and engineering geology in the Hunter Valley*, Conference Publications Springwood, Australia, 251-265.

Fox, A. J. (1972). "Outliers in time series." *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(3), 350-363.

Fredlund, D., and Houston, S. (2013). "Interpretation of soil-water characteristic curves when volume change occurs as soil suction is changed." *Advances in unsaturated soils*, 1, 15.

Galloway, D. L., Hudnut, K. W., Ingebritsen, S., Phillips, S. P., Peltzer, G., Rogez, F., and Rosen, P. (1998). "Detection of aquifer system compaction and land subsidence using interferometric synthetic aperture radar, Antelope Valley, Mojave Desert, California." *Water Resources Research*, 34(10), 2573-2585.

García-Davalillo, J. C., Herrera, G., Notti, D., Strozzi, T., and Álvarez-Fernández, I. (2014). "DInSAR analysis of ALOS PALSAR images for the assessment of very slow landslides: the Tena Valley case study." *Landslides*, 11(2), 225-246.

Gardner, W. (1956). "Calculation of capillary conductivity from pressure plate outflow data." *Soil Science Society of America Journal*, 20(3), 317-320.

Gardner, W. (1962). "Note on the separation and solution of diffusion type equations." *Soil Science Society of America Journal*, 26(4), 404-404.

Gallant (1975). "Nonlinear regression." *The American Statistician*, 29(2), 73-81.

Gay, D. A. (1995). “Development of a predictive model for pavement roughness on expansive clay.”

Ghanbarian-Alavijeh, B., Liaghat, A., Huang, G.-H., and Van Genuchten, M. T. (2010). "Estimation of the van Genuchten soil water retention properties from soil textural data." *Pedosphere*, 10.1016/S1002-0160(10)60035-5.

Ghiglia, D. C., and Pritt, M. D. (1998). "Two-dimensional phase unwrapping: theory, algorithms, and software." *A Wiley Interscience Publication*.

Goldstein, R. M., Engelhardt, H., Kamb, B., and Frolich, R. M. (1993). "Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream." *Science*, 262(5139), 1525-1530.

Gu, F., Sahin, H., Luo, X., Luo, R., and Lytton, R. L. (2014). "Estimation of resilient modulus of unbound aggregates using performance-related base course properties." *Journal of Materials in Civil Engineering*, 27(6), 04014188.

Guymon, G. L., Berg, R. L., and Hromadka, T. V. (1993). “Mathematical model of frost heave and thaw settlement in pavements.” Cold Regions Research And Engineering Lab Hanover NH.

Hallema, D. W., Périard, Y., Lafond, J. A., Gumiere, S. J., and Caron, J. (2015). "Characterization of water retention curves for a series of cultivated Histosols." *Vadose Zone J.*, 10.2136/vzj2014.10.0148.

Han, Z., Vanapalli, S. K., and Zou, W.-l. (2017). "Integrated approaches for predicting soil-water characteristic curve and resilient modulus of compacted fine-grained subgrade soils." *Canadian Geotechnical Journal*, 54(5), 646-663.

Hanssen, R. F. (2001). *Radar interferometry: data interpretation and error analysis*, Springer Science & Business Media.

Hallema, D. W., Périard, Y., Lafond, J. A., Gumiere, S. J., and Caron, J. (2015). "Characterization of water retention curves for a series of cultivated Histosols." *Vadose Zone Journal*, 14(6), vzj2014. 2010.0148.

Hertaeg, M. J., Tabor, R. F., Berry, J. D., and Garnier, G. (2020). "Radial wicking of biological fluids in paper." *Langmuir*, 36(28), 8209-8217.

Hooper, A. (2008). "A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches." *Geophysical Research Letters*, 35(16).

Hooper, A., Bekaert, D., Ekbal, H., and Spaans, K. (2018). "StaMPS/MTI manual: Version 4.1 b." *School of Earth and Environment, University of Leeds. Retrieved October*, 15, 2019.

Hooper, A., Bekaert, D., Spaans, K., and Arıkan, M. (2012). "Recent advances in SAR interferometry time series analysis for measuring crustal deformation." *Tectonophysics*, 514, 1-13.

Hooper, A., Zebker, H., Segall, P., and Kampes, B. (2004). "A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers." *Geophysical research letters*, 31(23).

Hooper, A., and Zebker, H. A. (2007). "Phase unwrapping in three dimensions with application to InSAR time series." *JOSA A*, 24(9), 2737-2747.

Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: principles and practice*, OTexts.

Huang, S., Barbour, S., and Fredlund, D. (1998). "Development and verification of a coefficient of permeability function for a deformable unsaturated soil." *Canadian Geotechnical Journal*, 35(3), 411-425.

Illston, B. G., Basara, J. B., Fiebrich, C. A., Crawford, K. C., Hunt, E., Fisher, D. K., Elliott, R., and Humes, K. (2008). "Mesoscale monitoring of soil moisture across a statewide network." *J. Atmos. Oceanic Technol*. 25(2), 167-182.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, Springer. 10.1007/978-1-4614-7138-7.

Jarvis, N., Bergström, L., and Dik, P. (1991). "Modelling water and solute transport in macroporous soil. II. Chloride breakthrough under non-steady flow." *Journal of Soil Science*, 42(1), 71-81.

Javid, A. H., and Bulut, R. (2019). "Evaluating Equilibrium Matric Suctions under Pavement System Based on Thornthwaite Moisture Index (TMI)." *Airfield and Highway Pavements 2019: Testing and Characterization of Pavement Materials*, American Society of Civil Engineers Reston, VA, 511-521.

Karunarathne, A., Gad, E., Disfani, M., Sivanerupan, S., and Wilson, J. (2016). "Review of calculation procedures of Thornthwaite Moisture Index and its impact on footing design." *Australian Geomechanics*, 51(1), 85-95.

Khoury, N. N., and Zaman, M. M. (2004). "Correlation between resilient modulus, moisture variation, and soil suction for subgrade soils." *Transportation Research Record*, 1874(1), 99-107.

Kohonen, T. (2012). *Self-organization and associative memory*, Springer Science & Business Media.

Lambe, T. W., and Whitman, R. V. (1991). *Soil mechanics*, John Wiley & Sons.

Legates, D. R., and Junghenn, K. T. (2018). "Evaluation of a simple, point-scale hydrologic model in simulating soil moisture using the Delaware environmental observing system." *Theoretical and applied climatology*, 132(1-2), 1-13.

Liang, R. Y., Rabab'ah, S., and Khasawneh, M. (2008). "Predicting moisture-dependent resilient modulus of cohesive soils using soil suction concept." *Journal of Transportation Engineering*, 134(1), 34-40.

Luo, X., Gu, F., Zhang, Y., Lytton, R. L., and Zollinger, D. (2017). "Mechanistic-empirical models for better consideration of subgrade and unbound layers influence on pavement performance." *Transportation Geotechnics*, 13, 52-68.

Lytton, R. L. "Prediction of movement in expansive clays." In *Proc., Vertical and horizontal deformations of foundations and embankments*, ASCE, 1827-1845.

Ma, D., Wang, Q., and Shao, M. (2009). "Analytical method for estimating soil hydraulic parameters from horizontal absorption." *Soil Science Society of America Journal*, 73(3), 727-736.

Mabirizi, D., and Bulut, R. (2010). "Unified testing method for measuring diffusion coefficients for unsaturated soil drying and wetting in laboratory." *Transp. Res. Rec.* 2170(1), 109-118.

Mabirizi, D., and Bulut, R. (2011). "Moisture Flow in Unsaturated Subgrade Soils." *Advances in Unsaturated Soil, Geo-Hazard, and Geo-Environmental Engineering*, 42-49.

Madhulatha, T. S. (2012). "An overview on clustering methods." *arXiv preprint arXiv:1205.1117*.

Massonnet, D., Briole, P., and Arnaud, A. (1995). "Deflation of Mount Etna monitored by spaceborne radar interferometry." *Nature*, 375(6532), 567-570.

Massonnet, D., and Feigl, K. L. (1998). "Radar interferometry and its application to changes in the Earth's surface." *Reviews of geophysics*, 36(4), 441-500.

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaute, T. (1993). "The displacement field of the Landers earthquake mapped by radar interferometry." *nature*, 364(6433), 138-142.

Massonnet, D., and Sigmundsson, F. (2000). "Remote sensing of volcano deformation by radar interferometry from various satellites." *Washington DC American Geophysical Union Geophysical Monograph Series*, 116, 207-221.

McKeen, R. G., and Johnson, L. D. (1990). "Climate-controlled soil design parameters for mat foundations." *Journal of Geotechnical engineering*, 116(7), 1073-1094.

Mikhailov, V., Kiseleva, E., Dmitriev, P., Golubev, V., Smolyaninova, E., and Timoshkina, E. (2014). "On reconstruction of the three displacement vector components from SAR LOS displacements for oil and gas producing fields." *Procedia Technology*, 16, 385-393.

Mitchell, P. W. "The structural analysis of footings on expansive soil." *Proc., Expansive Soils*, ASCE, 438-447.

Nasta, P., Vrugt, J. A., and Romano, N. (2013). "Prediction of the saturated hydraulic conductivity from Brooks and Corey's water retention parameters." *Water Resources Research*, 49(5), 2918-2925.

Ngoc, T. P., Fatahi, B., Khabbaz, H., and Sheng, D. (2020). "Impacts of matric suction equalization on small strain shear modulus of soils during air drying." *Canadian Geotechnical Journal*, 57(12), 1982-1997.

Olaiz, A. H., Singhar, S. H., Vann, J. D., and Houston, S. L. (2018). "Comparison and applications of the Thornthwaite moisture index using GIS." *PanAm Unsaturated Soils 2017*, 280-289.

Oloo, S. "The application of unsaturated soil mechanics theory to the design of pavements." *Proc., Proc. 1st Int. Conf. on the Bearing Capacity of Roads and Airfields, BCRA'98*, 1419-1428.

Pasha, A. Y., Khoshghalb, A., & Khalili, N. (2016). “Pitfalls in interpretation of gravimetric water content–based soil-water characteristic curve for deformable porous media.” *Int. J. Geomech*. 16(6), D4015004.

Pasha, A. Y., Khoshghalb, A., & Khalili, N. (2017). “Hysteretic model for the evolution of water retention curve with void ratio.” *J. Eng. Mech*. 143(7), 04017030. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001238

Parker, J., Kool, J., and Van Genuchten, M. T. (1985). "Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: II. Experimental studies." *Soil Science Society of America Journal*, 49(6), 1354-1359.

Parlange, J.-Y. (1971). "Theory of water-movement in soils: I. One-dimensional absorption." *Soil science*, 111(2), 134-137.

Passioura, J. (1977). "Determining soil water diffusivities from one-step outflow experiments." *Soil Research*, 15(1), 1-8.

Peltzer, G., and Rosen, P. (1995). "Surface displacement of the 17 May 1993 Eureka Valley, California, earthquake observed by SAR interferometry." *Science*, 268(5215), 1333-1336.

Penman, H. L. (1963). "Vegetation and hydrology." *Soil Science*, 96(5), 357.

Perera, Y. Y., Zapata, C. E., Houston, W. N., and Houston, S. L. (2004). "Long-term moisture conditions under highway pavements." *Geotechnical Engineering for Transportation Projects*, 1132-1143. https://doi.org/10.1061/40744(154)102.

Potter, K., Hagen, H., Kerren, A., and Dannenmann, P. (2006). "Methods for presenting statistical information: The box plot." *Visualization of large and unstructured data sets*, 4, 97-106.

PTI (2004). *Design and construction of post-tensioned slabs-on-ground*, Post-Tensioning Institute, Phoenix, Arizona.

Puppala, A. J., Manosuthikij, T., Hoyos, L., and Nazarian, S. (2009). "Moisture and Suction in Clay Subgrades Prior to Initiation of Pavement Cracking."

Quan-Jiu, W., Zhang, J.-H., and Jun, F. (2006). "An analytical method for relationship between hydraulic diffusivity and soil sorptivity." *Pedosphere*, 16(4), 444-450.

Richards, L. A. (1931). "Capillary conduction of liquids through porous mediums." *Physics*, 1(5), 318-333.

Rignot, E., Gogineni, S., Krabill, W., and Ekholm, S. (1997). "North and northeast Greenland ice discharge from satellite radar interferometry." *Science*, 276(5314), 934-937.

Rosen, P. A., Hensley, S., Joughin, I. R., Li, F. K., Madsen, S. N., Rodriguez, E., and Goldstein, R. M. (2000). "Synthetic aperture radar interferometry." *Proceedings of the IEEE*, 88(3), 333-382.

Russam, K., and Coleman, J. (1961). “The effect of climatic factors on subgrade moisture conditions.” *Geotechnique*. 11(1), 22-28. https://doi.org/10.1680/geot.1961.11.1.22.

Saha, S., Hariharan, N., Gu, F., Luo, X., Little, D. N., and Lytton, R. L. (2019). "Development of a mechanistic-empirical model to predict equilibrium suction for subgrade soil." *Journal of Hydrology*, 575, 221-233. https://doi.org/10.1016/j.jhydrol.2019.05.035.

Sandwell, D. T., and Price, E. J. (1998). "Phase gradient approach to stacking interferograms." *Journal of Geophysical Research: Solid Earth*, 103(B12), 30183-30204.

Schaap, M. G., Leij, F. J., and Van Genuchten, M. T. (1998). "Neural network analysis for hierarchical prediction of soil hydraulic properties." *Soil Science Society of America Journal*, 62(4), 847-855.

Scott, B. L., Ochsner, T. E., Illston, B. G., Fiebrich, C. A., Basara, J. B., and Sutherland, A. J. (2013). "New soil property database improves Oklahoma Mesonet soil moisture estimates." *Journal of Atmospheric and Oceanic Technology*, 30(11), 2585-2595. https://doi.org/10.1175/JTECH-D-13-00084.1.

Sommer, R., and Stöckle, C. (2010). "Correspondence between the Campbell and van Genuchten soil-water-retention models." *Journal of irrigation and drainage engineering*, 136(8), 559-562.

Struyf, A., Hubert, M., and Rousseeuw, P. (1997). "Clustering in an object-oriented environment." *J. Stat. Softw.*, 10.18637/jss.v001.i04.

Sun, X., Li, J., and Zhou, A. (2017). “Evaluation and comparison of methods for calculating Thornthwaite Moisture Index.” *J. Aust. Geomech*. 52(2), 61-75.

Thornthwaite, C., and Mather, R. (1955). "The water balance publication in climatology, 8 (1) DIT." *Laboratory of Climatology, Centerton, NJ 104p*.

Thornthwaite, C. W. (1948). "An approach toward a rational classification of climate." *Geographical review*, 38(1), 55-94. https://doi.org/10.2307/210739.

Tsay, R. S. (1986). "Time series model specification in the presence of outliers." *Journal of the American Statistical Association*, 81(393), 132-141.

Tukey, J. W. (1977). *Exploratory data analysis*, Reading, MA.

Van Genuchten, M. T. (1980). "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils." *Soil Sci. Soc. Am. J.*, 10.2136/sssaj1980.03615995004400050002x.

Vanapalli, S., Fredlund, D., Pufahl, D., and Clifton, A. (1996). "Model for the prediction of shear strength with respect to soil suction." *Canadian geotechnical journal*, 33(3), 379-392.

Vann, J. D. (2019). “A soil suction-oedometer method and design soil suction profile recommendations for estimation of volume change of expansive soils.” Arizona State University.

Vann, J., and Houston, S. (2021). "Field Soil Suction Profiles for Expansive Soil." *Journal of Geotechnical and Geoenvironmental Engineering*, 147(9), 04021080.

Veci, L., Prats-Iraola, P., Scheiber, R., Collard, F., Fomferra, N., and Engdahl, M. "The sentinel-1 toolbox." *Proc., Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 1-3.

Wang, Q., Shao, M., and Horton, R. (2004). "A simple method for estimating water diffusivity of unsaturated soils." *Soil Science Society of America Journal*, 68(3), 713-718.

Willmott, C. J., and Feddema, J. J. (1992). "A more rational climatic moisture index." *The Professional Geographer*, 44(1), 84-88.

Witczak, M. W., Zapata, C. E., and Houston, W. N. (2006). "Models incorporated into the current enhanced integrated climatic model: NCHRP 9-23 project findings and additional changes after version 0.7." *Final Report, Project NCHRP*.

Wolfe, W., and Butalia, T. (2004). "Continued monitoring of SHRP pavement instrumentation including soil suction and relationship with resilient modulus."

Wright, T., Parsons, B., and Fielding, E. (2001). "Measurement of interseismic strain accumulation across the North Anatolian Fault by satellite radar interferometry." *Geophysical Research Letters*, 28(10), 2117-2120.

Yagüe-Martínez, N., Prats-Iraola, P., Gonzalez, F. R., Brcic, R., Shau, R., Geudtner, D., Eineder, M., and Bamler, R. (2016). "Interferometric processing of Sentinel-1 TOPS data." *IEEE Transactions on Geoscience and Remote Sensing*, 54(4), 2220-2234.

Yu, C., Li, Z., Penna, N. T., and Crippa, P. (2018). "Generic atmospheric correction model for Interferometric Synthetic Aperture Radar observations." *Journal of Geophysical Research: Solid Earth*, 123(10), 9202-9222.

Yue, E., and Bulut, R. “Evaluation of climatic factors for the classification of oklahoma pavement regions.” In *Proc., Geo-Congress 2014: Geo-characterization and Modeling for Sustainability*, 4037-4046.

Zapata, C. E., Andrei, D., Witczak, M. W., and Houston, W. (2007). “Incorporation of environmental effects in pavement design.” *Road Mater. Pavement Des.* 8(4), 667-693.

Zhang, J., Peng, J., Zheng, J., Dai, L., and Yao, Y. (2019). “Prediction of resilient modulus of compacted cohesive soils in South China.” *Int. J. Geomech.* 19(7), 04019068.

Zhang, Y., Ochsner, T. E., Fiebrich, C. A., and Illston, B. G. (2019). “Recalibration of sensors in one of the world's longest running automated soil moisture monitoring networks.” *Soil Sci. Soc. Am. J.* 83(4), 1003-1011.

Zebker, H. A., Rosen, P. A., and Hensley, S. (1997). "Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps." *Journal of geophysical research: solid earth*, 102(B4), 7547-7563.

VITA

Amir Hossein Javid

Candidate for the Degree of

Doctor of Philosophy

Dissertation: VARIATION OF SOIL SUCTION AND APPLICATION OF REMOTE SENSING IN EVALUATING UNSATURATED SOIL BEHAVIOR WITHIN VADOSE ZONE

Major Field: Civil Engineering

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Civil Engineering at Oklahoma State University, Stillwater, Oklahoma in Spring, 2023.

Master of Science in Statistics at Oklahoma State University, Stillwater, Oklahoma in 2022.

Master of Science in Civil Engineering at Texas Tech University, Lubbock, Texas in 2017.

Master of Science in Geotechnical Engineering at Amirkabir University of Technology, Tehran, Iran in 2014.

Bachelor of Science in Civil Engineering at Amirkabir University of Technology, Tehran, Iran in 2011.

Professional Memberships:

Member of American Society of Civil Engineers (ASCE)

Member of American Statistical Association (ASA)