

The Pennsylvania State University

The Graduate School

College of Agricultural Sciences

**CATCHMENT-SCALE SOIL WATER RETENTION
CHARACTERISTICS AND DELINEATION OF
HYDROPEDOLOGICAL FUNCTIONAL UNITS IN THE
SHALE HILLS CATCHMENT**

A Thesis in

Soil Science

by

Douglas C. Baldwin

© 2011 Douglas C. Baldwin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2011

The thesis of Douglas C. Baldwin was reviewed and approved* by the following:

Henry Lin

Associate Professor of Hydropedology/Soil Hydrology
Thesis Advisor

Patrick Drohan

Assistant Professor of Pedology

Christopher Duffy

Professor of Civil Engineering

Jack Watson

Professor of Soil Science
Interim Department Head of the Crop and Soil Science Department

*Signatures are on file in the Graduate School.

Abstract

Quantification of soil hydraulic functions is essential to catchment studies. While soil water retention has been widely studied, a catchment-wide characterization of soil water retention parameters has not yet been commonly done. In this study, I report the spatial patterns of soil water retention parameters, obtained through in situ monitoring data, as a function of soil type, landform unit, and soil depth in the forested Shale Hills catchment in central Pennsylvania. Soil water matric potential and volumetric moisture content were collected in 2005-2010 at 61 sites throughout the catchment and at depths of 10, 20, 40, 80, and 100 cm. These data were fitted with the van Genuchten, Campbell, and Gardner soil water retention equations. Parameters from these curves were then analyzed in connection to soil-terrain attributes. Based on various statistical analyses, topographic wetness index, depth to bedrock, and curvature were found to have significant influence on soil water retention parameters across this catchment.

The spatial relationship of the van Genuchten parameters across the catchment at 10, 20, 40, 80 and 100 cm depths were quantitatively analyzed and compared. An increase in spatial variance from near surface 20 cm to the deeper 80 cm was evident in saturated moisture content as the semi-variogram range increased from 14.3 meters to 31.2 meters. Moisture retention parameters were estimated across the catchment at all available depths with regression kriging using Bayesian statistics to optimize spatial model parameters. These maps would be used to inform hydrologic modeling and ecological studies for the Shale Hills catchment.

A foundation for delineating functional units of the catchment with similar hydrological, pedogenic, and topographic properties (called Hydropedological Functional Units or HFUs) was established in this study through analyzing: 1) soil moisture profile storage maps spanning 2008 through 2010, 2) catchment-wide electromagnetic induction surveys from wet and dry seasons, 3) maps of topographic variables, and 4) landscape-scale soil retention parameters and other basic soil properties. Five apparent HFUs were identified based on the spatially and temporally extensive datasets: 1) Regression-kriged maps of depth to bedrock and solum total moisture storage at saturation were used in a combined principal component and fuzzy c-means clustering to quantitatively delineate

three hillslope HFUs; and 2) Two additional HFUs were delineated using slope value, elevation, and upslope contributing area. Total five HFUs were then compared with soil series map and landform units delineated using Park and vande Giesen's method (2004). All three delineation methods were compared with observed soil moisture data. Results showed that the HFUs out-performed the soil series map and the landform units in depicting soil solum moisture storage in general linear models.

This study shows that catchment-wide characterization of saturated moisture content can be integrated with topography and depth to bedrock to delineate Hydropedological Functional Units. Delineated HFUs generally predict soil moisture patterns more accurately than soil series and landform units. This research confirms that the delineation of sub-catchment units with soil depth, topographic variables, and soil properties can sufficiently separate a catchment into units with similar topography, soil properties, and hydrologic functions.

TABLE OF CONTENTS

	Page
List of Figures	viii
List of Tables	xii
Acknowledgements	xiii
Chapter 1: Introduction.....	1
1.1 Thesis Objectives	4
1.2 Literature Review on Catchment-Scale Soil Water Retention Characteristics	4
1.3 Literature Review of the Delineation of Hydropedological Functional Units	8
Chapter 2: Materials and Methods.....	10
2.1 The Shale Hills Catchment	10
2.2 Soil Moisture Retention Data Collection	13
2.3 Soil Moisture Retention Modeling	14
2.4 Maximum Likelihood Estimation of van Genuchten Parameters and Moisture Retention Model Diagnostics	17
2.5 Statistical Analysis of Moisture Retention Parameters	18
2.6 Geo-statistical Analysis of Moisture Retention Parameters	19
2.7 Topographic Wetness Index and Kriging of Total Moisture Storage in Profile	22
2.8 GIS Processing of Topographic Raster Maps and Depth to Bedrock Map	24

2.9 Refinement of Catchment-scale Depth to Bedrock Interpolation	29
2.10 Delineated Landform Units in the Shale Hills Using Park and vande Giesen (2004) Method	31
2.11 Analysis of Topographic and Hydrologic Datasets for Informing Hydropedological Functional Unit concept and generation	34
2.12 Quantitative Analysis of Raster Datasets to delineate HFUs	42
2.13 Validation of HFUs	47
Chapter 3: Results and Discussion	51
3.1 Landscape Factors Controlling Soil Moisture Retention	51
3.2 Spatial Distribution of the van Genuchten Parameters across the Shale Hills Landscape	62
3.3 General Terrain and Soil Characteristics of Hydropedological Functional Units	66
3.4 Validation and Cross-Validation of Hydropedological Functional Units with Soil Moisture Data	74
Chapter 4: Summary and Future Work	89
4.1 Summary	89
4.2 Future Work	92
References	94

Appendix A: Maximum Likelihood Optimization Procedure	99
Appendix B: Maps of the Van Genuchten Soil Water Retention Parameters	102
Appendix C: Maps of Basic Soil Properties	107
Appendix D: Additional Tables and Correlation Matrices	111

LIST OF FIGURES

Figure 2.1: Schematic showing the thickness of the solum and soil horizons as observed in soil pits dug for each of the five soil series in the Shale Hills and the precision soil map for Shale Hills. The Rushtown, Blairton, and Ernest soil profiles are deeper than the scale indicated as the soil pits could not go further down due to safety issue and the limitation of the backhoe used. Colored points indicate the exact depths that tensiometers were installed: red = 10 cm; blue = 20 cm; green = 40 cm; yellow = 80 cm; and cyan = 100 cm depth. Horizons are color coded indicating approximate real soil color **11**

Figure 2.2: Topographic Wetness Index map for the Shale Hills, with TDR-tensiometer site locations colored according to designated soil series. Numbers indicate site locations in a toposequence within the catchment that correspond to sites featured in Figure 3.2. The dotted box indicates the approximate boundary of the focused toposequence reported **23**

Figure 2.3: 0.5 meter resolution DEM for the Shale Hills generated from LiDAR flyover in February 2011. LiDAR data was preprocessed in UC Merced before subsequent processing in this study **25**

Figure 2.4: Maps of slope, log(upslope contributing area), surface curvature and the initial depth to bedrock for the Shale Hills. TDR site locations are plotted with point size on slope value, upslope contributing area and surface curvature indicating the magnitude of the map raster value at a site. Depth to bedrock shows the 223 augur measurement locations plotted with point size showing relative magnitude of soil depth **28**

Figure 2.5: Refined depth to bedrock map with all the available data in the Shale Hills (385 total). Included is the semi-variogram depicting spatial variance of depth to bedrock data fitted with a Spherical covariance function **30**

Figure 2.6: Scatterplot of log(A_s) and C_s raster data for the Shale Hills. Colored lines demarcate thresholds set for the delineation of 5 landform units. See Fig. 2.7 for actual spatial distribution of these 5 units for the Shale Hills **32**

Figure 2.7: Map of landform units (LFUs) in the Shale Hills delineated with the method of Park and vande Giesen (2004) **33**

Figure 2.8: Map of aggregated θ_s (solum storage at saturation). Regression kriging was used to interpolate θ_s with depth to bedrock being the significant regressor **35**

Figure 2.9: Maps of total moisture storage in solum during intermediate wetness conditions. The black oval in the center of the black circle indicates the origin of the ephemeral stream. Notice darker blue colors occur in areas lower in elevation than stream origin, while only light blue colors appear at elevations higher than stream origin **40**

Figure 2.10: Catchment-wide Electromagnetic Induction (EMI) surveys performed in the Shale Hills for the wet (April) and intermediate (October) moisture conditions. High EMI values indicate a higher concentration of electrically conductive material, such as soil water or saline material, within the soil. Higher EMI values appear over soil directly adjacent to the stream and abruptly decrease beyond the stream for both surveys. Maps were created using ordinary kriging interpolation **41**

Figure 2.11: Flow diagram depicting the sequence of processes (yellow trapezoids) with inputs and subsequent outputs (green boxes). The final HFU product is in light blue color **45**

Figure 3.1: Box plots of coefficient of determination (R^2 values) for the three soil moisture retention models for 232 depth-locations at 61 sites across the catchment. R-squared values of the three retention models are grouped according to landform unit where depth locations reside. The small letters at the top of each bar indicate significant differences and the numbers indicate sample size **52**

Figure 3.2: Series of total storage moisture maps from May to November in 2010. Moisture data was collected at 61 sites (indicated by red dots) across the catchment. Moisture storage values were interpolated across the catchment using regression kriging. Histograms show distribution of observed total storage values at all sites for each date. γ values for semi-variograms have been binned, and Spherical models were used to fit semi-variograms **53**

Figure 3.3: Soil Moisture Retention Curves using the van Genuchten retention model for sites along a toposequence in Shale Hills (see Figure 2). Soil moisture retention curves were fitted for all available depths at each site. Parameters and R-square values for each function are listed in the plot. The curves and observed data points are color-coded according to depth **54**

Figure 3.4: Boxplots of θ_s , n , α and ψ_e parameters for all depth-locations grouped by soil series. Letters indicate significant differences of means for parameter values among different soil series within each depth interval according to a Tukey HSD test at $p < 0.05$. Numbers above each box indicate sample size. α and ψ_e were natural log-transformed and n log10-transformed to maintain normality for the Tukey HSD test **57 - 58**

Figure 3.5: Series of semi-variograms showing semi-variance of α , n and θ_s at shallow (20 cm) and deep solum (80 cm) depths across the catchment. The γ values for each semi-variogram have been binned. The covariance models in each semi-variogram are Matern functions, and parameters for the covariance functions were optimized using a Bayesian estimation procedure. Parameters were transformed using a λ value (Table 3.3) optimized by a maximum likelihood algorithm during spatial model fitting to maintain independence and isotropy. Dotted lines indicate 95% Confidence Interval envelope of the data cloud **64**

Figure 3.6: Maps of θ_s at four different depths over Shale Hills. Regression kriging with significant topographic variables was used for interpolation. A Bayesian optimization was employed to optimize spatial model parameters. Areas with soil shallower than featured depth have been masked **65**

Figure 3.7: Bar plot depicting the amount of variance explained by PCI and PCII. PCI explains the vast majority of the variance and was used for fuzzy c-means cluster analysis 68

Figure 3.8: Raster of 3 clusters based upon multivariate analysis and clustering of Depth to Bedrock and θ_s Storage raster datasets. TDR-Tensiometer site locations are shown and symbolized according to ascribed soil series 69

Figure 3.9: Map of slope class delineations for Shale Hills. Soil Series boundaries are demarcated in bold. Slope Class D (0.15 – 0.25) does exist within the Ernest soil series boundary 71

Figure 3.10: Final HFU map with a DSHSFV unit (blue) separated from the DSHSCH (cyan) and ISMSCH (green). A SSLSFS (red) was separated from the SSLSPH (yellow). The ephemeral stream is demarcated and is shown to reside within the DSHSFV unit 72

Figure 3.11: Hydropedological Functional Units with 2nd Order Soil Series delineation and catchment ephemeral stream demarcation 73

Figure 3.12: Performance of 3 landscape characterizations in predicting total moisture storage within a linear model. The y-axis shows dates of moisture collection with average total storage from observations shown directly above. Lines indicate adjusted R² values indicating performance of each predictor in predicting total moisture storage in a linear model for each date 76

Figure 3.13: Adjusted R² diagnostic for linear modeling of volumetric moisture content at 10, 20, 40, 60, 80 and 100 cm depths with HFUs (blue), soil series (red), and LFUs (green) as predictors. Data were individually modeled for 58 dates from 2006 – 2010. The average volumetric moisture content in a specific depth from data collected at all sites in the catchment during a given date is shown on the bottom on each graph 78 - 80

Figure 3.14: Boxplots of all solum moisture storage data points from 58 dates spanning 2006 – 2010 for individual categories in HFUs, soil series and LFUs. Numbers above the boxplot indicate sample size. Letters show significant difference between categories of average solum moisture storage collected from the same aforementioned 58 dates from all TDR-tensiometer sites in the catchment 82

Figure 3.15: Boxplots of total profile texture storage of clay, silt, sand and organic matter (OM) collected from 58 different sites between categories of HFUs, soil series and LFUs. A Tukey HSD test was applied to find significant differences among categories for each texture class at $p < 0.05$. Numbers above boxplots indicate sample size 84

Figure 3.16: Three-dimensional rendition of Hydropedological Functional Units depicting areas of similar soil, terrain and hydrologic properties delineated for the Shale Hills. The stream is pictured, colored white 88

Figure B.1: Series of maps depicting the α soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate α values. Areas with depth to bedrock less than featured depth have been masked for each map 104

Figure B.2: Series of maps depicting the n soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate n values. Areas with depth to bedrock less than featured depth have been masked for each map **105**

Figure B.3: Series of maps depicting the θ_s soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate θ_s values. Areas with depth to bedrock less than featured depth have been masked for each map **106**

Figure C.1: Maps of total rock fragment, clay, silt and sand storage generated from regression kriging analysis on respective total rock fragment, clay, silt and sand storage values at 58 sites across Shale Hills. Although validation diagnostics showed R^2 values > 0.830 for all spatial models, nugget and trend effects may cause estimated data values to be different from observed data at some site locations **108**

Figure C.2: Series of maps depicting estimated clay content in Shale Hills from sample locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. Areas with depth to bedrock less than featured depth have been masked for each map **109**

Figure C.3: Series of maps depicting estimated sand content in Shale Hills from sample locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. Areas with depth to bedrock less than featured depth have been masked for each map **110**

LIST OF TABLES

Table 2.1: Identified units representing the general patterns of important landscape variables and selected soil moisture storage maps representing three different wetness conditions (wet, intermediate, dry, and saturation) 37

Table 3.1: Three-Way ANOVA of each soil water retention parameter against soil series, landform unit and depth factors. Significant categorical controls on the variance of parameter values were revealed through a generalized linear model. Some parameters were log-transformed before analysis to satisfy the normality requirement for the test 56

Table 3.2: Spearman correlation matrix of soil hydraulic parameters and soil-terrain attributes from all depth-locations in the Shale Hills. Values that are significant at $p < 0.05$ are highlighted in red and the p-value is indicated in parentheses 60

Table 3.3: Semivariogram parameters for Figure 8. The van Genuchten parameter values were transformed with the *Lambda* parameter to satisfy spatial isotropy. The range is the distance on the x-axis at the inflection point of the curve. The nugget is the intercept of the curve on the y-axis, and the partial sill is the distance on the y-axis between the nugget and maximum height of the curve. Correlated topographic covariates used in spatial modeling are listed with significance indicated (**** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$) 63

Table 3.4: Correlation matrix of raster datasets from a Spearman non-parametric correlation test. Red emboldened correlation coefficients indicate a significant correlation between variables at $p < 0.05$, and blue emboldened correlation coefficients represent a significant correlation at $p < 0.10$ 68

Table 3.5: Summary statistics (mean and standard deviation) of raster variables for each HFU category 73

Table 3.6: Results of linear and generalized linear models for each categorical predictor to 31 moisture total storage responses from dates in 2007 – 2010, a wet date (3/27/2007) and a dry date (7/30/2010). The best diagnostics are highlighted in grey. ANOVA and Coefficient of Variation results of each model are shown to the left of each predictor 75

Table D.1: Summary of alpha values for each soil series and landform unit separated by depth location. Letters beside the means indicate a significant difference of alpha values among depths within a landform unit or soil series according to a Tukey HSD test with a p-value < 0.0 111

Table D.2(a-e): Correlation matrices of soil water retention parameters (cyan) with topography (green) and soil properties (orange) at 5 different depths. A Spearman correlation was used to test significance. Values in red are significant at $p < 0.05$, with the p-value in parentheses next to the correlation coefficient 112 - 113

ACKNOWLEDGEMENTS

Much gratitude is extended to my adviser Dr. Henry Lin for all the time and effort spent in offering useful guidance throughout the course of my Master's program. This academic endeavor would not have been successful without Dr. Lin's support and advice.

The time and effort my committee members Dr. Patrick Drohan and Dr. Chris Duffy offered in reviewing my thesis work plan and critiquing this product is greatly appreciated.

This study was partially supported by the U.S. National Science Foundation through the Shale Hills Critical Zone Observatory grant (EAR-0725019)

Last but certainly not least, I am grateful for all the kind support that my family and friends have offered me throughout the course of my program.

Chapter 1: Introduction

Hydropedology endorses scientific pursuits of integrating soil, hydrology, and landscape across scales (Lin et al, 2008, Lin, 2010). The use of spatial-temporal patterns of soil moisture in relation to terrain attributes and soil properties is an effective means of deciphering the complex interactions and dynamics of landscape-soil-hydrology relationships. A better understanding of how soil properties and terrain attributes control hydrologic processes will facilitate improved understanding of hydrologic cycle, natural resource management, and pollution assessment and control.

Effective land use management is becoming increasingly imperative as human population increases and natural resources diminish. General land management practices do not adequately characterize the heterogeneity of soil properties and hydrologic functions across the landscape and consequently do not effectively protect environmental quality and conserve natural resources. Precision mapping of soil-hydrological properties across the landscape will facilitate more efficient and sustainable land management decisions (Lin et al., 2005; Ticehurst et al, 2007). For example, detailed maps of topography, soil moisture, and soil properties could be used to quantitatively delineate a watershed into sub-units with similar landscape features, soil properties, and hydrologic function. Such sub-units are termed “Hydropedological Functional Units (HFUs)” in this study, which provide building blocks for enhanced spatially-distributed soil and hydrologic modeling.

Soil science research has moved beyond the classical factorial analysis framework of Jenny’s (1941) by incorporating flow and transformation of matter and energy as a means to quantitatively distinguish soil-related natural processes and formation

(Misnasny et al, 2008; Rasmussen et al, 2005). Hydropedology, for example, recognizes the importance of hydrologic processes in soil formation and ecological functions and seeks to quantitatively understand how landscape features and soil properties exert a 1st-order control on hydrologic and ecological processes and, at the same time, how hydrologic and ecological processes feedback to soil formation, evolution, and functions (Lin et al., 2005, 2006, 2008). Hydropedology also emphasizes continuous monitoring and realistic modeling based on *in situ* field data (Baggaley et al. 2009; Lin, 2006; Lin et al., 2008). Hydropedology is also an important contributor to the holistic study of the Earth's Critical Zone. The Critical Zone is that portion of Earth's near-surface crust from the vegetation top down to the aquifer bottom, in which complex interactions involving rock, soil, water, air, and living organisms regulate the natural habitat and determine the availability of nearly-every life sustaining resources (NRC, 2001).

Hydropedologic research utilizes mapping to effectively portray soil properties in a landscape context. Hydropedological Functional Units (HFUs) are defined as landscape units with similar topographic, soil structural and soil-hydrologic properties (Lin et al, 2006b; Lin et al, 2008). HFUs represent divisions on the landscape with similar soil water retention, soil depth, soil moisture patterns, and soil morphologic features; they may greatly enhance soil mapping and hydrologic modeling since they represent divisions in the landscape with optimally low variability in soil moisture patterns, soil texture, and soil hydrologic characteristics (Lin et al., 2008). The delineation of HFUs by using topographic variables is possible, because soil moisture distribution variation across a catchment has been correlated with topographic variables such as elevation, slope, aspect, and depth to bedrock (Baggaley et al, 2009; Takagi, 2009).

Baggaley et al. (2009) have shown in their study that 80 % of soil moisture variation across a landscape in the UK was explained with topographic indices and soil map units.. Zhu et al (2010b) were able to integrate ECa values collected during EMI surveys, depth to bedrock and terrain indices through using geo-statistics to obtain the most optimum soil mapping accuracy for an agricultural landscape. Research at Shale Hills has also found correlations between soil properties and topographic indices with soil moisture patterns (Takagi, 2009). Takagi (2009) also found that topographic properties, such as slope and depth to bedrock, are useful regressors in predicting soil moisture in Shale Hills. Research from Winter (2001) has shown that landscapes may be partitioned into landscape-hydrologic units by defining areas of similar topography. As the HFUs express information regarding quantitative soil hydrologic functions across a landscape, a necessary task in the delineation of HFUs is to derive at least one of the two major soil-hydrologic functions: soil water retention or soil hydraulic conductivity. Mapping hydrologic soil properties would then relate soil hydrologic characteristics to a landscape- or catchment-scale. If topography and soil properties are correlated with soil moisture in a given landscape, maps of soil depth and topographic indices must be integrated with maps of soil hydrologic characteristics to delineate HFUs for that landscape.

The objective in this thesis research is to characterize soil water retention across the Shale Hills catchment at multiple depths, and to delineate Hydropedological Functional Units based on a combination of landscape features, terrain attributes, soil properties (including basic soil properties and soil water retention parameters), and soil moisture storage change over a two year time period (2008-2010) for the Shale Hills Catchment – a recently established national Critical Zone Observatory.

1.1 THESIS OBJECTIVES

More specifically, the objectives of this study include the following:

- 1.** Investigate soil-water retention function across the 7.9-ha Shale Hills Catchment and its spatial patterns and controls by soil types, soil depths, soil properties, terrain attributes, and landform features;
- 2.** Delineate Hydropedological Functional Units for the entire Shale Hills Catchment using all the available geospatial data of soil-water retention parameters, basic soil properties, topographic attributes, and other landscape features; and
- 3.** Validate the delineated HFUs in explaining the observed soil moisture spatial-temporal patterns, and compare the HFUs with soil series map and landform units delineated with the method of Park and Vande Giesen (2004).

1.2 Literature Review on Catchment-Scale Soil Water Retention Characteristics

Soil moisture retention is one of two fundamental soil-hydrologic processes and has received much attention in soil physics, engineering and ecological modeling. The moisture retention function quantitatively describes a soil's moisture holding capacity, the relative influence of gravimetric vs. capillary moisture on water retention in a soil, and the hydrologic characteristics of soil pore space. Soil moisture retention parameters have been used in various applications such as hydrologic models depicting runoff, streamflow, and infiltration rates. While soil water retention function has been

investigated extensively in various soils under diverse climates, topography, and vegetation (e.g., Fredlund and Xing, 1994; Hodnett and Tomasella, 2002; Pachepsky et al, 2006; Vereecken et al, 2007), a catchment-wide characterization of soil water retention has not been commonly reported.

Past catchment-wide soil moisture retention studies have been based on indirect modeling through developing pedo-transfer functions (PTFs) from basic soil property data obtained from soil cores (Fieke et al, 1996; Romano and Santini, 1997; Herbst and Diekkruger, 2001; Herbst et al, 2006) or indirect modeling from saturated hydraulic conductivity data made from soil core samples (McDonnell, 1990; Porebska et al, 2006; Botros et al, 2009). Indirect modeling from in situ hydraulic conductivity data collected on field plots (Kool and Parker, 1988) or point locations (Inoue et al, 1998) has been accomplished, but these experiments did not span an entire catchment, nor were they temporally extensive.

The use of PTFs has been extensively employed to predict soil hydraulic parameters based on soil texture and structure data (Dashtaki et al, 2010; Vereecken et al, 2010). Direct modeling for moisture retention from data collected in situ has been accomplished for field plot scale experiments (Jabro et al, 2009). Jabro et al, 2009 used six 117 cm x 117 cm frames with TDR and Water Mark sensors placed in two opposing corners to directly collect soil moisture content and soil water pressure data, respectively. The plots were saturated and then measurements were taken up to 450 hours after infiltration of all moisture. Although it was concluded that this method is accurate in collecting soil moisture retention data directly, the field experiments were not temporally extensive and did not span an entire catchment. Comprehensive direct modeling of soil

moisture retention for an entire catchment at multiple depths requires a temporally- and spatially-extensive field monitoring campaign.

Acquiring data for direct retention modeling is time-consuming and costly (Minasny et al, 1999; Romano and Palladino, 2002) Monitoring campaigns that conduct temporally-extensive collection of volumetric moisture content and matric potential facilitate the direct modeling of water retention curves. Should resources and time be available for the establishment of a dataset with a temporally-extensive account of volumetric soil moisture and soil matric potential, then direct modeling of soil water retention is possible. Hydrologic retention parameters obtained through direct monitoring would offer a direct analysis of hydraulic parameters to soil texture, soil depth and topographic variables without having to account for various sources of uncertainty that inevitably arise from indirect modeling or using PTFs (Minasny et al, 1999).

Romano and Palladino, 2002 found that terrain indices are correlated with hydraulic retention parameters obtained through PTFs during research in a catchment located in Basilicata Region, Italy. Herbst et al, 2006 used geostatistics in predicting soil water retention parameters obtained from PTFs across the Berrensiefen catchment, east of Cologne, Germany. They concluded that retention parameters should be modeled at point locations first followed by interpolation across space with terrain variables, rather than using correlated terrain variables to predict retention parameters alone. Lin et al, 2006 found that terrain indices exert a direct control on soil moisture patterns in a temperate forested catchment in Pennsylvania. Baggaley et al, 2009 confirmed that terrain does control soil moisture processes on a landscape scale in three different fields in Bedfordshire, UK.

Past soil spatial modeling has grouped landscape positions based on terrain and soil properties (McBratney et al, 2003). Hydrologic research has grouped locations based upon hydrologic data into similar landform groupings using ordinal analysis (Kumar and Duffy, 2009). Point locations at different depths across a catchment with defined soil water retention parameters, terrain information and soil properties may therefore be clustered into similar hydrologic-soil groupings as well, using variables that exert 1st order controls on hydrologic processes. Identifying 1st order controls on processes is a key step for classifying groupings of similar soil-hydrologic point locations.

As catchment-wide characterization of directly modeled soil water retention parameters in situ is lacking, spatial modeling and interpolation of directly modeled soil water retention parameters at different depths across a catchment would shed light on how directly modeled retention parameters are spatially distributed at the near-surface and deep subsurface. Interpolating and mapping retention parameters at different depths across Shale Hills is possible after spatial modeling through universal or regression kriging.

Examining soil-hydrologic interactions at multiple spatial scales is a staple in hydropedologic research (Lin, 2010). Hydropedology emphasizes continuous monitoring as a means to better understand landscape-soil-hydrologic relationships. The Shale Hills Critical Zone observatory has been established partly as a soil-hydrologic watershed-scale laboratory, where matric potential and volumetric moisture content have been collected weekly or bi-weekly since 2005 at sites that span the watershed. This has facilitated the direct modeling of soil hydraulic parameters across the catchment and the subsequent analysis.

This study attempted to address the following questions using an extensive monitoring dataset: 1) Are soil water retention parameters at different depths and sites across the Shale Hills controlled by soil types, soil depths, soil properties, terrain attributes, and landform units? 2) How are soil water retention parameters spatially distributed in the Shale Hills at near-surface and subsurface?

1.3 Literature Review of the Delineation of Hydropedological Functional Units

Precision and digital soil mapping have become increasingly important in pedologic and hydrologic studies (McBratney et al, 2003; Moller et al, 2008; Behrens et al, 2009; Hodza, 2010; Zhu et al, 2010; Yang et al, 2011). As soil series map units provide information about multiple attributes such as soil depth, textural class, slope value, parent material, and many others, more precise and functional soils information are needed for improving the accuracy of hydrologic models that utilize soil property inputs for storage and discharge calculations (Duffy, 1996), broadening the breadth and capability of ecologic models utilizing soil hydrologic inputs to help explain vegetation dynamics (Seyfried et al, 2005), and optimizing digital soil map creation for precision agricultural studies and applications (Zhu et al, 2010b).

As hydrologic models may use soil property information, such as saturated soil moisture storage, as parameters (Duffy, 1996); delineating soil map units that embody quantitative soil-hydrologic information for a catchment may benefit catchment-scale hydrological models. Ecological variables, such as vegetation coverage, correlate with soil depth and hydrologic data (Thompson et al, 2011), suggesting that delineating areas

of similar soil-hydrologic characteristics would also benefit ecohydrological models. Land management and planning decisions concerning agricultural fertilizer applications, stormwater drainage allocation, building development layout and landscaping design may also be better informed with a concise representation of soil functional characteristics represented by enhanced map units.

Raster maps of topographic variables have been used for delineating map units representing areas of similar terrain characteristics (Park and vande Giesen, 2004) and areas with similar mass balance index (Moller et al, 2008). To date, raster maps of directly modeled soil retention characteristics have not been utilized with terrain maps to produce similar topographic-soil-hydrologic units. Furthermore, validation of map units' performance in protecting soil moisture content with a temporally extensive dataset of manually collected soil moisture data has not been commonly reported (Park and vande Giesen, 2004).

The basic building blocks of HFUs consist of topographic indices and soil properties that are well correlated with moisture. Catchment- or landscape-scale characterization of soil water retention or soil hydraulic conductivity parameters is also necessary to attribute soil-hydrologic functional information to delineated landscape units. The integration of these diverse datasets into similar units is a process that may consist of geo-statistics (Zhu et al, 2010b), multi-variate statistics with maximum likelihood optimization (Moller et al, 2008), or combination of these techniques. As there are advantages to both of these GIS-related methodologies, this study seeks a combined approach of geo-statistics with multi-variate statistics using maximum likelihood optimization algorithms.

Chapter 2: Materials and Methods

2.1 The Shale Hills Catchment

The Shale Hills field research laboratory spans an entire 7.9 hectare catchment in the Pennsylvania Ridge and Valley physiographic region. The parent material throughout the watershed is continuously Rose Hill Shale. Shale Hills lies within a udic moisture regime, and the catchment has been covered by a natural forest ecosystem common to the Pennsylvania Ridge and Valley physiographic region. The outlet in Shale Hills is located in the western end of catchment with elevation approximately 256 m above sea level. The highest point in the catchment is in the north-eastern corner with elevation at 310 m a.s.l. Deciduous trees are prevalent throughout most of the landscape, apart from the low-elevation area in the west of the catchment, where mainly coniferous tree species are concentrated.

Steeply contrasting topography, persistent year-round precipitation patterns, annual snowmelts, seasonal temperature variation and the constant activity of forest-dwelling biota throughout the catchment have contributed to the weathering and formation of five soil series from the 200 m thick shale bedrock (Lin et al, 2006). Previous research has found that topography is a major control on soil-hydrological processes at Shale Hills (Lin et al, 2006). Research conducted in Lin et al, 2006 derived landform unit classifications and relative wetness conditions for soil moisture sampling sites in Shale Hills. Following multi-variate analysis with soil properties and terrain variables, four distinct landform units were delineated: valley, concave hillslopes, planar and convex hillslopes, and summit.

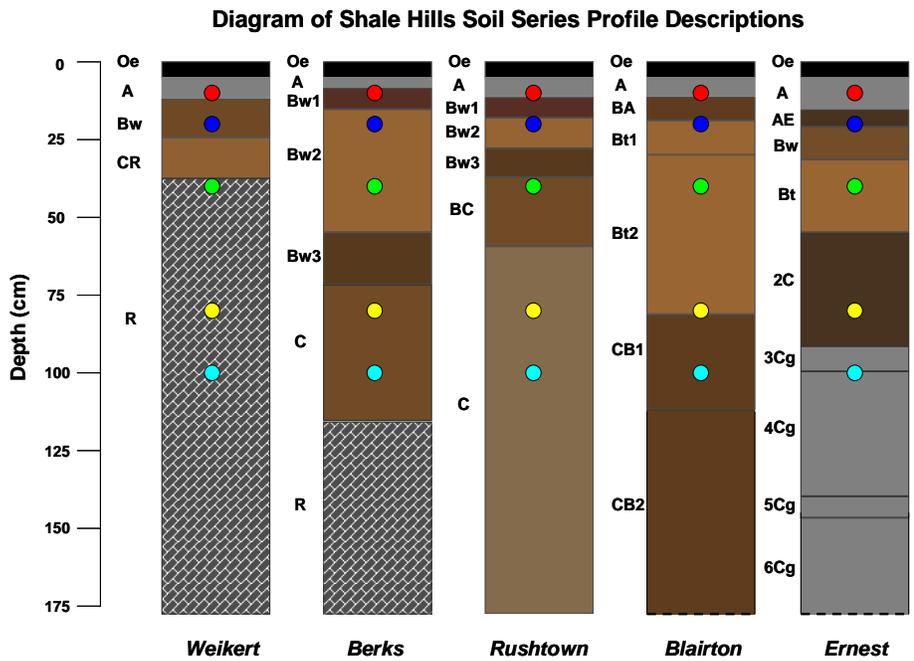
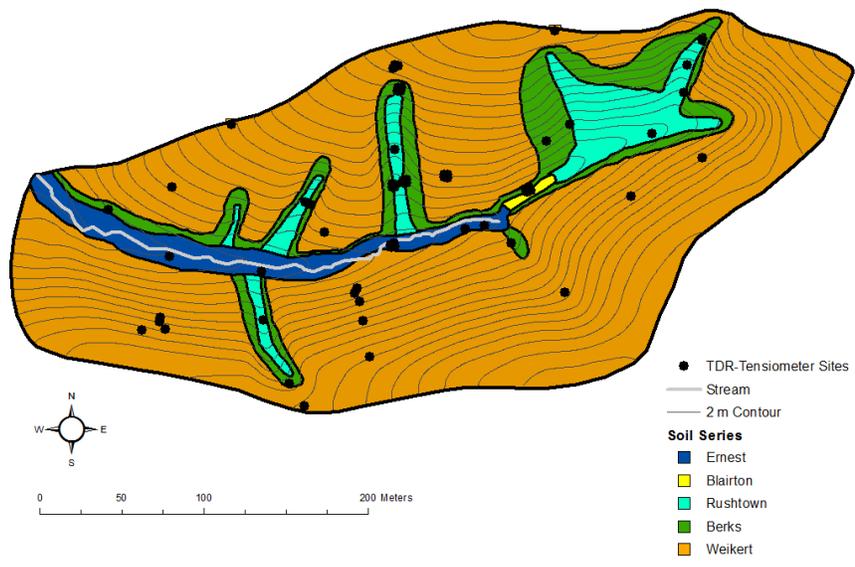


Figure 2.1: Schematic showing the thickness of the solum and soil horizons as observed in soil pits dug for each of the five soil series in the Shale Hills and the precision soil map for Shale Hills. The Rushtown, Blairton, and Ernest soil profiles are deeper than the scale indicated as the soil pits could not go further down due to safety issue and the limitation of the backhoe used. Colored points indicate the exact depths that tensiometers were installed: red = 10 cm; blue = 20 cm; green = 40 cm; yellow = 80 cm; and cyan = 100 cm depth. Horizons are color coded indicating approximate real soil color.

Figure 2.1 depicts general soil profile depths and horizon delineations for each soil series. An USDA-NRCS Order 2 soil survey conducted in Shale Hills concluded that

available soil series designations do not completely fit all soils in the catchment; rather, the soil series designations in Shale Hills are the closest approximations of currently available official soil series (Lin et al, 2006).

Ernest (Fine-loamy, mixed, superactive, mesic Aquic Fragiudults) and Blairton (Fine-loamy, mixed, active, mesic Aquic Hapludults) soils lie on the valley floor. They are the deepest soil series in the catchment with profiles that extend deeper than the bottoms of soil pits dug during the survey. Fragic properties are associated with Ernest soil's Btg horizon, which is a major structural differentiation from Blairton soils. Prismatic structure, high clay content, gleying and strong presence of redoximorphic features in the Btg horizon indicate that Ernest soil possesses a layer retarding vertical moisture percolation. A horizon with similar soil properties was not identified in the soil pit of the slightly higher-elevated Blairton soil.

Rushtown (Loamy-skeletal over fragmental, mixed, active, mesic Typic Dystrudepts) is found in the centers of seven swales that interrupt the surface of hillslopes around the catchment. It is moderately deep compared to the two other non-valley soils, but its B horizons are weakly developed. Berks (Loamy-skeletal, mixed, active, mesic Typic Dystrudepts) appear on the steep sides of swales in the catchment and appear as a transition between the convex and planar hillslopes to the concave swales. Weikert (Loamy-skeletal, mixed, active, mesic Lithic Dystrudepts) soil occupies approximately 70% of the area in Shale Hills. It dominates convex and planar hillslope and summit landform units.

2.2 Soil Moisture Retention Data Collection

Volumetric Soil Moisture Content (θ) has been collected in Shale Hills weekly or bi-weekly typically from mid-April to late November since 2004. A TRIME-FM (IMKO) Time Domain Reflectometry (TDR) device was used to collect soil moisture at depth intervals 0 to 0.20, 0.10 to 0.30, 0.30 to 0.50, 0.50 to 0.70, 0.70 to 0.90, and 0.90 to 1.1 m by inserting the probe into a PVC access tube buried at each site location across the catchment. These intervals correspond to depth points at 10, 20, 40, 60, 80, and 100 cm from the surface (further details in Lin et al, 2006). As the depths are recorded in cm, the unit for volumetric moisture content data for this research in Shale Hills is cm^3/cm^3 .

Soil matric potential (ψ_m) has been collected simultaneously with soil moisture content at 61 TDR sites consistently since 2005. Nests of tensiometers were installed about 0.15 m away from 61 TDR access tubes at 10, 20, 40, 80, and 100 cm depth points. As soil thickness varies among these sites from 21.4 to 196.7 cm, the number of tensiometers available for each site is dependent upon soil depth. The tensiometers are filled with de-aerated water to prevent erroneously high pressure measurements resulting from air bubbles. A tensimeter device (SMS, Arizona) was used to collect total potential (mbars) from the tensiometers during each data collection. The tensimeter device can record total potential values to a lower limit of approximately -850 mBars.

As water flows in the direction towards lower potential, water will move inside or outside a porous ceramic cup located at the bottom of the tensiometer until the difference of potential between water inside the tensiometer and water located in soil directly outside the porous cup is zero. Equilibrium is reached when the difference in potential is

zero (Young and Sisson, 2002). Matric potential is obtained from total potential measurements from the tensiometers in the field using the following equation:

$$\psi_m = \psi_t - \psi_g \quad (1)$$

where ψ_m is matric potential (mBars), ψ_t is total potential (mBars) reading from a tensiometer, and ψ_g is gravitational potential (cm) that is the distance between a reference elevation and the location of the tensiometer ceramic cup midpoint. If the reference elevation is set at the depth of the tensiometer cup, then the burden of the water column inside the tensiometer above the porous cup must be considered when calculating matric potential from manually collected tensiometer data. A centimeter of water in the tensiometer provides 1 mbar of burden on the ceramic cup; therefore the length of the water column inside of the tensiometer was recorded and added to field measurements to obtain matric potential values for a tensiometer during each collection date (Young and Sisson, 2002).

2.3 Soil Moisture Retention Modeling

The Gardner (Eq. 2), Campbell (Eq. 3) and van Genuchten (Eq. 5) moisture retention models were used to fit retention curves to volumetric moisture content vs. matric potential data from a total of 232 location-depth combinations across the Shale Hills. The Gardner (1970) model is expressed as

$$\psi = a\theta^{-b} \quad (2)$$

where ψ is matric potential, θ is volumetric moisture content and a and b are curve-fitting parameters .

The Campbell (1974) model is:

$$\psi = \psi_e \left(\frac{\theta}{\theta_s} \right)^{-b} \quad (3)$$

Where θ_s is volumetric moisture content at saturation, ψ_e is air-entry tension, and b is a curve-fitting parameter. The advantage to this model is the estimation of ψ_e , which theoretically represents the matric potential value where moisture in soil macropores have drained and air may infiltrate into the soil (Kosugi et al, 2002).

Equations (2) and (3) were curve-fit by first log-transforming both sides of the equation to obtain linearity and finding the slope (S) and intercept (I) of the resultant linear curve. The $-b$ variable of equation (2) was determined by S and the a variable in equation (2) was determined by $Exp^{(I)}$. The $-b$ variable of equation (3) is S and ψ_e is $Exp^{(I)}$. θ_s for equation (3) was estimated from a maximum likelihood algorithm and is the same θ_s value used for the van Genuchten equation. The slope ($-b$) for both Gardner and Campbell log-transformed linear curves were exactly the same, because equations (2) and (3) may be directly related via equation (4):

$$\psi_e = a * \theta_s^{-b} \quad (4)$$

The van Genuchten soil water retention model is expressed as (van Genuchten, 1984):

$$\theta = \theta_r + \frac{(\theta_s - \theta_r)}{[1 + (\alpha\psi)^n]^m} \quad (5)$$

$$m = 1 - \frac{1}{n} \quad (6)$$

where θ_r is residual moisture content, while α , n , and m are curve-fitting parameters. Directly relating m to n by (6) eliminates one curve-fitting parameter and theoretically allows the physical meanings attributed to α and n to be better defined (Kosugi et al, 2002). m was directly related to n using equation (6) for this study.

Residual moisture content has been defined as the moisture content at -15 Bar matric potential (van Genuchten, 1980), implying that the permanent wilting point of plants has been reached at this value. Since field data collection methods for this research allow lower limit of matric potential to be about -850 mBar, θ_r is treated as a curve-fitting parameter, as it has been in past research (Hodnett and Tomasella, 2002; Kosugi et al, 2002; van Genuchten, 1980).

The value of α (cm^{-1}) is approximately the inverse value of matric potential at the inflection point in the van Genuchten retention curve (Fredlund and Xing, 1994). The location of the inflection point on the van Genuchten curve indicates the relative influence of macropores on soil moisture flow. Macropore flow has been related to texture properties (Vereecken et al, 2010), but it has been shown that soil structure in field settings also contribute to macropore dominated flow (Hodnett and Tomasella, 2002; Lin et al, 1998; Zhou et al, 2008). Cracking between clay aggregates may induce macropore flow, for instance (Zhou et al, 2008).

The values of n influence the overall shape of a retention curve. Large n values create a steeper slope of the curve at its inflection point (Fredlund and Xing, 1994), possibly indicating a more effective release of moisture (Hodnett and Tomasella, 2002).

Low n values produce a gradual slope of the retention curve, and indicate that the release of moisture is more deliberate as ψ_m becomes more negative.

Fitting van Genuchten curves without parameter optimization has been accomplished in past research with knowledge of a soil's saturated hydraulic conductivity and water infiltration rate over time (Shao and Horten, 1998). Since saturated hydraulic conductivity and water infiltration data was not available for any depth locations, a maximum likelihood optimization procedure was applied with the van Genuchten model to estimate θ_s , θ_r , n , and α parameters.

2.4 Maximum Likelihood Estimation of the van Genuchten Parameters and Moisture Retention Model Diagnostics

The Maximum Likelihood Estimation (MLE) optimization of the van Genuchten parameters was performed using an algorithm coded in the R statistical computing software. A common method of optimizing van Genuchten parameters to fit $\theta(\psi_m)$ data is the Levenberg-Marquardt least-squares error algorithm (Marquardt, 1963). This algorithm has been implemented in software 'RETC' for optimization of van Genuchten model-fitting (van Genuchten, 1991). MLE optimization has considerable advantages over simple least squares estimation methods on manually collected θ and ψ_m field data (Hollenbeck and Jensen, 1998).

The MLE optimization is advantageous first because uncertainty or variance in the data is used during the optimization process to define parameter space; the data distribution of observed data informs parameter estimation. Another advantage of MLE is that parameters, given sufficiently wide bounds, will converge upon a global minimum

in an objective function. Successful completion of an MLE procedure also produces an invertible Hessian matrix, which is computed from the sensitivity of the model to the parameters. The proper inversion of the Hessian matrix is a diagnostic showing the estimated parameters have converged upon a finite space and are reliable estimates.

Convergence was achieved and the Hessian matrices were properly inverted during MLE optimization for van Genuchten model parameters at all 232 tesniometer depth locations in Shale Hills. All sites were modeled using the same starting conditions and parameter bounds. The L-BFGS method (Byrd et al, 1995) was used, because it allows box constraints. The α and n parameters had constraints imposed for effective global convergence, which were chosen to be wide, but not too wide to make convergence impossible. θ_s was constrained to be no less than the highest collected soil moisture value, and θ_r was constrained to be lower than the lowest collected soil moisture value.

2.5 Statistical Analysis of Moisture Retention Parameters

The Gardner, Campbell and van Genuchten parameters from all 232 depth locations were analyzed in a 3-way ANOVA test. The ANOVA was performed with a generalized linear model using soil series, landform units and depth and their interactions as categorical predictors and retention parameter as a response in order to discover factors that place a significant control on parameter variance. θ_r was not analyzed, as its physical meaning was nullified by the range of matric potential values being no greater than approximately (-) 850 mbars. Natural log-transformations were used on the a , b and

α values, and a \log_{10} - transformation was performed on n values to accommodate normality for the generalized linear model and subsequent ANOVA test.

The van Genuchten and Campbell ψ_e parameters at all depth locations were summarized according to soil series and landform units for separate depth intervals at 10, 20, 40, 80 and 100 cm. A Tukey HSD test was performed to find significant differences in means of each parameter between different soil series and landform units for each depth interval at $p < 0.05$.

Analyzing the correlation between texture, topographic and soil moisture retention variables is necessary for determining which topographic and soil properties place a significant control on soil water retention in Shale Hills. A correlation matrix was derived by performing a Spearman correlation analysis between hydraulic parameters, topographic variables and soil texture variables at all depth locations. The number of observations used for texture correlation differs from topography and moisture retention parameters, as texture information is incomplete at some depth locations.

2.6 Geostatistical Analysis of Moisture Retention Parameters

To assess the general spatial structure of hydraulic retention parameters across the catchment at the near-surface and deep solum, semi-variograms of hydraulic retention parameters were modeled for all depths. As there are a maximum of 61 site locations across the catchment for retention parameters and major topographic controls on the

retention parameters in Shale Hills, significant topographic covariates were used in constructing the spatial models (Zhu and Lin, 2009).

To estimate spatial model parameters, a maximum likelihood algorithm was used (Ribeiro and Diggle, 2001). Maximum likelihood is advantageous over weighted least squares in that it takes the uncertainty of observed data into account during the optimization procedure. Parameters optimized for spatial models include the partial sill, range, nugget, anisotropy ratio, anisotropy angle, and lambda. The nugget is a variable indicating the intercept of the semi-variance curve to the y-axis. The range parameter indicates the x-axis coordinate of the inflection point of the semi-variogram curve. A high range indicates a large variance of a parameter across space. The partial sill is the difference between the maximum semi-variance value on the curve and the nugget (Ribeiro and Diggle, 2001). For asymptotic theoretical semi-variogram models, such as the Matern covariance function, the partial sill is difference between the 95th percentile semi-variance value on the curve and the nugget. Lambda is a transformation parameter used on observation values to maintain normality, thus satisfying an important assumption of isotropy.

The basic set-up of the Gaussian random field spatial model used for spatial modeling of moisture retention parameters is (Ribeiro and Diggle, 2001):

$$Y(x, y) = \mu(x, y) + S(x, y) + e \quad (7)$$

where x, y are Euclidean coordinates, Y is the observed variable, $\mu(x)$ is the mean component or trend of the model, $S(x)$ is a stationary Gaussian process defined by a

spatial covariance model with parameters described above (sill, range, nugget, ...) and e is the residual term. If there were no significant topographic covariates to use as an external trend in a spatial model for a moisture retention parameter, Bayesian optimization with a constant trend was performed (Ribeiro and Diggle, 2001).

The Matern covariance function was used as $S(x)$ in equation (7). This function is a modified version of the exponential covariance function, and has been utilized to predict soil properties across landscapes (Minasny and McBratney, 2007). Cross-validation diagnostics were performed on each spatial model to test for independence, normality of predictions, and auto-correlation. Independence and normality of predictions were maintained for all retention parameter spatial models. Overt auto-correlation stemming from topographic covariates was not detected in any model, since there were no patterns present in model residuals.

After calculating optimized spatial models for retention parameters with maximum likelihood, a Bayesian spatial kriging method was utilized to interpolate parameter values across the catchment. This method uses the optimized spatial model parameters as prior information, while also iteratively testing a range of other spatial parameter values located within a probability distribution based around the prior information. The technique calculates samples with each parameter set and compares them to the probability of observed data values (Diggle and Ribeiro, 2002).

The parameter set with the highest frequency of predicted samples within the probability distribution of observed data are taken as the spatial parameters for kriging. The Matern covariance function with Bayesian optimized spatial parameters is used as $S(x)$ along with significant topographic covariates or site coordinates as $\mu(x)$ in the

Gaussian random field model (7) to predict values across the catchment All spatial modeling of moisture retention parameters was performed with the geoR package in R (Ribeiro and Diggle, 2001).

2.7 Topographic Wetness Index and Kriging of Total Moisture Storage in Profile

A Topographic Wetness Index (TWI) was calculated for Shale Hills using 1 meter resolution LiDAR derived elevation data. A D-inf algorithm (Tarboton, 1997) was used to delineate the upslope contributing area component of TWI (Böhner et al, 2006):

$$TWI = \ln \frac{(A_s)}{\tan(\beta)} \quad (8)$$

where A_s is “catchment upslope contributing area” and β is the slope value at a particular grid cell. Figure 2.2 displays a Topographic Wetness Index (TWI) map for Shale Hills with the 61 site locations used in this study displayed, symbolized by the soil series within which each site is located.

Topographic Wetness Index with TDR-Tensiometer Site Locations

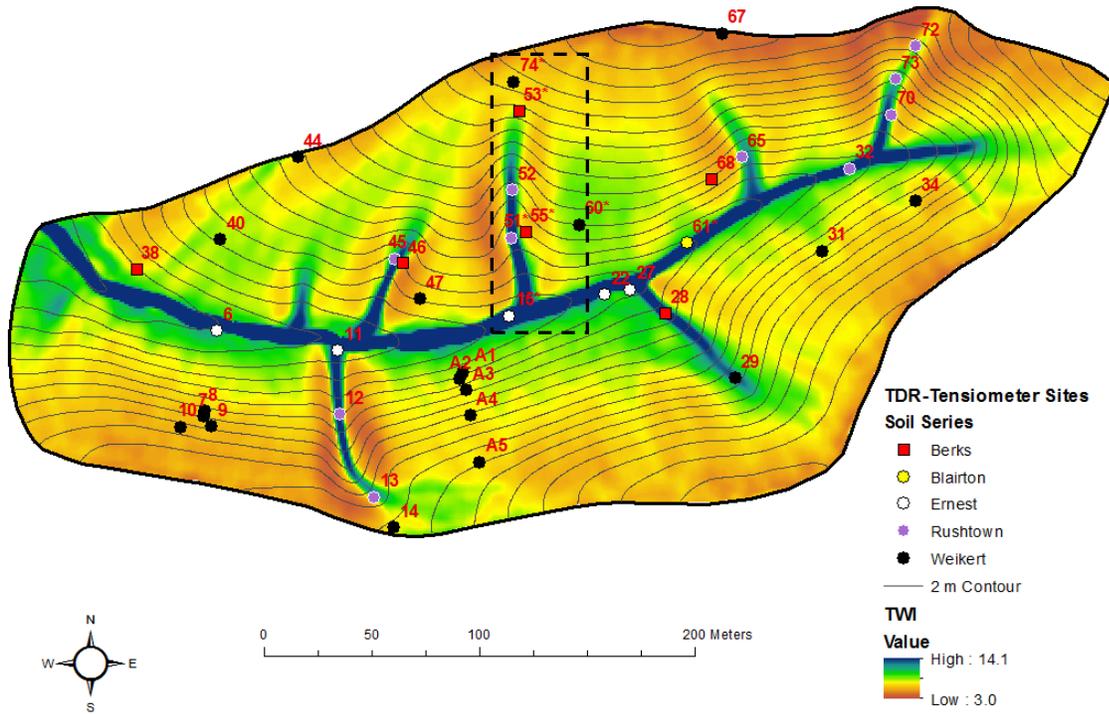


Figure 2.2: Topographic Wetness Index map for the Shale Hills, with TDR-tensiometer site locations colored according to designated soil series. Numbers indicate site locations in a toposequence within the catchment that correspond to sites featured in Figure 3.2. The dotted box indicates the approximate boundary of the focused toposequence reported.

Time-series soil moisture profile storage maps are also useful for identifying profile-scale soil moisture patterns over time. Profile soil moisture storage is calculated by the following equation (Zhu et al, 2010b):

$$S = \sum_i^n \theta_i * d \quad (9)$$

where S is profile soil moisture storage for a particular site (m), n is the number of depth points available at a site, θ_i is the volumetric moisture content at the i^{th} depth and d_i is the representative length of the i^{th} depth interval. The depth interval length (d) was 0.15 m

for 10 and 20 cm depths, and 0.20 m for 40, 60, 80 and 100 cm depths. S represents profile soil moisture storage within a 1.1 m soil profile for the deepest sites. Regression kriging with the package Gstat was used to interpolate total moisture storage for each date (Pabesma, 2004).

2.8 GIS Processing of Topographic Raster Maps and Depth to Bedrock Map

Topographic variables are cornerstones in the foundation of Hydropedological Functional Units and topographic raster datasets are necessary predictors during regression kriging of soil moisture at all depths and total moisture storage across Shale Hills. Topographic variables are correlated with soil moisture for all dates in the Shale Hills manual database (Takagi, 2009). A LiDAR flyover in February 2011 gathered elevation data to generate a high-resolution 0.5-meter DEM raster dataset for Shale Hills. The LiDAR data was preprocessed at UC Merced before processing occurred for this study. A Gaussian filter was applied with a 9 x 9 smoothing window to eliminate residual noise in the DEM. From the filtered LiDAR DEM, three topographic raster maps were created through GIS processing and were used to predict soil moisture and the delineation of HFUs during this study.

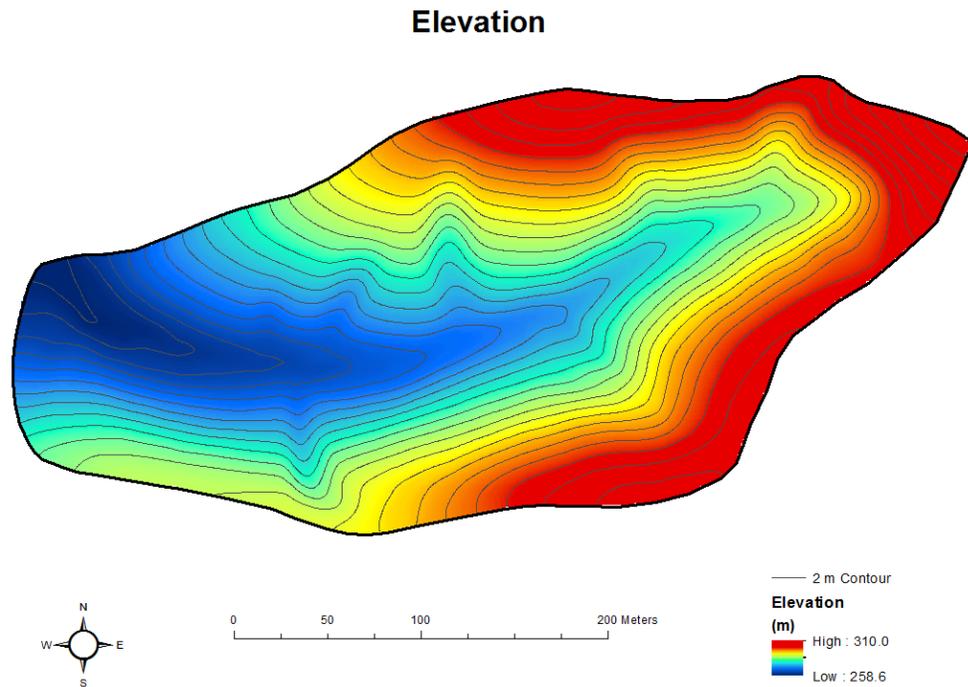


Figure 2.3: 0.5 meter resolution DEM for the Shale Hills generated from LiDAR flyover in February 2011. LiDAR data was preprocessed in UC Merced before subsequent processing in this study

Slope value is a ratio of a raster cell’s vertical distance “rise” over horizontal distance “run” based upon elevation data. Slope value displays the maximum change of elevation value for a cell based upon elevation values of neighboring cells. Higher slope values indicate a greater elevation change for a cell relative to a 3 x 3 cell window. As slope value is a general topographic indicator of the rate of soil moisture flux, it was found to be correlated with profile soil moisture storage distribution across Shale Hills in the vast majority of soil moisture collection dates for time-series profile soil moisture storage regression kriging. Previous research has also established slope as a major control on soil moisture patterns in Shale Hills (Lin et al, 2006; Takagi, 2009).

Surface Curvature is another topographic variable calculated from the LiDAR DEM raster data. Surface Curvature (C_s) is calculated from the following expression (Park and vande Giesen, 2004):

$$C_s = \left(\sum_{i=1}^n (z_i - z_n) / d_{in} \right) / n \quad (10)$$

where z_i is the elevation of the raster cell i , z_n is the average elevation of the surrounding area within a window of a given size, d is the horizontal distance between two cells, and n is the total number of surrounding cells.

A third topographic variable calculated from the LiDAR DEM is Upslope contributing area. Upslope Contributing Area (A_s) is calculated by (Park and vande Giesen, 2004):

$$A_s = (1/b) \sum_{i=1}^n \rho_i A_i \quad (11)$$

where A_i is the area of the grid cell i , n is the number of flow routes draining into A_i , ρ_i is a weight related to runoff mechanisms and b is the contour width as function of the raster resolution. A D-inf flow algorithm was used in this calculation, which bases the angle of each flow route into each cell upon the steepest downwards slope on 8 triangular facets centered within neighboring cells rather than 45° from the center of higher elevated neighboring cells (Tarboton, 1997). Applying a log-transformation to A_s map grids is generally useful, as the distribution of A_s values is heavily skewed in topographically

contrasting landscapes and $\log(A_s)$ has been shown to be a good predictor of soil moisture values across landscapes (Park and vande Giesen, 2004).

Depth to bedrock is a soil property variable placing a significant control on soil moisture patterns and soil water retention characteristics at Shale Hills (Lin et al, 2006). 233 depth to bedrock data points were gathered during an auguring campaign. These data points were used to derive a depth to bedrock map by performing universal kriging with the 233 augur measurements. However, further analysis with depth to bedrock data points collected during shallow well installations and GPR data has yielded a refined depth to bedrock map for Shale Hills.

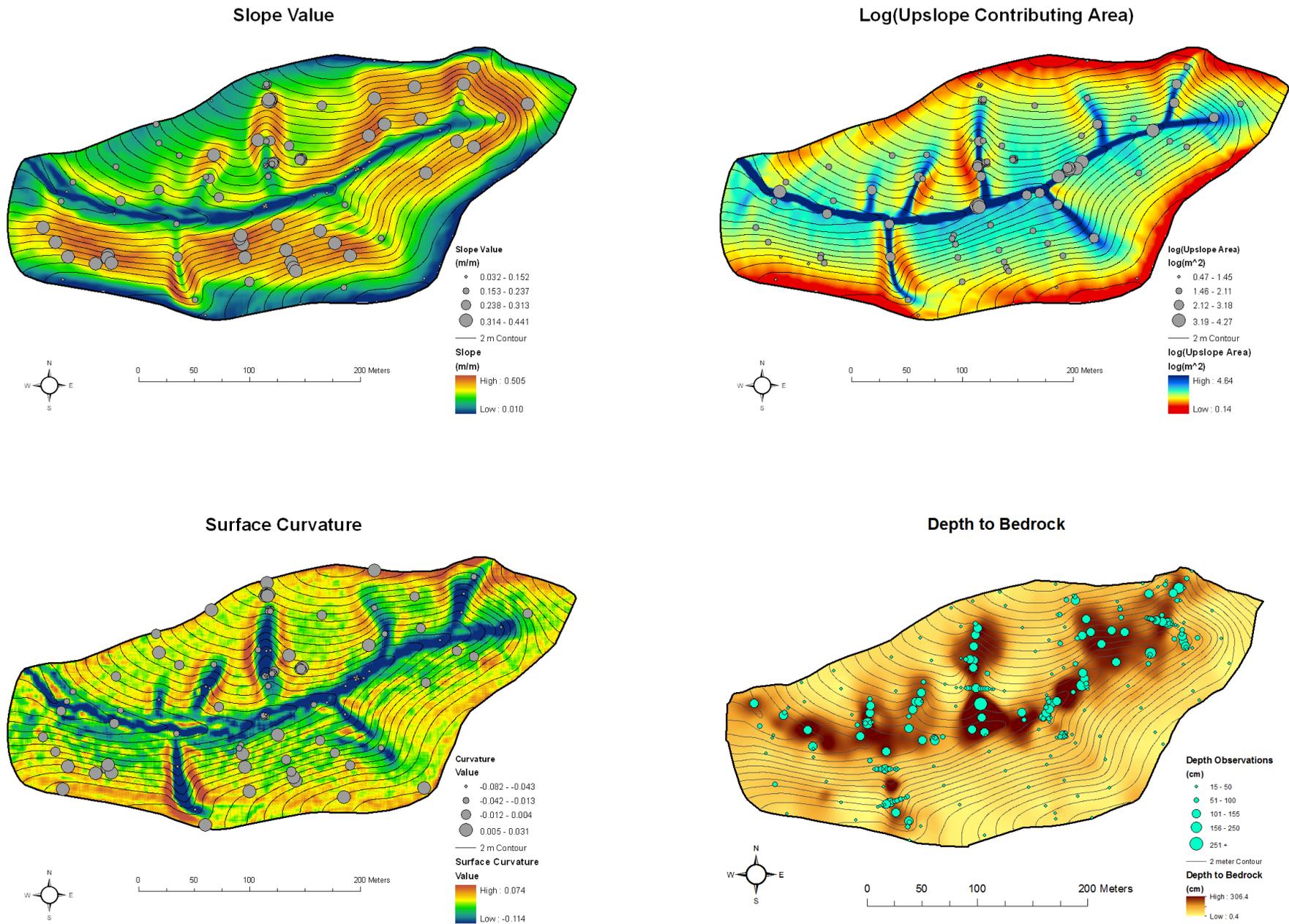


Figure 2.4: Maps of slope, log(upslope contributing area), surface curvature and the initial depth to bedrock for the Shale Hills. TDR site locations are plotted with point size on slope value, upslope contributing area and surface curvature indicating the magnitude of the map raster value at a site. Depth to bedrock shows the 223 augur measurement locations plotted with point size showing relative magnitude of soil depth

2.9 Refinement of Catchment-scale Depth to Bedrock Map

A fundamental dataset used in deriving HFUs and predicting catchment-wide soil moisture distribution is depth to bedrock. Catchment-wide depth to bedrock data is notoriously hard to obtain and map for catchments and landscapes the size of Shale Hills and greater. Advances in geophysical technology have shown great potential in being able to obtain depth to bedrock data non-invasively in the field.

Ground Penetrating Radar (GPR) has emerged as a geophysical tool used to study the subsurface. It is dependable enough to perform evaluations on the accuracy of seismic geophysical technology (Baker et al, 2001, Herbert, 2005). GPR studies have been carried out in Shale Hills across several transects in the catchment. GPR wave amplitude measurements are processed to acquire depth measurements to an interface that significantly alters the GPR wave's amplitude. An interface where significant changes in a wave's amplitude occur is the border between materials of drastically different densities (Daniels, 2004). A subsurface interface such as this may be considered the soil-bedrock interface if a noticeable trend exists in the data taken along the transect line.

The typical frequency of the radar used in Shale Hills is 400 mHz. This frequency is low enough to penetrate at a depth sufficient for recording the depth to bedrock in the catchment in several transects. A GPR instrument with 200 mHz was also available since soil profile depths in the catchment were discovered to be too deep for the 400mHz frequency. The lower frequency is useful for recording information about greater depths, but the lower resolution potentially makes interpolation about the soil-

bedrock interface more difficult, especially if the data are collected in sub-optimum environmental conditions (Daniels, 2004).

Processed GPR data surveys taken in Shale Hills with a GPR unit using a frequency of 200 MHz showed clear trends among wave amplitude changes along several transects, which in turn indicated the corresponding soil-bedrock interfaces. GPR data along two transects at the east end of the catchment were available for processing, and a total of 65 depth-to-bedrock data points at 1 meter spacing were calculated from these data. The depth to bedrock measurements from the GPR survey data were included with previously collected depth measurements from auguring and shallow well installments in Shale Hills to generate a refined depth to bedrock map for Shale Hills by using regression kriging with 385 total depth measurements. Surface curvature and TWI were regressors used in predicting spatial autocorrelation of depth to bedrock, and the cross-validation of the spatial model exhibited an R^2 of 0.87.

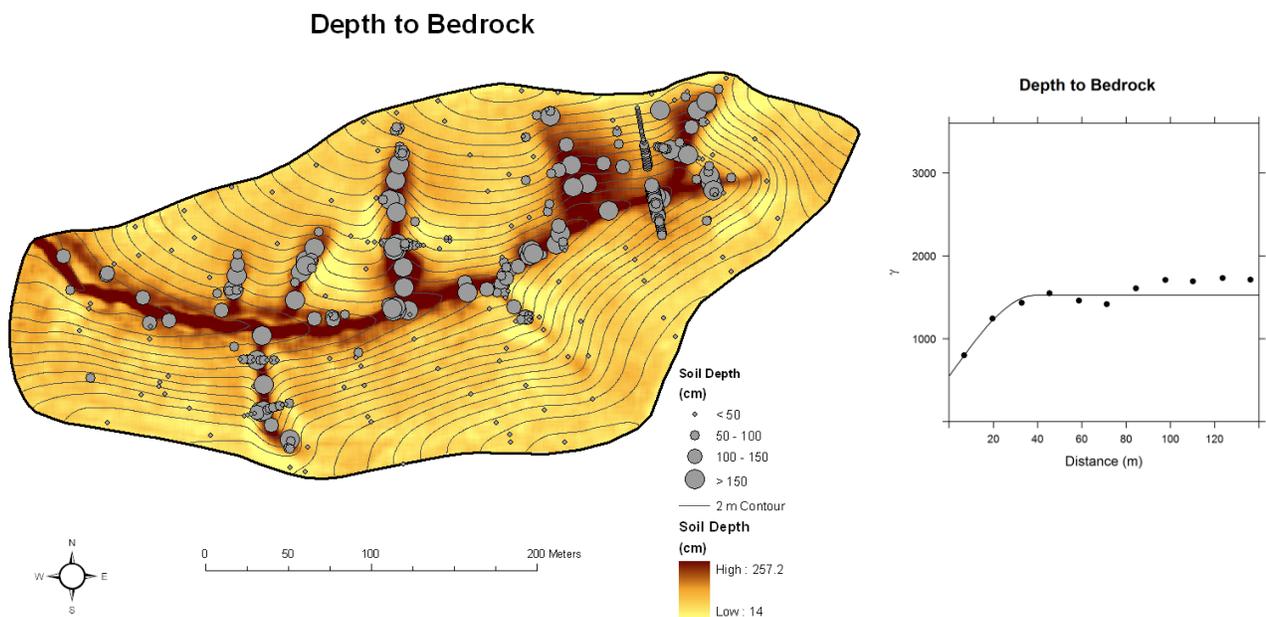


Figure 2.5: Refined depth to bedrock map with all the available data in the Shale Hills (385 total). Included is the semi-variogram depicting spatial variance of depth to bedrock data fitted with a Spherical covariance function.

Catchment-scale depth to bedrock maps may be potentially generated for other catchments without the need for extensive auguring campaigns by using GPR data acquired from spatially wide-ranging surveys, while assuring sound measurements with proper validation techniques. Such catchment-wide GPR surveys should be conducted during periods of the year with the best environmental conditions for GPR data collection at a given catchment (NRCS). Auguring may be required to validate and calibrate GPR depth measurements. Depth to bedrock data from other GPR surveys in Shale Hills may be utilized to further refine the depth to bedrock map, if needed.

2.10 Delineated Landform Units in the Shale Hills Using Park and vande Giesen (2004) Method

Park and vande Giesen (2004) developed a method for quantitative delineation of landform units in a landscape. A procedure for delineating landform units (LFUs) using a raster DEM dataset was developed and used in the Tarrawarra Catchment, Australia (Park and van de Giesen, 2004). The method first requires a scatterplot of log(Upslope Contributing Area) [A_s] and Surface Curvature [C_s] raster data for an area to be plotted against one another. From this scatterplot, the user may use both reasoning and familiarity with a study area to manually establish threshold values in the C_s and $\log(A_s)$ data to delineate landforms such as “shoulder”, “interfluvium”, “backslope”, “toeslope”, etc., from the scatterplot. The amount of landform units that may be delineated in an area depends upon the nature of the scatterplot.

Figure 2.5 shows the scatterplot of C_s versus $\log(A_s)$ raster values for Shale Hills. Five landform delineations were found reasonable for this data:

- i) Summit = $C_s > -0.05 \ \& \ \log(A_s) < 0.94$
- ii) Convex Hillslope = $C_s > 0.01 \ \& \ \log(A_s) > 0.94 \ \& \ \log(A_s) < 3.25$
- iii) Planar Hillslope = $C_s > -0.01 \ \& \ C_s < 0.01 \ \& \ \log(A_s) > 0.94 \ \& \ \log(A_s) < 3.25$
- iv) Concave Hillslope = $C_s < -0.01 \ \& \ \log(A_s) > 0.94 \ \& \ \log(A_s) < 3.25$
- v) Valley = $C_s < 0.01 \ \& \ \log(A_s) > 3.25$

These delineation thresholds are demarcated on Figure 2.6 with lines color-coded for each landform. Colors corresponding to each landform are indicated in parentheses next to each landform unit name above.

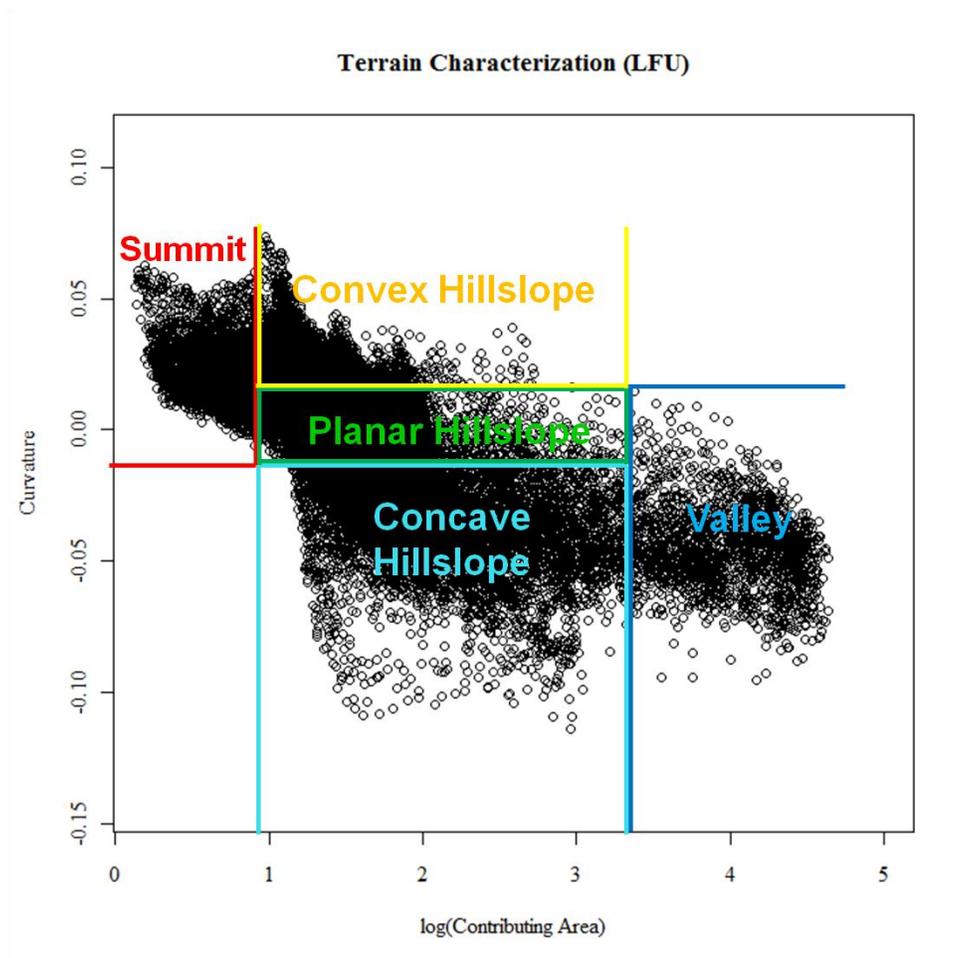


Figure 2.6: Scatterplot of $\log(A_s)$ and C_s raster data for the Shale Hills. Colored lines demarcate thresholds set for the delineation of 5 landform units: spell out these landform unit names here. See Fig. 2.7 for actual spatial distribution of these 5 units for the Shale Hills.

After raster cells have been assigned a landform unit designation, a raster map of delineated landform units may be plotted. The raster dataset itself may be used as a factor predictor variable in models. Figure 2.7 shows LFUs for Shale Hills delineated with Park and vande Giesen's 2004 method. The procedure caused 8 swales to be derived from the C_s versus $\log(A_s)$ data. The summit does not wrap completely around the catchment, as raster cells on edges of the northwestern slope and south-central slope hold $\log(A_s)$ values greater than the $0.94 \log(\text{m}^2)$ threshold. This is possible, because the D-inf algorithm allocates less upslope contributing area to cell neighboring steep slope gradients. The Slope Value map in Figure 2.4 reveals a more gradual change in Slope Values downslope from catchment edges without a delineated Summit LFU.

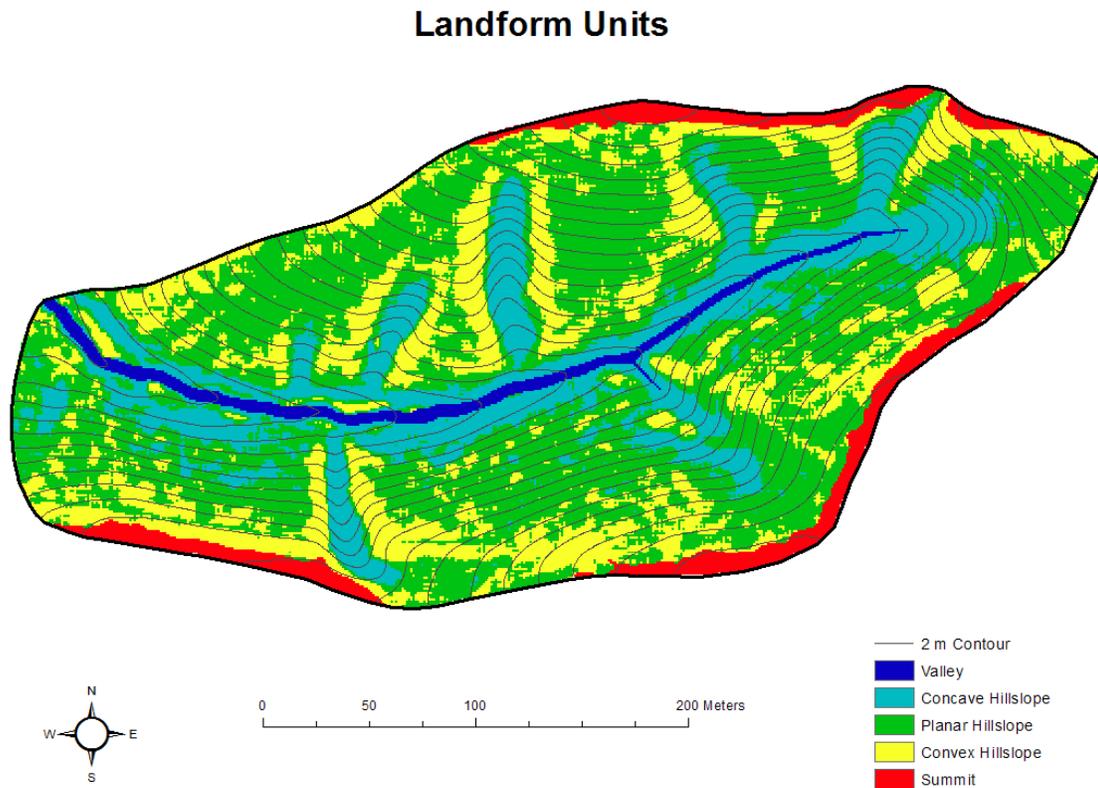


Figure 2.7: Map of landform units (LFUs) in the Shale Hills delineated with the method of Park and vande Giesen (2004)

2.11 Analysis of Topographic and Hydrologic Datasets for Informing Hydropedological Functional Unit Concept and Generation

Proper delineation of map units representing zones of similar soil-hydrologic function must be constituted by a set of coherent procedural steps in order to be both reproducible for other landscapes and catchments and successful in yielding physically meaningful geographic entities. HFUs are a convergence of pedologic, hydrologic and topographic information. They represent unique consociations assembled from variables that control soil-hydrologic processes. Although approaches using process-based topographic properties (Park and vande Giesen, 2004) and the mass-balance index (Moller et al, 2008) have been developed to characterize similar soil-landform units, aggregating soil hydrologic function parameters, topographic data and soil properties, such as depth to bedrock, into similar functional units so far has not been reported.

Both scaling and synthesizing diverse datasets meaningfully are two obstacles confronting the delineation of HFUs. Topographic properties and soil moisture patterns have been found to be well related at Shale Hills (Lin et al, 2006 and Lin, 2006), yet the integration of many soil moisture data maps across an extensive time-period into a single index that may be directly related to or integrated with topographic maps has not been accomplished. However, soil-water retention parameters obtained through direct modeling have been calculated at the point-depth scale and aggregated to the site-profile scale. Profile-scale soil water retention parameters were mapped using geostatistics (details in Section 2.6), consequently allowing the representation of soil water retention characteristics at the catchment- or terrain-scale (Figure 2.8).

Solum Saturated Moisture Storage

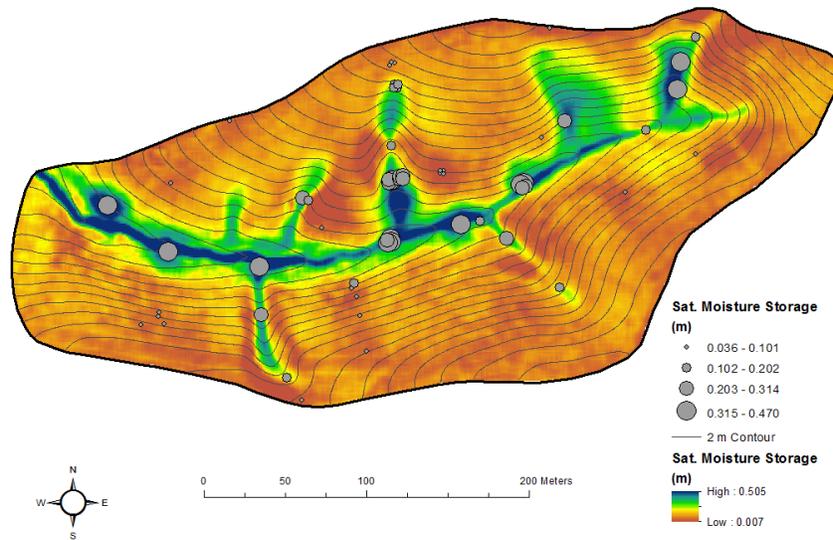


Figure 2.8: Map of aggregated θ_s (solum storage at saturation). Regression kriging was used to interpolate θ_s with depth to bedrock being the significant regressor.

Preceding the aggregation of topographic attributes, soil hydrologic parameters and depth to bedrock, a thorough analysis of available raster datasets for Shale Hills was performed to delineate potential units within each individual dataset. Total profile moisture storage was calculated and interpolated across the catchment for dates of data collection from May 2008 to November 2010. These maps were analyzed in a time-series sequence to discern similar spatial patterns of profile soil moisture storage over time, and to distinguish general soil wetness conditions that occur each year in the catchment. Three general soil wetness conditions were found to recur each year: wet (December to May), intermediate (May to end of June and end of September to November) and dry (July to end of September).

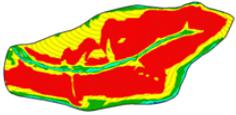
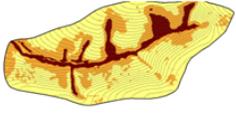
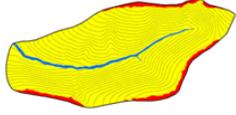
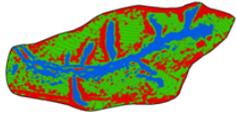
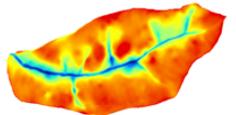
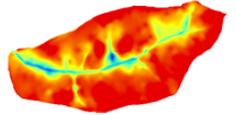
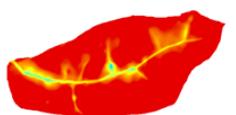
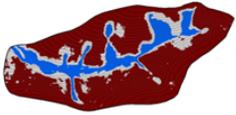
Table 2.1 exhibits the results of possible unit delineations for raster datasets in Shale Hills. Units were delineated differently for each raster dataset; however these delineations each followed a general scientific or quantitative reasoning. Rasters without a clear connection to moisture patterns or that frustrated an understandable methodology

for delineating units were not analyzed. Terrain rasters that had a published method in discerning separate units included slope, surface curvature and upslope contributing area. Separate ranges of depth to bedrock values have a clear connection to the delineation of official soil series in Shale Hills. Solum θ_s storage and profile soil moisture storage maps were analyzed in a straight-forward statistical method, as these rasters have a direct connection to depth to bedrock and terrain via regression kriging. Table 2.1 shows that a maximum of five clusters are possible by observing topographic, depth to bedrock, and θ_s storage maps.

Surface curvature and $\log(\text{upslope contributing area})$ both exhibited three units defined by threshold values instituted from the Park and vande Giesen-directed analysis of the scatterplot in Figure 2.6. This analysis carves surface curvature into concave, planar and convex terrain surface units. The $\log(\text{upslope area})$ map shows small summit and valley units which are distinctly separated from the hillslope. The convex and planar hillslope and concave hillslope cells have moderate upslope areal influence compared to the little to no upslope influence at the summit and the abruptly high upslope influence found on the valley floor.

Slope was separated according to USDA-NRCS soil slope class designations. The catchment is relatively steep in general, as very little of the catchment fits within slope classes A and B. Slope values near the extent of class C and lower show a distinct valley unit which envelopes the extent of Ernest and Blairton soil series. Slope class C is also evident around the summit of the catchment. The hillslope shows a division between a steep slope class F and less steep class D.

Table 2.1: Identified units representing the general patterns of important landscape variables and selected soil moisture storage maps representing three different wetness conditions (wet, intermediate, dry, and saturation).

Variable	Number of Units	Classes	Data Range		
Slope	5	Class A	0 -- 0.03 m/m		
		Class B	0.03 -- 0.08 m/m		
		Class C	0.08 -- 0.15 m/m		
		Class D	0.15 -- 0.25 m/m		
		Class F	> 0.25 m/m		
Depth to Bedrock	3	Shallow	0 -- 50 cm		
		Moderate	50 -- 100 cm		
		Deep	> 100 cm		
Log(A_s)	3	No or Little Upslope Area	0 -- 0.94 log(m ²)		
		Some Upslope Area	0.94 -- 3.25 log(m ²)		
		High Upslope Area	> 3.25 log(m ²)		
Curvature	2	Concave	< -0.01		
		Planar	-0.01 -- 0.01		
		Convex	> 0.01		
Profile Storage Wet Date: 12/6/2008	5	Low Storage Summit			
		Low Storage Hillslope			
		Medium Storage Hillslope			
		High Storage Hillslope			
		High Storage Valley			
Profile Storage Transition Date: 6/17/2010	4	Low Storage Hillslope			
		Medium Storage Hillslope			
		High Storage Swale			
		High Storage Valley			
Profile Storage Dry Date: 7/08/2010	3	Low Storage Hillslope			
		High Storage Hillslope			
		High Storage Valley			
Solum θ_s Storage	3	Low θ _s Storage	0.006 -- 0.124 m		
		Medium θ _s Storage	0.124 -- 0.186 m		
		High θ _s Storage	0.186 -- 0.247 m		

As with slope, curvature and log(upslope contributing area), the three units amidst the depth to bedrock raster were delineated according to empirical knowledge. The high resolution soil survey for the Shale Hills separated the Berks soil series from the Rushtown series by a depth to bedrock value of 100 cm, and depth to bedrock measurements in the Rushtown, Ernest and Blairton soil series have a minimum limit of

100 cm. The Berks and Weikert are separated according to depth to bedrock by a value of 50 cm. These specific soil attributes prompt the delineation of three depth units of relatively shallow, intermediate and deep soil in the Shale Hills using break values of 50 and 100 cm.

Curvature and depth to bedrock show three distinct units that have arisen in part from dynamic terrain-hydrologic processes. The soil residing on convex hillslope areas are actively weathered by fluid movement along a topographic gradient. This fluid movement in turn shifts materials into concave hillslope and valley positions. The continuous weathering allows convex hillslope soil to be deeper than much of the planar hillslope areas, but the flux of material from fluid movement inevitably feeds into the concave hillslope and valley locations, making these areas the deepest within the catchment. The depth to bedrock map neatly shows these three depth distinctions.

The solum-scale θ_s raster is unique as it was generated from θ_s values directly modeled from data collected *in situ* and significant topographic variables during regression kriging. It carries information regarding maximum profile soil water retention capacity, which is controlled by both topography and soil structure. The θ_s storage data was correlated with topographic wetness index and depth to bedrock. Additional information unique to these two spatial regressors is embodied within this raster, as it shows a uniquely wide pattern of high values in the valley close to the catchment outlet, which is not present in any topographic or depth to bedrock map.

Three units were derived from this map as it has close association to depth to bedrock. The data distribution of solum θ_s cell values is skewed towards zero. The mean raster value served as one break point, and the θ_s value at the distance of the

standard deviation of the raster dataset (0.06) away from zero served as another break point in demarcating three units. This method was used because of 1) the statistical straight-forwardness and 2) profile θ_s values between the mean and maximum values represent transition areas between typically low moisture holding hillslope and high moisture holding concave hillslope and valley soil. A portion of the footslope area close to the catchment outlet fits within the high moisture capacity unit. Site 38, which lies on the footslope adjacent to the catchment outlet, contains relatively high θ_s values within its solum, thereby contributing to the broadness of the wet θ_s unit at the catchment outlet.

The range of observed profile soil moisture storage values contained in maps for dates spanning 2008 to 2010 was 0.008 to ~ 0.50 m. From the sequence of total moisture maps, a maximum five soil profile moisture storage units were expressed according to general patterns. The wet conditions showed the maximum units, and the dry conditions showed the least amount of units. Based upon the total moisture storage range of values, a custom-made color spectrum was created and set as standard for each total moisture storage map. This color scheme was designed by using graduated red colors for total soil moisture values at ~ 0 to 0.10, orange for 0.10 to 0.20, yellow for 0.20 to 0.30, cyan for 0.30 to 0.40 and blue for 0.40 to 0.50. Prominent areas that exhibited a similar band of colors were in turn considered a general moisture unit for a date.

The total moisture storage map sequence consistently exhibited the same apparent valley unit expressed in the Slope map amidst low-elevation, < C class slope values. However, in a substantial number of dates during all moisture conditions, a partitioning in moisture storage patterns appeared within the valley unit, dividing it roughly at the origin of the ephemeral stream. This division was more prevalent during dry and

intermediate conditions. Figure 2.9 shows three dates from intermediate wetness conditions that clearly show a divide in valley moisture patterns at the elevation where the ephemeral stream appears above ground.

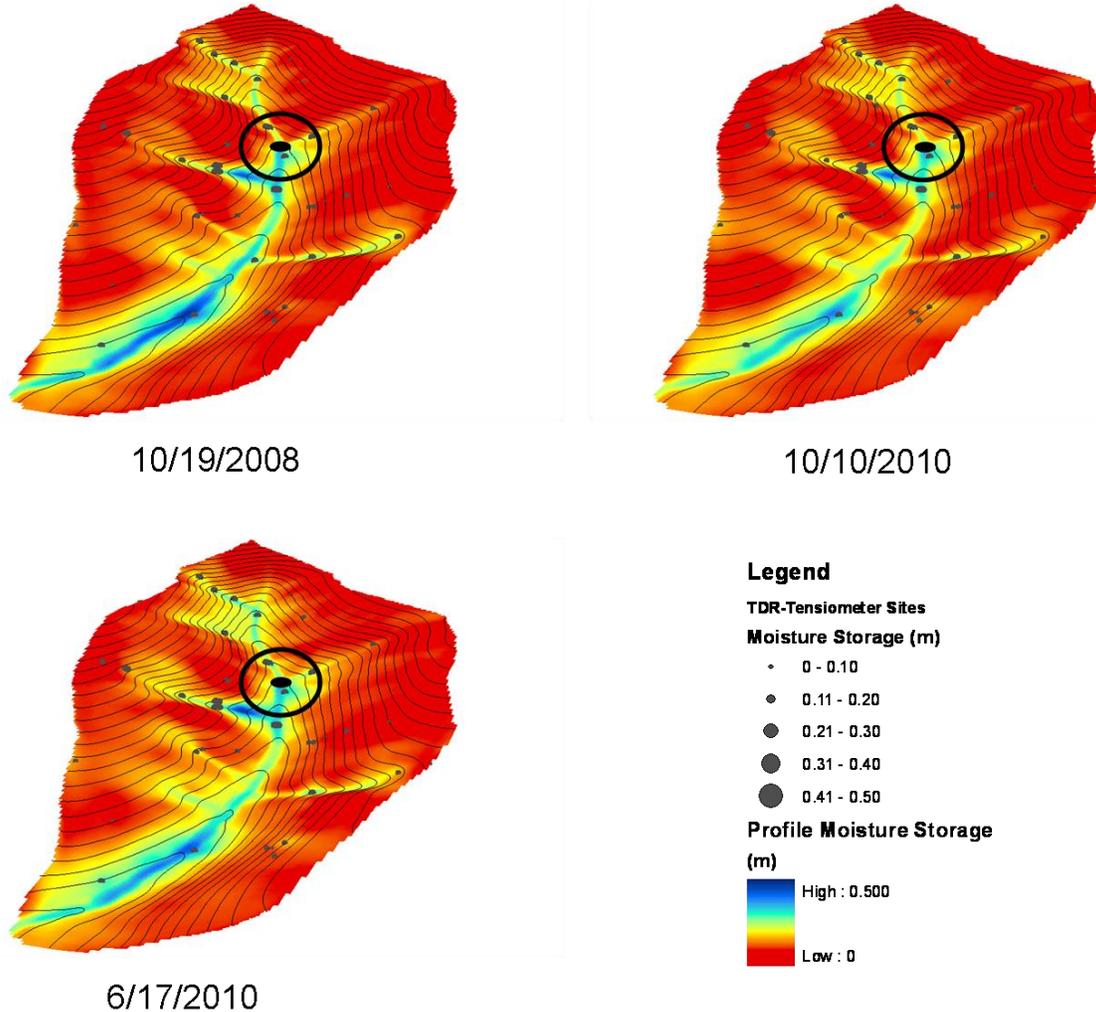


Figure 2.9: Maps of total moisture storage in solum during intermediate wetness conditions. The black oval in the center of the black circle indicates the origin of the ephemeral stream. Notice darker blue colors occur in areas lower in elevation than stream origin, while only light blue colors appear at elevations higher than stream origin.

The profile soil storage maps of intermediate wetness conditions are not conclusive by themselves, but catchment-wide Electro-Magnetic Induction (EMI) data from multiple surveys also clearly distinguishes higher EMI patterns in the valley area

holding the ephemeral stream compared to the higher elevated valley (Figure 2.10). This indicates a difference in not only general soil moisture patterns, but also pedologic properties between the valley soil encompassing the ephemeral and the valley soil at a higher elevation than the stream.

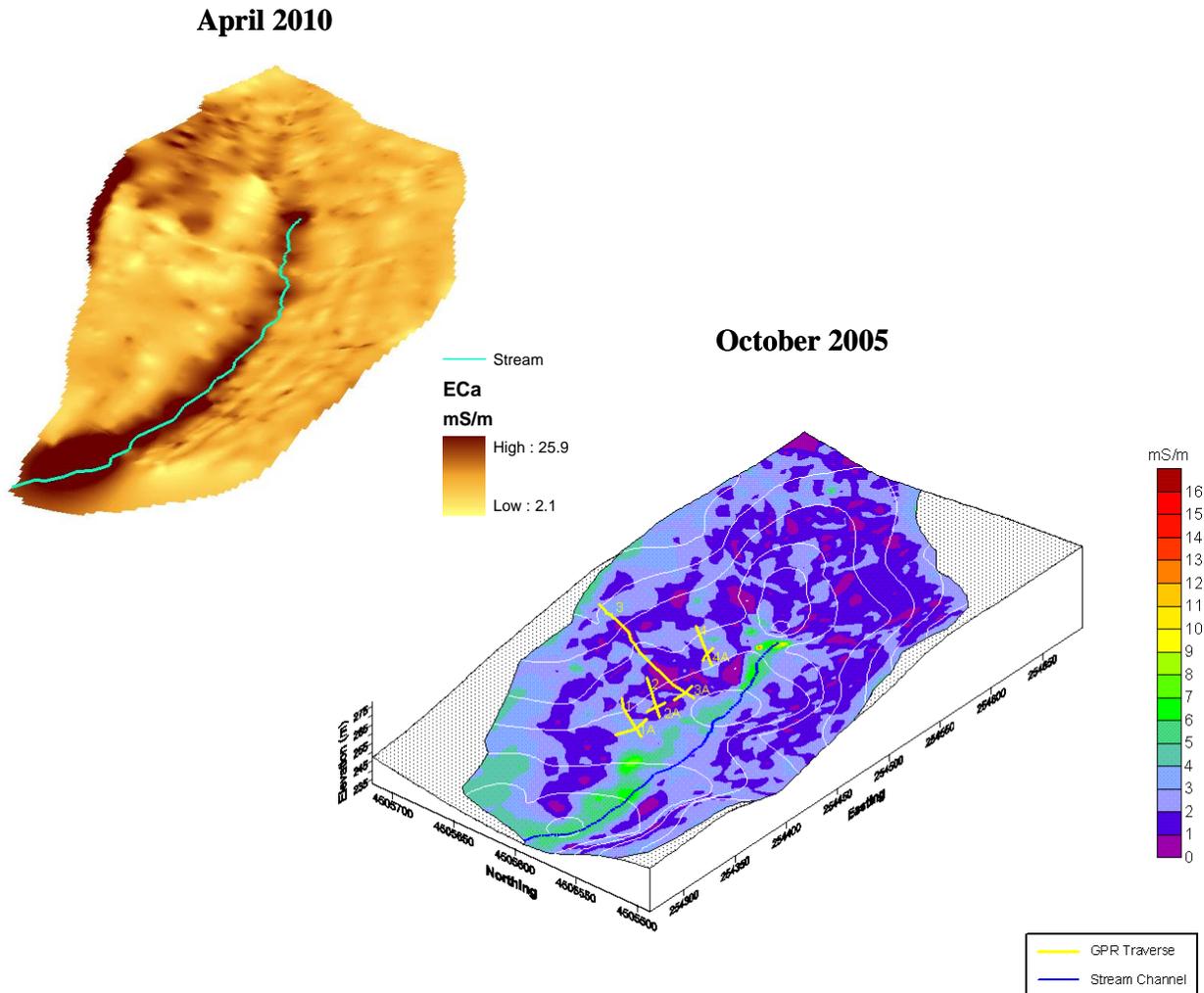


Figure 2.10: Catchment-wide Electromagnetic Induction (EMI) surveys performed in the Shale Hills for the wet (April) and intermediate (October) moisture conditions. High EMI values indicate a higher concentration of electrically conductive material, such as soil water or saline material, within the soil. Higher EMI values appear over soil directly adjacent to the stream and abruptly decrease beyond the stream for both surveys. Maps were created using ordinary kriging interpolation.

2.12 Quantitative Analysis of Raster Datasets to Delineate HFUs

All raster maps observed in the preliminary analysis held values at the same spatial scale, so the remaining obstacle lay in synthesizing raster datasets with physically disparate data into a single map. The preliminary analysis of available topographic, soil property and soil hydrologic data provided guidance in deciding how many units should be delineated, where these generally exist in the landscape, and what rasters should be used during the delineation procedure. Research of catchment-scale soil water retention characteristics for Shale Hills utilized a Principal Component Analysis (PCA) to transform topographic properties, soil depth and soil retention characteristics into a reduced set of uncorrelated variables. Each of the new principal components describes a portion of variance that exists within the dataset of the original variables. The first PC holds the maximum amount of information, and each subsequent PC holds a decreased share of variance proffered by the manifest variables.

To maximize efficiency of a PCA, a group of variables should be selected that do not share an excessive amount of redundant information in the dataset and avoid multicollinearity (Mardia et al, 1979). A PCA may expose variables that explain the maximum amount of information within a set of variables by observing the magnitude of scores the variables individually place on each PC. The manifest variable with the highest score on the first PC controls much of the variance held amongst the collection of manifest variables. Besides locating important drivers of variance in a dataset, a PCA is useful for obtaining coordinates for each point of observation relative to each PC's axis in rotated space. In the case of raster datasets, the observation points are grid cells holding different variable values. The quantity of coordinates, or rotated values, associated for

each grid cell is equal to the number of PCs defined. This method allows disparate variables to be related to each other through their relationship to uncorrelated variables. Prior to PCA, all input datasets were scaled with the root mean square error of their observations. To judge the extent of correlation between raster datasets representing data at the landscape-scale, a Spearman non-parametric correlation test was performed on all rasters representing topographic, soil property and hydrologic data in the Shale Hills geospatial database.

The rotated subspace defined after a PCA of a group of variables is indistinguishable to the centroid cluster subspace delineated during K-means cluster analysis (Ding and He, 2004). This allows for k- or c-means clustering of coordinates assigned to each grid cell in Shale Hills from a PCA, essentially serving as a means to quantitatively characterize similar patches of raster cells based upon different sets of topographic, soil and hydrologic variables. A fuzzy c-means clustering was used for this purpose, as it allows for a non-discrete, proportion-based assignment of grid cells to similar groupings calculated through a maximum likelihood algorithm. The fuzzy c-means clustering algorithm is based off the following equation (Pal et al, 1996):

$$\sum_i \sum_j \mu_{ij}^m * d_{ij} \quad (12)$$

where u_{ij} is the membership of observation i in cluster j , and d_{ij} is the distance or dissimilarity between observation i and center j . The dissimilarities used are the sums of squares. m is the degree of “fuzziness”, in this case set to 2, or lowest fuzziness. The algorithm ceases when it is unable to reduce the value of the objective function by

$T_{rel} * (|V| * T_{rel})$, where T_{rel} is the relative convergence tolerance, and V is the value of the objective function.

The output includes $n_{cluster}$ membership values for each variable location. These are the proportion (0 to 1) of a location's membership for each of the clusters. A location's cluster membership is based upon the maximal membership value for a cluster. Since soil and topography is heterogeneous and continuous, membership of a soil to a topographic-depth group should also not be discrete. Clustering involving rotated values from PCA at a grid cell using fuzzy logic accomplishes 1) objective grouping of sites into similar clusters based upon their relationship to topography, soil properties and moisture retention in linear orthogonal space, 2) provides non-distinct membership of sites into a designated cluster, and 3) avoids multi-collinearity by using rotated data values from PCA (Ding and He, 2004).

Since an outcome of a fuzzy c-means cluster analysis is that each raster cell contains proportional values depicting the cell's probable association to each delineated cluster. These proportional values may be used in a more in-depth neighborhood analysis for possible up- or down-scaling of delineated clusters into "parent" or "children" clusters, respectively (Moller et al, 2008). Such an analysis was not performed for this study because of time constraint.

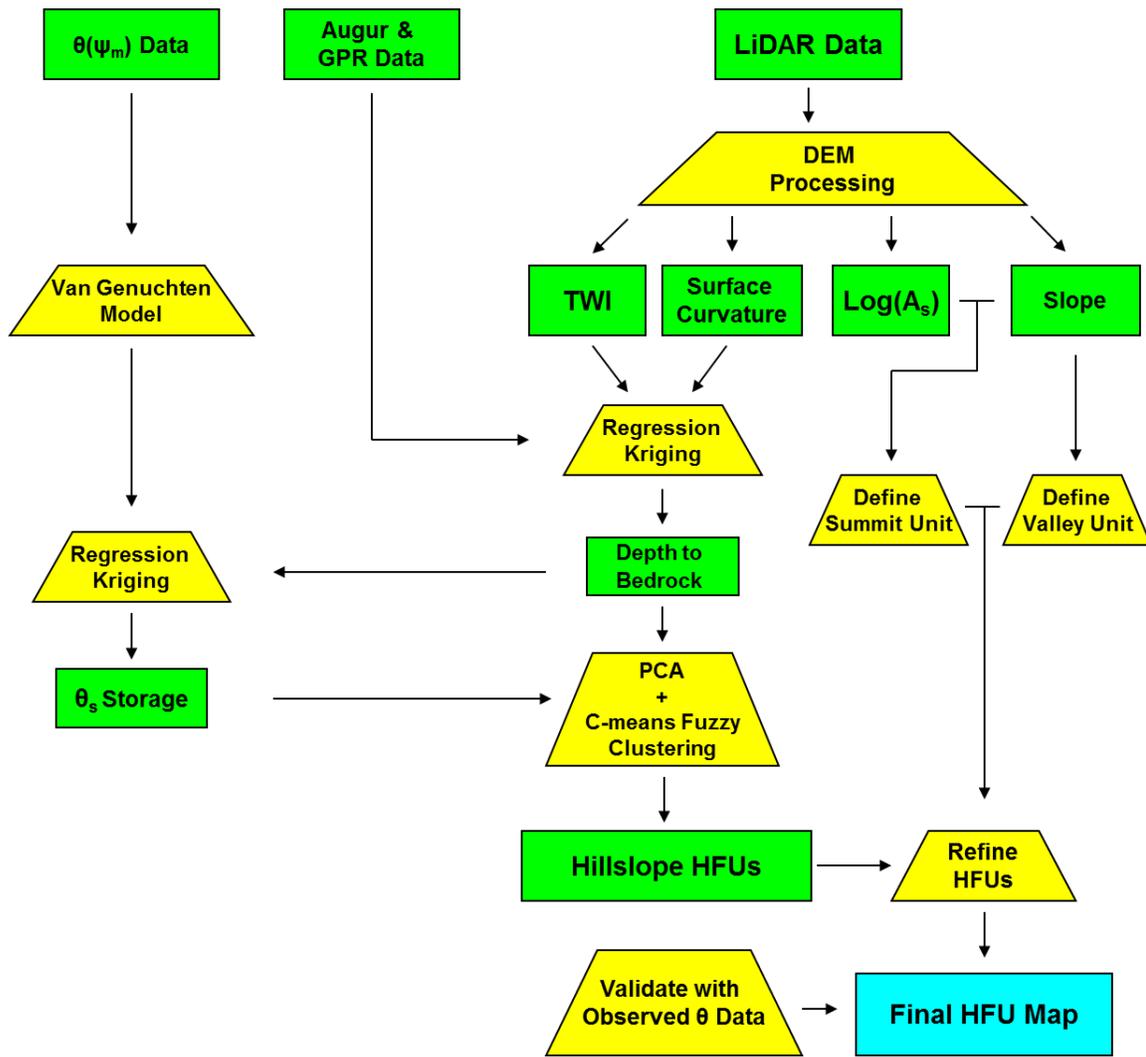


Figure 2.11: Flow diagram depicting the sequence of processes (yellow trapezoids) with inputs and subsequent outputs (green boxes). The final HFU product is in light blue color

Figure 2.11 outlines the flow of specific processes involved in the delineation of HFUs. By observing the total moisture storage sequence, it was determined that 5 HFUs may be meaningfully delineated from Shale Hill's raster data. The depth to bedrock, surface curvature and solum θ_s storage maps contained three units that represented similar terrain-hydrologic processes. Through the Park and vande Giesen analysis, the $\text{log}(A_s)$ map showed a distinct summit discharge zone at the edge of the catchment. The slope map had five separate classes, but three combined slope units at a relatively low

elevation derived a valley unit more akin to the apparent valley unit found in the total moisture map sequence. From observing EMI surveys and profile moisture storage moisture maps, it is clear that the portion of the valley encompassing the ephemeral stream is a distinct unit from the higher elevated valley, and that the higher elevated valley generally functions more similarly to concave hillslope landscape positions.

The depth to bedrock and solum θ_s storage raster datasets were chosen to be inputs for the PCA according to the following justifications:

i) *Depth to bedrock* offers three units depicting distinct formed from subsurface hydrologic and soil weathering processes. Depth to bedrock incorporates general patterns from surface curvature and topographic wetness index from regression kriging, so both moisture flux direction and the general topographic gradient is incorporated into depth to bedrock. Depth to bedrock distinguishes shallow versus intermediate depth hillslopes accurately.

ii) *θ_s Storage* grants three general units that represent varying levels of estimated maximum moisture retention in the soil profile across the catchment. General patterns from TWI and depth to bedrock are incorporated in θ_s , but the valley and footslopes landscape positions adjacent to the catchment outlet exhibit High moisture holding capacity, resulting partly from alluvial processes and fluid movement towards the catchment outlet, which is not displayed in depth to bedrock.

Slope value was not included in the initial process, as its flat valley unit is not distinctive enough with other maps to be delineated through clustering. Moller et al (2008) faced a similar issue when attempting to delineate a Floodplains unit from terrain data using clustering, and opted to define their Floodplain unit with a separate process.

Total moisture storage data was used as validation data and therefore could not be quantitatively utilized in delineating HFUs. The minimum or limiting number of clusters with the four aforementioned maps is 3, and this was the amount of clusters designated to the fuzzy c-means clustering algorithm following a PCA of the four maps.

After 3 initial hillslope HFUs were delineated, the slope was then used to define a valley from the Deep Soil, High Storage, Concave Hillslope cluster. Slope is occasionally used during soil surveys to draw rough boundaries of soil series in the landscape, so applying the slope map as a quantitative refinement tool is somewhat consistent with field methods. The Flat Valley unit was delineated only to the elevation where the ephemeral stream originates. Following the delineation of the Valley unit slope, the Summit unit was defined by applying the threshold $\log(A_s)$ value for Summit obtained from the Park and vande Giesen analysis on the Hillslope cluster.

2.13 Validation of HFUs

Although physical commensurable soil-landscape-hydrologic functional units may be delineated using raster datasets, the capability of these units to correlate to profile moisture storage patterns was tested using both a linear model . The linear model set up consisted of:

$$\theta \text{ Storage}_i \sim \text{HFU}$$

where $\theta \text{ Storage}_i$ is solum soil moisture storage data collected at a TDR site for a given date i as a response, with the predictor being the categorical Hydropedological Functional

Units (HFU) as factors. Data from soil moisture collected during 54 dates spanning December 2006 – November 2010 were used for validation. Volumetric moisture content at different depths for dates across the same span of time was also used as response variables predicted by HFUs. The linear model provided the coefficient of determination (R^2) as a diagnostic to judge the performance of HFUs in predicting general moisture patterns for each date. The linear model provided the standard errors and p-value corresponding from an ANOVA of HFUs to moisture data for each date.

Multiple responses to one predictor variable are permitted in linear models, which prompted the use of the following model set up:

$$\theta \text{ Storage}_i + \theta \text{ Storage}_{i-1} + \dots + \theta \text{ Storage}_{i-n} \sim \text{HFU}$$

where $\theta \text{ Storage}_i$ is profile soil moisture storage at a date i , with the following responses being total moisture storage from subsequent dates of data collection. Dates from January 2007 to November 2010 were used as multiple responses in this set up, however dates were selectively placed in the model if data for at least 60 sites across the catchment were non-missing. A missing data point would disqualify an entire site's data across all dates in this model set up. 31 dates were used as multiple responses in a linear model to measure HFU performance in predicting moisture patterns across a 3 year time period for at least 60 sites, which span the breadth of the catchment and represent all HFUs. An ANOVA of HFUs to moisture data from all 31 dates used in the linear model was also performed.

To assess overall moisture storage variability within soil and landform categorizations in the catchment, a coefficient of variation (CV) was calculated for each HFU, soil series and landform unit. The CV is determined by:

$$\frac{StdDev(Observations)}{Mean(Observations)} \quad (13)$$

where StdDev is the standard deviation, mean is the arithmetic mean and the Observations are all moisture measurements taken from TDR locations within each HFU from available dates from December 2006 to November 2010 (102 dates total).

Tukey HSD multiple comparison tests were performed on the average solum moisture storage values collected from the same 102 dates used in the linear model across 60 sites with HFUs, soil series and landform units as categories for three separate tests. The data distribution of solum moisture storage values corresponding to each soil, landform and HFU was normal. This was performed to detect significant difference of means of average solum storage values across a 3 year time period between individual HFUs, soil series and landform units with a p-value < 0.05. A Tukey HSD multiple comparison test was also performed on total profile texture storage (see Appendix C for details) collected from 58 different locations with HFUs as categories. Significant differences were detected with a p-value < 0.05.

The linear modeling validation procedure applied to the delineated HFUs was also performed on soil series and LFUs delineated with Park and vande Giesen's 2004 method. Diagnostics from the linear model were compared to assess which landscape

categorization best predicted general patterns of total storage and volumetric moisture content from different depths.

Chapter 3: Results and Discussion

3.1 Landscape Factors Controlling Soil Moisture Retention

The coefficient of determination (R^2) was used as a model performance diagnostic for all soil water retention models (Figure 3.1). As equations' (1) and (2) coefficients could be directly related to each other (Equation 3), their R^2 values were the same. The van Genuchten model performed better than the Gardner and Campbell models. An extra parameter in the model and maximum likelihood estimation optimization offered greater flexibility and accuracy in predicting moisture than the power functions. Gardner and Campbell models performed relatively better for locations in the hilltop and deep valley depth locations.

A time-series map sequence of total profile moisture storage (Figure 3.2) shows a general pattern for spring, summer and autumn 2010. Areas with low elevation and deep depth to bedrock retain significant amounts of moisture during the dry period. Semi-variogram sills were highest in the wet months of November and May, but lowest in the dry months of September and July, which indicates that the spatial variance of total moisture storage is higher in wet periods. This may related to valley soil retaining higher amounts of moisture during wet periods than dry periods.

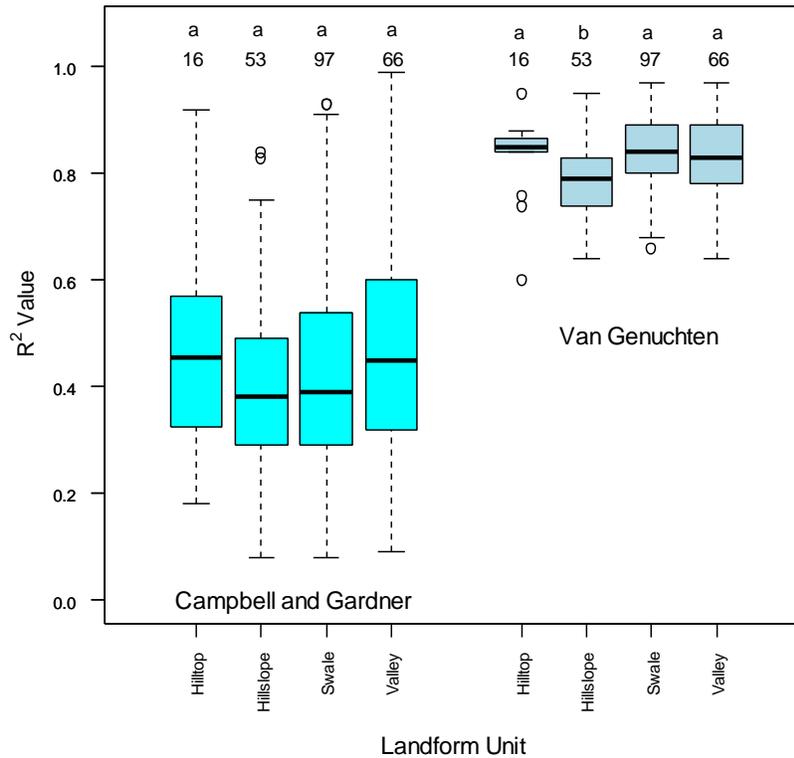


Figure 3.1: Box plots of coefficient of determination (R^2 values) for the three soil moisture retention models for 232 depth-locations at 61 sites across the catchment. R -squared values of the three retention models are grouped according to landform unit where depth locations reside. The small letters at the top of each bar indicate significant differences and the numbers indicate sample size

The dashed rectangle in Figure 2.2 delineates a catena in Shale Hills that has been featured in past hydrogeologic research in Shale Hills (Lin and Zhou, 2008). The toposequence of soils here is hilltop Weikert, hillslope Weikert, swale Berks, swale Rushtown and valley Ernest. Sommer and Schlichting (1997) discussed different archetypes of catenas based on the overall hydrologic or im(mobilization) processes occurring in a catena. This catena fits neatly in the *translocation* type of catena (Sommer and Schlichting, 1997), as there are sub-areas with leaching processes and sub-areas with accumulation processes that coexist within the same catena. The catena concept may be visualized when qualitatively analyzing the soil hydrologic retention function at multiple depths from sites located in diverse combinations of soil and landforms along the sequence.

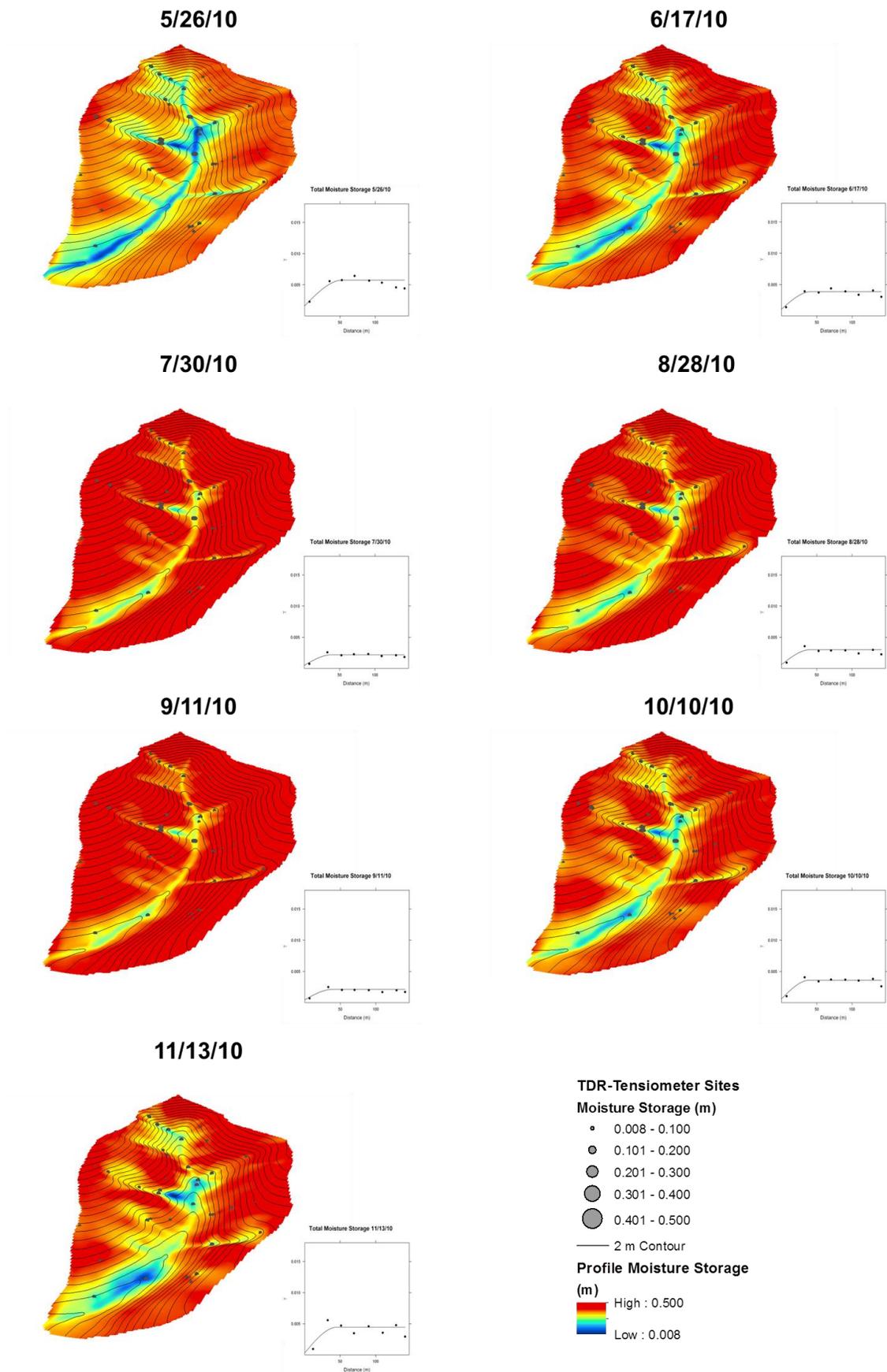


Figure 3.2: Series of total storage moisture maps from May to November in 2010. Moisture data was collected at 61 sites (indicated by red dots) across the catchment. Moisture storage values were interpolated across the catchment using regression kriging. Histograms show distribution of observed total storage values at all sites for each date. γ values for semi-variograms have been binned, and Spherical models were used to fit semi-variograms.

Figure 3.3 demonstrates van Genuchten water retention curves for available depth locations at six sites within the catena delineated in Figure 2.2. Each site corresponds to a different landform and soil combination.

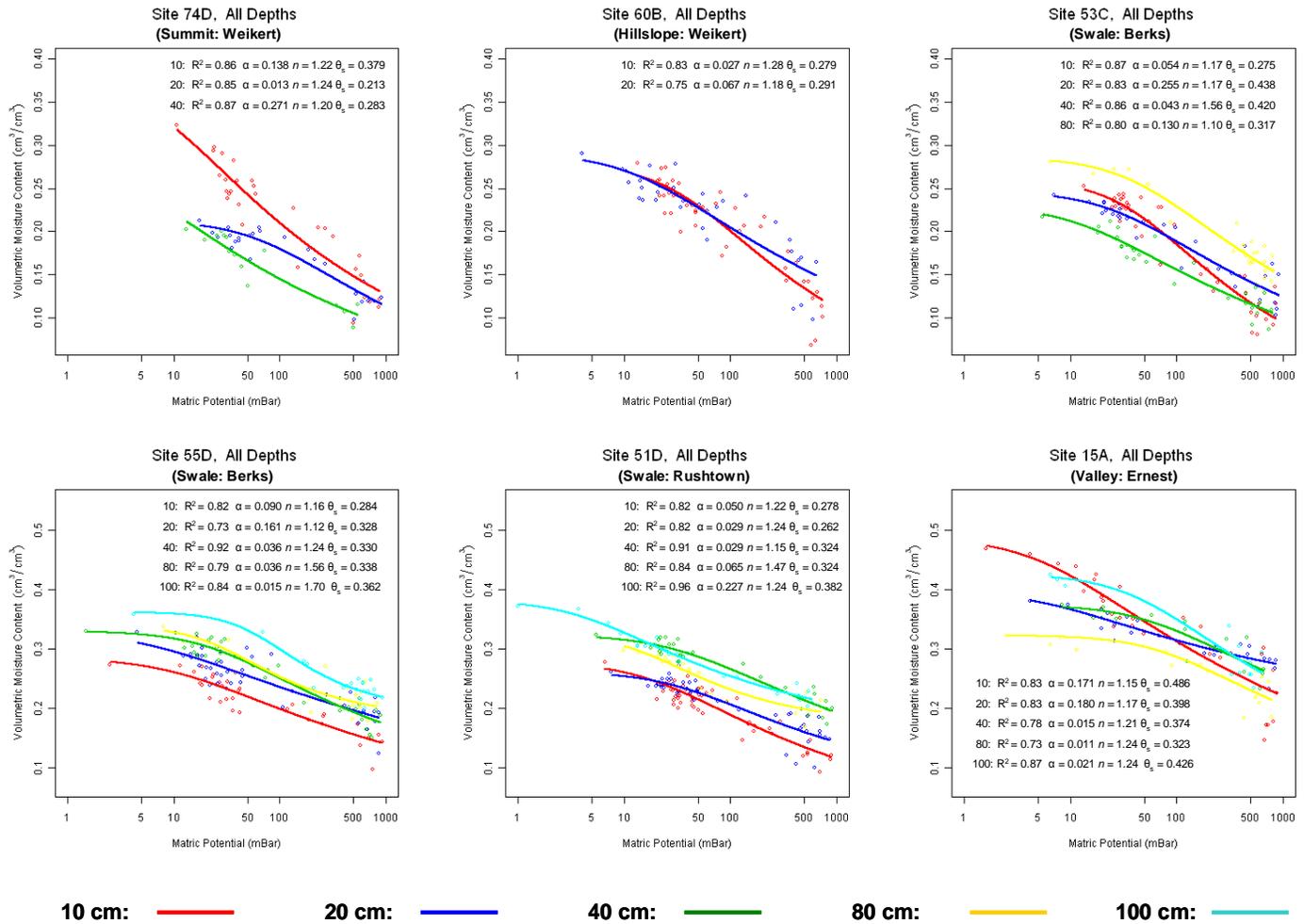


Figure 3.3: Soil Moisture Retention Curves using the van Genuchten retention model for sites along a topequence in Shale Hills (see Figure 2). Soil moisture retention curves were fitted for all available depths at each site. Parameters and R-square values for each function are listed in the plot. The curves and observed data points are color-coded according to depth.

The curves for Weikert soil at the hilltop are similar at 10 and 40 cm depths in that they both have a high air-entry value and an effective release of moisture as matric potential becomes more negative. The topographic location of the hilltop soil exhibits a strong control on the shape of these curves, since moisture at the highest topographic

position may more easily flow down-gradient laterally. The curves for Weikert soil at 60B overlap, as the data distributions themselves are similar; this indicates that soil profiles on this planar hillslope position are relatively homogeneous with depth.

The swale soils' curves show a marked contrast with the hilltop and hillslope curves in that the deepest curve has "flipped" positions with the 10 cm curve by being distinctly higher on the y-axis. Site 53B is a medium depth site on a steep-sloped, side-swale location. Site 55D is a deeper site located on a side swale, further downslope from 53B. The slope is less steep at its location and the upslope contributing area greater, which may explain the more gradual release of moisture at 10 cm.

The 100 cm curve for the center-swale Rushtown soil is more gradual, likely from a significant flux of vertical and lateral moisture. Gravimetric moisture influence is high in this location, indicating the possible presence of large cracks in the C horizon. All Ernest soil valley curves indicate a gradual release of moisture. This is physically expected, as the deep soil in the valley is a focal point of lateral moisture fluxes from upslope areas on both North and South hillslopes.

The differences between retention curves for locations in different soil series and landform units at two different depths suggest that soil series, landform units, and depth factors control variance of retention parameters across the Shale Hills catchment. Table 3.1 reveals 3-way ANOVA results of each moisture retention parameter compared to depth, soil series and landform unit factors. In all parameters at least one of the soil series, landform unit and depth factors or associated interactions significantly controlled variance for each parameter.

Table 3.1: Three-Way ANOVA of each soil water retention parameter against soil series, landform unit and depth factors. Significant categorical controls on the variance of parameter values were revealed through a generalized linear model. Some parameters were log-transformed before analysis to satisfy the normality requirement for the test.

	ln(α)		log(n)		θ_s	
	Mean Sq Error	Pr - value	Mean Sq Error	Pr - value	Mean Sq Error	Pr - value
Depth	0.285	0.625	0.0000143	0.927	0.057306	<0.001 ***
Soil Series	0.223	0.945	0.0042813	0.040 *	0.205408	<0.001 ***
Landform	3.914	0.021 *	0.0003228	0.902	0.040478	<0.001 ***
Depth*Soil Series	1.806	0.197	0.0027081	0.173	0.018857	0.001 **
Depth*Landform	1.327	0.342	0.0024987	0.219	0.00854	0.104
Soil Series*Landform	2.988	0.114	0.0013615	0.369	0.105304	<0.001 ***
Depth*Soil Series*Landform	0.914	0.381	0.0001259	0.785	0.006848	0.198
Residuals	1.186		0.0016809		0.004109	

	ln(α)		Gardner & Campbell ln(b)		ψ_e	
	Mean Sq Error	Pr - value	Mean Sq Error	Pr - value	Mean Sq Error	Pr - value
Depth	601.05	<0.001 ***	16.4064	<0.001 ***	423.69	0.077 .
Soil Series	84.79	<0.001 ***	0.3455	0.031 *	106.57	0.530
Landform	38.27	0.016 *	0.0929	0.535	295.92	0.088 .
Depth*Soil Series	58.38	<0.001 ***	0.2522	0.098 .	75.76	0.689
Depth*Landform	53.58	0.002 **	0.2955	0.076 .	112.73	0.473
Soil Series*Landform	0.04	0.950	0.051	0.527	1001.75	0.007 **
Depth*Soil Series*Landform	32.06	0.087 .	0.5324	0.042 *	91.04	0.411
Residuals	10.87		0.1271		134.14	

Significance Codes: '****' 0.0001 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Value ranges are similar for all the van Genuchten parameters among different depths (Figure 3.4). θ_s has significant differences between soil series and landform units at each depth. The low-laying Ernest valley soils have significantly higher saturation values than other soil series for each horizon. This is physically consistent with both the potentially high moisture accumulation in the catchment valley and the higher matrix pore volume in valley soils resulting from relatively greater clay content.

The α parameter was significantly higher at 40 cm depth in hilltop landform locations and significantly lower in valley locations. As hilltop sites are among the shallowest in the catchment, unconsolidated shale bedrock or rocky C horizons may contribute to macropores resulting from cracks and spaces between coarse material. A relatively impermeable layer appears in some valley sites at 40 cm depth, which would restrict macropore flow, as this layer is dense and prismatic.

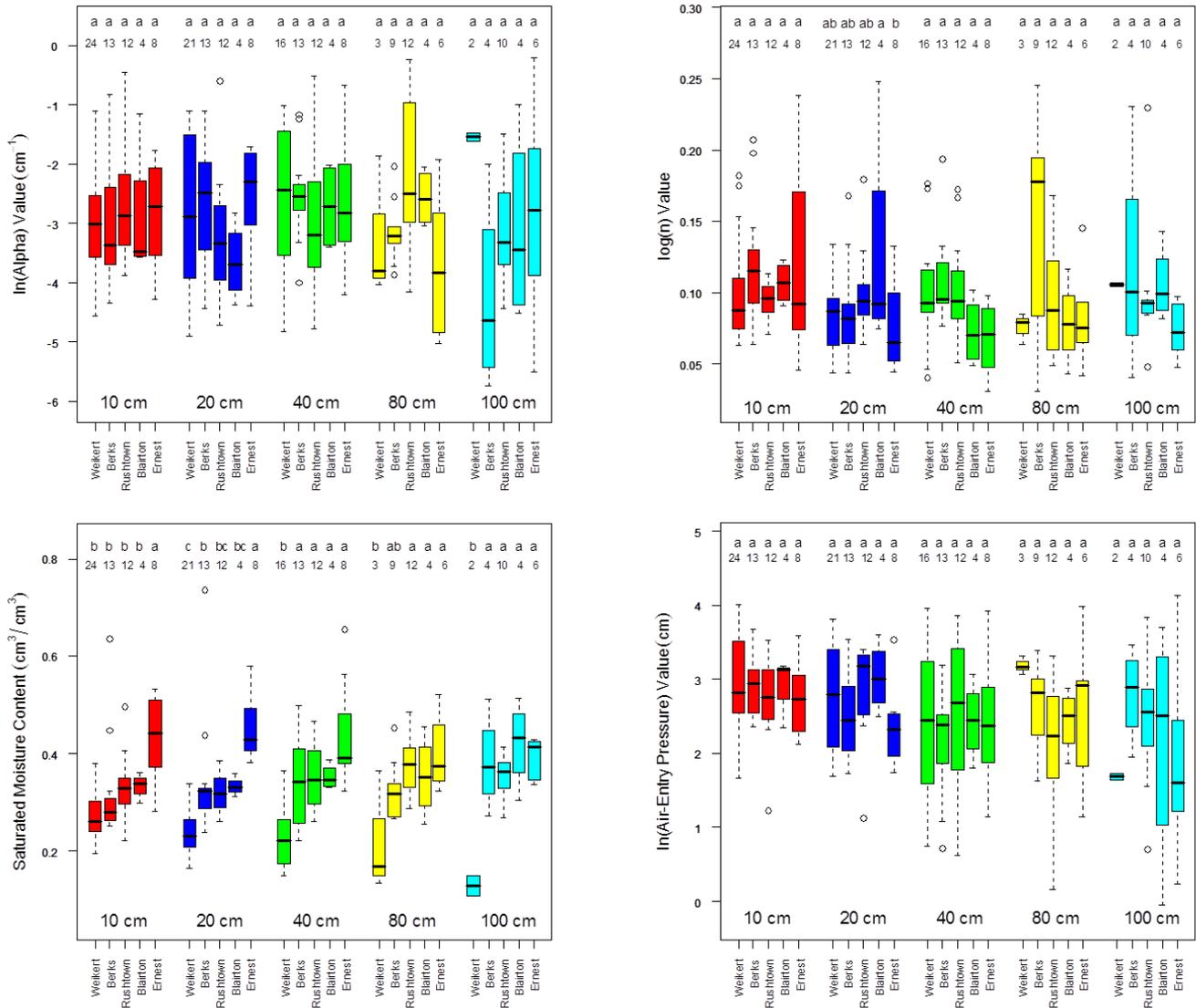


Figure 3.4: Boxplots of θ_s , n , α and ψ_e parameters for all depth-locations grouped by soil series. Letters indicate significant differences of means for parameter values among different soil series within each depth interval according to a Tukey HSD test at $p < 0.05$. Numbers above each box indicate sample size. α and ψ_e were natural log-transformed and n log₁₀-transformed to maintain normality for the Tukey HSD test.

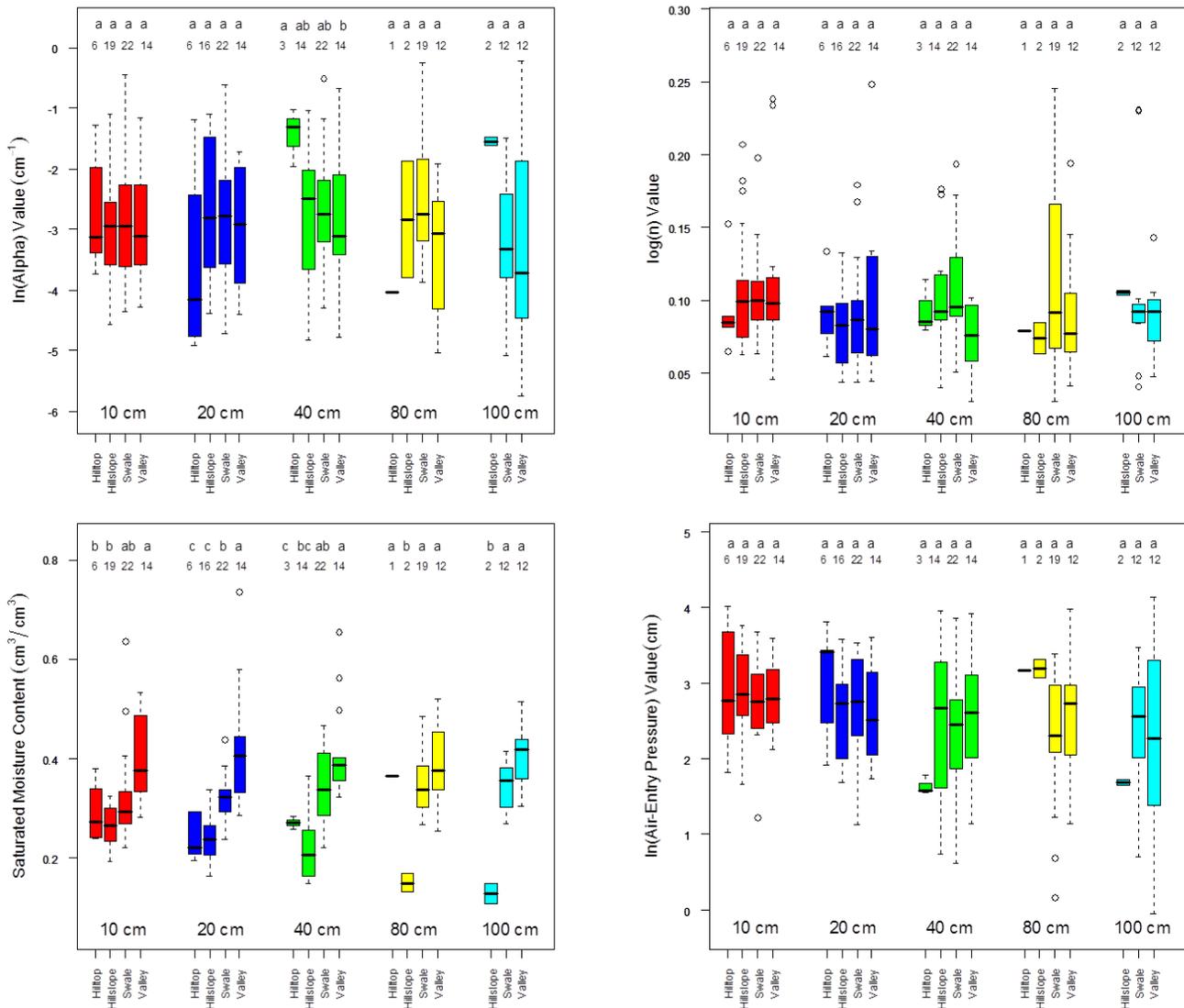


Figure 3.4 (continued): Boxplots of θ_s , n , α and ψ_e parameters for all depth locations grouped by landform units. Letters indicate significant differences of means for parameter values among different landform unit within each depth interval according to a Tukey HSD test at $p < 0.05$. Numbers above each box indicate sample size. α and ψ_e were natural log-transformed and n log10-transformed to maintain normality for the Tukey HSD test.

The significantly higher mean n value for Blairton soil at 20 cm is attributed its deep, well-structured profile that allows moisture to percolate vertically from a near-surface depth effectively. The lowest n value at 20 cm in Ernest soils is realistic considering the soil's lower elevation and water restricting layer below 20 cm that stunts vertical moisture movement. The mean n value for Ernest is lower at 40 cm, which is a typical depth where the fragic horizon occurs for this soil in Shale Hills.

Table 3.2 shows a correlation matrix of topographic variables, hydraulic parameters and soil texture content. The α parameter is well correlated with other hydraulic parameters, however the only topographic variable it is correlated with is topographic wetness index. The α parameter is negatively correlated with ψ_e , which reinforces the empirically derived notion that α is approximately the inverse of air-entry pressure (van Genuchten, 1980). The negative correlation with n implies that macropore flow is generally higher in locations with less effective release of moisture or greater pore size distribution. This is realistic in valley and center swale depths where cracks exist in the massive or sub-angular soil structure.

Since α is significantly different between landform units at the 40 cm depth (Figure 3.4), the significantly negative association with TWI in Table 3.2 may stem from generally higher macropore flow occurring at 40 cm depth in landform units with substantially lower potential wetness. The lack of correlation between α and texture variables further indicates that this is a *structure* as well as a textural parameter. It may be high in clayey soil with cracks between aggregates, but also high in hilltop and hillslope soil with large spaces between coarse material. Low α values may exist in unstructured hillslope sites, which have silty particles between coarse fragments and valley locations with high clay content and little ped aggregation. Texture generally becomes a control on α when soil is unstructured (Hodnett and Tomasella, 2002).

Table 3.2: Spearman correlation matrix of soil hydraulic parameters and soil-terrain attributes from all depth-locations in the Shale Hills. Values that are significant at $p < 0.05$ are highlighted in red and the p-value is indicated in parentheses.

Parameters	α	n	θ_s	a	ψ_e	$-b$	Slope Value	TWI	Curvature	Elevation	Upslope Area	Depth to Bedrock	Sand %	Silt %	Clay %	Organic Matter %	Rock Fragment %
α	1																
n	-0.28 (<0.001)	1															
θ_s	0.15 (0.026)	0	1														
a	-0.24 (<0.001)	0.35 (<0.001)	0.25 (<0.001)	1													
ψ_e	-0.84 (<0.001)	0.17 (0.008)	-0.26 (<0.001)	0.43 (<0.001)	1												
$-b$	-0.12	0.36 (<0.001)	-0.11	0.89 (<0.001)	0.36 (<0.001)	1											
Slope Value	0.1	-0.01	-0.07	-0.08	-0.14 (0.034)	-0.03	1										
TWI	-0.14 (0.030)	0.03	0.2 (0.002)	0.1	0.14 (0.028)	-0.02	-0.64 (<0.001)	1									
Curvature	0.07	-0.13 (0.050)	-0.25 (<0.001)	-0.11	-0.06	0	0.35 (0.005)	-0.75 (<0.001)	1								
Elevation	0.02	0.09	-0.28 (<0.001)	0.03	0.03	0.16	0.39 (0.002)	-0.50 (<0.001)	0.32 (0.011)	1							
Upslope Area	-0.06	-0.14 (0.033)	0.48 (<0.001)	0.02	-0.03	-0.19 (0.017)	-0.07	0.65 (<0.001)	-0.54 (<0.001)	-0.32 (0.011)	1						
Depth to Bedrock	-0.05	-0.04	0.55 (<0.001)	0.05	-0.07	-0.18 (0.004)	-0.07	0.38 (0.003)	-0.58 (<0.001)	-0.23	0.66 (<0.001)	1					
Sand %	0.13	-0.11	-0.07	-0.17 (0.045)	-0.09	-0.13 (0.005)	0.27 (0.002)	-0.08	-0.17 (0.046)	0.37 (<0.001)	0.01	0.19 (0.028)	1				
Silt %	-0.11	0.14	-0.03	0.18 (0.038)	0.1	0.19 (0.028)	-0.22 (0.010)	0.07	0.22 (0.009)	-0.32 (<0.001)	-0.04	-0.24 (0.005)	-0.85 (<0.001)	1			
Clay %	-0.1	0	0.34 (<0.001)	0.12	0.05	-0.04	-0.25 (0.003)	0.17 (0.041)	-0.13	-0.27 (0.001)	0.18 (0.032)	0.22 (0.012)	-0.53 (<0.001)	0.11	1		
Organic Matter %	-0.04	0.16	-0.22 (0.010)	0.35 (<0.001)	0.15	0.49 (<0.001)	-0.08	-0.05	0.1	-0.06	-0.19 (0.026)	-0.20 (0.017)	-0.05	0.05	0.03	1	
Rock Fragment %	0.15	-0.17 (0.044)	-0.50 (<0.001)	-0.29 (<0.001)	0.01	-0.08	0.01	-0.08	0.11	0.1	-0.28 (<0.001)	-0.40 (<0.001)	0.22 (0.009)	-0.19 (0.024)	-0.23 (0.007)	0.08	1

It is clear that surface curvature and upslope contributing area place a relatively large control on n . These topographic correlations indicate that i) soils with high upslope accumulation areas release moisture less effectively ii) areas with negative, or concave, curvature generally release moisture more effectively. While the negative correlation with upslope contributing area makes sense, the n parameter's association with surface curvature is somewhat counter-intuitive. Areas with negative curvature generally have deeper and more weathered soil. If n is negatively correlated with surface curvature or in soil with negative curvature near the surface, it may be suggested that moisture effectively drains vertically from the surface in deep soil. Figure 3.4 shows n is significantly higher at 20 cm in deep and concave Blairton soil, while correlation matrices in Appendix D show stronger negative

correlations with n at near-surface depths of 10 and 20 cm. The n parameter is negatively correlated with rock fragments. This follows the logic that moisture is generally released more effectively at the near surface in deeper, well weathered soil, which would contain less rock fragments.

The θ_s parameter is influenced by both pore space volume and amount of moisture flux occurring in a depth location, and correlation results show that θ_s is controlled by both topographic and texture variables. The significant topographic correlations indicate that i) areas with deep depth to bedrock contain depths with higher moisture saturation, ii) more concave landscape positions have higher saturated moisture contents, iii) lower elevation increases saturated moisture contents, iv) locations with extensive upslope contributing areas tend to have higher saturated moisture contents, and v) areas that are potentially wetter have greater saturated moisture contents.

A strong positive correlation exists with clay, which is reasonable in that soils with high clay contents tend to hold more moisture. The strong negative correlation with rock fragments is reasonable, because higher rock fragment content allows for less space to be filled with moisture at saturated field conditions. The negative control organic matter places on saturated moisture content may be counter-intuitive at first, since organic matter tends to increase a soil's capacity to hold moisture. In undisturbed forest settings, such as Shale Hills, higher organic matter may indicate a greater presence of roots at a location. An increased amount of roots would decrease the maximum capacity of a soil moisture content, since roots occupy space and extract moisture from soil.

3.2 Spatial Distribution of the van Genuchten Parameters across the Shale Hills Landscape

To assess the spatial variability and patterns of soil moisture characteristics across the Shale Hills catchment at near-surface and deep depths, spatial models for 10, 20, 40, 80 and 100 cm depths were created for the α , n and θ_s van Genuchten parameters. Table 3.3 displays parameters for each spatial model and any significant topographic covariates used as external trends for each model. Figure 3.5 shows a series of semi-variogram curves for near-surface 20 cm depth and deep profile 80 cm depth. The standard deviation of each moisture retention parameter across space is represented by envelopes in the semivariogram figures.

The n parameter at the near-surface 20 cm has a relatively long range of 62 meters with a positive spatial trend with depth to bedrock. Areas with deeper soil may drain moisture from 20 cm depths relatively more effectively, because of greater profile extent for vertical moisture release. The range dramatically dropped to 3 meters at the 80 cm depth for n , indicating that the effective drainage characteristic is more similar in the deep profile across the catchment. The negative relationship between topographic wetness index and n at 80 cm indicates that soil 80 cm below the surface at sites with higher potential wetness drain moisture more gradually.

Table 3.3: Semivariogram parameters for Figure 8. The van Genuchten parameter values were transformed with the *Lambda* parameter to satisfy spatial isotropy. The range is the distance on the x-axis at the inflection point of the curve. The nugget is the intercept of the curve on the y-axis, and the partial sill is the distance on the y-axis between the nugget and maximum height of the curve. Correlated topographic covariates used in spatial modeling are listed with significance indicated (**** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Parameter	Depth (cm)	Range (m)	Partial Sill	Nugget	Lambda	Correlated Topographic Coefficients
n	10	31.2	0.0008	0.0007	-3.0	None
n	20	62.0	0.0005	0.0012	-3.0	Depth to Bedrock **
n	40	12.8	0.0007	0.0020	-3.0	Elevation **
n	80	3.0	0.0029	0.0007	-3.0	- TWI *
n	100	6.3	0.0024	0.0008	-3.0	- TWI *
θ_s	10	37.4	0.182	0.262	-1.0	TWI *
θ_s	20	14.3	0.083	0.044	-0.28	- Elevation **
θ_s	40	10.5	0.047	0.008	0.23	- Slope *
θ_s	80	31.2	0.002	0.001	1.32	- Slope *
θ_s	100	62.2	0.002	0.001	1.34	- Slope * - Curvature *
α	10	45.2	0.851	3.7529	-0.30	None
α	20	60.8	1.77	1.13	-0.07	None
α	40	48.3	0.47	1.1439	-0.05	None
α	80	33.9	1.35	1.82	-0.14	Elevation *
α	100	45.5	0.53	1.21	0.00	log(Upslope Area) *

A disparity in range values also exists in θ_s at 20 and 80 cm, suggesting that significant differences exist in the soil porosity and field saturation across the catchment at 20 and 80 cm. The smaller 14.3 meter range at 20 cm suggests that soil porosity is more homogeneous at this depth. Saturated moisture at 80 cm has a higher range at 31.2 meters, which is a function of disparate soil porosities. Tensiometers at 80 cm are submerged in either BC horizon or C horizon.

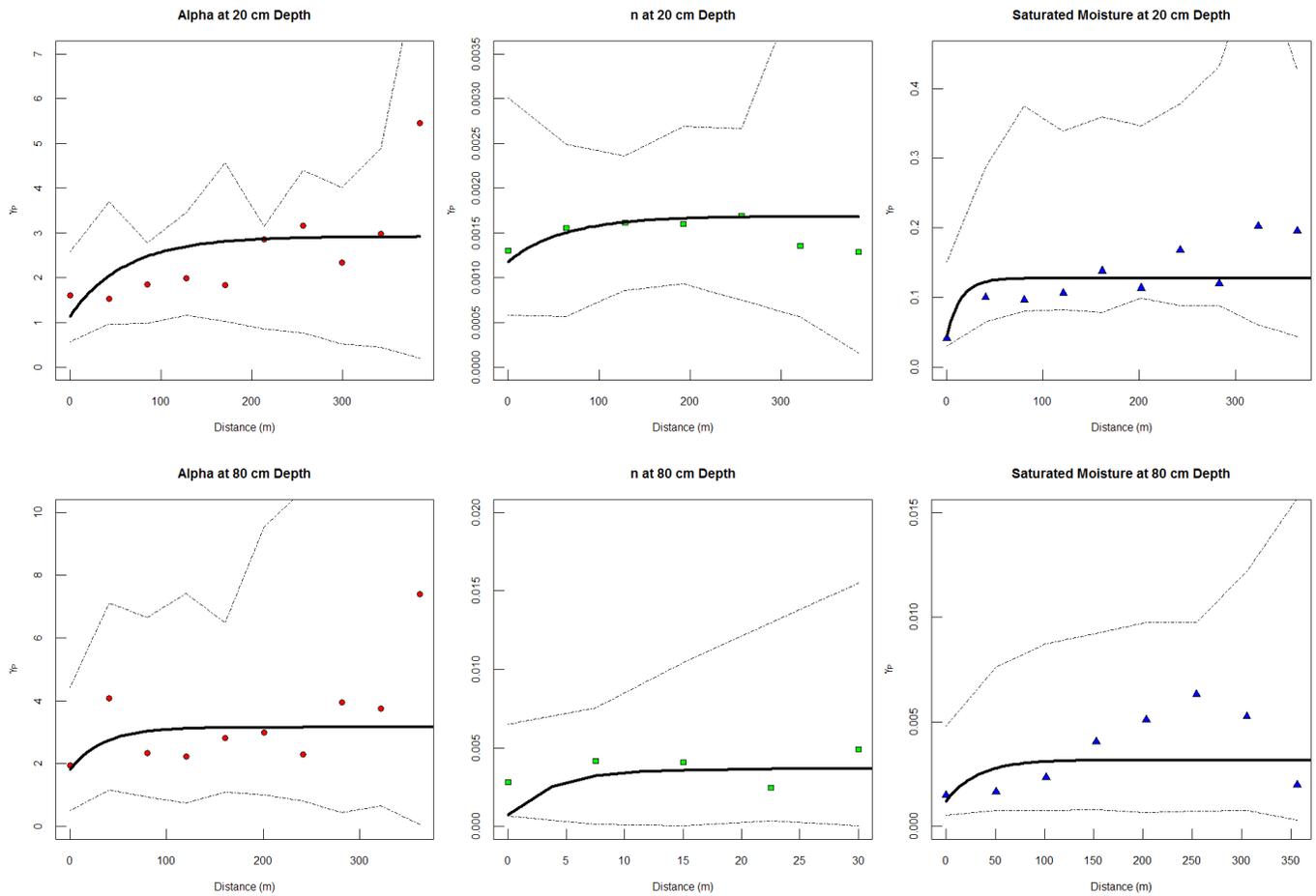


Figure 3.5: Series of semi-variograms showing semi-variance of α , n and θ_s at shallow (20 cm) and deep solum (80 cm) depths across the catchment. The γ values for each semi-variogram have been binned. The covariance models in each semi-variogram are Matern functions, and parameters for the covariance functions were optimized using a Bayesian estimation procedure. Parameters were transformed using a λ value (Table 3.3) optimized by a maximum likelihood algorithm during spatial model fitting to maintain independence and isotropy. Dotted lines indicate 95% Confidence Interval envelope of the data cloud.

There is also a notable difference in ranges for α values between 20 and 80 cm.

The higher α range at 20 cm is an indication of disparate soil structure at the near surface across the catchment. The 20 cm depth may be home to a dense layer as in some Ernest sites, aggregated clayey soil in Blairton sites, or coarse, unstructured soil at Weikert sites.

The lower range for α at 80 cm suggests that the differences in soil structure at this depth are less pronounced among different sites across the catchment. The positive correlation with elevation at 80 cm indicates that sites with greater elevation tend to have

more macropore controlled flow at 80 cm. This makes sense when considering the higher amount of rock fragments that are present in locations within shallower soil or C horizon.

Figure 3.6 displays a series of maps depicting θ_s values at 10, 20, 40, and 80 cm depths in Shale Hills created through regression kriging. Patterns of θ_s on maps at each depth consistently show the highest degree of field saturation is located in sites close to

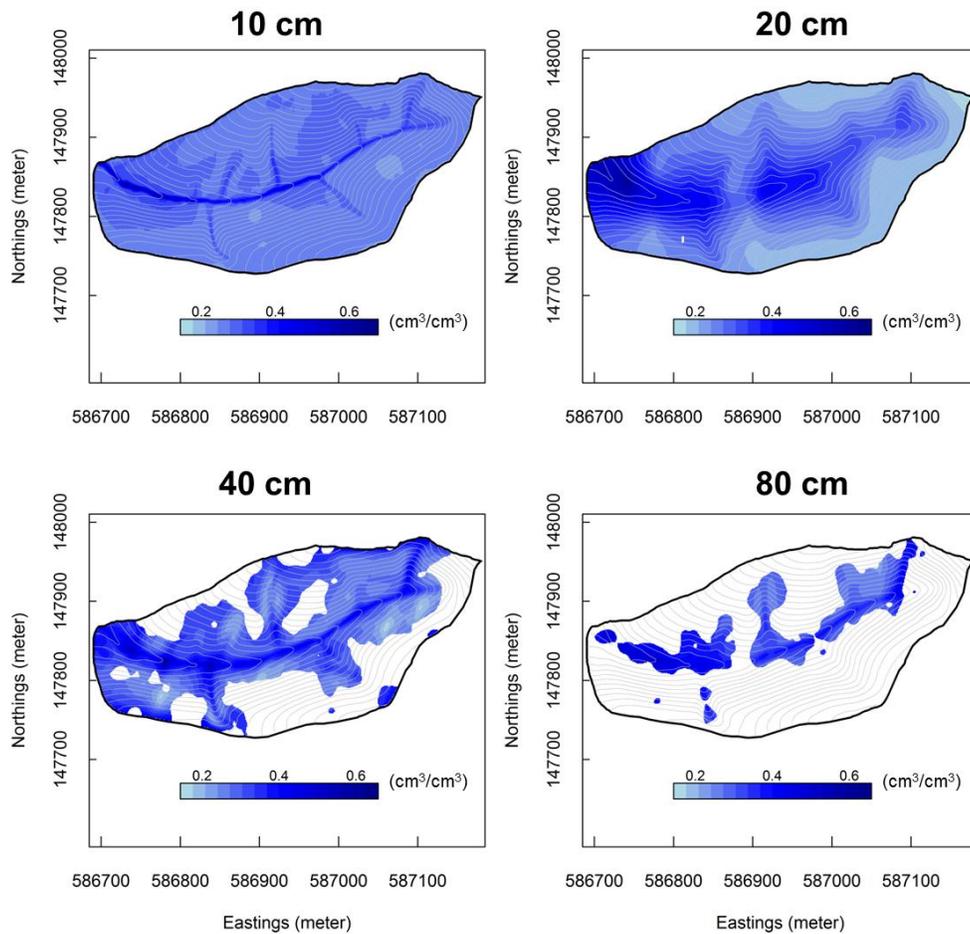


Figure 3.6: Maps of θ_s at four different depths over Shale Hills. Regression kriging with significant topographic variables was used for interpolation. A Bayesian optimization was employed to optimize spatial model parameters. Areas with soil shallower than featured depth have been masked.

the catchment outlet. Lower field saturation values are consistently found on convex hillslopes and side-slope areas at each depth.

Interpreting field saturation throughout the catchment at different depths has ecological modeling significance as different tree species can be related to varying degrees of wetness present in the root zone. As field saturation indicates a soil's maximum moisture holding capacity and is related to both topography and porosity, catchment-wide θ_s maps at multiple depths may also have contributions to catchment-scale hydrological outflow modeling, as well. Furthermore, mapping of field normalized moisture content ($\frac{\theta}{\theta_s}$) is possible with maps of θ_s . This is useful in determining the relative degree of moisture content that is present within a soil at a given time for a certain depth.

3.3 General Terrain and Soil Characteristics of Hydropedological Functional Units

Abundant correlation among raster variables is evident from the correlation matrix in Table 3.4. Solum θ_s storage and depth to bedrock were chosen for delineation of hillslope HFUs. Although depth to bedrock was used in interpolating solum θ_s storage, depth to bedrock shows more accurate distinctions between shallow hillslope soil and intermediate deep hillslope soil than θ_s and thus was used in tandem with θ_s storage. The θ_s storage suggests unique information relative to depth to bedrock in that it shows higher soil moisture holding capacity in the footslope areas near the catchment outlet, which is

not exhibited distinctly in the depth to bedrock. This is important for maintaining a comprehensive incorporation of solum scale soil water retention characteristics for soil close to the catchment outlet.

Even though depth to bedrock and θ_s storage were the only two raster datasets selected for PCA, they carry general patterns from multiple inter-correlated topographic variables as a consequence of their generation from regression kriging. The PCA was necessary in reducing the complex and correlated data structure of the two compound raster datasets. Another reason to perform a PCA on the two raster datasets was the PCA's capability of relating a dataset of soil depth information with a dataset of saturated moisture storage. Although both rasters appear similar, the physical attributes each dataset expresses is not easily relatable. Depth to bedrock expresses a static soil property, but θ_s storage is derived from the aggregation of modeled parameters calculated from dynamic $\theta(\psi_m)$ data. The scaling of both raster datasets before PCA prevented depth to bedrock data from dominating the unique information found within the θ_s storage data.

Figure 3.7 displays a barplot showing the resulting variances explained by each PC derived from the analysis. As expected, one PC dominates the variance, by explaining 97.1 % of the total variance among depth to bedrock and θ_s storage with 2.9 % left as noise. Both variables have a high positive score loading of ~ 0.70 each on PC I, indicating that PC I explains soil weathering and maximum field profile-scale soil moisture holding capacity.

Spearman Correlation	α	θ_s	n	Slope	TWI	Curvature	Elevation	Upslope Area	Soil Depth	Sand	Silt	Clay
α Geometric mean	1											
θ_s Storage	-0.01	1										
n Geometric mean	-0.21 (0.097)	0.1	1									
Slope	0.09	-0.50 (<0.001)	0.25 (<0.001)	1								
TWI	-0.07	0.70 (<0.001)	-0.19	-0.59 (<0.001)	1							
Curvature	-0.04	-0.60 (<0.001)	-0.03	0.37 (0.003)	-0.64 (<0.001)	1						
Elevation	-0.11	-0.50 (<0.001)	0.30 (0.021)	0.42 (<0.001)	-0.52 (<0.001)	0.26 (0.045)	1					
Upslope Area	-0.04	0.52 (<0.001)	-0.18	-0.46 (<0.001)	0.81 (<0.001)	-0.67 (<0.001)	-0.43 (<0.001)	1				
Soil Depth	-0.01	0.80 (<0.001)	0.04	-0.46 (<0.001)	0.64 (<0.001)	-0.66 (<0.001)	-0.41 (<0.001)	0.55 (<0.001)	1			
Profile Sand Storage	0.21	0.47 (0.003)	-0.03	0.01	0.34 (0.038)	-0.55 (<0.001)	0.06	0.2	0.59 (<0.001)	1		
Profile Silt Storage	0.01	0.79 (<0.001)	0.22	-0.35 (0.034)	0.57 (<0.001)	-0.50 (0.001)	-0.42 (0.008)	0.33 (0.046)	0.82 (<0.001)	0.55 (<0.001)	1	
Profile Clay Storage	0.02	0.79 (<0.001)	0.16	-0.40 (0.013)	0.60 (<0.001)	-0.63 (<0.001)	-0.34 (0.039)	0.40 (0.013)	0.85 (<0.001)	0.55 (<0.001)	0.87 (<0.001)	1

Table 3.4: Correlation matrix of raster datasets from a Spearman non-parametric correlation test. Red emboldened correlation coefficients indicate a significant correlation between variables at $p < 0.05$, and blue emboldened correlation coefficients represent a significant correlation at $p < 0.10$

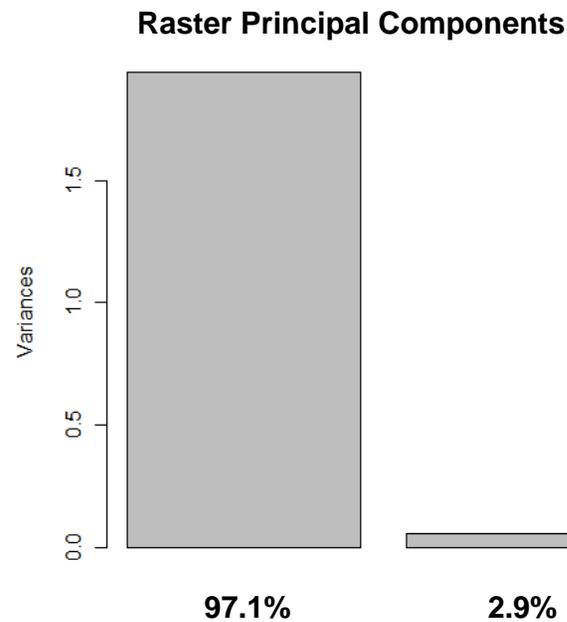


Figure 3.7: Bar plot depicting the amount of variance explained by PCI and PCII. PCI explains the vast majority of the variance and was used for fuzzy c-means cluster analysis

PC I was used to obtain a rotated data value relative to rotated component space for each grid cell for input into fuzzy c-means clustering. Based upon preliminary analysis of raster datasets and the limiting number of units expressed in the two raster datasets analyzed in the PCA, three clusters were designated for the c-means cluster algorithm. Figure 3.8 shows a raster of 3 delineated clusters resulting from the cluster analysis. Cells were assigned a cluster based upon the maximum proportion value associated with the potential clusters.

Hillslope Hydropedological Functional Units

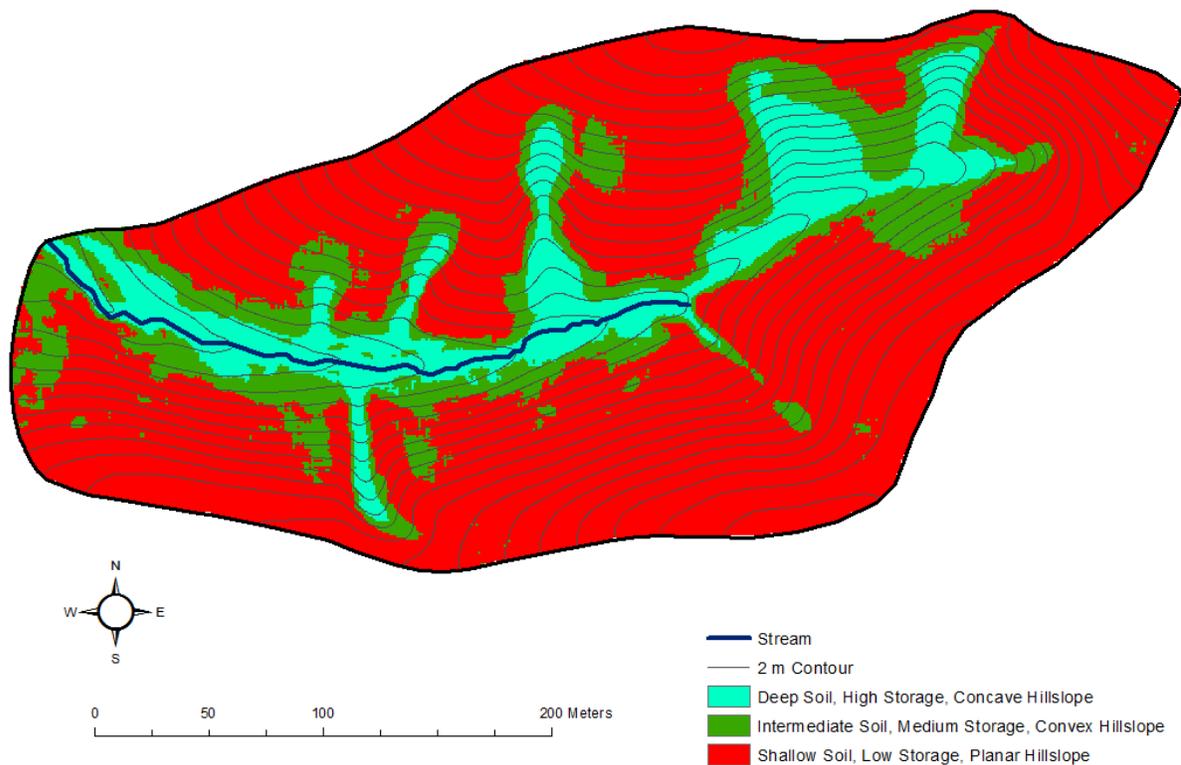


Figure 3.8: Raster of 3 clusters based upon multivariate analysis and clustering of Depth to Bedrock and θ_s Storage raster datasets. TDR-Tensiometer site locations are shown and symbolized according to ascribed soil series.

The influence of each contributing raster may be seen in Figure 3.9. The three clusters were named “Deep Soil, High Storage, Concave Hillslope” (DSHSCH), “Intermediate Soil, Medium Storage, Convex Hillslope” (ISMSCH) and “Shallow Soil, Low Storage, Planar Hillslope” (SSLSPH). The SSLSPH cluster encompasses planar areas with shallow soil and the DSHSCH cluster occurs in concave areas with deep soil. The ISMSCH designation for the green colored cluster occurs predominately on the convex relief between the concave hillslope and planar hillslope landform areas and encompasses both Berks and deep Weikert soil series. The ISMSCH acts as a transition zone of fluid and material flux between the SSLSPH and the DSHSCH units. The PCA and subsequent fuzzy c-means clustering delineated three main hillslope units out of the depth to bedrock and θ_s storage raster data that were very similar to units delineated for both datasets in the preliminary analysis (Table 2.1). The incorporation of solum θ_s storage did contribute unique catchment outlet information, while accurate catchment-wide soil depth information was apparent preserved with the depth to bedrock raster dataset.

The threshold slope value used to separate a Valley unit from the DSHSCH and ISMSCH clusters was informed by analyzing the slope map. From Figure 3.9, it is apparent that slope values up to class C (green) seem to separate a valley unit from the remainder of both DSHSCH and ISMSCH clusters. In order to delineate a distinct valley unit, the threshold slope value was increased slightly from 0.150 to 0.185m/m. This threshold slope value allowed the inclusion of almost all the low elevation raster cells typically associated with the Ernest soil series into a Valley unit.

Slope Classes with Soil Series Delineation

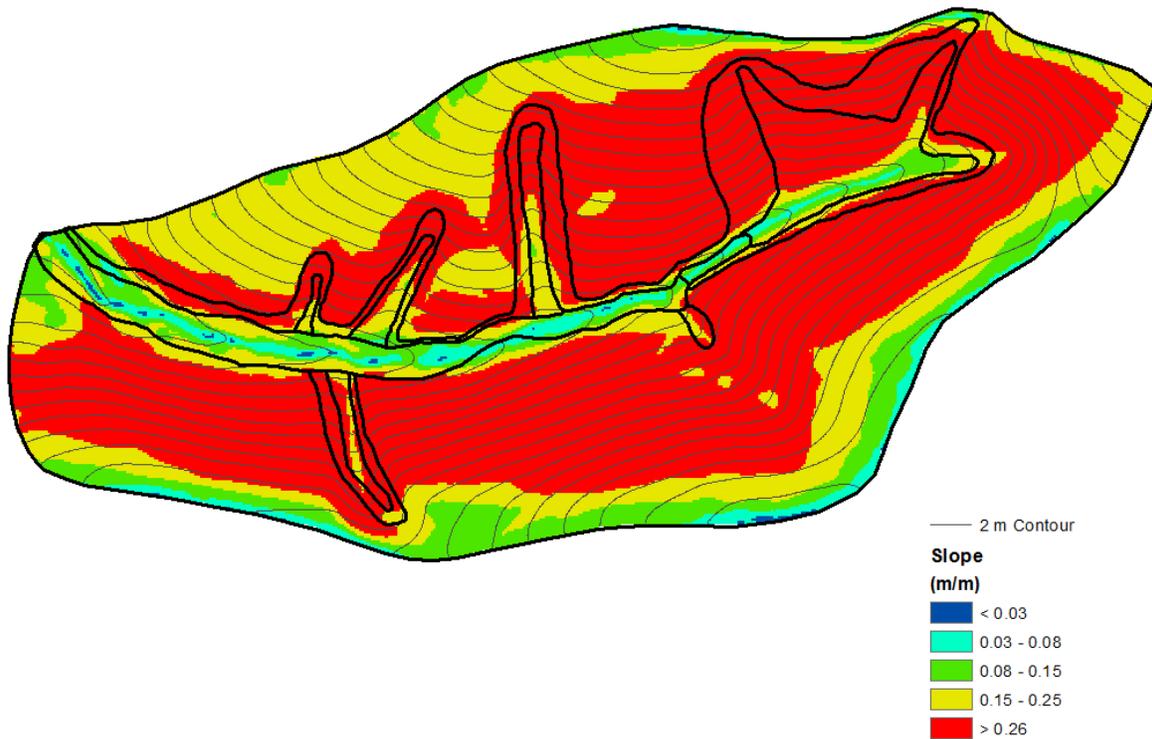


Figure 3.9: Map of slope class delineations for Shale Hills. Soil Series boundaries are demarcated in bold. Slope Class D (0.15 – 0.25) does exist within the Ernest soil series boundary

As EMI surveys and profile moisture storage maps indicate a separate function for the valley area from the origin of the ephemeral stream to the catchment outlet, the elevation of the stream origin was used to limit the Valley unit. The elevation of the stream origin is approximately 272 meters above sea level, and this was used with the LiDAR elevation data to abbreviate the Valley, which slope value was not capable of doing by itself.

Park and vande Giesen's analysis procured a $\log(A_s)$ value for a summit landform, which was then in turn used with slope to concisely delineate a Flat Summit unit from the SSLSPH cluster. The threshold slope value for the Flat Summit is slightly less than slope class C at 0.130 m/m, so that a muddled over-reach of the Flat Summit into the SSLSPH

unit could be avoided. To extend a consistent naming convention for all HFUs, the Valley unit is officially termed Deep Soil, High Storage, Flat Valley (DSHSFV) and the Summit unit is termed Shallow Soil, Low Storage, Flat Summit (SSLSFS).

At the conclusion of the process, 5 finalized HFUs were delineated. These final units represent a precision map of similar soil-terrain-hydrologic functional units created from raster sets embodying datasets that represent a variety of natural processes including soil weathering, topographic hydrologic gradient, surface-induced moisture flux direction, soil water retention capacity, and soil-stream dynamics Table 3.5 displays summary statistics of raster data for each HFU category. Figure 3.11 clearly shows the apparent similarity between the precision soil survey and HFU map.

Hydropedological Functional Units

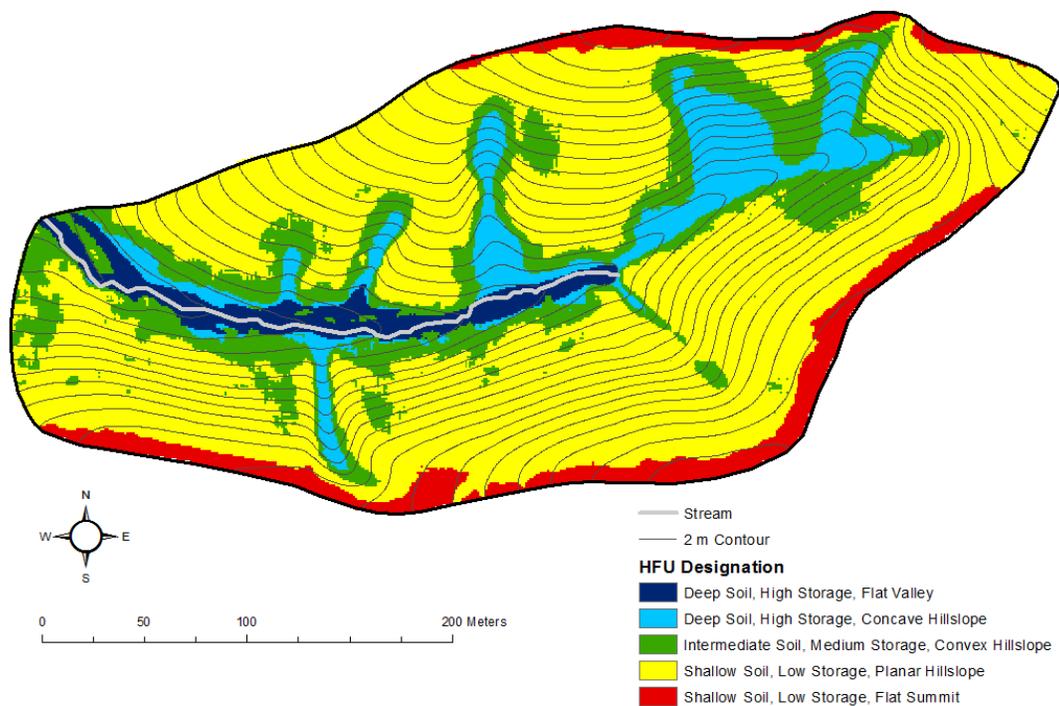


Figure 3.10: Final HFU map with a DSHSFV unit (blue) separated from the DSHSCH (cyan) and ISMSCH (green). A SSLSFS (red) was separated from the SSLSPH (yellow). The ephemeral stream is demarcated and is shown to reside within the DSHSFV unit.

Table 3.5: Summary statistics (mean and standard deviation) of raster variables for each HFU category

Attribute	DSHSFV		DSHSCH		ISMSCH	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Elevation (m)	264.9	3.6	277.8	8.4	276.5	10.3
Slope Value (m/m)	0.104	0.044	0.261	0.090	0.294	0.071
Surface Curvature	-0.032	0.020	-0.028	0.022	-0.007	0.017
Topographic Wetness Index	8.96	2.66	5.99	1.70	5.02	0.72
Depth to Bedrock (cm)	134.4	32.6	121.4	27.9	72.0	12.0
Solum θ_s Storage (m)	0.286	0.066	0.236	0.045	0.150	0.021

Attribute	SSLSPH		SSLSFS	
	Mean	Std Dev	Mean	Std Dev
Elevation (m)	285.1	10.1	295.1	6.8
Slope Value (m/m)	0.283	0.076	0.110	0.049
Surface Curvature	0.007	0.011	0.017	0.012
Topographic Wetness Index	4.65	0.49	4.06	0.52
Depth to Bedrock (cm)	39.8	10.0	32.8	9.1
Solum θ_s Storage (m)	0.095	0.018	0.084	0.016

Hydropedological Functional Units and Soil Series Delineation

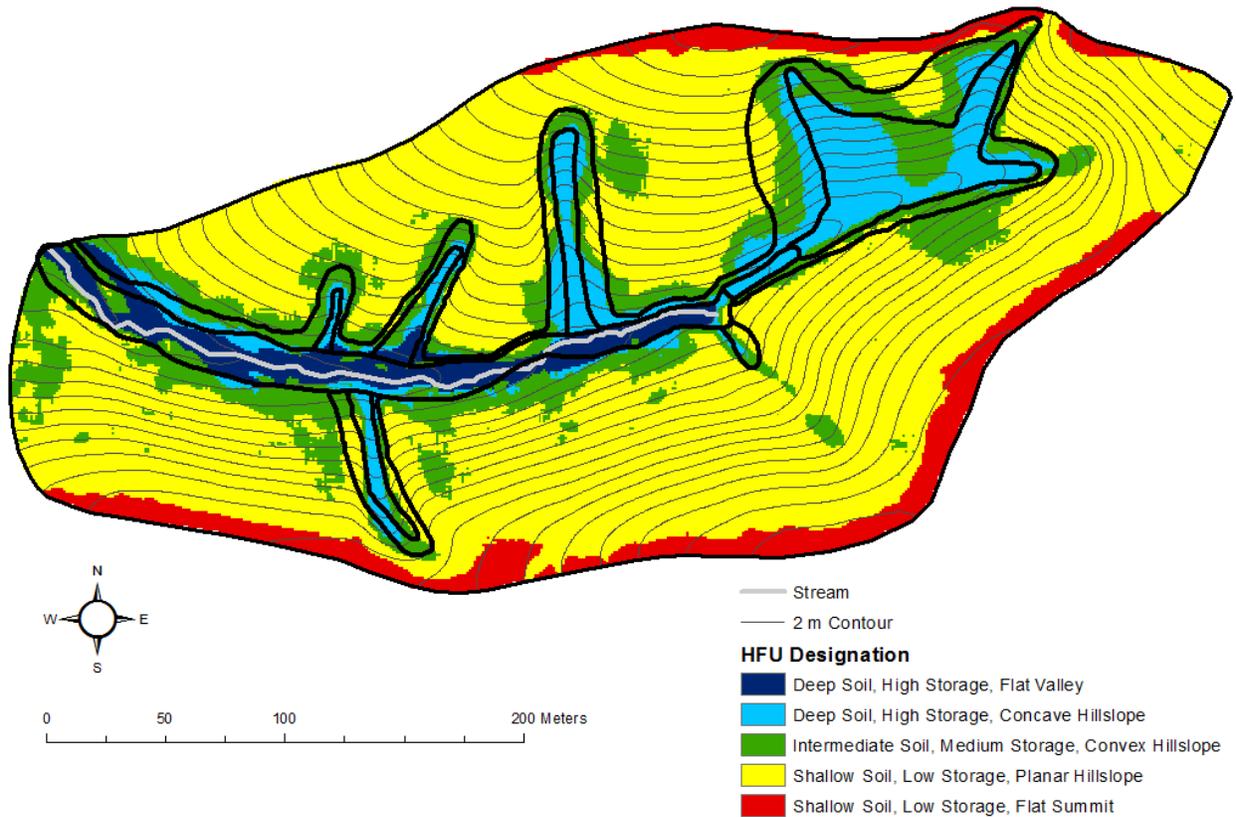


Figure 3.11: Hydropedological Functional Units with 2nd Order Soil Series delineation and catchment ephemeral stream demarcation

3.4 Validation and Cross-validation of Hydropedological Functional Units with Soil Moisture Data

The final HFU raster (Figures 3.10 and 3.11) represents a quantitative synthesis of similar soil-terrain-hydrologic information, and as such, the HFUs reflect consociations of major controls driving soil-hydrologic processes. Soil series designations reflect a synthesis of soil properties and landscape patterns, as well. Park and vande Giesen (2004) attempted to delineate purely terrain-based units (LFUs) in Australia that represent topographic functions. As HFUs represent more information than LFUs, it is expected that soil moisture data across time would be well correlated with HFUs and would perhaps out-perform LFUs in a validation procedure involving observed soil moisture data.

Table 3.6 shows the outcome of the linear model with HFUs, LFUs and soil series each as predictors to solum moisture storage collected at 31 different dates across multiple years and wetness conditions. The date with the highest average observed moisture storage and the date with lowest average observed solum moisture storage were used as individual responses. The predictor with the highest R^2 and lowest residual standard error more accurately predicts total moisture storage patterns for each time period. ANOVA results of each model and the Coefficient of Variation (CV) for each category are also posted.

Table 3.6: Results of linear and generalized linear models for each categorical predictor to 31 moisture total storage responses from dates in 2007 – 2010, a wet date (3/27/2007) and a dry date (7/30/2010). The best diagnostics are highlighted in grey. ANOVA and Coefficient of Variation results of each model are shown to the left of each predictor.

Time Period:	4 Year (31 Dates) 2007 - 2010			Extreme Wet Date 3/27/2007		Extreme Dry Date 7/30/2010	
HFU	CV	Std. Error	P-value	Std. Error	P-value	Std. Error	P-value
High Storage Flat Valley	0.257	0.462	2.00E-16 ***	0.028	2.00E-16 ***	0.011	2.00E-16 ***
High Storage Concave Hillslope	0.358	0.527	4.80E-12 ***	0.034	1.29E-09 ***	0.013	2.41E-09 ***
Medium Storage Planar Hillslope	0.419	0.632	2.00E-16 ***	0.044	7.06E-11 ***	0.016	8.35E-16 ***
Low Storage Convex Hillslope	0.387	0.536	2.00E-16 ***	0.035	2.00E-16 ***	0.013	2.00E-16 ***
Low Storage Flat Summit	0.410	0.979	6.93E-15 ***	0.066	1.63E-09 ***	0.024	1.09E-12 ***
	Residual SE:	1.22		Residual SE:	0.08	Residual SE:	0.03
	Adj R²	0.883		Adj R²	0.787	Adj R²	0.861
Soil	CV	Std. Error	P-value	Std. Error	P-value	Std. Error	P-value
Ernest	0.292	0.660	1.26E-13 ***	0.054	9.50E-05 ***	0.016	1.91E-11 ***
Blairton	0.312	0.801	0.0004 ***	0.061	0.5136	0.020	0.0005 ***
Rushtown	0.337	0.555	0.0035 **	0.053	0.8717	0.014	0.0050 **
Berks	0.492	0.400	2.00E-16 ***	0.043	1.02E-07 ***	0.010	1.20E-14 ***
Weikert	0.547	0.490	3.53E-09 ***	0.049	0.0004 ***	0.012	3.50E-08 ***
	Residual SE:	1.39		Residual SE:	0.11	Residual SE:	0.03
	Adj R²	0.850		Adj R²	0.670	Adj R²	0.817
LFU	CV	Std. Error	P-value	Std. Error	P-value	Std. Error	P-value
Valley	0.371	1.150	2.11E-07 ***	0.061	4.44E-05 ***	0.025	3.41E-08 ***
Concave Hillslope	0.474	0.768	2.05E-06 ***	0.049	0.0009 ***	0.017	4.20E-07 ***
Planar Hillslope	1.305	0.597	0.0003 ***	0.039	0.0036 **	0.013	0.0035 **
Convex Hillslope	0.275	1.431	0.5313	0.084	0.5290	0.031	0.5641
Summit	0.415	2.669	0.5664	0.154	0.6112	0.058	0.6656
	Residual SE:	2.60		Residual SE:	0.15	Residual SE:	0.06
	Adj R²	0.471		Adj R²	0.333	Adj R²	0.501

Between HFUs, LFUs and soil series, HFUs showed a significantly better ability to predict total moisture storage than soil series and LFUs. Soil series, which takes into account soil depth, slope and other soil properties, did better than LFUs. Purely terrain-based landscape characterization does not directly represent subsurface processes or soil water retention, leading to a relative misalignment with respect to representing soil moisture storage patterns compared with soil series and HFUs.

Figure 3.12 displays R^2 diagnostics of HFUs, LFUs and soil series as predictors to solum moisture storage for separate dates from December 2006 to November 2010. HFUs and soil series outperform LFUs consistently at predicting solum moisture storage across all wetness conditions across several years. As solum moisture storage may be directly related to soil weathering and soil properties, soil series and HFUs display the importance of characterizing subsurface processes and soil properties when delineating landscape scale soil units. HFUs also consistently outperform soil series in this context, as well.

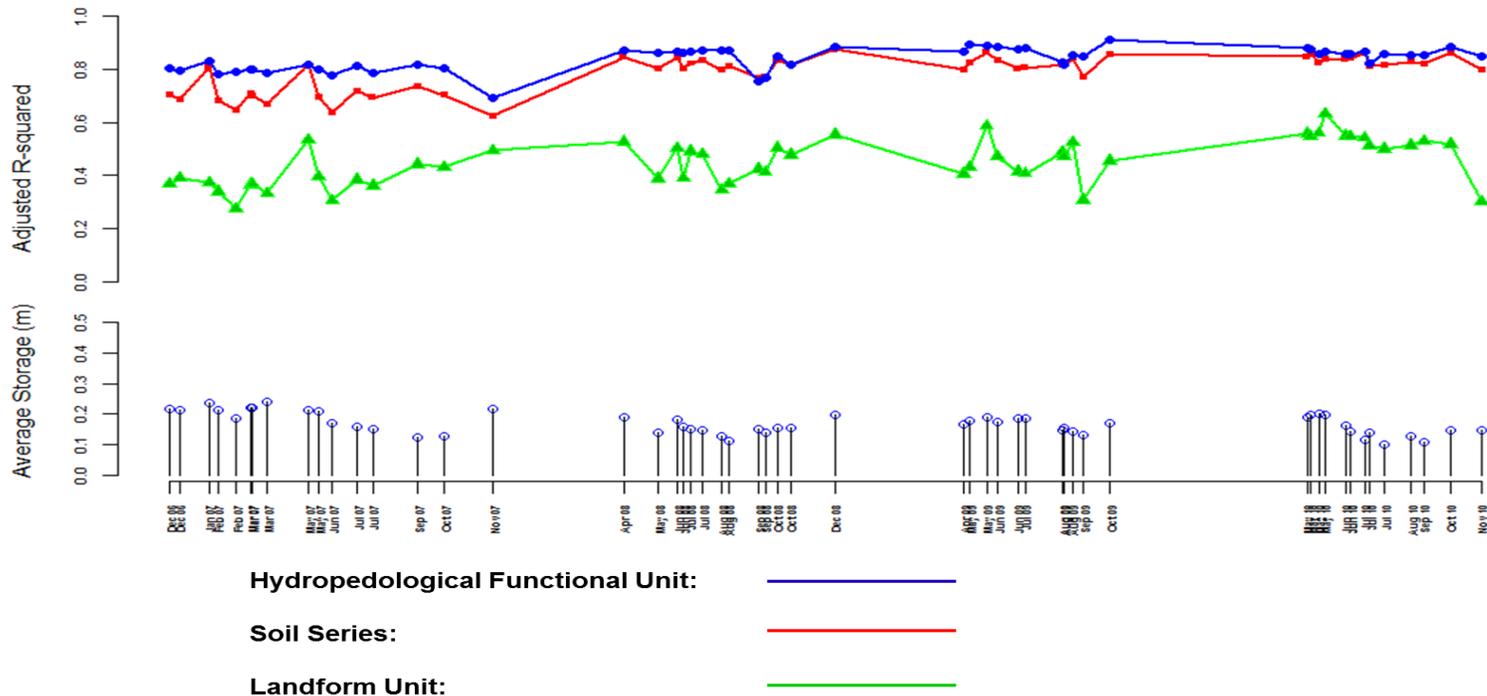
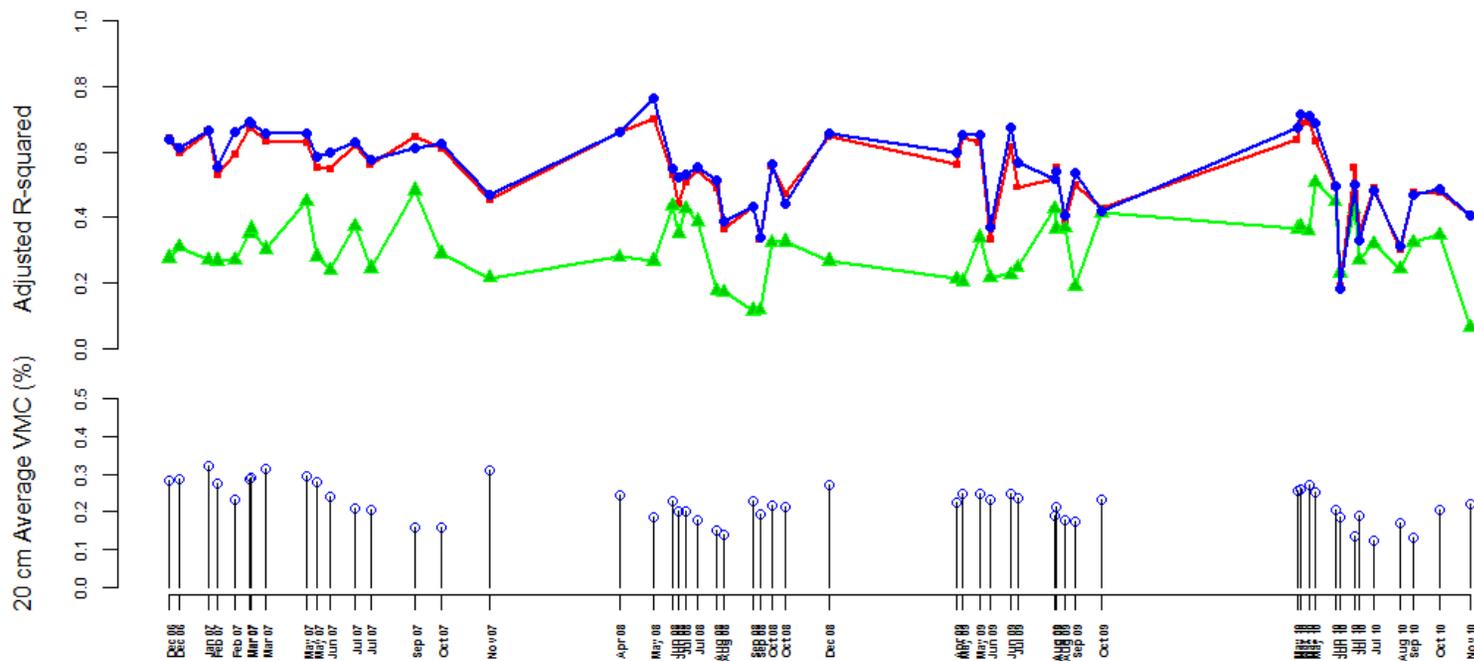
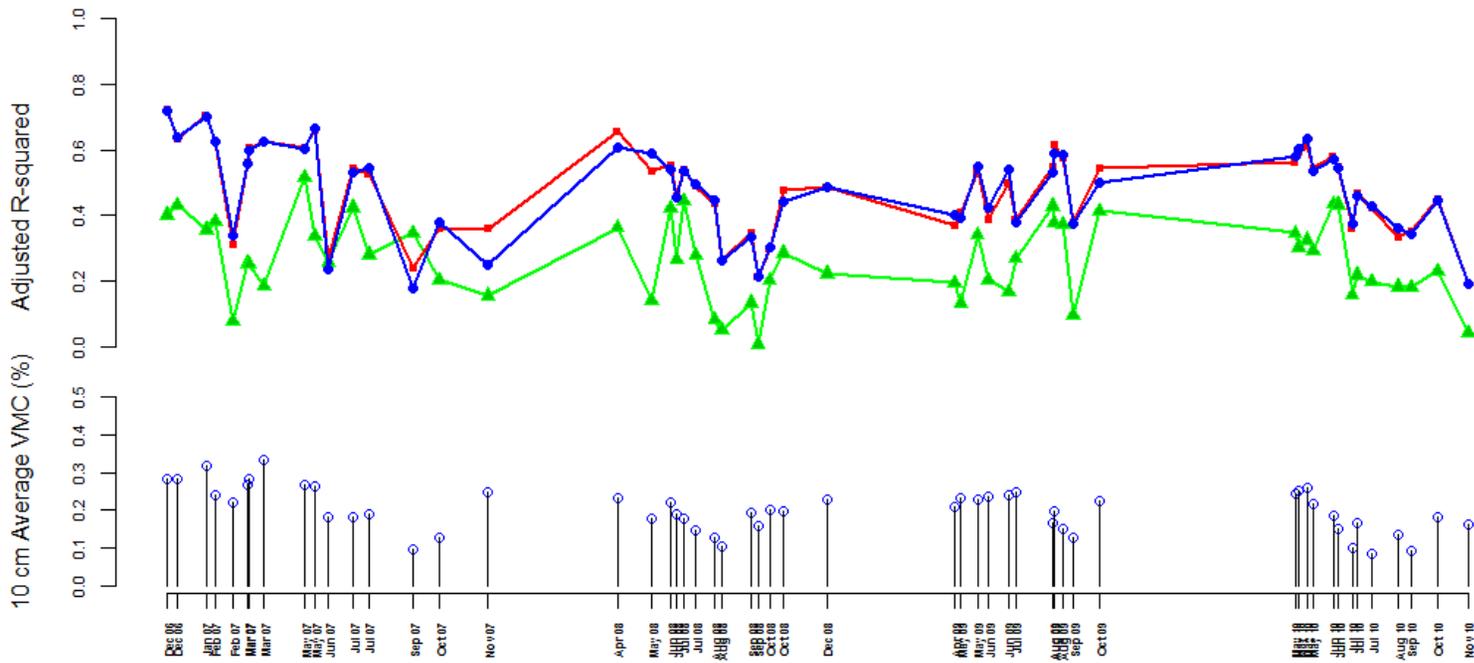


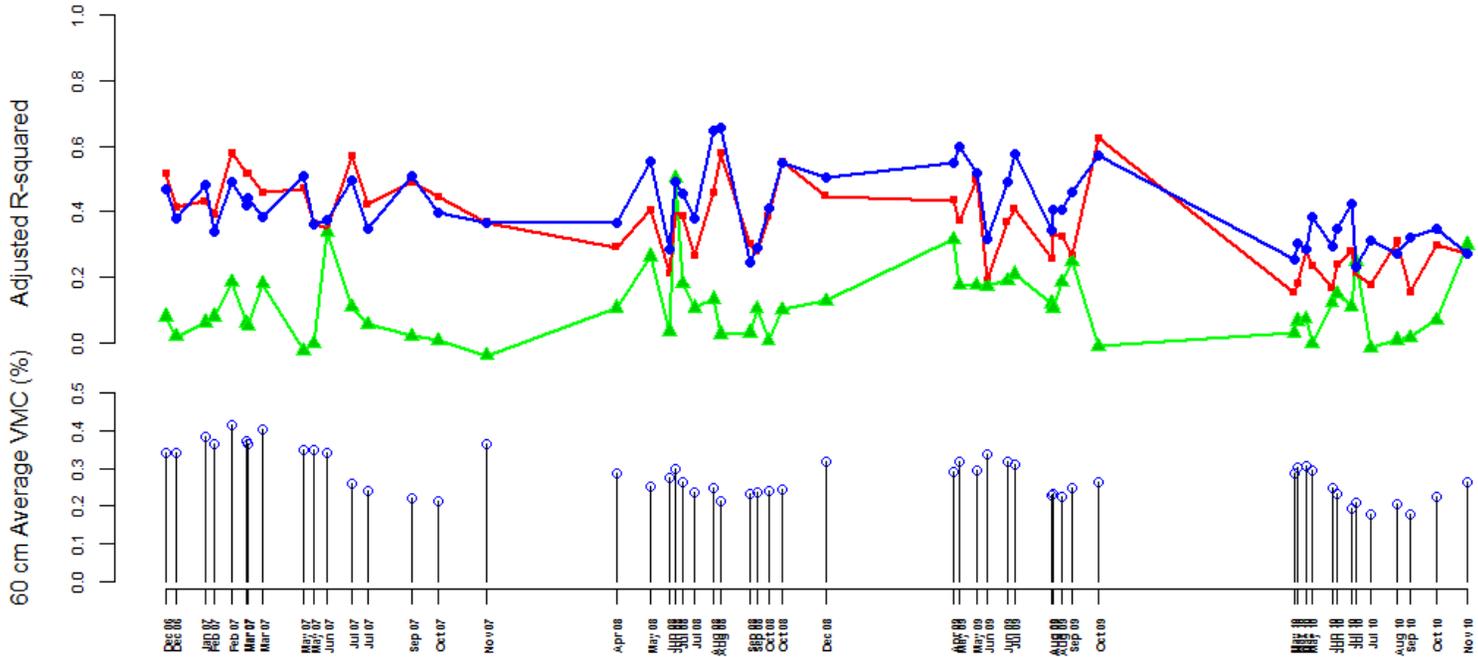
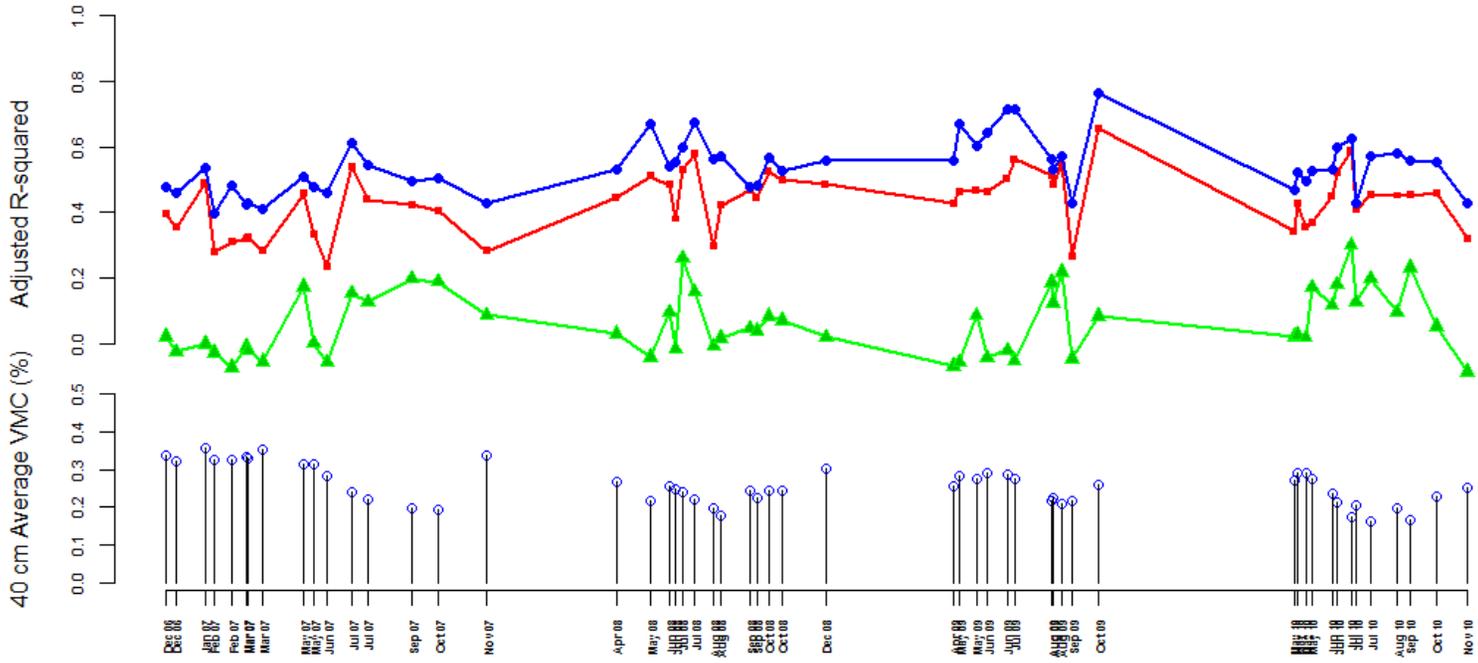
Figure 3.12: Performance of 3 landscape characterizations in predicting total moisture storage within a linear model. The y-axis shows dates of moisture collection with average total storage from observations shown directly above. Lines indicate adjusted R^2 values indicating performance of each predictor in predicting total moisture storage in a linear model for each date.

Soil water retention is a fundamental soil-hydrologic function that was integrated into HFUs with the solum θ_s storage map. The precision of LiDAR raster maps, integration of solum moisture retention capacity of a soil, and refinement of depth to bedrock instill further advantages within HFUs in predicting solum moisture storage over soil series. In this sense, HFUs are describing the same attributes as traditional soil map units but at a more precise and quantitative scale.

Volumetric moisture patterns at 10, 20, 40, 60, 80 and 100 cm depths were analyzed with HFUs, soil series and LFUs over the same time period as solum moisture storage. From Figure 3.13, it is apparent that soil series and HFUs both predict volumetric moisture content with higher R^2 values more often than LFUs. The consideration of subsurface processes appears to aid the prediction of soil moisture at a point-depth scale over consideration of topography alone. LFUs do better at predicting moisture at the 80 cm depth than other depths; this may be realistic when considering how LFUs may well characterize areas of distinct moisture accumulation function in the landscape.

HFUs perform similarly to soil series at 10, 20, and 80 cm depths. HFUs do noticeably better than soil series overall at 40 and 60 cm depths, but soil series does better than HFUs at 100 cm depth. The lower sample size of sites with solums at 100 cm somewhat nullifies the advantage soil series has at this depth towards overall performance in predicting solum storage. HFUs increased performance at 40 and 60 cm depth may stem from differentiating sites by their solum's maximum moisture field capacity. A site's general soil structure and texture content, storage of organic matter and degree of soil weathering influences its solum saturated moisture storage.





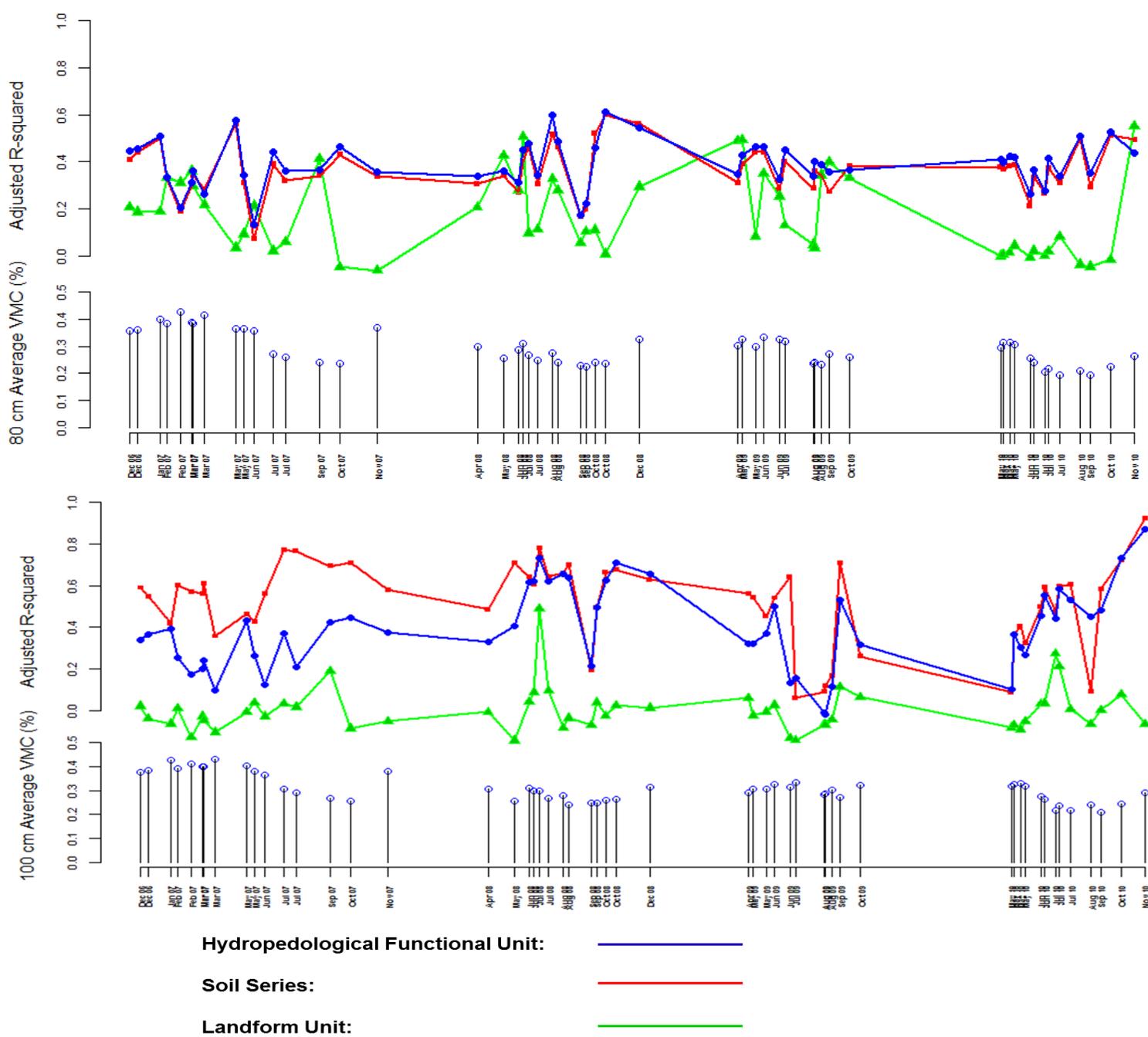


Figure 3.13: Adjusted R^2 diagnostic for linear modeling of volumetric moisture content at 10, 20, 40, 60, 80 and 100 cm depths with HFUs (blue), soil series (red), and LFUs (green) as predictors. Data were individually modeled for 58 dates from 2006 – 2010. The average volumetric moisture content in a specific depth from data collected at all sites in the catchment during a given date is shown on the bottom on each graph.

Results of a Tukey HSD multiple comparisons of average soil moisture solum storage across 2006 – 2010 act as a further assessment on the HFUs propensity to describe similar acting soil-hydrologic zones (Figure 3.14). It is clear that HFUs show four significantly distinct groups of average moisture storage over a 3 year period, while soil series and LFUs show 3. Soil series was more precise in showing 3 groups than LFUs, however. The SSLSFS HFU and Summit LFU clearly store the least amount of moisture as these units discharge precipitation inputs relatively effectively. The DSHSFV HFU and Ernest soil store significantly more moisture in their solums than the rest of the catchment. It is apparent that defining Ernest soil extent based upon effective soil depth and redoximorphic features, while also defining the DSHSFV unit extent based upon higher observed profile moisture storage and higher ECa values adjacent to the ephemeral stream have aided both soil series and HFUs in predicting soil moisture patterns.

The DSHSCH and ISMSCH HFUs are significantly different, while the Rushtown and Berks soil series are not differentiated significantly from average solum moisture storage. This implies HFUs have an advantage over soil series of expressing areas of distinct average solum moisture storage across the concave and convex hillslope areas. The ISMSCH and the SSLSPH units are also significantly differentiated, and these units are well correlated with solum moisture storage patterns as shown by the ANOVA in Table 3.6. The SSLSPH and SSLSFS units are not significantly different between each other, but do hold significantly less moisture than the other units. Both of these units fall within the Weikert soil series extent. Weikert soil was found to hold significantly less moisture than other soil.

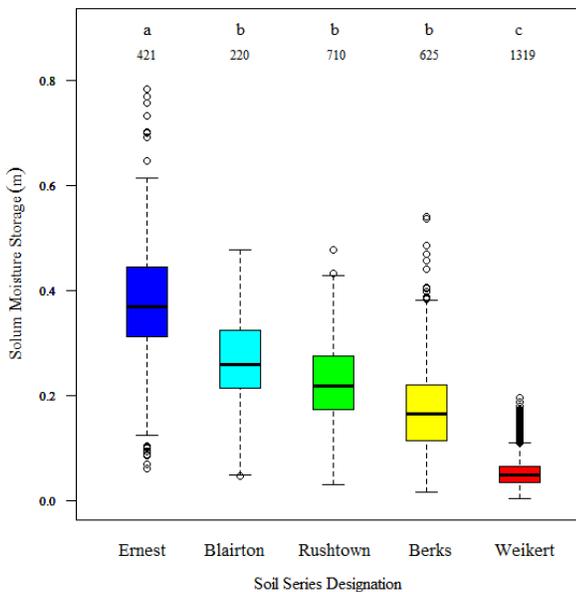
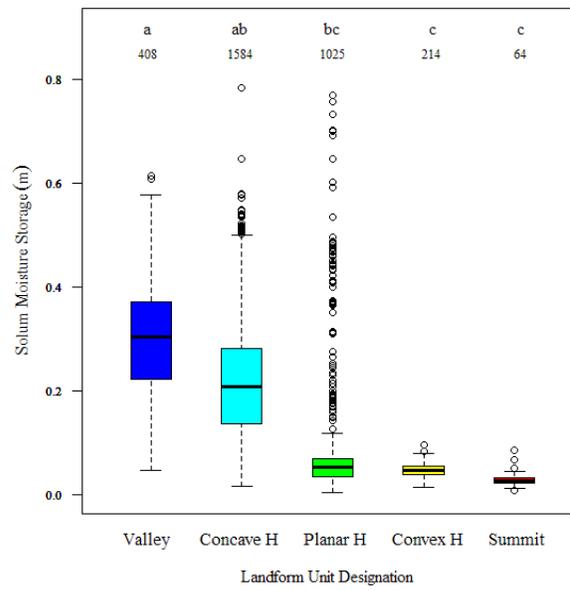
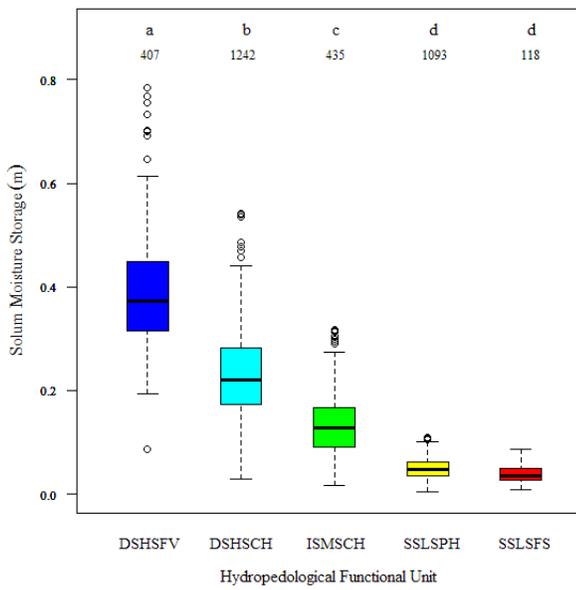


Figure 3.14: Boxplots of all solum moisture storage data points from 58 dates spanning 2006 – 2010 for individual categories in HFUs, soil series and LFUs. Numbers above the boxplot indicate sample size. Letters show significant difference between categories of average solum moisture storage collected from the same aforementioned 58 dates from all TDR-tensiometer sites in the catchment

HFUs do show a distinction of soil moisture patterns between units, but this study also determined whether they show a similar distinction with soil textural properties as soil series. Figure 3.15 show a series of boxplots expressing significant difference among the HFUs in total profile texture storage of sand, silt, clay and organic matter. Boxplots

for soil series and LFUs are also shown. No significant difference was found for rock content between any categories in each classification system.

From Figure 3.15, it is apparent that HFUs do show a distinction of soil profile texture storage for each class. HFUs display more significant differences in sand storage than soil series, however greater amounts of sand seem to lie in the Concave Hillslope locations and the Valley, implying that Rose Hill shale weathering processes may not follow a generally straightforward pattern. Silt Content is predictably lower in Weikert soil and the Shallow Soil, Low Storage HFUs, because silt displacement is related to the alluvial processes. Areas with greater moisture discharge tend to have less silt content, and areas of moisture accumulation tend to receive silt particles. Organic matter is somewhat consistent across all coverage schemes, with the valley soils and landform registering with significantly higher organic matter storage and the DSHSCH and DSHSFV HFUs also showing significantly higher organic matter storage.

The soil survey delineation does distinguish separate areas of clay storage more effectively than HFUs. The Berks soil series has a distinct intermediate storage of clay compared to Weikert soil and Rushtown, Blairton and Ernest. The HFU classification did not capture the subtle and significant difference in clay storage; however this is a difficult feat since clay does weather away from shale rock fragments across the entire catchment relatively well. Since HFUs are generally comparable with soil series in showing distinctions in texture properties among different units, it may be stated that the delineated HFUs express sub-catchment units that can characterize relatively distinct basic soil properties close or at the precision of a soil survey.

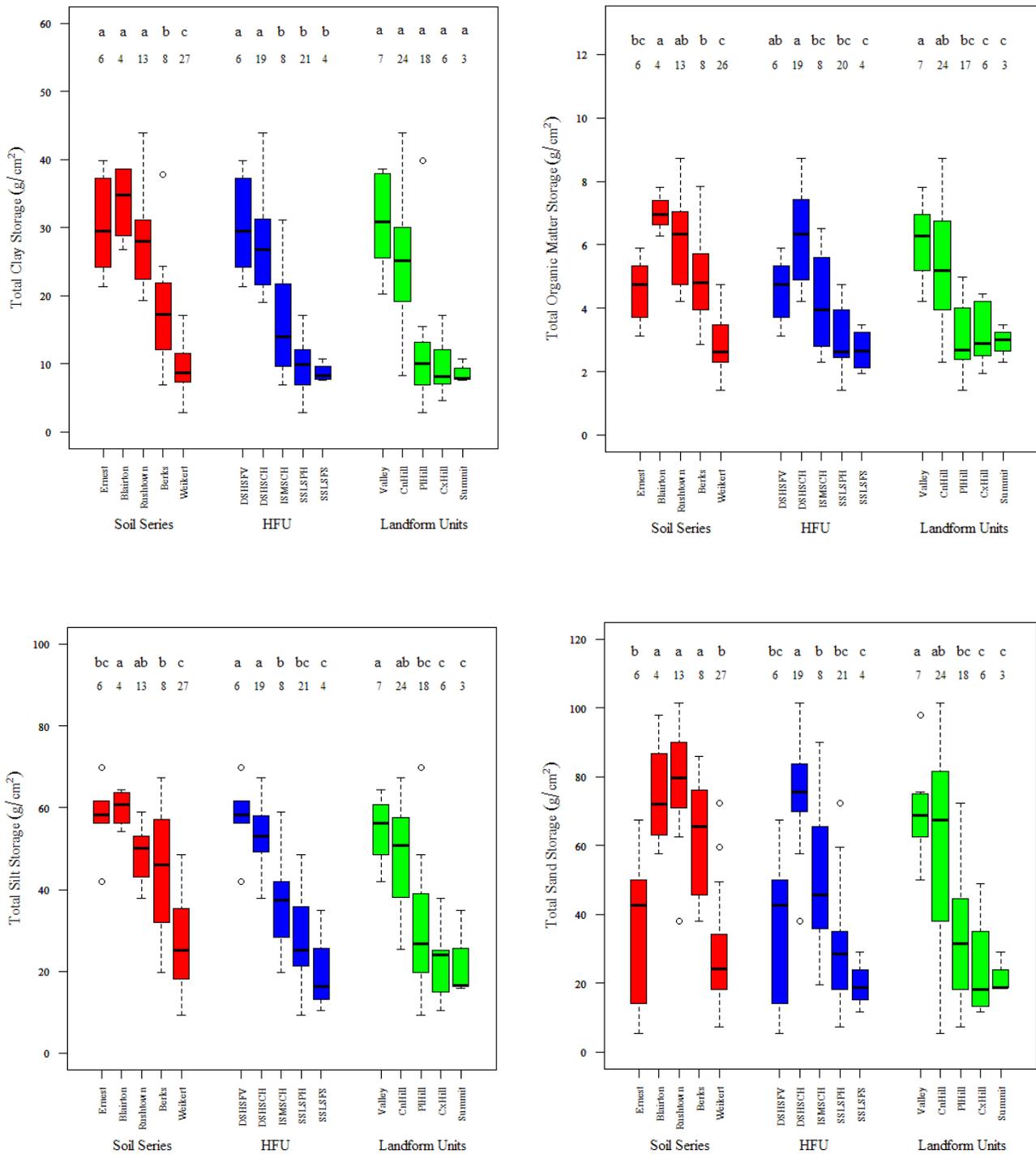


Figure 3.15: Boxplots of total profile texture storage of clay, silt, sand and organic matter (OM) collected from 58 different sites between categories of HFUs, soil series and LFUs. A Tukey HSD test was applied to find significant differences among categories for each texture class at $p < 0.05$. Numbers above boxplots indicate sample size.

The 3-dimensional map in Figure 3.16 of HFUs assists in intuitively grasping the HFUs inter-related soil-hydrologic function as sub-catchment units. The SSLSFS unit clearly acts mostly as a moisture dispersion zone, with soil moisture discharging from the shallow profiles of this unit towards the surrounding SSLSPH unit after precipitation events. Some areas of edge of the catchment are not SSLSFS, as these areas have an upslope contributing area higher than the acceptable threshold for the SSLSFS unit. The SSLSFS stores the least amount of moisture and has the least amount of silt, as much of the silt is discharged downslope out of the SSLSFS unit.

The SSLSPH unit directly receives moisture and material that is discharged from SSLSFS areas. The SSLSPH is comparable to Weikert soil in that it covers a broad expanse of the catchment and has predominately planar surface curvature. The soil in the SSLSPH unit is likely to be < 50 cm, owing to its generation partially from the refined depth to bedrock map. Besides having shallow soil, the SSLSPH holds the least amount of clay along with the SSLSFS, and this implies that little relatively little soil weathering has taken place in this unit. SSLSPH soil holds the least amount of moisture in the catchment, along with the SSLSFS unit.

The next unit after the SSLSPH in the catchment-wide gradient of moisture and material flux is the ISMSCH. These areas are generally convex and have over 50 cm deep soil thickness, which differentiates it from the SSLSPH. The best adjective for this unit is “transition”, as it is a transition between shallow, unstructured SSLSPH to deeper, more structured DSHSCH and DSHSFV units. Every topographic and soil property attribute about the ISMSCH is relatively intermediate besides mean slope value (Table 3.5). The ISMSCH has the highest mean slope in the catchment, while also having

intermediate soil thickness and θ_s storage. Steep slopes are present on swale side-slopes, where material that directly flushes off the SSLSPH towards the concave DSHSCH area is held. Pockets of ISMSCH do interrupt the SSLSPH, which correspond to scattered convexities amidst the SSLSPH, occurring either from tree-throw, tree stands or non-linear weathering processes. The soils in these small ISMSCH convexities tend to be deeper than 50 cm and hold relatively more moisture than the surrounding SSLSPH.

After moisture flows out of the ISMSCH, it typically flushes directly into the DSHSCH. The DSHSCH acts a channel of moisture and materials from the surrounding hillslope into much of the valley, and shares characteristics of both the DSHSFV and ISMSCH units. The DSHSCH unit is home to the deepest recorded soil in the catchment, and this soil is located at the bottom of the large swale in the south-facing hillslope, just north of the DSHSFV unit. This area acts as a bottleneck of material as matter moves down the Concave Hillslope unit towards the DSHSFV unit. The DSHSCH holds significantly less moisture storage than the DSHSFV and significantly more than the ISMSCH unit does. Rushtown soil and the Swale LFU are the closest categorical comparisons to the DSHSCH, however the DSHSCH unit is better correlated with solum storage over time than either of these other units (Table 3.6). The DSHSCH unit holds significantly higher silt content in its soil profile than other units along with the DSHSFV, and it holds the most overall sand content in within its soil relative to all other units in the catchment.

The DSHSCH unit also encompasses Blairton soil and low-elevation Rushtown that have < 0.185 m/m slope value, because the DSHSFV unit begins at the origin of the ephemeral stream. This low-laying area of the DSHSCH does not incur continuously

standing water over the wet season from the ephemeral stream, indicating that any soil morphological properties that result in annual solum-to-surface saturation in the DSHSFV are absent or function differently in the DSHSCH unit. Rushtown and Blairton soils both have higher standard errors of predicting multiple solum moisture storage values than the DSHSCH HFU indicating that the combination of Rushtown and Blairton soils within the DSHSCH unit does not majorly affect the DSHSCH unit's propensity to predict soil moisture patterns. As the DSHSCH unit contains deeper depth to bedrock on average, but holds significantly less observed total moisture storage than the DSHSFV unit, it is apparent that the soil in the DSHSFV unit functions differently hydrologically than soil in the low-elevation DSHSCH unit.

The final destination for moisture naturally inputted into the catchment is the DSHSFV unit, before ultimate exit at the catchment outlet. This unit is very well correlated with moisture storage response over time and holds a significantly higher amount of moisture storage, indicating that the DSHSFV is a truly distinct unit from a hydrologic process standpoint. DSHSFV soil has significantly less overall sand content in the solum than the DSHSCH unit. Both fragic properties and gleying were found in soil located within the DSHSFV unit during the soil survey, indicating that the substantial, solum-spanning saturation this soil undergoes every year has had morphological impacts over time. The DSHSFV has the highest consistent TWI than all other units, but has a lower mean soil thickness than the DSHSCH unit. This could be explained by the moisture flux that occurs in the DSHSFV, which may carry a disproportionate amount of soil material directly out of the catchment relative to material moving in from adjacent DSHSCH or ISMSCH areas. The significantly high clay and

silt content found in DSHSFV soil points to multiple soil accumulation processes occurring in the DSHSFV. Both continuous residual shale weathering and alluvial material flux occur simultaneously in this unit. The DSHSFV also has the highest significant Organic Matter storage in its soil, alluding to a concentrated deposition of leaves and other organic residue.

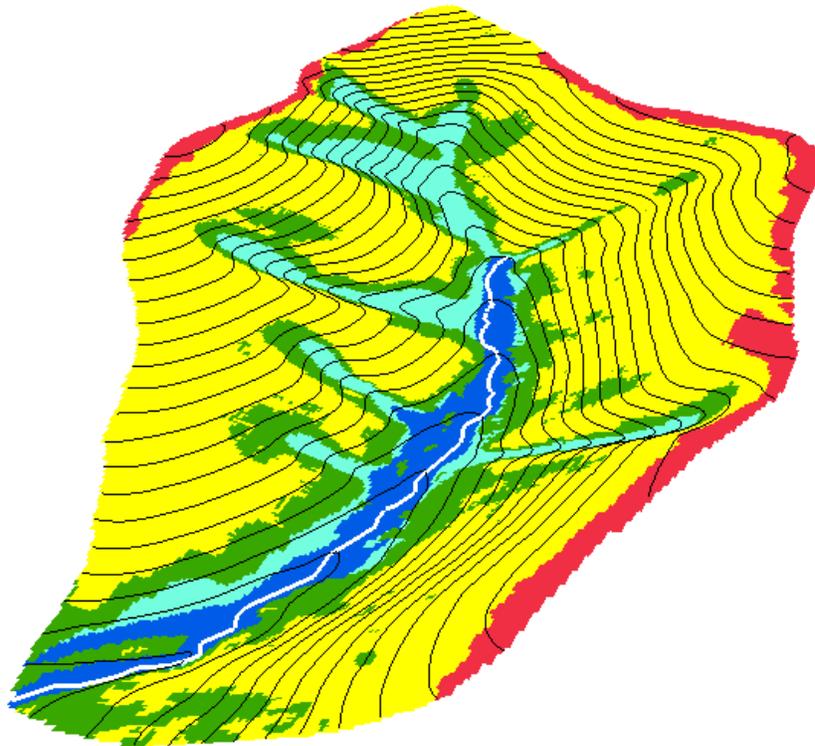


Figure 3.16: Three-dimensional rendition of Hydropedological Functional Units depicting areas of similar soil, terrain and hydrologic properties delineated for the Shale Hills. The stream is pictured, colored white.

Chapter 4: Summary and Future Work

4.1 Summary

Extensive soil moisture retention data collected across the Shale Hills catchment at multiple depths showed that soil water retention is mainly controlled by soil type and landscape position in the catchment. The ANOVA results and Tukey difference of means tests indicate that θ_s is significantly different between soil series and landform units at all depths. This stems from the physical influence of soil structure, soil pore space, and general landscape position on the depth-scale θ_s parameter. The n parameter is significantly different across separate soil series at the 20 cm depth, and this was deemed a function of the relative ease for moisture to effectively drain vertically or laterally from 20 cm depth in each soil. The α parameter is significantly different between landform units at 40 cm depth, which is related to pronounced differences of soil structure between landform units at 40 cm depth across the catchment. Topographic Wetness Index, depth to bedrock, and curvature are significant variables that control soil moisture retention parameters at the landscape-scale.

Catchment-wide spatial modeling of retention parameters revealed differences in ranges between three key van Genuchten parameters at the near-surface 20 cm depth and deeper subsurface at 80 cm depth. Moisture content at field saturation was found to be more variable across space in the deeper part of soil profile than in the near-surface because of soil structure and porosity differences among sites with deep θ_s measurements. The α and n parameters had a higher spatial variance at the near-surface. The higher spatial variance for α at the near-surface is related disparities of soil structure in the B

horizon among sites at 20 cm, while the higher spatial variance for n at the near-surface is generated by differences in effective soil depth between sites across the catchment, which controls effective vertical moisture release from the near surface.

Mapping soil moisture retention parameters was accomplished through the use of Bayesian regression kriging with Gaussian random field spatial modeling. Maps of continuously interpolated soil retention parameter values across the catchment can be utilized in hydrological and ecological modeling, and may inform future sampling locations for soil moisture retention data collection.

Through integrated analysis of topographic, soil depth, and landscape-scale soil water retention parameters, five HFUs were identified that exhibit contrasting landscape-soil-hydrology features. A sequence of 41 maps (from April 2008 to November 2010) of catchment-wide total soil moisture storage in solum showed clear pattern of the five HFUs.

The 5 rasterized HFUs showed distinctions in average total moisture storage and described soil moisture patterns slightly better than the detailed soil series map developed for the Shale Hills, but much better than the rasterized Landform Units (LFUs) delineated with Park and vande Giesen's (2004) method. The HFUs showed comparable distinctions in soil textural properties as the soil series map. According to an ANOVA test, HFUs were better correlated with total soil moisture storage as categories than soil series and LFUs across 31 dates from 2007 to 2010. The HFUs also performed better in predicting total moisture storage categorically than the soil series and LFUs for 54 separate dates from 2006 to 2010 according to diagnostics from linear modeling.

To obtain all topographic, soil property and soil water retention parameter maps for creating HFUs in a catchment, the following procedures must be undertaken:

- 1) Topographic data: LiDAR flyover for high resolution (1 x 1 meter) DEM
- 2) Depth to Bedrock data: Auguring campaign to adequately span an entire catchment or an extensive GPR survey campaign. Regression kriging should be used in interpolating depth to bedrock with topographic covariates obtained from 1
- 3) Soil water retention data (θ_s parameter): TDR-tensiometer sites to span a catchment Regression kriging should be used in interpolating soil water retention data with topographic covariates or depth to bedrock obtained from 1 and 2

The above procedures together are time-consuming and expensive. To characterize soil water retention relatively straightforwardly for another catchment or landscape, soil cores may be collected to obtain θ_s data in lab analysis or obtain cores for KuPF analysis.

Through this study it has been shown that catchment-wide soil water retention is possible to characterize, and catchment-wide soil water retention data may be used in tandem with topography and soil depth to delineate soil-terrain-hydrological functioning units that accurately represent general total moisture storage patterns when compared with a soil survey. This study should indicate that additional efforts could be directed to refine the delineation of sub-catchment soil-terrain-hydrologic units and assess how sub-catchment units may individually control higher scale processes, such as total catchment discharge and catchment-scale tree-water uptake.

This study is also serves as a positive reinforcement for methods used by soil scientists in segregating areas of similar soil functional characteristics in the landscape.

This research used digital topographic and soil property information to delineate HFUs, while soil scientists in the field also use topography and local soil properties to guide their understanding of how soil series are delineated in the landscape. Pedologists are also concerned with the functional relationship soil has with the landscape (Jenny, 1941), and the characterization of landscape-soil functional relationships is reliant upon topographic and soil property information used in this research. Both soil series and HFUs performed considerably well in predicting soil moisture patterns in Shale Hills. The computer delineation of HFUs closely resembled the 2nd order soil survey performed in the field.

4.2 Future Work

Saturated soil hydraulic conductivity (K_{sat}) is routinely obtained through lab analysis. As this represents one point on the curve, unsaturated hydraulic conductivity could render a more complete hydraulic conductivity curve up to potentials more negative than -850 mBar. This would allow the characterization of a soil's hydraulic conductivity to go beyond the upper limit for the tensiometers used in this study for the field monitoring of soil matric potential.

Unsaturated soil hydraulic parameters obtainable through devices such as KuPF may work in the same manner as K_{sat} or other soil hydraulic parameters investigated in this study. Parameters obtainable from KuPF analysis may also be used to refine HFUs delineated in this study. However, although KuPF analysis is theoretically able to provide data for a conductivity curve to very negative potentials and real θ_r conditions, it

is obtained through a soil core, and does not characterize data collected *in situ* in natural settings. Furthermore, edge effects and soil disturbance are potential issues that must be handled carefully during soil core collection. Therefore, *in situ* monitoring of soil matric potential beyond the upper limit for the tensiometers is highly desirable, which is currently on-going in the Shale Hills Critical Zone Observatory study.

References

- Baker, G., Steeples, D. and Feroci, M. (1997) "The time dependence of shallow reflection data", *The Leading Edge*, **16**, 1663-1666
- Baggaley, N., Mayr, T. and Bellamy, P. (2009) "Identification of key soil and terrain properties that influence the spatial variability of soil moisture throughout the growing season", *Soil Use and Management*, **25**, 262-273
- Böhner, J., McCloy, K.R., Strobl, J. [Eds.] (2006): *SAGA – Analysis and Modelling Applications*. Göttinger Geographische Abhandlungen, Vol.115
- Botros, F., Harter, T., Onsoy, Y., Tuli, A and Hopmans, J. (2009) "Spatial Variability of Hydraulic Properties and Sediment Characteristics in a Deep Alluvial Unsaturated Zone", *Vadose Zone Journal*, **8**, 276–289
- Behrens, T., Zhu, A-X., Schmidt, K., and Scholten, T. (2010) "Multi-scale digital terrain analysis and feature selection for digital soil mapping", *Geoderma*, **155**, Issues 3-4, 175-185
- Byrd, R., Lu, P. and Nocedal, J. (1995) "A limited memory algorithm for bound constrained optimization", *SIAM Journal on Scientific and Statistical Computing*, **16**, Issue 5, 1190-1208
- Campbell, Gaylon (1974) "A Simple Method for Determining Unsaturated Hydraulic Conductivity From Moisture Retention Data", *Soil Science*, **117**, Issue 6
- Dashtaki, S., Homaei, M. and Khodaverdiloo, H. (2010) "Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data", *Soil Use and Management*, **26**, Issue 1, 68-74
- Daniels, D.J. (2004) *Ground Penetrating Radar, 2nd Ed.*, Institution of Electrical Engineers, London. 726
- Diggle, P.J. and Ribeiro Jr, P.J. (2002) "Bayesian inference in Gaussian model-based geostatistics", *Geographical and Environmental Modelling*, **6**, No. 2, 129-146.
- Ding, C., and He, X. (2004) "K-means Clustering via Principal Component Analysis", *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada
- Duffy, C.J. (1996) "A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain", *Water Resources Research*, **32**, No. 8, 2421-2423
- Feike, L., Alves, W., van Genuchten, M.T. and Williams, J. (1996) "The UNSODA Unsaturated Soil Hydraulic Database." *U.S. Salinity Laboratory, USDA*. Riverside, California
- Fredlund, D.G. and Xing, A. (1994) "Equations for the Soil-Water Characteristic Curve", *Canadian Geotechnical Journal*, **31**, No 3

- Gardner, W.R., (1970) "Field measurement of soil water diffusivity", *Soil Sci. Soc. Am. Proc.*, **34**, 832-833
- Herbert, Gabriel (2005) "A Geophysical Investigation of Hydraulic Pathways at the Panola Mountain Research Watershed", Thesis, Georgia Institute of Technology
- Herbst, M., Diekkruger, B., Vereecken, H. (2006) "Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes", *Geoderma*, **132**, 206– 221
- Herbst, M. and Diekkruger, B (2002) "The influence of the spatial structure of soil properties on water balance modeling in a microscale catchment", *Physics and Chemistry of the Earth, Parts A/B/C*, **27**, Issues 9-10, 701-710
- Hodnett, M. and Tomasella, J. (2002) "Marked Differences Between van Genuchten Soil Water-Retention Parameters for Temperate and Tropical Soils: A New Water-Retention Pede-Transfer Functions Developed for Tropical Soils", *Geoderma*, **108**, 155-180
- Hollenbeck, K and Jensen, K. (1998) "Maximum Likelihood Estimation of Unsaturated Hydraulic Parameters", *Journal of Hydrology*, **210**, 192-205
- IMKO Micromoduletechnik, <http://www.imko.de/en/home-navbar>
- Inoue, M., Šimunek, J., Hopmans, J. and Clausnitzer, V. (1998) "In situ estimation of soil hydraulic functions using a multistep soil-water extraction technique", *Water Resources Research*, **34**, Issue 5, 1035–1050
- Jabro, J., Evans, R., Kim, Y. and Iversen, W. (2009) "Estimating in situ soil-water retention and field water capacity in two contrasting soil textures", *Irrigation Science*, **27**, Issue 3, 223-229
- Jenny, H., (1941) *Factors of Soil Formation: A System of Quantitative Pedology*, Dover Publications, Inc., New York
- Johnson, E and Kiyoko, M. (2008) "Testing the assumptions of chronosequences in succession", *Ecology Letters*, **11**, 419-431
- Kool, J. and Parker, J. (1988) "Analysis of the inverse problem for transient unsaturated flow", *Water Resources Research*, **24**, Issue 6, 817-830
- Kosugi, K., Hopmans, J. and Dane, J. (2002) "Soil Solution Phase: Parametric Models", *Methods of Soil Analysis, Part 4: Physical Methods*, 739 - 758
- Kumar, M. and Duffy, C. (2009) "Detecting hydroclimatic change using spatio-temporal analysis of time series in Colorado River Basin", *Journal of Hydrology*, **374**, 1-15
- Lin, H. (2010) "Earth's Critical Zone and Hydropedology: Concepts, Characteristics, and Advances", *Hydrology and Earth System Sciences*. **14**

- Lin, H., Brooks, E., McDaniel, P., and Boll, J. (2008) “Hydropedology and Surface/Subsurface Runoff Processes”, *Encyclopedia of Hydrologic Sciences*
- Lin, H. (2006) “Temporal Stability of Soil Moisture Spatial Pattern and Subsurface Preferential Flow Pathways in the Shale Hills Catchment”, *Vadose Zone Journal*, **5**, Issue 1, 317-340
- Lin, H., Kogelmann, W., Walker, C. and Bruns, M. (2006) “Soil moisture patterns in a forested catchment: A hydropedological perspective”, *Geoderma*, **131**, 345-368
- Lin, H., Bouma, J., Pachepsky, Y., Western, A., Thompson, J., Genuchten, R., Vogel, H., and Lilly, A. (2006b) “Hydropedology: Synergistic Integration of Pedology and Hydrology”, *Water Resources Research*, **42**, 1-13
- Lin, H., Bouma, J., Wilding, L., Richardson, J., Kutilek, M., and Nielsen, D. (2005) “Advances in Hydropedology”, *Advances in Agronomy*, **85**, 1 - 89
- Lin, H., McInnes, K., Wilding, L. and Hallmark, C. (1998) “Macroporosity and Initial Moisture Effects on Infiltration Rates in Vertisols and Vertic Intergrades”, *Soil Science*, **163**, Issue 1, 2-8
- Mardia, K., Kent, J., and Bibby, J. (1979) *Multivariate Analysis*, London: Academic Press.
- Mayr, T. and Jarvis, N. (1999) “Pedotransfer functions to estimate soil water retention parameters for a modified Brooks–Corey type model”, *Geoderma*, **91**, Issues 1-2, 1-9
- McBratney, A., Mendonca Santos, M. and Minasny B. (2003) “On digital soil mapping”, *Geoderma*, **117**, 3-52
- McDonnell, Jeffrey (1990) “A Rationale for Old Water Discharge Through Macropores in a Steep, Humid Catchment”, *Water Resources Research*, **26**, Issue 11, 2821-2832
- Minasny, B., McBratney, A. and Salvador-Blanes, S (2008) “Quantitative models for pedogenesis — A review”, *Geoderma*, **144**, 140-157
- Minasny, B. and McBratney, A. (2007) “Spatial prediction of soil properties using EBLUP with the Matérn covariance function”, *Geoderma*, **140**, 324–336
- Moller, M., Volk, M., Friedrich, K., and Lymburner, L. (2008) “Placing soil genesis and transport processes in a landscape context: A multi-scale terrain analysis approach.” *Journal of Plant Nutrition and Soil Science*, **171**, 419-430
- National Research Council [NRC] (2001) “Basic research opportunities in earth science”, *National Academy Press*, Washington DC, USA
- Natural Resources Conservation Service, *Ground Penetrating Radar Suitability Maps*, <http://soils.usda.gov/survey/geography/maps/GPR/index.html>

- Pal, N., Bezdek, J., and Hathaway, R. (1996). Sequential competitive learning and the fuzzy c-means clustering algorithms. *Neural Networks*, **9**, Issue 5, 787–796
- Pachepsky, Y., Rawls, W. and Lin, H. (2006) “Hydropedology and Pedotransfer Functions”, *Geoderma*, **131**, 308-316
- Park, S. and vande Giesen, N. (2004) “Soil-landscape delineations to define spatial sampling domains for hillslope hydrology,” *Journal of Hydrology*, **295**, 28-46
- Pebesma, E.J., (2004) “Multivariable geostatistics in S: the gstat package”, *Computers & Geosciences*, **30**, 683-691
- Porebska, D., Slawinski, C., Lamorski, K., and Walczak, R. (2006) “Relationship Between van Genuchten’s Parameters of the Retention Curve Equation and Physical Properties of Soil Solid Phase”, *International Agrophysics*, **20**, 153-159
- Rasmussen, C., Southard, R., and Horwath, W. (2005) “Modeling Energy Inputs to Predict Pedogenic Environments using Regional Environmental Databases”, *Soil Science Society of America Journal*, **69**, 1266 - 1274
- Ribeiro, P. and Diggle, P. (2001) “geoR: a package for geostatistical analysis”, *R-NEWS*, **1**, Issues 1609-3631, 15-18, <http://cran.R-project.org/doc/Rnews>
- Romano, N. and Palladino, M. (2002) “Prediction of soil water retention using soil physical data and terrain attributes”, *Journal of Hydrology*, **265**, Issues 1-4, 56-75
- Romano, N. and Santini, A. (1997) “Effectiveness of using pedo-transfer functions to quantify the spatial variability of soil water retention characteristics”, *Journal of Hydrology*, **202**, Issue 1-4, 137-157
- Seyfried, M., Schwinning, S., Walvoord, M., Pockman, W., Newman, B., Jackson, R., and Phillips F. (2005) “Ecohydrological Control of Deep Drainage in Arid and Semiarid Regions”, *Ecology*, **86**, Issue 2, 277-287
- Shao, M. and Horten, R. (1998) “Integral Method for Analyzing Soil Hydraulic Properties”, *Soil Science Society of America Journal*, **62**, Issue 3, 585-592
- Sommer, M., and Schlichting, E. (1997) “Archetypes of catenas in respect to matter a concept for structuring and grouping catenas”, *Geoderma*, **76**, 1-33
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at <http://websoilsurvey.nrcs.usda.gov/>
- Takagi, Kenneth (2009) “Static and dynamic controls on soil moisture spatial variability in the Shale Hills catchment”, *Masters Sc Thesis*, Pennsylvania State University
- Tarboton, D. (1997) “A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models,” *Water Resources Research*, **33**, No. 2, 309-319

- Thompson, S., Harman, C., Troch, P., Brooks, P., and Silvapalan, M. (2011) "Spatial scale dependence of ecohydrologically mediated water balance partitioning: A synthesis framework for catchment ecohydrology", *Water Resources Research*, **47**, doi:10.1029/2010WR009998
- Van Genuchten, M. Th., (1980) "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils", *SSSAJ*, **44**, 892-898.
- Venables, W. N. and B. D. Ripley (2002) *Modern Applied Statistics with S*, Springer-Verlag
- Vereecken, H., Kasteel, R., Vanderborght, J. and Harter, T. (2007) "Upscaling Hydraulic Properties and Soil Water Flow Processes in Heterogeneous Soils: A Review", *Vadose Zone Journal*, **6**, 1 - 28
- Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. and van Genuchten, M.Th. (2010) "Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review", *Vadose Zone Journal*, **9**, Issue 4, 795 - 820
- Winter, T. (2001) "The Concept of Hydrologic Landscapes", *American Water Resources Association*, **37**, Issue 2, 335 - 349
- Yang, L., Jiao, Y., Fahmy, S., Zhu, A-X., Hann, S., Burt, J., and Qi, F. (2011) "Updating Conventional Soil Maps through Digital Soil Mapping", *SSSAJ*, **75**, No. 3, 1044 - 1053
- Young, M. and Sisson, J. (2002) "Soil Solution Phase: Tensiometry", *Methods of Soil Analysis, Part 4: Physical Methods*, 575 - 608
- Zhou, X., Lin, H. and White, E. (2008) "Surface soil hydraulic properties in four soil series under different land uses and their temporal changes", *Catena*, **73**, Issue 2, 180-188
- Zhu, Q., Lin, H., and Doolittle, J. (2010) "Repeated Electromagnetic Induction Surveys for Determining Subsurface Hydrologic Dynamics in an Agricultural Landscape", *Soil Science Society of America Journal*, **74**, No. 5, 1750 - 1762
- Zhu, Q., Lin, H., and Doolittle, J. (2010b) "Repeated Electromagnetic Induction Surveys for Improved Soil Mapping in an Agricultural Landscape", *SSSAJ*, **74**, 1763–1774
- Zhu, Q. and Lin, H. (2009) "Combining Sample Size, Spatial Structure, and Auxiliary Variables to Determine Optimal Kriging for Soil Properties in Contrasting Landscapes", *Hydrology and Earth System Sciences*, **13**, 1503–1518
- Zhu, Q. and Lin, H. (2009) "Simulation and validation of concentrated subsurface lateral flow paths in an agricultural landscape", *Hydrology and Earth Systems Sciences*, **13**, 1503-1518

Appendix A: Maximum Likelihood Optimization Procedure

The optimization of van Genuchten parameters was performed using a Maximum Likelihood Estimation (MLE) algorithm in R (package: bbmle). The standard method of optimizing van Genuchten parameters to fit $\theta(\psi_m)$ data in hydrologic literature is the Levenberg-Marquardt least-squares error algorithm. This algorithm has been implemented in software ‘RETC’ for straight-forward optimization of van Genuchten curve-fitting problems (van Genuchten, 1991). MLE is another optimization procedure in hydrologic research. MLE optimization has considerable advantages over simple least squares estimation methods for data with substantial uncertainty derived from measurement error and natural conditions, such as the manually collected $\theta(\psi_m)$ data for Shale Hills (Hollenbeck and Jensen, 1998).

Hollenbeck and Jensen (1998) did a series of water retention and outflow experiments, while giving details about optimizing van Genuchten parameters with MLE. The following are advantages of using MLE optimization:

- Uncertainty in the data, obtained from the distribution of **observed data**, is used during the optimization process to define the parameter space of parameters being optimized (data distribution informs parameter estimation, not arbitrary weights)
- With sufficiently wide bounds, parameters will converge upon a “global” minimum in the objective function. This indicates that additional data samples will not alter the parameter estimate.
- A Hessian matrix is computed from the sensitivity of the model to the parameters. Inversion of the Hessian matrix is a diagnostic that the estimated parameter has converged upon a finite space
- An parameter confidence region may be calculated from the inverse of the Hessian matrix as an ellipsoid:

$$(\mathbf{p} - \mathbf{p}^*)' \sum_{\mathbf{p}}^{-1} (\mathbf{p} - \mathbf{p}^*) = \Delta O(\mathbf{p})_{CR}$$

Where \mathbf{p} is a parameter estimate, \mathbf{p}^* is the optimized parameter and $\Delta O(\mathbf{p})_{CR}$ is the calculated parameter region (region) based upon the *absolute* difference of the change in the Objective function to the minimum of the Objective function. This confidence region calculation works for models that are characteristically linear; however, Hollenbeck and Jensen (1998) found that the non-linear van Genuchten model will behave locally linear, meaning that the ellipsoid calculation still gives a good estimate of the confidence region. The parameter samples in the calculated confidence region may be used as a prior probability distribution for a Bayesian model.

The retention data for all depths at 61 sites (232 total depths) across Shale Hills was fitted with a van Genuchten model using MLE optimization of θ_s , θ_r , α and n . In all 232 locations, the parameters **converged** and the Hessian matrices **could be inverted**. All sites were modeled using the same starting conditions and parameter bounds:

1) Density function: Moisture was chosen to be predicted with the model, because well temporally-spanned volumetric moisture data has been empirically shown to be typically normally distributed (Fredlund and Xing, 1994). Histograms were plotted for observed Volumetric Moisture Content at each location, and normal distribution functions were successfully fitted for each location. The density function for the predicted Volumetric Moisture data

2) Negative log-likelihood formula: The formula used for Volumetric Moisture Content

prediction is van Genuchten's equation: $\theta = \theta_r + \frac{(\theta_s - \theta_r)}{[1 + (\alpha \psi_m)^n]^m}$, where $m = 1 - \frac{1}{n}$

3) MLE optimization method: L-BFGS; a quasi-Newtonian method modified by Byrd et al (1995), which allows box constraints. The constraints were used, as the α and n parameters can be extremely sensitive and must have constraints imposed for effective global convergence. The bounds were chosen to be wide, but not too wide to make convergence impossible (see below). θ_s must also be constrained to not be lower than the highest moisture value, which would cause model failure, and θ_r must not be higher than

the lowest moisture value for the same reason. It also does not make physical sense for θ_s to be lower than the highest moisture point and θ_r to be higher than the lowest moisture point. Since it is quasi-Newtonian (or a variable metric algorithm), the method uses function values and gradients to “construct” a quantitative image of the surface to be optimized.

4) Upper bounds:

$\theta_s = 1.0$ (indicating 100% space filled with moisture)
 $\theta_r =$ minimum observed moisture value
 $\alpha = 1 \text{ cm}^{-1}$
 $n = 10$

5) Lower bounds:

$\theta_s =$ maximum observed moisture value
 $\theta_r = 0$
 $\alpha = 0.001 \text{ cm}^{-1}$
 $n = 1.01$

6) Starting value (~standard initial values for silt loam in RETC):

$\theta_s =$ maximum observed moisture value + 0.05
 $\theta_r = 0.05$
 $\alpha = 0.02 \text{ cm}^{-1}$
 $n = 1.25$

These upper and lower bounds and initial values provided the best curve fits and allowed inversion of Hessian matrices for **all depths**. Many other parameter ranges were tested with no convergence or non-invertable Hessian matrices for a number of depths.

Therefore, the bounds and initial values indicated represent the most **robust** and **effective** starting conditions for MLE optimization of retention parameters for Shale Hills soils.

Appendix B: Van Genuchten Soil Water Retention Parameter Maps

During the study of catchment-scale soil water retention characteristics at Shale Hills, maps of van Genuchten soil water retention parameters were created for the catchment at 10, 20, 40, 80 and 100 cm depth locations. Bayesian regression kriging and Bayesian kriging with a constant trend were used for the interpolation of each soil water retention parameter across Shale Hills at each depth. Catchment-wide characterization of soil water retention properties for different depths may be used to inform or refine hydrologic and ecologic models designed for processes controlled in part by soil water retention, such as catchment discharge or the intake of soil water by plants.

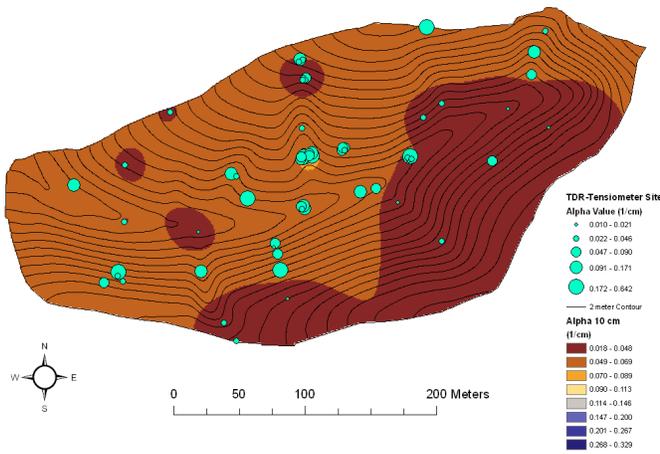
Bayesian kriging is advantageous in estimating data with a relatively weak spatial structure, since it uses the probability distribution of observed data to optimize spatial model parameters and takes into account spatial model parameters' uncertainty (Diggle and Ribeiro, 2002). Posterior probability distributions are available for each spatial model parameter after Bayesian kriging. As α and n exhibited complicated spatial structure for individual depth locations, Bayesian kriging was well suited for estimating these parameters and θ_s , even though θ_s showed relatively strong spatial structure and was correlated with topographic indices at each depth location.

The following steps outline the fundamental process of Bayesian kriging:

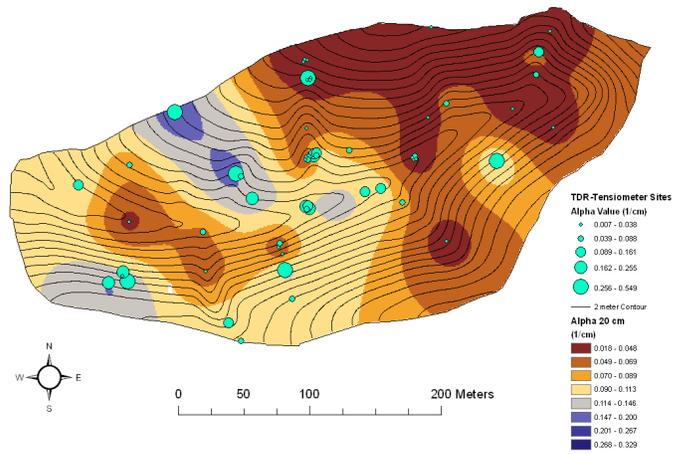
- 1) Test for significant relationship between topographic variables and van Genuchten soil water retention parameter using a Stepwise Regression. If significant relationships exist, use correlated topographic variables in regression kriging. If no significant relationships exist, use kriging with a constant trend
- 2) Utilize a Gaussian random field spatial model with a Matern covariance function to define spatial model parameters (further details in Section 2.6). Prior estimates of spatial model parameters are obtained from using a maximum likelihood algorithm
- 3) Apply a validation diagnostic test to assess model performance with prior spatial parameters in predicting soil water retention parameters at observed locations.

- 4) Apply a cross-validation test to determine any auto-correlation patterns in preliminary spatial model residuals and check for normality in prediction distribution (necessary conditions for isotropy)
- 5) Input preliminary spatial parameters obtained by maximum likelihood as priors into a Bayesian kriging procedure. This procedure optimizes spatial parameters based upon their ability to predict values lying within the observation value's probability distribution. The upper and lower bounds ascribed to the range of possible spatial parameter values are based upon uncertainty ascribed to each spatial parameter
- 6) Use Bayesian optimized spatial parameters (mean or mode of a parameter's posterior probability distribution) to predict soil water retention parameters across the catchment

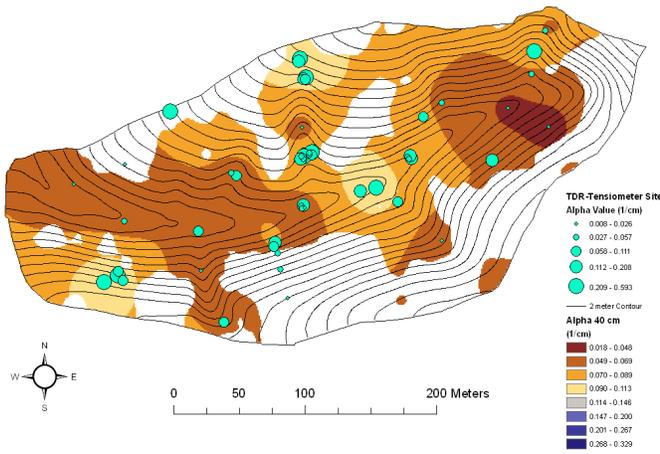
Alpha Value at 10 cm



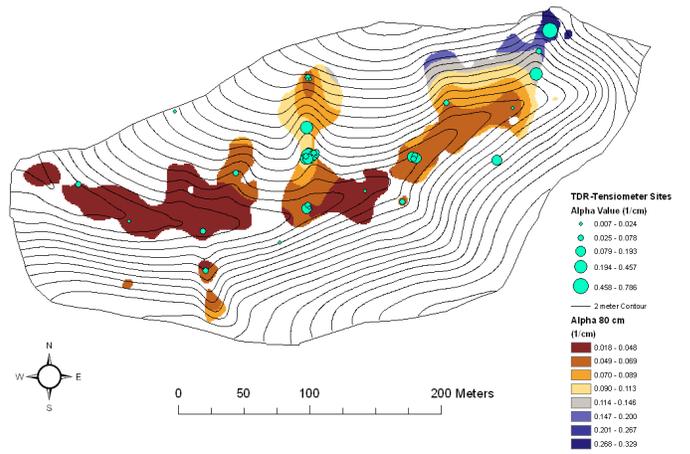
Alpha Value at 20 cm



Alpha Value at 40 cm



Alpha Value at 80 cm



Alpha Value at 100 cm

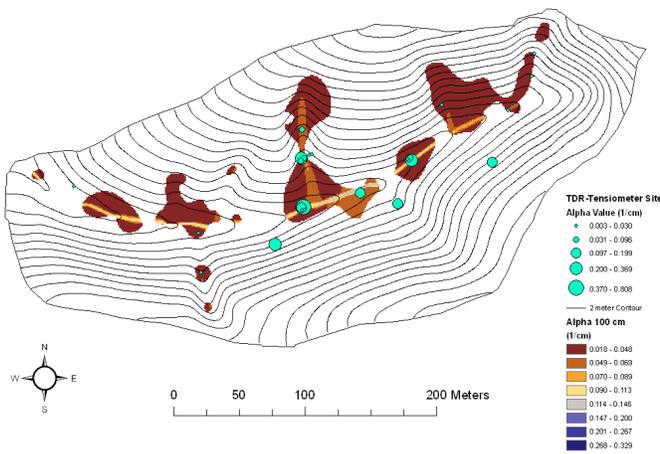


Figure B.1: Series of maps depicting the α soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate α values. Areas with depth to bedrock less than featured depth have been masked for each map.

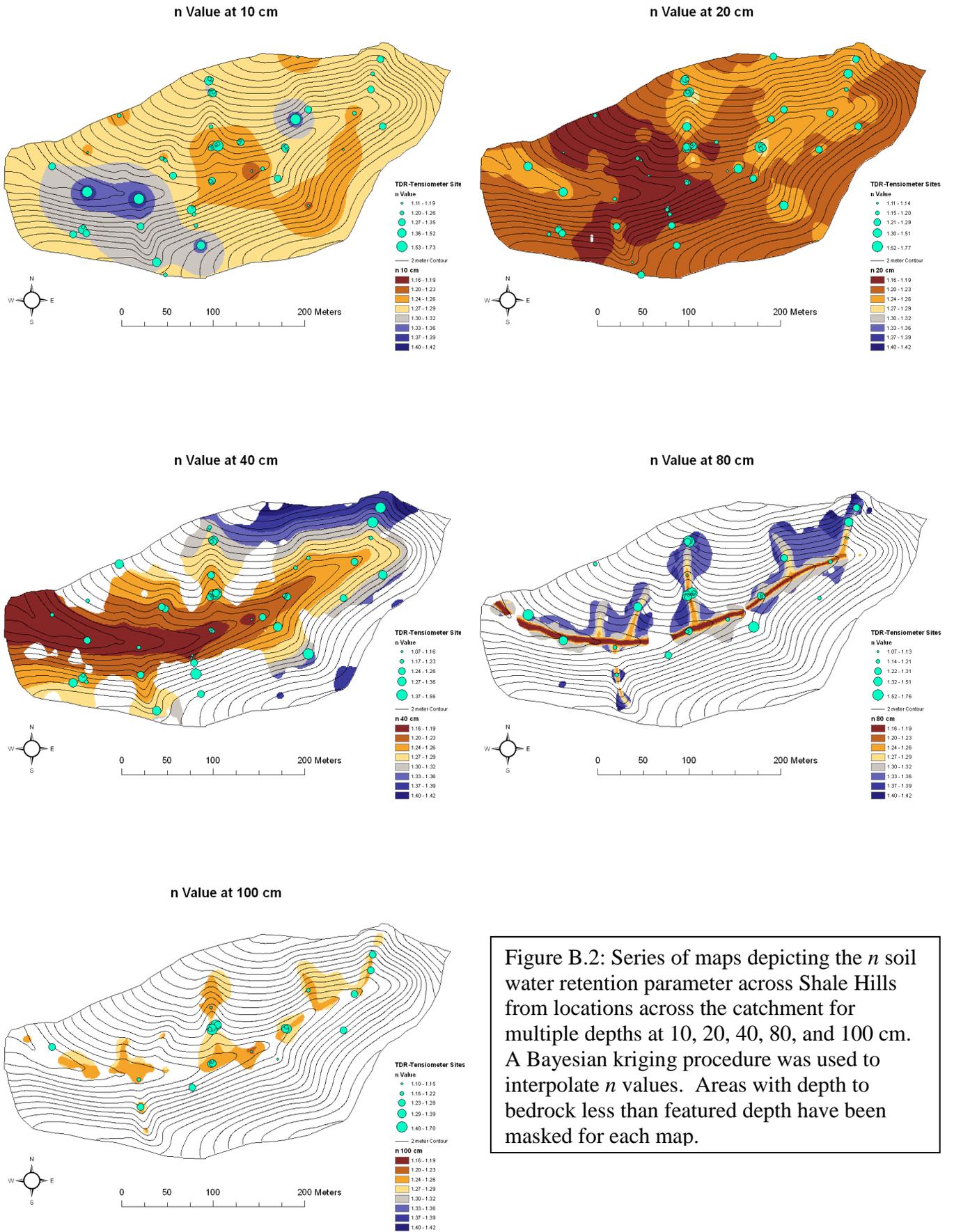
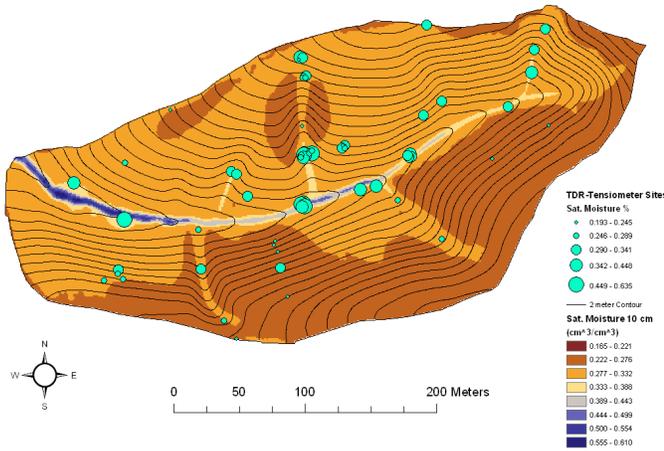
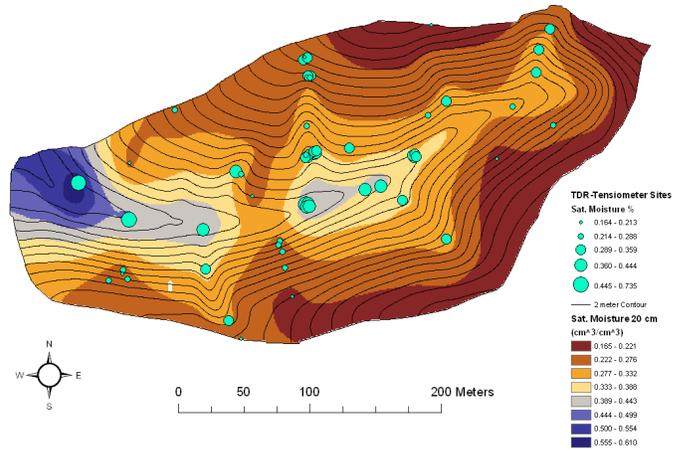


Figure B.2: Series of maps depicting the n soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate n values. Areas with depth to bedrock less than featured depth have been masked for each map.

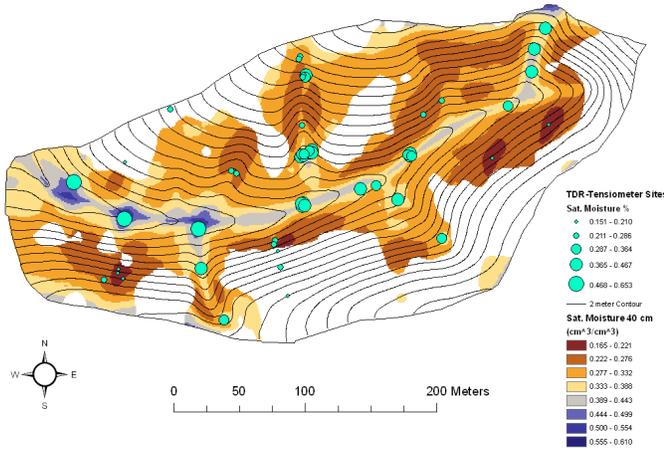
Saturated Moisture Content at 10 cm



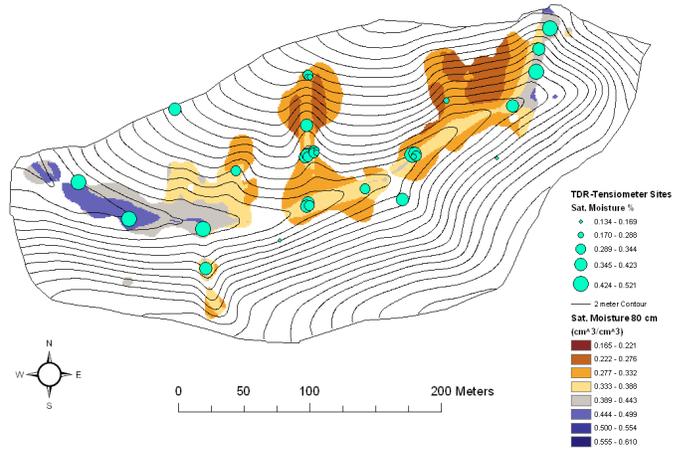
Saturated Moisture Content at 20 cm



Saturated Moisture Content at 40 cm



Saturated Moisture Content at 80 cm



Saturated Moisture Content at 100 cm

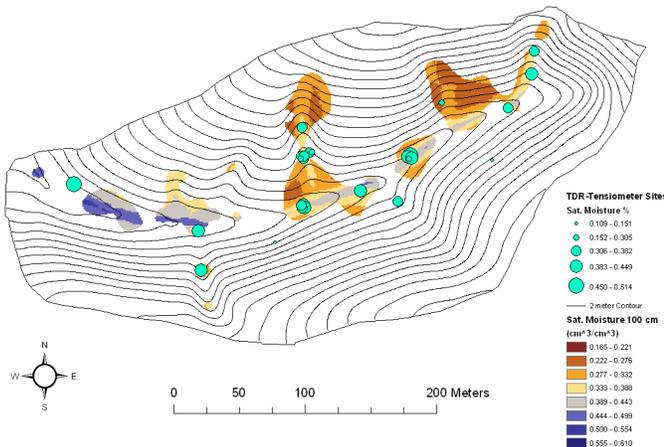


Figure B.3: Series of maps depicting the θ_s soil water retention parameter across Shale Hills from locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. A Bayesian kriging procedure was used to interpolate θ_s values. Areas with depth to bedrock less than featured depth have been masked for each map.

Appendix C: Texture Maps

Texture content is a basic soil property, indicating the ratio of the mass of a soil textural class to the mass of soil within a given volume. Soil texture classes analyzed at Shale Hills include clay, silt, sand, rock fragments and organic matter (OM). Texture content data was obtained for A, B, and C horizons at 58 TDR site locations across the catchment. As the thickness and boundaries for A, B, and C horizons are not consistent across sites at Shale Hills, texture content data was reorganized by recording texture values present at available 10, 20, 40, 80 and 100 cm depth locations for all 58 sites.

Total storage (g/cm^2) of each texture class within the soil profile at all 58 sites was also calculated. The following procedure yields total texture storage for a soil profile:

- 1) Calculate texture storage for each horizon at a given site

$$HS = TC\left(\frac{\text{g}}{\text{g}}\right) * HT(\text{cm}) * \rho_d\left(\frac{\text{g}}{\text{cm}^3}\right)$$

where HS is horizon storage, TC is texture content at a horizon, HT is horizon thickness, and ρ_d is bulk density of the soil for a given site. Bulk density data for each soil series was acquired from Lin, 2006.

- 2) Calculate total texture storage for the soil profile at a given site

$$TS = HS_A + HS_B + HS_C$$

where TS is total texture storage for a site and $HS_{A,B,C}$ is horizon storage at available A, B or C horizons at a site.

With texture storage data and texture content data available at sites spanning the catchment, spatial statistics were applied to estimate texture content values at five depths and texture storage values across Shale Hills allowing the subsequent creation maps of texture data. Regression kriging and kriging with a constant trend were used to interpolate texture values over the catchment, as kriging provides a smooth, but accurate, estimation of values across space. If topographic variables were correlated with texture values, regression kriging was performed. If no topographic variables were correlated with texture, then kriging with a constant trend was performed using maximum likelihood optimization. All geostatistical analysis for texture data was performed in geoR (Ribeiro, and Diggle, 2001).

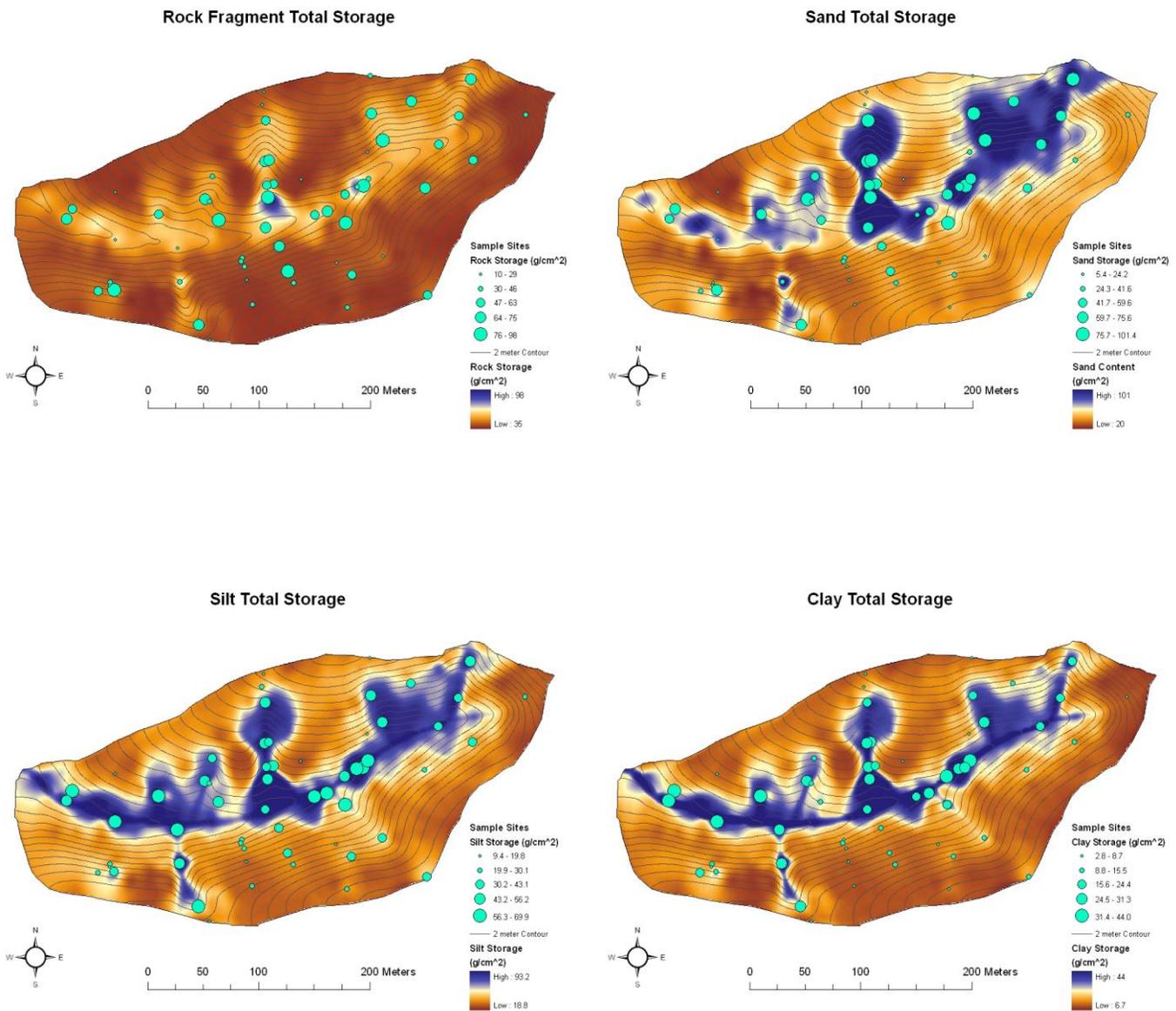


Figure C.1: Maps of total rock fragment, clay, silt and sand storage generated from regression kriging analysis on respective total rock fragment, clay, silt and sand storage values at 58 sites across Shale Hills. Although validation diagnostics showed R^2 values > 0.830 for all spatial models, nugget and trend effects may cause estimated data values to be different from observed data at some site locations

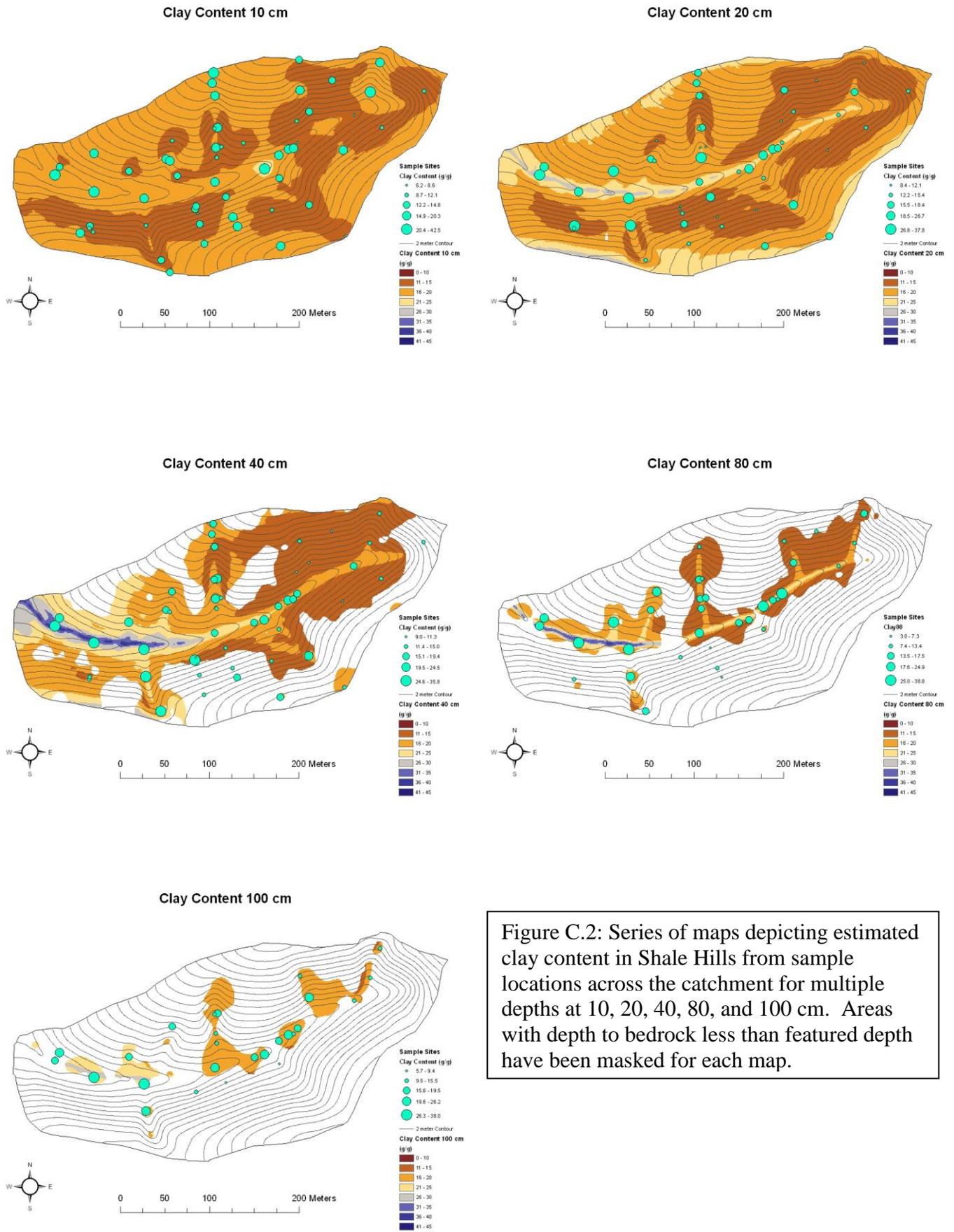


Figure C.2: Series of maps depicting estimated clay content in Shale Hills from sample locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. Areas with depth to bedrock less than featured depth have been masked for each map.

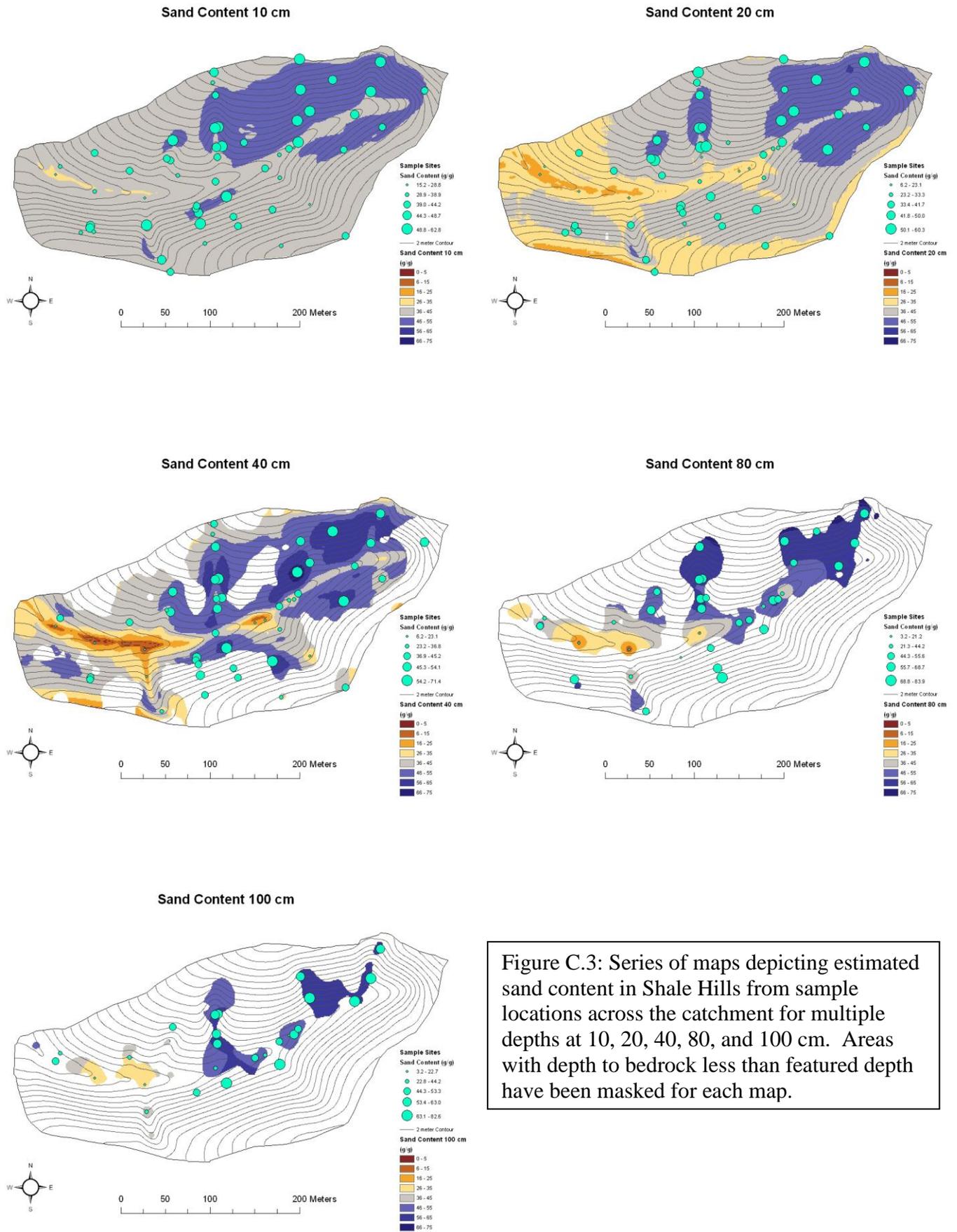


Figure C.3: Series of maps depicting estimated sand content in Shale Hills from sample locations across the catchment for multiple depths at 10, 20, 40, 80, and 100 cm. Areas with depth to bedrock less than featured depth have been masked for each map.

Appendix D: Additional Tables and Correlation Matrices

Throughout the course of this research, additional information about soil water retention parameters was recorded. The following are a collection of tables and correlation matrices representing this information.

Table D.1: Summary of alpha values for each soil series and landform unit separated by depth location. Letters beside the means indicate a significant difference of alpha values among depths within a landform unit or soil series according to a Tukey HSD test with a p-value < 0.05

Alpha Values for different Depths within Soil Series

Weikert					Berks					Rushtown				
Depth (cm)	SE	Alpha Mean		Sampel Size	Depth (cm)	SE	Alpha Mean		Sampel Size	Depth (cm)	SE	Alpha Mean		Sampel Size
10	0.202	0.055	a	24	10	0.261	0.048	ab	13	10	0.269	0.068	a	12
20	0.283	0.057	a	21	20	0.271	0.075	a	13	20	0.315	0.040	a	12
40	0.315	0.076	a	16	40	0.212	0.078	a	13	40	0.359	0.051	a	12
80	0.688	0.039	a	3	80	0.187	0.043	ab	9	80	0.363	0.123	a	12
100	0.071	0.213	a	2	100	0.818	0.014	b	4	100	0.274	0.043	a	10

Ernest					Blairton				
Depth (cm)	SE	Alpha Mean		Sampel Size	Depth (cm)	SE	Alpha Mean		Sampel Size
10	0.313	0.058	a	8	10	0.593	0.054	a	4
20	0.323	0.078	a	8	20	0.327	0.026	a	4
40	0.382	0.071	a	8	40	0.375	0.066	a	4
80	0.513	0.024	a	6	80	0.245	0.076	a	4
100	0.765	0.060	a	6	100	0.813	0.045	a	4

Alpha Values for different Depths within Landform Units

Summit					Hillslope					Swale				
Depth (cm)	SE	Alpha Mean		Sampel Size	Depth (cm)	SE	Alpha Mean		Sampel Size	Depth (cm)	SE	Alpha Mean		Sampel Size
10	0.385	0.063	a	6	10	0.231	0.051	a	19	10	0.203	0.059	a	22
20	0.606	0.027	a	6	20	0.282	0.072	a	16	20	0.225	0.059	a	22
40	0.275	0.241	a	3	40	0.320	0.061	a	14	40	0.203	0.073	a	22
80	NA	0.018	a	1	80	0.968	0.058	a	2	80	0.256	0.088	a	19
					100	0.071	0.213	a	2	100	0.292	0.041	a	12

Valley				
Depth (cm)	SE	Alpha Mean		Sampel Size
10	0.246	0.055	a	14
20	0.269	0.053	a	14
40	0.290	0.054	a	14
80	0.308	0.036	a	12
100	0.510	0.037	a	12

Table D.2.(a-e): Correlation matrices of soil water retention parameters (cyan) with topography (green) and soil properties (orange) at 5 different depths. A Spearman correlation was used to test significance. Values in red are significant at $p < 0.05$, with the p-value in parantheses next to the correlation coefficient

a

Parameters 10 cm	α	n	θ_s	a	ψ_e	- b
α	1					
n	-0.44 (<0.001)	1				
θ_s	0.50 (<0.001)	0	1			
a	-0.26 (0.040)	0.64 (<0.001)	0.24	1		
ψ_e	-0.91 (<0.001)	0.40 (0.001)	-0.47 (<0.001)	0.42 (<0.001)	1	
- b	-0.28 (0.030)	0.68 (<0.001)	-0.06	0.90 (<0.001)	0.43 (<0.001)	1
Slope Value	0.05	0.07	0.05	-0.1	-0.15	-0.03
TWI	-0.21	0.06	0.06	0.19	0.24	0.1
Curvature	0.12	-0.14	-0.15	-0.09	-0.1	-0.04
Elevation	-0.12	-0.09	-0.19	-0.16	0.07	-0.1
Upslope Area	-0.11	-0.02	0.38 (0.003)	0.18	0.08	-0.01
Depth to Bedrock	0.09	0.12	0.52 (<0.001)	0.24	-0.05	0.04
Sand %	0.35 (0.033)	-0.09	0.21	-0.1	-0.35 (0.035)	-0.07
Silt %	-0.39 (0.018)	0.1	-0.41 (0.011)	-0.02	0.35 (0.033)	0.05
Clay %	-0.04	0.06	0.26	0.33 (0.046)	0.09	0.18
Organic Matter %	0.19	-0.11	-0.05	0.16	-0.11	0.21
Rock Fragment %	0.18	-0.3	-0.27	-0.24	-0.13	-0.14

b

Parameters 20 cm	α	n	θ_s	a	ψ_e	- b
α	1					
n	-0.48 (<0.001)	1				
θ_s	0.13	-0.01	1			
a	-0.32 (0.015)	0.45 (<0.001)	0.54 (<0.001)	1		
ψ_e	-0.92 (<0.001)	0.34 (0.010)	-0.22	0.34 (0.008)	1	
- b	-0.16	0.54 (<0.001)	0.09	0.79 (<0.001)	0.24	1
Slope Value	0.24	-0.2	0.04	-0.05	-0.24	-0.02
TWI	-0.16	0.18	0.26 (0.050)	0.21	0.17	0.07
Curvature	0.02	-0.18	-0.32 (0.013)	-0.25	-0.05	-0.11
Elevation	-0.26 (0.050)	0.18	-0.32 (0.016)	0.03	0.29 (0.028)	0.14
Upslope Area	-0.1	-0.04	0.59 (<0.001)	0.27 (0.040)	0.05	-0.06
Depth to Bedrock	-0.14	0.11	0.65 (<0.001)	0.36 (0.005)	0.05	-0.02
Sand %	-0.04	-0.09	-0.33 (0.046)	-0.27	0.03	-0.18
Silt %	0.02	0.13	0.27	0.18	-0.05	0.1
Clay %	-0.04	0.05	0.37 (0.023)	0.36 (0.027)	0.07	0.22
Organic Matter %	0	0.04	-0.23	-0.18	0.12	-0.1
Rock Fragment %	0.33 (0.044)	-0.17	-0.69 (<0.001)	-0.57 (<0.001)	-0.23	-0.24

c

Parameters 40 cm	α	n	θ_s	a	ψ_e	- b
α	1					
n	0.1	1				
θ_s	-0.01	0	1			
a	-0.30 (0.031)	0.25	0.58 (<0.001)	1		
ψ_e	-0.87 (<0.001)	-0.07	0	0.39 (0.003)	1	
- b	-0.14	0.42 (0.002)	0.2	0.86 (<0.001)	0.27 (0.050)	1
Slope Value	0.14	0	-0.1	-0.22	-0.32 (0.021)	-0.19
TWI	-0.35 (0.009)	-0.01	0.15	0.19	0.48 (<0.001)	0.08
Curvature	0.28 (0.042)	-0.17	-0.24	-0.31 (0.024)	-0.27 (0.050)	-0.25
Elevation	0.31 (0.023)	0.24	-0.12	0.08	-0.38 (0.005)	0.27 (0.050)
Upslope Area	-0.24	-0.24	0.42 (0.002)	0.11	0.2	-0.14
Depth to Bedrock	-0.17	-0.33 (0.017)	0.58 (<0.001)	0.31 (0.024)	0.04	0.04
Sand %	-0.04	-0.17	-0.45 (0.009)	-0.37 (0.035)	-0.06	-0.23
Silt %	-0.02	0.12	0.34	0.28	0.16	0.12
Clay %	-0.05	0.12	0.47 (0.006)	0.47 (0.005)	0.06	0.41 (0.019)
Organic Matter %	0.05	0.26	-0.36 (0.036)	-0.26	-0.06	-0.12
Rock Fragment %	-0.05	0.03	-0.68 (<0.001)	-0.47 (0.005)	0.17	-0.29

d

Parameters 80 cm	α	n	θ_s	a	ψ_e	- b
α	1					
n	-0.33	1				
θ_s	0.18	0.07	1			
a	0.03	0.18	0.37 (0.032)	1		
ψ_e	-0.70 (<0.001)	-0.07	-0.46 (0.006)	0.12	1	
- b	0.18	0.24	0.02	0.83 (<0.001)	0.02	1
Slope Value	0.02	0.11	-0.29	-0.03	0.13	0.14
TWI	0.16	-0.06	0.35 (0.044)	0.17	-0.31	0.13
Curvature	-0.44 (0.010)	-0.04	-0.2	-0.17	0.40 (0.018)	-0.19
Elevation	0.11	0.02	-0.27	0.11	0.2	0.25
Upslope Area	0.14	-0.25	0.27	0	-0.27	-0.03
Depth to Bedrock	0.34 (0.050)	-0.06	0.31	-0.05	-0.38 (0.025)	-0.04
Sand %	0.38	-0.11	-0.17	0.17	-0.13	0.32
Silt %	-0.39	0.2	0.16	-0.32	0.04	-0.51 (0.038)
Clay %	-0.33	-0.01	0.08	-0.08	0.15	-0.16
Organic Matter %	-0.02	-0.08	0.09	0.36	0.02	0.44
Rock Fragment %	0.34	-0.15	-0.44	-0.33	0.14	-0.08

e

Parameters 100 cm	α	n	θ_s	a	ψ_e	- b
α	1					
n	-0.25	1				
θ_s	-0.05	-0.18	1			
a	-0.49 (0.012)	0.24	0.3	1		
ψ_e	-0.86 (<0.001)	0.18	-0.12	0.58 (0.002)	1	
- b	-0.21	0.48 (0.014)	-0.03	0.80 (<0.001)	0.43 (0.030)	1
Slope Value	-0.05	-0.09	-0.22	-0.23	0.09	-0.19
TWI	0.1	-0.09	0.13	0.33	-0.12	0.19
Curvature	0.19	0.02	-0.22	-0.66 (<0.001)	-0.25	-0.39 (0.048)
Elevation	0.32	0.1	-0.50 (0.009)	-0.01	-0.01	0.36
Upslope Area	0.19	-0.28	0.2	0.05	-0.22	-0.19
Depth to Bedrock	-0.14	-0.07	0.13	0.3	0.11	0.11
Sand %	0	-0.29	-0.51	0.41	0.55	0.42
Silt %	0.27	0.26	0.5	-0.38	-0.75 (0.005)	-0.29
Clay %	-0.46	0.01	0.43	-0.27	-0.03	-0.52
Organic Matter %	-0.37	-0.1	0.05	0.1	0.19	0.07
Rock Fragment %	0.14	-0.19	-0.72 (0.006)	-0.32	0.14	-0.08