GEOSTATISTICAL METHODS FOR PREDICTING SOIL MOISTURE
CONTINUOUSLY IN A SUBALPINE BASIN

A Technical Paper
Presented to
the Faculty of Natural Sciences and Mathematics
University of Denver

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Geographic Information Science

by
Katherine E. Williams
June 3, 2011
Advisor: Dr. Sharolyn J. Anderson
# TABLE OF CONTENTS

TABLE OF FIGURES & TABLES ........................................................................................................................................... 2

ABSTRACT ........................................................................................................................................................................ 3

BACKGROUND ................................................................................................................................................................. 4

METHODS .......................................................................................................................................................................... 7

- Site Description .......................................................................................................................................................... 7
- Data ............................................................................................................................................................................. 8
- Field Collection .......................................................................................................................................................... 9
- Data Processing ......................................................................................................................................................... 9
- Geoprocessing ............................................................................................................................................................ 10
  - Regression ............................................................................................................................................................ 10
  - Interpolation ........................................................................................................................................................ 11

RESULTS ........................................................................................................................................................................... 13

- Regression Model ..................................................................................................................................................... 13
- Derived Layers .......................................................................................................................................................... 13
- Interpolation ............................................................................................................................................................ 14
- Hillslope Comparison ............................................................................................................................................... 17

DISCUSSION ................................................................................................................................................................. 19

CONCLUSIONS ............................................................................................................................................................... 22

ACKNOWLEDGEMENTS .................................................................................................................................................. 23

REFERENCES ................................................................................................................................................................. 24

APPENDIX A—FIELD GUIDE & SAMPLING PROTOCOL ................................................................................................. 26

APPENDIX B—GEODATABASE CONSTRUCTION ........................................................................................................... 31

APPENDIX C—GEOSTATISTICS EQUATIONS, DEFINITIONS, & IMPLICATIONS ............................................................. 33

- Ordinary Least Squares (OLS) Regression .................................................................................................................. 33
- Coefficient of Determination (Adjusted R²) .................................................................................................................. 33
- Variance Inflation Factor (VIF) ..................................................................................................................................... 33
- Corrected Akaike’s Information Criterion (AICc) ........................................................................................................ 34
- Jarque-Bera (JB) Statistic .......................................................................................................................................... 34
- Koenker’s Studentized Bruesch-Pagan (Koenker’s BP) Statistic .................................................................................. 34
- Moran’s Index (MI) .................................................................................................................................................... 35
- Spatial Weights Matrix (SWM) .................................................................................................................................. 35
- Geographically Weighted Regression (GWR) ............................................................................................................... 35
- Inverse Distance Weighting (IDW) .......................................................................................................................... 36
- Ordinary Kriging (OK) ............................................................................................................................................. 37
- Cross Validation .......................................................................................................................................................... 37

APPENDIX D—DERIVING CONTINUOUS SURFACES FOR VEGETATION TYPE & PERCENT COVER .......................... 39

- Data Derivatives ......................................................................................................................................................... 39
- Vegetation Type Classification ...................................................................................................................................... 39
- Percent Cover ............................................................................................................................................................ 40
TABLE OF FIGURES & TABLES

FIGURE 1. OVERVIEW OF THE GEOSTATISTICAL PROCESS .......................................................................................... 6
FIGURE 2. OVERVIEW OF STUDY AREA ..................................................................................................................... 7
FIGURE 3A. IKONOS-2 VISIBLE-NEAR INFRARED SATELLITE IMAGE ........................................................................ 8
FIGURE 3B. DIGITAL ELEVATION MODEL DERIVED FROM LIDAR DATA ..................................................................... 8
FIGURE 4. VWC SAMPLING PATTERN .......................................................................................................................... 9
FIGURE 5. SAMPLE POINTS USED IN GEOSTATISTICAL METHODS ........................................................................... 11
TABLE 1. DIAGNOSTIC STATISTICS FOR OLS AND GWR MODELS .......................................................................... 13
FIGURE 6. CONTINUOUS SURFACES OF EXPLANATORY VARIABLES .................................................................... 14
FIGURE 7. INTERPOLATED VWC SURFACES USING INVERSE DISTANCE WEIGHTING AND ORDINARY KRIGING METHODS (FIELD POINTS) ........................................................................................................... 15
FIGURE 8. INTERPOLATED VWC SURFACES USING INVERSE DISTANCE WEIGHTING AND ORDINARY KRIGING METHODS (10 METER & 5 METER POINT GRIDS) ......................................................................................................................... 16
TABLE 2. CROSS VALIDATION STATISTICS FOR INTERPOLATION SURFACES ............................................................... 17
FIGURE 9. COMPARISON OF VWC ESTIMATED BY THE SMSS INTERPOLATION SURFACES TO FIELD MEASUREMENTS MADE DURING THE HILLSLOPE STUDY IN 2010 ......................................................................................... 18
FIGURE A1. FIELD SAMPLING MAP ........................................................................................................................ 30
FIGURE A2. SMSS GEODATABASE SCHEMA ............................................................................................................... 32
TABLE A1. CLASSIFIED VEGETATION TYPES AND CORRESPONDING CODES ............................................................ 39
TABLE A2. KNOWLEDGE BASE CLASSIFICATION RULES .......................................................................................... 40
TABLE A3. VEGETATION CLASS CODES AND LOGARITHMIC RECLASSIFICATION CODES ........................................ 40
FIGURE A3. CLASSIFICATION OF VEGETATION TYPE IN LOCH VALE WATERSHED FROM IKONOS-2 IMAGERY ...... 40
FIGURE A4. PROCESS FLOW TO DERIVE VEGETATION TYPE AND PERCENT VEGETATION COVER ..................................................... 41
TABLE A4. ALLOCATION OF CLASSES INTO VEGETATION DENSITY CATEGORIES ......................................................... 42
FIGURE A5. PYTHON SCRIPT DESIGNED TO COUNT THE NUMBER OF MAJORITY VALUES IN A THREE BY THREE NEIGHBORHOOD .................................................................................................................. 43
FIGURE A6. MODEL DESIGNED TO CALCULATE PERCENT COVER BASED ON HIGH COUNT VALUE AND CLASSIFIED VEGETATION TYPE ............................................................................................................. 43
ABSTRACT

This soil moisture spatial survey (SMSS) study demonstrates the power of geostatistical methods for quantifying physical variables spatially. The goals of this study were to explore and compare geostatistical predictive models and interpolation methods and to produce a continuous soil moisture surface throughout the sub-alpine forested basin of a watershed in Rocky Mountain National Park, Colorado. Through a combination of regression, prediction, variable derivation, and interpolation, a high resolution, continuous surface was produced. The final Geographically Weighted Regression (GWR) model used to predict volumetric water content based on derived Slope, Vegetation Type, and Percent Cover had an Adjusted R² value of 0.71 and showed no significant spatial autocorrelation of the residuals. The Vegetation Type and Percent Cover surface were classified and geoprocessed from IKONOS-2 Near Infrared-Visible Imagery and 1 meter LiDAR digital elevation model data. Both field measurements and predicted measurements from the GWR model were used to create interpolation surfaces using Ordinary Kriging (OK) and Inverse Distance Weighting (IDW) methods. The IDW surface outperformed the OK surface by showing lower error and modeling the desired microtopographical features in greater detail. Using field measurements only, the soil moisture surface captured the average soil moisture trend along transects in an existing Hillslope Study, but did not portray the variability along the slope. Using the GWR model to predict soil moisture at points in a 5 meter grid improved the texture of the surface and captured more of the local variability on the slope. This study demonstrates a unique combination of data and methods for quantifying the spatial distribution of an environmental variable.
BACKGROUND

In the field of geostatistics, many innovative techniques and methods have been developed with the capacity to inform studies of physical variables. These tools can be powerful additions to field collection and traditional research methods by extending the amount of information drawn from limited samples. This study focuses on soil moisture, an important variable in hydrological and ecological studies that is incompletely understood and characterized. The geostatistical techniques applied enabled a limited number of field measurements to predict soil moisture continuous throughout the study area. The approach was to first identify a properly specified Ordinary Least Squares (OLS) model from which to justify a Geographically Weighted Regression (GWR) model. This model was then used to predict soil moisture as volumetric water content (VWC) where the explanatory variables were known. From these predicted values, interpolation methods were used to complete the continuous surface and estimate values everywhere in the study area.

While this approach is perhaps novel for soil moisture studies, the utility of GWR and interpolation techniques have been demonstrated in studies of other environmental variables. Relationships between precipitation and other physical variables have been refined using GWR (Brunsdon, McClatchey and Unwin 2001, Lynch 2003, Lloyd 2010). GWR has been used in studies of spatial distributions of the normalized difference vegetation index (NDVI) in relation to other variables (Foody 2003, Propastin 2006). These studies demonstrate how the increased ability of GWR to model local relationships between dependent and explanatory variables can give the model flexibility to emulate real variations in relationships (Foody 2003, O’Sullivan and Unwin 2003, Raty and Kangas 2007). GWR allows the weights of the explanatory variables to vary by creating a matrix of spatially varying weights throughout the study area. This also allows the dependent variables to be predicted based on explanatory variables and the weights assigned on the weights surface created from measured data (Foody 2003). GWR should only be used as an extension of OLS regression if nonstationarity is statistically significant in the OLS model (Foody 2003, Rosenshein and Scott 2010). OLS is not only most robust to diagnostic statistics, but is a global model and is thus a simpler representation of the data relationships (Scott and Pratt 2009).

Spatial interpolation techniques predict continuous surfaces for the variable of interest, but are not set up to deal with regression type relationships between predictive and dependent variables. This study estimates a continuous soil moisture surface based on field measurements alone, but also looks at interpolations based on GWR predictions. The ability to use GWR first to predict the dependent variable greatly simplified the type of interpolation required. Kriging with an external drift (KED) or Co-Kriging are examples of other methods that attempt to deal with variables that change predictability in space (Bardossy and Lehmann 1998, Propastin 2006). Other methods like KED and Co-Kriging introduce added complexity into the interpolation procedure which is unnecessary when the relationships of explanatory variables are already accounted for by a proper OLS or GWR predictive model. The interpolation methods compared here are Inverse Distance Weighting (IDW) and Ordinary Kriging (OK) which
are two of the simplest and most common interpolation techniques. IDW uses distance as the
primary determinant in assigning weights to neighboring values while OK determines weights
based on existing relationships in the data (O'Sullivan and Unwin 2003, Propastin 2006). These
simple interpolation methods performed well in estimating the distribution of VWC once the
VWC values were predicted by the GWR model at point locations uniformly spaced within the
study area.

VWC is an important environmental variable and has been extensively studied in the field. As
temperatures increase worldwide, mountain environments are gaining attention as regions that
may be most affected. These systems are already hung in delicate balance between short
growing seasons, harsh environmental extremes, and often dynamic processes related to steep
verticality (Funnell and Parish 2001). Soil moisture is an important part of mountain
ecosystems with implications for vegetation distributions, habitat boundaries, erosion
potential, and change over time. Soil facilitates water transport and vegetation structures that
are sensitive to change. Soil moisture regulates plant community distribution and vigor,
nutrient turnover in the soil, and response to precipitation events. In high elevation
ecosystems the maintenance of soil moisture becomes more critical because the soils are often
shallow with poorly developed organic horizons (Reuth, Baron and Allstott 2003, Molotch, et al.
2009). Water tends to evaporate more readily in these soils due to limited organic material on
the surface, leading to a disruption of normal soil processes. As air temperatures increase, the
evaporation rate will also increase. Studies of the relationship between vegetation, seasonal
snow accumulation and melt, ground and air temperatures, topography and soil moisture
Understanding the distribution and dynamics of soil moisture will help researchers and park
managers identify areas susceptible to affects of climate change, such as drought stress and
increased erosion potential. Soil moisture is an important variable in studies of hydrologic
processes but is difficult to characterize due to extreme local, spatial, and temporal variability
(Kampf and Burges 2007).

The Loch Vale Watershed is located in Rocky Mountain National Park in Colorado. Loch Vale is
a Long-Term Ecological Research (LTER) site supervised by the United States Geological Survey.
Research at this location investigates the impact of atmospheric nitrogen deposition and
climate variability in high elevation ecosystems (Reuth, Baron and Allstott 2003). The
watershed is one of the most researched alpine/subalpine sites in the world and is used as an
indicator region to assess the condition of and establish standards for alpine water quality and
nitrogen deposition levels (Theobald, et al. 2009). Loch Vale has numerous research projects
including a Hillslope Study targeting soil moisture processes. The Hillslope study is in its third
year and monitors snow depth, soil moisture, and soil temperature on two transects in Loch
Vale. Preliminary findings from the Hillslope study show soil moisture is regulated by
topography and is not distributed on a gradient with upslope always less moist than downslope
locations. This finding contradicts assumptions about soil moisture which is often used as an
input variable in hydrological models for wetness. Hydrological models rely on assumptions
about soil moisture distribution that have not been fully verified in the field, especially at small
watershed scales. The Hillslope Study found that topographic features such as exposed
bedrock, seeps, and depressions alter the soil moisture gradient from a smooth uphill to downhill increase and introduce dry and wet patches along the slope. In addition, soil moisture measurements show extreme local variability making soil moisture predictions difficult without extensive ground data (Kampf and Markus 2009).

Soil moisture is currently measured on the ground with a hand-held instrument. While this method is accurate and widely used, it does not provide watershed-wide soil moisture information. Soil moisture can be measured remotely using radar pulses as demonstrated with success (Moran, et al. 2005). New methods are under evaluation to use cosmic-ray technology to measure soil moisture over large areas (Shuttleworth 2011). The technology is available, but the cost is high to obtain this data at high resolution. Therefore, this study focused on field based measurements, but sampled throughout the entire study site to inform a spatial geostatistical method of prediction. Additional satellite imagery products and airborne light detection and ranging (LiDAR) data were used to derive continuous vegetation type and percent ground vegetation cover. This classification was fine enough to distinguish small terrain features that are lost at coarser resolutions. This detailed land classification layer could be incorporated into other spatial analysis or modeling studies for this area. For the Hillslope Study in particular, these derived surfaces will be valuable because soil moisture at the local level is sensitive to vegetation cover as demonstrated in this study and by others (Rodriguez-Iturbe, et al. 1999, Molotch, et al. 2009).

The objectives of this study were to (1) investigate the relationships between soil moisture, micro-topography, and vegetation using regression techniques, (2) construct detailed vegetation type and percent cover surfaces for the Loch Vale Watershed study area, and (3) produce a continuous VWC surface that reflected the predictions of the regression model while comparing IDW and OK interpolation methods. The combination of regression, derived remote sensing layers, and interpolated surfaces form the geostatistical method used to investigate the distribution of VWC in the lower basin of Loch Vale. This surface was validated using cross validation measures and was verified against the Hillslope Study transects data.
METHODS

Site Description
Loch Vale Watershed is a subalpine/alpine basin located in Rocky Mountain National Park, Colorado, USA (Figure 2). The Loch Vale catchment consists of a variety of land cover types ranging from glacial ice and talus to alpine meadows and old-growth timber stands (Clow, et al. 2003). High elevation and position at the Continental Divide create extreme seasonal weather conditions to which the ecosystem is delicately attuned. High levels of atmospheric nitrogen are deposited on this location from Colorado’s Front Range and annual mean temperatures are rising consistently. The catchment is dominated by rock and talus but also sustains a thick old-growth Subalpine Fir (*Abies lasiocarpa*) and Engelmann Spruce (*Picea engelmanni*) forest (Reuth, Baron and Allstott 2003). The primary form of precipitation in the basin is snow accumulation during winter months. A sudden spring ablation event is characteristic of the hydrology followed by slow, steady melt and sporadic rain events throughout the growing season. The watershed consists of steep terrain with elevation ranging from 2974 meters at the outlet to 4010 meters on the southwest ridgeline.

The ongoing Hillslope Study in Loch Vale was begun in 2008 by researchers at Colorado State University. Two transects were established that begin in talus and end at or near streams. The Loch Vale catchment consists of two tributaries, Andrew’s Creek and Icy Brook, that flow into the main outlet. The transects were positioned such that they fell within different tributary basins. Transect 1 in the Andrew’s Creek basin has a northern aspect while Transect 2 on the Icy Brook side has a southeasterly aspect. This results in some difference in the local snow accumulation and melt regime. The transects were sampled in 2008, 2009 and 2010 at 10 meter intervals for soil moisture, soil temperature, and snow depth. Preliminary conclusions indicate that local topography (elevation) was the best estimator of soil moisture (Markus 2008, Kampf and Markus 2009).

Figure 2. Overview of Study Area. Loch Vale Watershed is located in Rocky Mountain National Park, Colorado. The watershed is shown with a hillshade and the study area for the Soil Moisture Spatial Survey is outlined in blue.
The focus area for the soil moisture spatial survey (SMSS) was the lower basin of the Loch Vale watershed where soil structure was deep enough for contiguous measurements over a large area. This study area included a range of vegetation cover types, percent slopes, aspects, and elevations. The upper basins of the watershed had patches of significant vegetation cover, but were eliminated from the SMSS because they were not continuous with the lower basin and posed difficulties for spatial statistical analysis and field collection practicality. The lower basin of Loch Vale ranges from 3100 meters to 3300 meters in elevation, covers an area of about 1.1 km², and is characterized by high relief and topographic variability.

**Data**

The data in this study consists of field measurements, a Near Infrared (NIR)-Visible IKONOS-2 image pansharpened to 1 meter spatial resolution (Image Grant, GeoEye Foundation), and a 1 meter digital elevation model (DEM) derived from LiDAR data collected in August 2010 (Figures 3A and 3B). Field measurements from the SMSS yielded 107 sample locations. Ten additional sites were defined in ArcGIS 10 (Esri, Inc.) to enable continuous prediction of VWC throughout the study area. Five of these points were located on water bodies and the other five were positioned on prominent rock outcroppings. These areas had no soil and could not be measured in the field. Water and Rock were assigned values of 1 and 0 respectively for VWC and zero for Percent Vegetation Cover. Interpolation methods assume the predicted variable is continuous, therefore these points were added to inform the model in areas without soil to enable a continuous surface interpolation. The IKONOS-2 image and LiDAR DEM were used to derive explanatory variables for the regression models.
Field Collection

Field samples were collected during a three day field campaign on August 9-11, 2010. Researchers collected data at 111 sample sites throughout the study area. Of these 111 sites, three were repeat samples used to calibrate between days and one was taken on a rock and contained no soil moisture information. The final number of sample points was 107.

At each sampling location researchers took fifteen VWC measurements with a Time Domain Reflectometer (TDR) instrument. Three measurements were taken at the center of the plot and three measurements were taken two meters from the center in each cardinal direction (Figure 4). This plus-sign sampling strategy was developed to capture the extreme local variability common to near surface soil moisture measurements in Loch Vale (Kampf and Burges 2007). Probe depth was 12 cm for most measurements, but a 5 cm probe was used on steep, rocky slopes where the soil layer was shallow and unreasonable to sample with a 12 cm probe.

In addition to VWC, researchers collected GPS positions and recorded site characteristics for each sampling location. A minimum of 120 GPS positions were recorded at the center point of each sampling site using a Trimble GeoXT or Trimble Juno handheld mapping grade GPS unit (Trimble, Inc.). The GPS units were pre-loaded with maps of the area and target sampling points used by researchers to navigate in the watershed. Collected site characteristics included dominant vegetation cover (Dry Meadow, Forest, GrassyRocky, or Wet Meadow) and approximate percent vegetation cover and percent canopy cover. Digital photos of each sampling site were taken to be a reference if needed. The guide used by the field crew can be found in Appendix A.

Data Processing

The data assimilation process entailed a multi-person effort to download and differentially correct GPS positions from the GPS units, extract information from the GPS data dictionaries, download data from the soil moisture probes, and enter data recorded in field books. The different data formats were consolidated into a flat table consisting of 107 records and 50 fields for each record. The table was highly repetitive, not spatially referenced, and difficult to query or analyze. For proper functionality the data needed to be loaded in a database.

A File Geodatabase in ArcGIS 10 (Esri, Inc) was chosen as the most suitable database. A geodatabase format is suitable for maintaining the spatial integrity of the sample points and providing a way to reduce storage volume by relating data in multiple tables. The geodatabase allows relationships and rules to be established between feature classes and tables. The geodatabase can also handle storage of other spatial information such as elevation grids, lidar point files, reference maps, and products from statistical analysis. The geodatabase is a
dynamic environment that accommodates data manipulation, exploration, analysis, and spatial display. For a detailed explanation of Geodatabase construction, see Appendix B.

**Geoprocessing**

**Regression**

Data from the SMSS field points were analyzed using a supplementary Exploratory Regression Tool available in the Supplementary Spatial Statistics Toolbox available through Esri, Inc. online resources (Esri, Inc 2010). This tool iterates through Ordinary Least Squares (OLS) models using all possible combinations of input explanatory variables within defined parameters. The spatial weights matrix used was an inverse distance squared conceptualization of the spatial relationship between points. A series of diagnostic statistics identified properly specified regression models from this process. These statistics are the Coefficient of Determination ($R^2$), Corrected Akaike’s Information Criterion (AICc), Jarque-Bera Statistic, Kroenker BP Statistic, Variance Inflation Factor (VIF), and Moran’s Index. These statistics test different aspects of the model and are explained in detail along with OLS in Appendix C.

Candidate explanatory variables included field collected variables (Dominant Vegetation Type, Percent Vegetation Cover, and Percent Canopy Cover), derived and calculated variables (Elevation, Slope, Solar Radiation, VWC Variance, Curvature, Distance to Ridgeline, Distance to Streams or Lakes, Surface Flow Direction, Surface Flow Accumulation), and field collector variables to test for collection bias (Date Collected, Collection Team, GPS Unit, TDR Probe, GPS Position, GPS Horizontal Precision, GPS Vertical Precision, Latitude, and Longitude).

Once a proper OLS model is identified, a Geographically Weighted Regression (GWR) model can be used if nonstationarity is observed. The same explanatory variables identified for the OLS model are used in GWR to ensure model validity. The diagnostic statistics are more robust for OLS than for GWR, so if possible, a proper model should be found using OLS before continuing to GWR (Scott and Pratt 2009).

Once validated, a regression model can be used to make predictions of the dependent variable where the explanatory variables are known (Scott and Pratt 2009). In this study, continuous surfaces for explanatory variables were needed. Derived variables were already continuous, but field collected variables were not. A continuous Vegetation Type was derived using the IKONOS-2 image and LiDAR DEM in an expert classification as described in Appendix D. This produced ten categorical classes which were recoded in an ordinal ranking scheme from least to most wet. These land cover types included non-soil regions to accommodate a continuous predictive surface. Regions of rock were regarded as zero VWC while water was most wet and assumed to have a VWC of 1. The GWR method treats this ordinal variable as continuous so results should be interpreted with caution. Percent Cover was derived from the Vegetation Type surface using an innovative reclassification and focal statistics method also described in Appendix D.
Interpolation
Once the continuous explanatory variable layers were produced, the values were extracted to the field sample points and the OLS model was re-run to assess whether or not the model was still valid with the derived data as opposed to the field estimations. Then two point grids at 10 meter and 5 meter intervals were created throughout the study area. The values of the explanatory variable surfaces were extracted to the grid points and the GWR model was used to predict VWC at each point. These predicted VWC values at the grid points, along with the field measured and GWR predicted values at the SMSS sampling points, were used to create and compare interpolated Ordinary Kriging (OK) and Inverse Distance Weighting (IDW) continuous VWC surfaces. This produced eight interpolation maps for comparison. OK was
chosen over other types of kriging methods because it is the simplest, most common, and this dataset had no special parameters that necessitated a more complex method. The OK model used a standard search neighborhood with a full sector and twenty neighbors. The semivariogram was calculated using 12 lags with different lag intervals depending on the input data and a Stable Model without anisotropy to fit the semivariogram. The IDW search radius was defined to match the range of the semivariogram from the OK method. The standard search neighborhood was a full sector with twenty neighbors. The cross validation statistics used to summarize and compare the different models were Predicted Mean Error, Predicted RMSE, Predicted Regression Function, and Error Regression Function. In addition to cross validation, predicted VWC was compared to field measurements from the Hillslope study using approximate locations of the hillslope sample points. All geostatistics were conducted in ArcGIS 10 (Esri, Inc.). Equations, definitions, and implications for the geostatistics are described in Appendix C.
RESULTS

Regression Model
The only passing OLS model from the Exploratory Regression Tool had three explanatory variables: Slope, Vegetation Type, and Percent Cover. These three variables formed a passing OLS model with both field collected and derived values for Vegetation Type and Percent Cover. Only the model statistics for the derived variables are reported since these were used in subsequent analysis while the field measurements were only used in comparing interpolated surfaces.

The OLS model had an Adjusted R\(^2\) of 0.61 which shows an intermediate level performance (Table 1). The Koenker BP p-value was significant, indicating nonstationarity which means the explanatory variables did not have a constant relationship to the predicted variable across the study area. This necessitated using GWR even though the Jarque-Bera and Moran’s Index on the residuals were not significant at the 95% confidence level. The Jarque-Bera test shows that the OLS model did not contain significant bias and the Moran’s Index value near zero means the residuals were spatially random.

The GWR model had an Adjusted R\(^2\) value of 0.71 which indicates an increase in model performance. The AICc is also about 19 units lower than for the OLS model which indicates that the overall fit is better with the GWR model. The residuals for the GWR model pass the Moran’s Index test with a value near zero and a non-significant p-value. The Jarque-Bera and Koenker BP are not suitable statistics for GWR and are not included in the output statistics. The GWR model was selected as the most appropriate for this dataset and was used to predict soil moisture for the field points, the 10 meter grid points, and the 5 meter grid points.

TABLE 1. DIAGNOSTIC STATISTICS FOR OLS AND GWR MODELS

<table>
<thead>
<tr>
<th></th>
<th>OLS (n=117)</th>
<th>GWR (n=117)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted R(^2)</td>
<td>0.607749</td>
<td>0.713158</td>
</tr>
<tr>
<td>AICc</td>
<td>-122.1762</td>
<td>-140.9467</td>
</tr>
<tr>
<td>Koenker BP (p-value)</td>
<td>28.90099 (0.000002*)</td>
<td>---</td>
</tr>
<tr>
<td>Jarque-Bera (p-value)</td>
<td>4.92341 (0.085289)</td>
<td>---</td>
</tr>
<tr>
<td>Residual Moran’s Index (p-value)</td>
<td>0.17524 (0.555171)</td>
<td>-0.061323 (0.865530)</td>
</tr>
</tbody>
</table>

* Indicates significance at 95% confidence.

NOTE: Adjusted R\(^2\) indicates model performance with higher values indicative of better performance, AICc (Corrected Akaike’s Information Criterion) measures model fit with lower values indicative of better fit, the Koenker BP Statistic measures stationarity with significant p-value indicating non-stationarity, the Jarque-Bera Statistic assess the distribution of the model residuals with significant p-value indicating non-normal distribution, and Moran’s Index measures spatial autocorrelation with significant p-value indicating non-random spatial distribution.

Derived Layers
The derived Slope, Vegetation Type, and Percent Cover maps characterize the lower basin of Loch Vale well (Figure 6). The Vegetation Type map corresponded well with aerial photography
of the basin in a visual inspection. The Vegetation Type map matched with the SMSS field designation only 53% of the time. However, the OLS model using the derived values instead of the field values was still valid. The Percent Cover in forested regions was restricted to a maximum of 50 percent because the understory vegetation is often sparse due to limited sunlight. The thickest vegetation cover should be in the meadow and tundra classes as they are low-lying dense vegetation types. The percent cover map reflects these attributes of vegetation type as well as uniformity of vegetation across an area. Areas of rock, snow, shadow, and water have no vegetation and thus received a Percent Cover designation of zero.

**Interpolation**

The VWC interpolated surfaces are clearly limited by the number of sample locations when the field data is used. The interpolated surfaces predict broad regions of wetter and dryer areas with little local variability. However when comparing the OK maps for the Field Measurements versus the GWR Predicted Measurements, the pattern is very similar with some smoothing in the GWR predictions (Figure 7A and 7B). The error maps are also similar and show higher error...
in areas near the edges of the study area and between transitions from dry to wet regions. The IDW maps of the predicted and measured field points show slightly finer contouring than either of the corresponding OK maps, but overall wet areas and dry areas are located in the same regions (Figure 7C and 7D). The difference between the measured and predicted VWC for the interpolated surface is negligible (Table 2).

Figure 7. Interpolated VWC Surfaces using Inverse Distance Weighting and Ordinary Kriging Methods. (A) IDW surface from 117 measured field points. (B) OK surface from measured field points. (C) IDW surface from GWR predicted VWC at the same 117 field point locations used in A. (D) OK surface from GWR predicted VWC at the same 117 field point locations used in A. Error maps are based on standard cross validation outputs in ArcGIS 10 Geostatistical Analyst.

The interpolated surfaces for the generated 10 meter and 5 meter point grids are rougher and follow the topography more closely (Figure 8). The error for these surfaces is very low because the study area was sampled uniformly and with small point distances compared to the field points. The increased resolution in the 5 meter grid shows the edges of the lake and the center ridge feature in greater detail. IDW again shows less smoothing than the OK method.
Comparing error between the interpolation methods, IDW has lower error in each case, though the values are very similar and all errors are low (Table 2). The predicted regression functions are nearly identical between comparisons using field points. However, as the number of input points increases, the regression function for the IDW method has a slope closer to one and an intercept closer to zero than the OK method. This indicates a slightly better fit for the IDW method. The error regression functions all have a noticeable negative slope, though as the point resolution increases, the magnitude of the value decreases.
**Table 2. Cross Validation Statistics for Interpolation Surfaces Based on Measured or GWR Predicted VWC at Either Field Sampling Points, 10 Meter Grid Points, or 5 Meter Grid Points.** An explanation of the cross validation statistics is provided in Appendix C.

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>Ordinary Kriging</th>
<th>Inverse Distance Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Field Points Measured</strong> n=117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted Mean Error</td>
<td>0.00632622</td>
<td>0.00467828</td>
</tr>
<tr>
<td>Predicted RMSE</td>
<td>0.20752391</td>
<td>0.20251504</td>
</tr>
<tr>
<td>Predicted Regression Function</td>
<td>(y = 0.2345 \times x + 0.2019)</td>
<td>(y = 0.2362 \times x + 0.1994)</td>
</tr>
<tr>
<td>Error Regression Function</td>
<td>(y = -0.7655 \times x + 0.2019)</td>
<td>(y = -0.7638 \times x + 0.1994)</td>
</tr>
</tbody>
</table>

| **Field Points GWR Predicted** n=117 |                  |                            |
| Predicted Mean Error         | 0.00613349       | 0.00540272                 |
| Predicted RMSE               | 0.17366084       | 0.17235140                 |
| Predicted Regression Function| \(y = 0.2661 \times x + 0.2002\) | \(y = 0.2470 \times x + 0.2157\) |
| Error Regression Function    | \(y = -0.7339 \times x + 0.2002\) | \(y = -0.7530 \times x + 0.2157\) |

| **10m Grid Points GWR Predicted** n=1057 |                  |                            |
| Predicted Mean Error         | -0.00005312      | 0.00006528                 |
| Predicted RMSE               | 0.12523941       | 0.12139920                 |
| Predicted Regression Function| \(y = 0.6253 \times x + 0.0762\) | \(y = 0.6603 \times x + 0.0692\) |
| Error Regression Function    | \(y = -0.3747 \times x + 0.0762\) | \(y = -0.3397 \times x + 0.0692\) |

| **5m Grid Points GWR Predicted** n=4253 |                  |                            |
| Predicted Mean Error         | 0.00002023       | 0.00000738                 |
| Predicted RMSE               | 0.10854328       | 0.10241228                 |
| Predicted Regression Function| \(y = 0.8153 \times x + 0.0409\) | \(y = 0.8419 \times x + 0.0353\) |
| Error Regression Function    | \(y = -0.1847 \times x + 0.0409\) | \(y = -0.1581 \times x + 0.0353\) |

**Hillslope Comparison**

Figures 9A and 9B show graphs of VWC along the two Hillslope Study transects. The average VWC for all measurements during the 2010 season are plotted with modeled VWC from the SMSS Study. The Hillslope Study measurements demonstrate the variability of soil moisture in the mid-slope region but showing an increasing trend overall with movement downhill. The IDW predictions with the 117 SMSS field points capture the mean of the water content, but do not show mid-slope variability. In contrast, the 5m grid GWR predictions do show the midslope variability while increasing in water content overall. The variability does not match precisely with the measured values from the Hillslope Study, but many of the ups and downs occur at the same locations, especially on the Andrew’s Creek transect. The 5 meter grid IDW modeled VWC over predicts on the Icy Brook transect except at two sample locations but displays similar trends.
FIGURE 9. COMPARISON OF VWC ESTIMATED BY THE SMSS INTERPOLATION SURFACES TO FIELD MEASUREMENTS MADE DURING THE HILLSLOPE STUDY IN 2010. THE IDW INTERPOLATION SURFACE FROM THE SMSS FIELD MEASUREMENTS CAPTURES THE MEAN TREND OF THE HILLSLOPE STUDY MEASUREMENTS WHILE THE IDW SURFACE INTERPOLATED FROM THE GWR PREDICTIONS AT THE 5 METER GRID POINTS IS ABLE TO DISTINGUISH THE MIDSLOPE VARIABILITY OBSERVED IN THE HILLSLOPE STUDY. (A) SHOWS MEASUREMENTS FROM TRANSECT 1 IN THE ANDREW’S CREEK DRAINAGE OCCUPYING THE NORTHWEST BRANCH OF LOCH VALE. (B) SHOWS MEASUREMENTS FROM TRANSECT 2 IN THE ICY BROOK DRAINAGE IN THE SOUTHWEST PART OF LOCH VALE.
The first goal of this study was to detect key topographic controls on soil moisture. The regression approach evaluated different combinations of physical and non-physical explanatory variables and came up with only one passing model. This model had three variables: Slope, Vegetation Type, and Percent Cover. These variables are meaningful in a physical sense as factors that relate directly to soil moisture. Slope influences the drainage rate as well as the accumulation of a soil layer and is a logical control of soil moisture. Vegetation Type and Percent Cover however are more likely consequences of existing soil conditions rather than controls. The distribution of vegetation types and their associated percent cover are likely heavily influenced by soil type, depth, and moisture content. However, this study shows that these variables can be used as indicators of other variables which may actually control moisture content, but are difficult to measure. These include soil texture, composition, porosity, depth to bedrock, and horizon characterization among others. Soil metrics such as these could not be derived from the remote sensing data available for this study nor were they available at the resolution needed to produce a detailed VWC surface. Given the extreme local variability of VWC measurements the 1 meter spatial resolution of the input variables was a necessity.

Measuring slope in the field is labor intensive, but easily acquired from a DEM. The LiDAR data provided an excellent DEM for deriving slope. Vegetation Type and Percent Cover on the other hand are more readily assessed in a field campaign, but difficult to quantify from digital data. The techniques used in this research required extensive knowledge of the study area and construction of a classification model including elevation and slope as ancillary data. This approach was time intensive and would not be directly repeatable in a different study area. The classification also had to be detailed in order to derive the level of information desired. Even though a vegetation classification was available for all of Rocky Mountain National Park, it was not at a fine enough scale to yield meaningful VWC predictions. Both the Vegetation Type and Percent Cover layers depended on the quality of the initial image classification.

The cost of producing these derived continuous layers should be considered in any plans to repeat this type of work. The final results support the importance of generating high quality continuous surfaces for the explanatory variables. These layers enabled prediction of the dependent variable anywhere in the study area based on the regression model determined by the field measurements. This method does not circumvent the collection of field data, but it does enhance the explanatory power of a small number of samples relative to the research area. In this study, the increased resolution and prediction of VWC throughout the study area were well worth the initial cost. The interpolated surfaces from the 117 field sampling sites were simply not sufficient for characterizing soil moisture distribution at a fine enough scale to pick out small topographic features such as rock outcroppings, sinks, and enclosed meadows. The ability to predict VWC every five meters allowed the production of a surface that actually approximated field measurements taken by an entirely different study. However, the
scalability of the results has not been demonstrated and future studies should look at the performance of these explanatory variables in nearby places.

The interpolated surfaces show similar results between the field measurements and the GWR predictions when the same 117 field points are used. This is encouraging because it means the derived layers produced meaningful estimations of the explanatory variables on the ground. This justifies continuing in the analysis to predict VWC at other locations in the study area where there were no field measurements. The 10 meter and 5 meter grids were chosen to evaluate the prediction at different levels of complexity. The smallest grid size possible in this study would be 1 meter since the explanatory variables are at this resolution. However, a 5 meter grid was sufficient to distinguish the edges of lakes and prominent boulder fields. This also allows some room for error that was included through extensive data processing and GPS positional accuracy of the initial field points. The median horizontal precision for the GPS points was 3.9 meters so a 5 meter sampling grid is still above the error of most GPS locations. The 10 meter grid interpolation showed clear improvement over the field point interpolations. The edges of features are more apparent in the surface and the texture shows more of the abrupt land cover change characteristic of Loch Vale. The 5 meter grid adds to the clarity and improves the predicted regression function by increasing the slope nearer to one and further decreases the magnitude of the negative slope in the error regression function. The negative trend in all of the interpolated surface error regression equations indicates that something is affecting the models uniformly. This error seems to be alleviated by increasing the number of initial points, but is still noticeable in the 5 meter point grid interpolation. This trend may be due to forcing the ordinal Vegetation Type variable to a continuous scale.

The OK to IDW comparison showed that IDW always predicted with lower error and better fit than OK for this study. OK tended to overly smooth variations in areas where variables changed over a short distance. The intent of this study was to model local terrain features that influenced soil moisture. IDW was more adept at capturing small variations in topography and vegetation distribution. This yielded a more complete picture of the microtopography and its impact on moisture distribution. The point grids were evenly spaced which did favor IDW although the smoothing quality of Kriging would still be a factor even if a random distribution of closely spaced points were generated. The difference in the information provided by the 117 field points versus the point grids shows that deriving the data layers as mentioned above was worthwhile. Though time intensive, the maps of the interpolations based on the derived layers are more informative and pertinent to the goals of this study. The IDW map from the 10 and 5 meter grids cover the entire study area and pick out soil moisture variations at fine scales. The IDW layer generated from the 5 meter point grid was selected as the highest quality. This surface should be used in other analyses to determine ground accuracy as well as utility in informing other studies of this site.

The comparison between the SMSS modeled VWC surfaces and the Hillslope Study transects was encouraging, but should be regarded with some caution. The GPS positions for the Hillslope Study sample points had low accuracy and some were missing. The points reported are a combination of measured GPS positions and sample points defined in a point layer. More
careful GPS measurements are needed to decisively determine whether or not the IDW surface correctly follows field measurements. With the current sample location coordinates, it is clear that the interpolated surface from the SMSS field points was too coarse to see mid-slope variability, but closely followed the mean VWC trend. This shows that the interpolated surface from the 117 SMSS points was reliable but just did not have the amount of detail needed to see small changes. The derived data layers and 5 meter point grid produced an interpolated surface that showed the level of variability seen in the Hillslope Study, although the modeled value for all locations does not always have the same magnitude or follow the same pattern as the measured values. This may indicate that the GPS positions for the Hillslope points do not accurately place them in projected geographic space with the IKONOS-2 and LiDAR data. The discrepancies may also demonstrate some of the remaining error in the model. However, the modeled surfaces performed well when estimating measurements that were not used to inform the development of the model in any way and were collected throughout the summer instead of on the exact dates of the SMSS.
CONCLUSIONS

This research demonstrates a valuable extension of geostatistical methods. The study shows how limited field data can be combined with other data products to yield results beyond the scope of either dataset. The modeled soil moisture surface is informed by an array of variables that capture unique properties of the underlying topography without requiring extensive surveying of the area. According to this study, soil moisture can be predicted as VWC using Slope, Vegetation Type, and Percent Cover, though the later two are more realistically consequences rather than causes of existing soil moisture patterns. The derived data layers for Vegetation Type and Percent Cover were successfully used as prediction explanatory variables and improved the interpolated surface over the field measurements along. IDW was found to out-perform OK in each interpolation scenario and produced a surface that was truer to the topographic variability of interest in this study because OK smoothed the surface undesirably. Otherwise, both methods performed with low error and strong predictive ability. The comparison of the final 5 meter grid IDW surface to the separate but related Hillslope Study measurements showed the interpolated surface produced the midslope variability as desired. These results are encouraging, but implications are limited by the scale and complexity of the study. These issues should be further explored to discover how this method may translate to other areas or variable measurements. However, this approach accomplished the goal of characterizing the distribution of soil moisture based on microtopography with minimal impact on the environment in Loch Vale. The methods used show the importance and utility of identifying a proper regression model that can be used to predict the dependent variable where explanatory variables are known. It is important to note that the success of this study was closely associated with choosing variables and methods appropriate for the research questions and data available. Future studies should consider carefully which methods and parameters are most meaningful and powerful given the limitations and goals of the project.
ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Dr. Sharolyn J. Anderson (University of Denver) for her insight and advice throughout this project. Also my readers, Dr. Paul C. Sutton (University of Denver), Dr. Hillary Hamann (University of Denver), Dr. Stephanie Kampf (Colorado State University), and Gary LaVanchy (University of Denver) for their feedback and input into this technical report. Also, Dr. Stephanie Kampf, Dr. Jill S. Baron, and Eric Richer (Colorado State University) for helping with research design, field preparation, and equipment acquisition. Also Dr. Jeff Niemann and Steve Midkauff (Colorado State University) for lending TDR probes and for technical assistance. In addition I would like to acknowledge the hard work and support of the field crew: Dr. Sharolyn J. Anderson, Dr. Paul C. Sutton, Joseph H. Hoover, and Brenton Wonders (University of Denver), Dr. Jill S. Baron, Eric Richer, Cara Moore, Sean Jacobson, Julia Spencer, Brooke Osborne, and Benjamin Reynolds (Colorado State University). Also Matthew Schoolmeester-Winnick (Colorado State University) for help with data entry. In addition, I would like to acknowledge Judy Visty and Jeff Connors with The National Park Service for evaluating my research permit and providing GIS data layers for the area, The American Society for Photogrammetry and Remote Sensing Rocky Mountain Region for funding this research in part, the GeoEye Foundation for providing the IKONOS-2 image through a student image grant, and Dr. Lauren Scott (Esri, Inc.) for providing support and guidance with the regression analysis.
REFERENCES


Shuttleworth, WJ. "At last it is possible to measure area-average soil moisture!" *Hydrology Days 2011*. Fort Collins: Colorado State University, 2011.

The following instructions were used by the field crew for taking soil moisture measurements and for navigating in the study area using the Trimble GPS unit. Every item on the checklist was required for each sample point. Refer to Figure A1 for an overview map of the study area and target sample locations.

Sample Point Checklist

- 120 GPS positions
- Defined ID in GPS
- 15 soil moisture measurements recorded in field book
  - 3 at center
  - 3 two meters N
  - 3 two meter E
  - 3 two meters S
  - 3 two meters W
- Record number from TDR recorded in field book
- Land cover recorded in field book
- Percent land cover recorded in field book
- Percent canopy cover recorded in field book
- Photo of sampling point taken from 4 meters to the south, facing north
- Photo number recorded in field book

Niemann Lab TDR100 Soil Moisture Probe Instructions

Instrument setup:

1. Plug in probe through Tupperware into TDR100.
2. Plug hand-held Campbell Scientific Sampler into cable connection labeled “C/S”.
3. Plug battery into converter box (green knob into matching slot).
4. Sampler should turn on. Press “enter” “enter” “enter” “enter” until screen reads “Sample” on top line and has a list of probes beneath.
5. Check color on attached probe cable and select the appropriate probe. To do this scroll to the probe that says “true”. Press “enter” “enter” to change that probe to “false”. Scroll to the desired probe color and press “enter” “enter” to change that probe to true.
Sampling:

1. To collect a sample, insert the probe fully into the ground, scroll to “Sample”, press “enter” “enter” and wait until the value changes from “true” to “false”. The probe is now done reading and has saved the value in a data table.
2. Record the VWC number in the field book along with the Record Number.
3. Collect 3 soil moisture measurements at the center of the sampling point, then 3 measurements 2 meters away from the center in each cardinal direction.
4. Move to next sample site.

**Vegetation and Percent Cover**

At each sampling point, the vegetation type, approximate percent land cover, and approximate percent canopy cover should be recorded. This task should be performed by the person handling the field book.

1. The dominant vegetation type should be chosen for each location. Percent cover will be approximate, but try to be consistent with your assignments. Types of vegetation are:
   a. Dry Meadow—Grassy without standing water
   b. Enclosed Meadow—Grassy area without overhead canopy cover, surrounded by forest
   c. Krummholz—Short, scrubby trees
   d. Rock/Talus—Flat rock or boulder field
   e. Snow
   f. Spruce/Fir Forest—Coniferous forest
   g. Tundra—High elevation, very short and turf-like
   h. Wet Meadow—Grassy with standing water
   i. Water
   j. Other—record description
2. Take a photo of the site. Take a picture standing 4 meters from the center of the plot and facing north. Include the entire plot (center, 2 meters N, S, W, and W). In the field book, record the number of the picture in the camera.

**Trimble GeoXT/Juno and TerraSync Operation Instructions**

*The GeoXT has a touch screen and a stylus for operation. Make sure to use only the plastic tip of the stylus and not the other pen tip.*

1. To begin using the GeoXT, lightly press the power button located on the front part of the device.
2. Load the TerraSync software through the “Start” menu or by pressing “F1” on the side panel.

*TerraSync has five pages that can be accessed from the dropdown menu in the upper left-hand corner of the screen. Different pages allow you to view or create data, check your location, and check satellite status. Below is a brief description of each page and how it will be used in this survey.*
The first page in the dropdown menu is “Map”. This page will display any layers within a loaded data file and will display a red x that denotes your current location.

The second page is “Data”. This page allows you to open files or create new files. Within the options on this page, you can modify features, add features, or add information to existing features.

The third page is “Navigation”. This page contains options to set a target location and assists the user in navigating to the point.

The fourth page is “Status”. In this page, you can see a map of current satellite positions and view the predicted satellite geometry for your location.

The fifth page is “Setup”. You will use this page to turn GPS on or off and check or change settings.

3. Go to the “Setup” page by selecting “Setup” from the upper left-hand dropdown menu. Press the “Coordinate System” button and make sure the system is set to UTM 13 North NAD83. Press the “GPS” button on the upper right-hand side to turn on the satellite receiver. It will take a few minutes for the instrument to establish satellite connection and orient. Make sure the front side of the unit is facing the sky.

4. Go to the “Status” page and select “Skyplot” from the dropdown menu below the “Status” menu button. This sub-page will show you the satellites that are currently in view and will allow you to adjust the acceptable PDOP level.

PDOP (Position Dilution of Precision) is a measure of the quality of a GPS position. For this survey, maximum allowable PDOP is 8. Preferred PDOP level is 6 which is the middle position on the slider bar between “Productivity” and “Precision”. If you are having difficulty measuring a point, move the slider one to two increments to the left (toward “Productivity). If this does not solve the problem, go to the “Plan” sub-page from the menu under the “Status” menu. This sub-page shows the predicted PDOP for your location. If the PDOP is over 10, you will have to wait until the satellite geometry improves. The “Plan” graph can help you determine your next low-PDOP window.

5. Next, load the ‘SMSurveyBaseLayers_All’ file. To do this, select “Data” from the upper left-hand menu, then select “Existing File” from the dropdown menu beneath “Data”, highlight the desired file from the listed files, and press the “Open” button on the upper right-hand side of the screen. Press “OK” in the window that pops up confirming an antenna height of 0.000 meters (i.e. no antenna).

6. Go to the “Map” page. You should see a map of the Loch Vale area with an array of points. These points are in a layer named “SMSurvey_TargetPts”. This is the only layer you will modify in this survey. The layer contains 151 points and is meant to be a loose guide for sampling locations. You will create positions for these points and give them a unique identifier.

7. To select a target point, zoom to your destination area and highlight a point in “SMSurvey_TargetPts”. This can be done by using the pointer tools dropdown menu located
just below the “Map” menu in the upper left-hand part of the screen. The pointer menu can be used to select, zoom in, zoom out, pan, digitize or measure.

8. Once you have a point selected, press the “Options” dropdown menu (to the right of the points menu), select “Set Nav Target”, and then chose your selected point that appears in the next menu window. You can either navigate to that point using the “Nav” page or by following the progress of your red “x” on the “Map” page.

You do not have to sample exactly where the target points fall on the map. You should aim for that general area, but you can modify based on what you see on the ground. You need an area about 2 meters by 2 meters to sample soil moisture. A variety of vegetation cover and topography is desired, so the collection points should be as close to the target points as possible.

9. Once you reach your target area and select your sampling site, mark the middle with a flag and write the ID number on the flag if you are separated from the soil moisture crew. The ID number is a 4-digit code and should be: Digit 1 = Trimble unit number, Digit 2 = Sampling day (1, 2, or 3), Digits 3 and 4 = sample location number for that day (e.g. 4110 would indicate that Trimble unit 4 collected this point on the first day of sampling and it was the 10th point collected that day).

10. Stand directly over your flag and from the “Map” page, highlight your target point and press the “Options” menu. Select “Update Selected Feature”. This will transfer you to the “Data” page.

11. Under “Options” on the “Data” page, press “Logging Interval” and make sure you are set to collect a position every 1 second (1s) and that accuracy is set to “Carrier”. If you are experiencing difficulty collecting a point, try switching accuracy to “Code”. Carrier is more accurate and preferred, but Code processing is acceptable.

12. Press the “Log” button on the upper right side of the screen and choose the “Update Feature” radial button then press “OK”. Log a minimum of 120 positions.

13. While collecting positions, check “Mark as updated” and enter the proper “Id” code (use the numeric pad located next to the word “TerraSync” in the blue toolbar at the bottom of the screen). If you are with the soil moisture crew, enter the record number on the TDR for the first soil moisture measurement taken at that sampling point.

14. Press the “OK” button to stop recording and save the new information for this point. The point will now relocate on the map and will have a known position rather than a question mark when you select that point. Be sure not to write over points that already have spatial survey information.
FIGURE A1. FIELD SAMPLING MAP. BLACK CROSSES INDICATE TARGET SAMPLING POINTS THAT WERE USED FOR RESEARCHER NAVIGATION. THIS MAP CONTAINS 151 TARGET POINTS. 107 SAMPLE POINTS WERE ACTUALLY COLLECTED AND 10 WERE ADDED DIGITALLY OVER WATER AND ON PROMINENT ROCK OUTFORIZINGS.
APPENDIX B—GEODATABASE CONSTRUCTION

Figure A2 shows the geodatabase schema including all aspatial tables, relationship classes, feature datasets, feature classes, and raster files. The tables were carefully designed to consolidate repetitive information while maintaining full data accessibility from the SMSS point feature class. The flat data table was normalized into ten aspatial tables and relationships were constructed to link tables based on unique identifiers. A relationship class was also made to link the main aspatial table to the point feature class that contained the spatial location of each sample point.

Three main tables were constructed that included unique record information in three categories: Raw Soil Moisture measurements, GPS Detailed Information, and Field Site Information. The Field Site Information table was related directly to the SMSS Points Feature Class. This table is the main hub table that is linked to all other tables. Because the Raw Soil Moisture and GPS Detailed Information tables contain unique entries for each record, these tables include the Unique Sample_ID for each record and can easily be related to the SMSS Point Feature Class at any time.

The other tables in the geodatabase contain consolidated information about vegetation classification, GPS unit, Sample Date, Soil Moisture Probe, and Research Team. The researchers collected sample points in teams and each researcher could be on multiple teams during any given sampling day. This resulted in a many-to-many relationship between the Research Team table and the Researcher Name table. This relationship was resolved by adding an additional table that contains information regarding which teams in which each researcher participated.

Once the conceptual framework was developed for the database design, tables were constructed in ArcCatalog (Esri, Inc) using built in functionality. Fields and data types were specified for each table to be compatible with data types found in the import table. Data types also had to be compatible between table fields that were used to link tables through relationship classes. Data were imported to populate each table directly from the flat table of compiled field data.

After populating the tables, relationships were built using ArcCatalog’s Relationship Class wizard. Tables were linked based on pre-determined primary and foreign keys and had either one-to-one or one-to-many cardinality. These relationships persist for the lifetime of the geodatabase and can be used to query records, select features, or run analysis within ArcGIS Desktop. The querying capability is somewhat limited to the ArcGIS framework and full SQL commands are not available. However, the geodatabase provides enough functionality to meet the needs of this dataset.

In addition to normalizing the field data, the geodatabase was loaded with basemaps and data layers to be used in future analyses and displays. These include aspect in the study area, site hydrography, target sampling points, study area boundary, and trails feature classes.
Additional raster files such as multiple aerial photos, a ten meter digital elevation model surface, and derived slope and aspect grids are included in the geodatabase.

**FIGURE A2. SMSS GEODATABASE SCHEMA.** ASPATIAL TABLES ARE DEPICTED WITH FIELD HEADINGS IN YELLOW BOXES, RELATIONSHIPS ARE DEPICTED BY GRAY BOXES, THE FEATURE DATASET IS DENOTED BY THE ORANGE BOX, FEATURE CLASSES ARE GREEN, AND RASTER FILES ARE REPRESENTED IN PINK.
**APPENDIX C—GEOSTATISTICS EQUATIONS, DEFINITIONS, & IMPLICATIONS**

**Ordinary Least Squares (OLS) Regression**

Ordinary Least Squares is one of the most common regression methods. The method defines a global linear equation that fits the data by minimizing the distance squared between data points and the regression line. The model can be used to evaluate relationships between dependent and predictor variables. The OLS model can also be used to predict unknown dependent values where explanatory variables are known. This approach generates a single equation for the entire dataset (Esri, Inc.).

General Equation:  \( y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \) where \( y \) is the dependent variable, \( \beta \) are the coefficients, \( X \) are the explanatory variables, and \( \epsilon \) is the random error term.

SMSS Outcome:  \( y = -0.000655 \cdot \text{PercentCover} + 0.094993 \cdot \text{VegType} - 0.002161 \cdot \text{Slope} - 0.232570 \)

Implications: Vegetation Type has a positive relationship with VWC while Percent Cover and Slope have a negative relationship with VWC. Diagnostic statistics are needed to assess this model. Regression models are powerful tools, but can lead to serious errors if the model is misspecified or the residuals contain significant patterns. Misspecification is a result of missing explanatory variables. Patterns in the residuals may appear as non-normal distribution of residuals, significant spatial clustering of residual, or non-linear relationships. Each of these conditions must be evaluated to assess the OLS model. The OLS Regression Tool in ArcGIS provides the following statistics to evaluate the model.

**Coefficient of Determination (Adjusted R\(^2\))**

The Adjusted R\(^2\) measure the ability of the model to explain the variability in the dependent variable. Values range from 0 to 1 and value closer to 1 model the dependent variable more closely. The Adjusted R\(^2\) is more robust than the Multiple R\(^2\) because it is adjusted for the number of variables in the regression model (Esri, Inc.).

The SMSS OLS model had an Adjusted R\(^2\) of 0.61. This value was greater than the 0.50 threshold desired and thus is a passing measure for evaluation of the model.

**Variance Inflation Factor (VIF)**

The VIF assesses redundancy between variables. If two variables are explaining the same portion of the variability in the dependent variable, this value will be high. The lower the value, the less explanatory variables interact (Esri, Inc.). This relationship can be visualized graphically by plotting the explanatory variables against one another and fitting a univariate regression model. The relationship should not show strong correlation. If it does, this demonstrates
multicollinearity and both variables should not be included in the multivariate regression model.

The SMSS OLS model had VIF values of 1.21, 1.46, and 1.23 for PercentCover, VegType, and Slope respectively. These values are well below the 7.5 cutoff value showing that the explanatory variables are not redundant and pass this test.

**Corrected Akaike’s Information Criterion (AICc)**

AICc measures model fit and is useful when comparing between differences models. This measure can also be used to determine smoothing parameters or search window radii for interpolation or GWR techniques. This statistic does not have a threshold value. Lower values indicate better model fit. A change of even three units is significant for this statistic (Hurvich, Simonoff and Tsai 1998, Kim and Kim 2007, Esri, Inc.).

The AICc value for the SMSS OLS model was -122.1762.

**Jarque-Bera (JB) Statistic**

The JB statistic tests the numeric distribution of the OLS residuals. The residuals are the distances squared between the final regression line and the data points. The residuals represent error not explained by the explanatory variables. When the test is significant (low p-value) the residuals are not normally distributed and some bias exists in the model. This bias can potentially be removed by checking for missing explanatory variables, transforming explanatory variable data, or evaluating the removal of outliers. The JB statistic should not be significant in order to identify a properly specified OLS model (Esri, Inc.).

The JB statistic for the SMSS OLS model was 4.9234 with a p-value of 0.085289. This is not significant at the 0.05 level. A higher value would be desirable, so this test should be accepted with caution since it falls so near the cutoff.

**Kroenker’s Studentized Bruesch-Pagan (Kroenker’s BP) Statistic**

The KB statistic evaluates the relationship of the explanatory variables and the dependent variable spatially throughout the study area. If this relationship remains consistent, then a global model is appropriate for the modeling the dependent variable. If the relationship changes significantly throughout different regions of the study area, then a local model should be used. If Kroenker’s BP statistics has a significant p-value, the relationship is not consistent and demonstrates nonstationarity. If this is the case, Geographically Weighted Regression should be evaluated (Esri, Inc.).

The Kroenker’s BP statistic for the SMSS OLS model was 28.9100 with a p-value of 0.000002. This indicates that the relationship between the explanatory variables and the dependent
variables is not consistent everywhere in the modeled study area. Because of this result, the global OLS model is not the most appropriate for this study. A Geographically Weighted Regression model was evaluated to incorporate local relationships.

**Moran’s Index (MI)**
The OLS residuals can be mapped in space to give some insight into the spatial distribution rather than the numeric distribution of the residuals. MI is an evaluation of spatial autocorrelation. It tests for clustering or dispersion against an expected distribution based on complete spatial randomness (CSR). Values for the MI values range from -1 to 1 with negative values indicative of dispersion, positive values related to clustering, and zero representing the expected CSR. A passing result of this model is a value near zero with a non-significant p-value indicating no statistically significant difference from CSR in the spatial distribution of the residuals (Propastin 2006, Esri, Inc.).

The MI for the SMSS OLS residuals was 0.175238 with a p-value of 0.555171. This shows the residuals exhibit no statistically significant spatial clustering or dispersion. These results support a properly specified OLS model.

**Spatial Weights Matrix (SWM)**
A SWM specifies the type of relationship that exists in the dataset. It provides the structure for certain statistics in which to operate. MI requires a spatial weights matrix as do other hot spot and clustering statistics. The conceptualization of the SWM should reflect the real spatial relationship of the data as much as possible. The SWM is built around the input data and can be specified as a neighborhood or distance relationship with many modifying parameters. The SWM used in this study was an inverse distance weighting conceptualization to the second power (Esri, Inc.).

**Geographically Weighted Regression (GWR)**
GWR is a local regression method that calculates explanatory variable coefficient for each input point based on the surrounding points in a specified Kernel and Bandwidth. The Kernel can be fixed or adaptive and the bandwidth can be determined based on statistical measures or a defined distance or number of neighbors. GWR models can have severe problems if not properly evaluated. The best way to evaluate a GWR model is to first run the dependent variable and explanatory variables through an OLS model and examine the model using the diagnostics statistics for each explanatory variable. If the OLS model is properly specified, the same model in GWR can be trusted. The output diagnostic statistics for GWR are not as robust in evaluating the model as those associated with OLS. MI should be run on the GWR residuals to ensure there is no statistically significant pattern in the residuals (Foody 2003, Propastin 2006, Esri, Inc.).
In the SMSS GWR model, a fixed kernel with an AICc defined bandwidth was used. An adaptive kernel was explored, but did not alter the outcome. The AICc method chooses the bandwidth by finding the lowest AICc associated with a search neighborhood. The bandwidth for the SMSS GWR model was 266.208 meters.

The AICc of the GWR model was -140.9467 which is lower compared to the AICc of -122.1762 for OLS. The GWR AICc is nearly 19 units lower than the OLS AICc showing the GWR model has a better fit. The Adjusted R² is also higher at 0.71 indicating higher performance.

The MI on the GWR residuals was -0.061323 with a p-value of 0.865530. This indicates that the residuals are statistically no different from the expected distribution based on CSR.

**Inverse Distance Weighting (IDW)**

Interpolation methods estimate values at unknown or un-sampled locations based on the information provided at known points. The IDW method is an exact interpolator meaning values fall within the minimum and maximum values of the input data. The method is one of the simplest and uses distance explicitly to define the weight a nearby value will have in determining the value at an unknown location. The nearer an input point, the more influential is will be in predicting the unknown value (O'Sullivan and Unwin 2003).

The equation for IDW is

\[
Z_{\text{ij}} = \frac{\sum_{i=1}^{n} \frac{Z_i}{d_i^r}}{\sum_{i=1}^{n} \frac{1}{d_i^r}}
\]

where \(Z_{\text{ij}}\) is the value at an unknown location, \(Z_i\) is the value at neighboring points, and \(d_i\) is the distance between the known and prediction location raised to the power of \(r\).

Important parameters for IDW are defining the power function and the search window. The distance relationship can be specified at different exponential powers to determine the rate distance decay. As the power \(r\) of the inverse distance increases, the rate of decay increases meaning farther values have less influence on the unknown value. Exponents of one or two are most common. The SMSS study used an exponent of two.

The number of neighboring input values that will be included in the calculation is determined by the shape and size of the search window. The search window can be circular or elliptical to deal with anisotropic relationships. The number of neighbors can be set to a minimum and maximum number where data values may be sparse such as near the edges. The radius of the window can be fixed or adaptive to the number of neighbors. The window can also be broken up into sectors to include data in all directions around the point to be estimated. This can help with directional bias (Esri, Inc.).
For the SMSS study, the search window was specified as a single sector, circular window with a minimum of 10 neighbors and a maximum of 20 neighborhoods. The bandwidth was specified to match the range of the semivariogram (see below). Variations on this method were explored, but differences were minor.

**Ordinary Kriging (OK)**

OK uses the underlying spatial relationship between data to build a model to predict a continuous surface. OK takes advantage of spatial autocorrelation and uses that relationship to define an appropriate way to predict unknown areas. The OK method first samples the input data and models the relationship between the variance in value and the distance between points. The parameters for this interpolation method are diverse and include metrics to define the expression of the semivariogram, the shape of the model fit to the data points, the search window shape and size, and the minimum or maximum number of neighbors to use. This information is used to fit a model from which to predict unknown locations (O'Sullivan and Unwin 2003, Esri, Inc.).

The OK function is given by

\[
\hat{Z}(s_o) = \sum_{i=1}^{n} \lambda_i Z(s_i)
\]

where \( \hat{Z}(s_o) \) is the value given to an estimated point \( s_o \), \( \lambda_i \) is an unknown weight, \( Z(s_i) \) is the value at location \( i \), and \( n \) is the sampled number of neighbors.

The parameters for the SMSS OK interpolations were a semivariogram with twelve lags, a Stable model type, a single circular sector search window with a range specified by the semivariogram, and a minimum of ten neighbors and a maximum of twenty neighbors.

**Cross Validation**

Cross validation is a part of the output of the Geostatistical Analyst Toolbar results in ArcGIS 10 when performing IDW or OK. Cross validation measures provide summary measures of error and allow comparison between interpolation methods. The Mean Error is the average difference between observed and predicted values. The Root Mean Square Error (RMSE) shows how well the model predicts observed values. Low values for both error statistics indicate a better model (Esri, Inc.).

The Mean Error is given by

\[
\frac{\sum_{i=1}^{n} (\hat{Z}(s_i) - z(s_i))}{n}
\]

and the RMSE is given by

\[
\sqrt{\frac{\sum_{i=1}^{n} (\hat{Z}(s_i) - z(s_i))^2}{n}}
\]

where \( \hat{Z}(s_i) \) is the measured value, \( z(s_i) \) is the predicted value, and \( n \) is the number of samples.

The Mean Error and RMSE for all interpolations in this study were very low (Table 2).
The Predicted Regression Function shows the type (negative or positive sign of slope) of relationship between predicted and observed values, the strength of the relationship (magnitude of slope), and the error (intercept). The Error Regression Function shows the same relationship for the residuals (Propastin 2006, Esri, Inc. n.d.). Both equations follow the form of a linear regression equation of \( y = mx + b \) where \( y \) is the measured value, \( x \) is the predicted or error value, \( m \) is the slope, and \( b \) is the intercept.

The equations for the SMSS interpolations can be found in Table 2.
APPENDIX D—DERIVING CONTINUOUS SURFACES FOR VEGETATION TYPE & PERCENT COVER

Data derivatives

Because the interpolation methods using the field measurements alone did not yield the kind of precision desired in the VWC surface, a new approach was explored. For both OLS and GWR, the dependent variable can be predicted based on the regression equation if the explanatory variables are known. In order to predict VWC at the grid points, the goal was to derive continuous slope, vegetation type, and percent cover surfaces for the study area. If these could be derived, VWC could be predicted at selected points throughout the watershed. The lidar data provided a 1 meter DEM surface. The DEM was used to generate a slope surface with 1 meter resolution using the Slope Tool in ArcGIS 10 (Esri, Inc.). The other two explanatory variables required a more involved classification process to derive continuous surfaces from IKONOS-2 imagery and the 1 meter DEM.

Vegetation Type Classification

The IKONOS-2 image was classified using a Maximum Likelihood Supervised Classification in ENVI (ITT, Inc.) with seven classes (Water/Shadow, Dry Meadow, Wet Meadow, Forest, Rock, Snow, and GrassyRocky). Post-processing of the classified image using a KnowledgeBase (Erdas IMAGINE) separated the classes further into 10 classes (Tables A1 and A2). Water and Shadow were separated using slope and elevation surfaces from the LiDAR 1 meter DEM to located flat areas at specified elevations. This resulted in some high shadow classified as water in the upper basins of the Loch Vale watershed, but the water bodies in the SMSS study area were distinguished with high accuracy. The Meadow classes were further refined using slope and elevation to pull out true wet meadows (less than 4% slope) from Dry Meadows, and to distinguish Tundra at elevations over 3300 meters from meadows in the lower basin. Similarly, the Forest class was separated into Krummholz (high elevation scrubby trees) from lower old-growth forest. The end product was a classified image with ten classes that corresponded well with other aerial photographs of the watershed as well as expert knowledge of the landscape (Figure A3). The classified image was further processed in ArcGIS 10 (Esri, Inc.) using the Focal Statistics Tool to determine the majority Vegetation Type in a 3x3 neighborhood. This approach emulated the field method where researchers determined the dominate vegetation type two meters in each direction from the center of the sampling site (Figure A4).

<table>
<thead>
<tr>
<th>Vegetation Type</th>
<th>Class Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1</td>
</tr>
<tr>
<td>Snow</td>
<td>2</td>
</tr>
<tr>
<td>Tundra</td>
<td>3</td>
</tr>
<tr>
<td>Wet Meadow</td>
<td>4</td>
</tr>
<tr>
<td>Dry Meadow</td>
<td>5</td>
</tr>
<tr>
<td>Krummholz</td>
<td>6</td>
</tr>
<tr>
<td>Forest</td>
<td>7</td>
</tr>
<tr>
<td>GrassyRocky</td>
<td>8</td>
</tr>
<tr>
<td>Shadow</td>
<td>9</td>
</tr>
<tr>
<td>Rock</td>
<td>10</td>
</tr>
</tbody>
</table>

TABLE A1. CLASSIFIED VEGETATION TYPES AND CORRESPONDING CODES
Percent Cover

Determining Percent Cover from the classified image required an innovative geoprocessing solution. Figure A4 shows an example of this process for a cell with its surrounding 3x3 neighborhood. The center raster cell had a value of 4 and surrounding neighborhood cells of 4, 4, 4, 2, 5, 5, and 4. The majority statistic returned for this center pixel would be 4 (Wet Meadow) and is the derived value for the vegetation type explanatory variable. In order to find the percent cover for the major vegetation type, the number of cells making up that majority is needed. The following solution addresses this need. First each vegetation type was recategorized to a logarithmic scale (Table A3). Then a Focal Statistics analysis was run to sum the cells in a 3x3 neighborhood. This sum acted as a counter for each class. In the example, the sum of the logarithmic values would be 25020 indicating two 5’s, five 4’s, and two 2’s in the neighborhood. The center point of each raster cell was then converted to a point in a shapefile and the point carried the value assigned by the sum focal analysis. The value field of the shapefile was converted to a string then to a list using a python script. The list was then sorted and the last item in the list was pulled into a new field. The last item in the list was the highest number in the string and represented the number of cells in a neighborhood that made up the majority. To deal with ties, another field was created to count the number of

<table>
<thead>
<tr>
<th>Class</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Slope==0</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==2, Elev&gt;3290, Elev&lt;3315</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==2, Elev&gt;3090, Elev&lt;3163</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==2, Elev&gt;3460, Elev&lt;3465</td>
</tr>
<tr>
<td>Snow</td>
<td>MaxL==1</td>
</tr>
<tr>
<td>Tundra</td>
<td>MaxL==3, Elev&gt;3400</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==7, Elev&gt;3300</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==5, Slope&lt;=10, Elev&gt;3300</td>
</tr>
<tr>
<td>Wet Meadow</td>
<td>MaxL==3, Elev&lt;=3300, Slope&lt;=5</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==7, Elev&lt;=3300, Slope&lt;=5</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==5, Slope&lt;=4</td>
</tr>
<tr>
<td>Dry Meadow</td>
<td>MaxL==3, Elev&lt;=3400</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==7, Elev&lt;=3400</td>
</tr>
<tr>
<td></td>
<td>OR MaxL==5, Slope&lt;=10, Slope&gt;4</td>
</tr>
<tr>
<td>Krummholz</td>
<td>MaxL==4, Elev&gt;3300</td>
</tr>
<tr>
<td>Forest</td>
<td>MaxL==4</td>
</tr>
<tr>
<td>GrassyRocky</td>
<td>MaxL==5, Slope&gt;10</td>
</tr>
<tr>
<td>Shadow</td>
<td>MaxL==2, Slope&gt;5</td>
</tr>
<tr>
<td>Rock</td>
<td>MaxL==6</td>
</tr>
<tr>
<td></td>
<td>OR MaxL&gt;0</td>
</tr>
</tbody>
</table>

Table A2. Knowledge Base rules designed to separate seven classes into ten from a Maximum Likelihood Supervised Classification (MaxL), a 1 meter Slope layer (Slope), and a 1 meter Elevation Surface (Elev).

Table A3. Vegetation Class Codes and Logarithmic Reclassification Codes.
high values in the string. The python script is shown in Figure A5. The Focal Statistics Tool output data type is signed 32-bit integer. This meant the method could only handle nine rather than 10 classes because 9,000,000,000 is above the range of values handled by the data type. A quick solution was to set non-vegetated cover types (Water, Snow, Rock, and Shadow) to zero so the counter values would not exceed the limit of the data type.

![Figure A4](image)

**FIGURE A4. PROCESS FLOW TO DERIVE VEGETATION TYPE AND PERCENT VEGETATION COVER FROM THE CLASSIFIED IKONOS-2 IMAGE. ARROWS INDICATE TOOLS USED IN ARCGIS 10 UNLESS DENOTED BY (PYTHON SCRIPT) IN WHICH CASE THESE STEPS WERE CARRIED OUT USING THE SCRIPT SHOWN IN FIGURE A6.**

The shapefile was converted back to a 1 meter raster with the high number in the VALUE field. To calculate Percent Cover, vegetation type needed to be considered. Some vegetation types like meadow and tundra are thick and cover the ground more completely than grasses growing in rocky areas (GrassyRocky class) and Forest which does not permit a full understory to develop. The ground cover in forest is sparser under a needle tree canopy than in the open. Therefore, both the Forest and GrassyRocky classes were only weighted half as much in the percent cover calculation. Cover types which do not accommodate vegetation (Water, Snow, Shadow, and Rock) were assigned a zero percent cover value. The percent cover calculation was carried out in Erdas IMAGINE using the Knowledge Engineer to create a binary mask for each of the groups requiring different calculations (Table A4). The classified image was multiplied by the mask to extract the desired cover types for the individual percent cover calculations, an equation to calculate percent was applied to the separate masked images, and then the outputs were summed back into a single raster surface (Figure A6). This Percent Cover raster was not a perfect estimation of percent cover, but still produced passing results in the OLS regression model with derived Slope and Vegetation Type when values were extracted to the field samplings points (Figure 6C).
# Import arc geoprocessing module
import arcpy

## Note: to rerun the script, paste the fullpath and file name
## of the shapefile or geodatabase feature class from ArcCatalog between the quotes below. The lower
## "r" is important so don't delete it. e.g. r"C:\Folder\filename.shp"

# Set variable to the point shapefile.
shp = r"C:\Files\Winter_2011\ROMO_Project\Lochvale_veg_points.gdb\expertclass_lochvale_log3x3sum"

# Add Mjr_Count and Tie_Count fields to the point shapefile.
arcpy.AddField_management (shp, "Mjr_Count", "LONG")
arcpy.AddField_management (shp, "Tie_Count", "LONG")

# Place an update cursor in the point shapefile table
# move to the first record in the table.
cur = arcpy.UpdateCursor(shp)
row = cur.next()

# Loop through each record in the table.
while row:
    # Extract summed grid code value
    sum = row.getValue("GRID_CODE")
    # Convert the grid code value to a string
    str_sum = str(sum)
    # Convert the string to a list
    sum_list = list(str_sum)
    # Sort the list smallest to largest
    srt_sum_list = sorted(sum_list)
    # Select the largest number in the list (last number of the sequence)
    mrj_count = srt_sum_list[-1]
    # Count the number of times the largest number occurs (number of ties)
    tie_count = sum_list.count(mrj_count)
    # Add the largest number to the Mjr_Count field
    row.setValue("Mjr_Count", mrj_count)
    # Add the number of ties to the tie count field
    row.setValue("Tie_Count", tie_count)

# Import arc geoprocessing module
import arcpy

## Note: to rerun the script, paste the fullpath and file name
## of the shapefile or geodatabase feature class from ArcCatalog between the quotes below. The lower
## "r" is important so don't delete it. e.g. r"C:\Folder\filename.shp"

# Set variable to the point shapefile.
shp = r"C:\Files\Winter_2011\ROMO_Project\Lochvale_veg_points.gdb\expertclass_lochvale_log3x3sum"

# Add Mjr_Count and Tie_Count fields to the point shapefile.
arcpy.AddField_management (shp, "Mjr_Count", "LONG")
arcpy.AddField_management (shp, "Tie_Count", "LONG")

# Place an update cursor in the point shapefile table
# move to the first record in the table.
cur = arcpy.UpdateCursor(shp)
row = cur.next()

# Loop through each record in the table.
while row:
    # Extract summed grid code value
    sum = row.getValue("GRID_CODE")
    # Convert the grid code value to a string
    str_sum = str(sum)
    # Convert the string to a list
    sum_list = list(str_sum)
    # Sort the list smallest to largest
    srt_sum_list = sorted(sum_list)
    # Select the largest number in the list (last number of the sequence)
    mrj_count = srt_sum_list[-1]
    # Count the number of times the largest number occurs (number of ties)
    tie_count = sum_list.count(mrj_count)
    # Add the largest number to the Mjr_Count field
    row.setValue("Mjr_Count", mrj_count)
    # Add the number of ties to the tie count field
    row.setValue("Tie_Count", tie_count)

# Import arc geoprocessing module
import arcpy

## Note: to rerun the script, paste the fullpath and file name
## of the shapefile or geodatabase feature class from ArcCatalog between the quotes below. The lower
## "r" is important so don't delete it. e.g. r"C:\Folder\filename.shp"

# Set variable to the point shapefile.
shp = r"C:\Files\Winter_2011\ROMO_Project\Lochvale_veg_points.gdb\expertclass_lochvale_log3x3sum"

# Add Mjr_Count and Tie_Count fields to the point shapefile.
arcpy.AddField_management (shp, "Mjr_Count", "LONG")
arcpy.AddField_management (shp, "Tie_Count", "LONG")

# Place an update cursor in the point shapefile table
# move to the first record in the table.
cur = arcpy.UpdateCursor(shp)
row = cur.next()

# Loop through each record in the table.
while row:
    # Extract summed grid code value
    sum = row.getValue("GRID_CODE")
    # Convert the grid code value to a string
    str_sum = str(sum)
    # Convert the string to a list
    sum_list = list(str_sum)
    # Sort the list smallest to largest
    srt_sum_list = sorted(sum_list)
    # Select the largest number in the list (last number of the sequence)
    mrj_count = srt_sum_list[-1]
    # Count the number of times the largest number occurs (number of ties)
    tie_count = sum_list.count(mrj_count)
    # Add the largest number to the Mjr_Count field
    row.setValue("Mjr_Count", mrj_count)
    # Add the number of ties to the tie count field
    row.setValue("Tie_Count", tie_count)

# Import arc geoprocessing module
import arcpy

## Note: to rerun the script, paste the fullpath and file name
## of the shapefile or geodatabase feature class from ArcCatalog between the quotes below. The lower
## "r" is important so don't delete it. e.g. r"C:\Folder\filename.shp"

# Set variable to the point shapefile.
shp = r"C:\Files\Winter_2011\ROMO_Project\Lochvale_veg_points.gdb\expertclass_lochvale_log3x3sum"

# Add Mjr_Count and Tie_Count fields to the point shapefile.
arcpy.AddField_management (shp, "Mjr_Count", "LONG")
arcpy.AddField_management (shp, "Tie_Count", "LONG")

# Place an update cursor in the point shapefile table
# move to the first record in the table.
cur = arcpy.UpdateCursor(shp)
row = cur.next()

# Loop through each record in the table.
while row:
    # Extract summed grid code value
    sum = row.getValue("GRID_CODE")
    # Convert the grid code value to a string
    str_sum = str(sum)
    # Convert the string to a list
    sum_list = list(str_sum)
    # Sort the list smallest to largest
    srt_sum_list = sorted(sum_list)
    # Select the largest number in the list (last number of the sequence)
    mrj_count = srt_sum_list[-1]
    # Count the number of times the largest number occurs (number of ties)
    tie_count = sum_list.count(mrj_count)
    # Add the largest number to the Mjr_Count field
    row.setValue("Mjr_Count", mrj_count)
    # Add the number of ties to the tie count field
    row.setValue("Tie_Count", tie_count)
# Update the row values
cur.updateRow(row)
# Delete temporary variables
del sum, str_sum, sum_list, srt_sum_list, mrj_count, tie_count
# Move to the next row
row = cur.next()

# Delete row and cursor objects, arcpy module.
del cur, row, arcpy

FIGURE A5. PYTHON SCRIPT DESIGNED TO COUNT THE NUMBER OF MAJORITY VALUES IN A THREE BY THREE NEIGHBORHOOD. THE INPUT DATA ARE POINTS GENERATED FROM THE CENTER OF ONE METER RASTER CELLS. THE VALUES ARE SUMMED LOGARITHMIC CLASSIFIED DATA SUCH THAT EACH PLACEHOLDER IN THE STRING REPRESENTS A DIFFERENT CLASS FROM THE ORIGINAL VEGETATION TYPE RASTER IMAGE. THE SCRIPT PULLS OUT THE HIGHEST VALUE IN THE STRING AS WELL AS TIES TO GIVE A COUNT OF THE MAJORITY VEGETATION CLASSIFICATION IN EACH CELL NEIGHBORHOOD.

FIGURE A6. MODEL DESIGNED TO CALCULATE PERCENT COVER BASED ON HIGH COUNT VALUE AND CLASSIFIED VEGETATION TYPE. THE HIGH COUNT RASTER IS THE PYTHON SCRIPT OUTPUT (FIGURE A6) AND THE VEGETATION TYPES ASSOCIATED WITH EACH MASK ARE LISTED IN TABLE A3. THE FINAL OUTPUT OF THIS MODEL WAS THE DERIVED PERCENT COVER CONTINUOUS SURFACE USED IN THE REGRESSION MODEL TO PREDICT VWC AT POINTS IN THE 10 METER AND 5 METER GRIDS (FIGURE 6C).