# PREDICTIVE SOIL MAPPING TO IMPROVE THE PHYSICAL BASIS OF DISTRIBUTED ECOHYDROLOGICAL MODELS IN ARID ENVIRONMENTS 

## NM WRRI Technical Completion Report No. 382

Colby W. Brungard<br>Mikayla J. Allan



Excavated soil profile showing soils and vegetation common in the study area. The soil in this picture was excavated until a root restricting petrocalcic horizon was encountered.

New Mexico Water Resources Research Institute
New Mexico State University
MSC 3167, P.O. Box 30001
Las Cruces, New Mexico 88003-0001
(575) 646-4337 email: nmwrri@.nmsu.edu

By<br>Colby W. Brungard, Assistant Professor Mikayla J. Allan, Graduate Assistant<br>Department of Plant and Environmental Sciences<br>New Mexico State University

TECHNICAL COMPLETION REPORT
Account Number 128377
Technical Completion Report \#382

December 2019

New Mexico Water Resources Research Institute in cooperation with the
Department of Plant \& Environmental Sciences New Mexico State University

The research on which this report is based was financed in part by the U.S. Department of the Interior, Geological Survey, through the New Mexico Water Resources Research Institute.


## DISCLAIMER

The purpose of the NM Water Resources Research Institute (NM WRRI) technical reports is to provide a timely outlet for research results obtained on projects supported in whole or in part by the institute. Through these reports the NM WRRI promotes the free exchange of information and ideas and hopes to stimulate thoughtful discussions and actions that may lead to resolution of water problems. The NM WRRI, through peer review of draft reports, attempts to substantiate the accuracy of information contained within its reports, but the views expressed are those of the authors and do not necessarily reflect those of the NM WRRI or its reviewers. Contents of this publication do not necessarily reflect the views and policies of the Department of the Interior, nor does the mention of trade names or commercial products constitute their endorsement by the United States government.

## Table of Contents

DISCLAIMER ..... iii
TABLE OF CONTENTS ..... iv
LIST OF FIGURES ..... v
LIST OF TABLES ..... vi
ABSTRACT ..... vii
INTRODUCTION ..... 1
Site Information ..... 1
Data Collection and Laboratory Analysis ..... 2
Analysis ..... 4
RESULTS AND DISCUSSION ..... 5
CONCLUSIONS ..... 14
REFERENCES ..... 15
APPENDIX A CODE FOR NESTED SPATIAL SAMPLING ..... 18
APPENDIX B SOIL PROFILE DESCRIPTIONS ..... 25
APPENDIX C DATA USED FOR GEOSTATISTICAL ANALYSIS ..... 124
APPENDIX D CODE USED FOR GEOSTATISTICAL MODELING OF SAND AND CLAY ..... 133
APPENDIX E CODE USED FOR GEOSTATISTICAL MODELING OF SOIL DEPTH ..... 187

## LIST OF FIGURES

Figure 1. Study area. The solid white line outlines the landform (ballena) used to define the study area. The Tromble Weir Watershed (TWW) boundary is the black line inside the larger study area. The circles are sampling locations. The star on the inset map shows the location of the study area in southern NM

Figure 2. A schematic of the modified balanced nested sampling design used to identify sampling locations. Numbers indicate the physical distance between hierarchical stages. Adapted from Webster et al. (2006).3

Figure 3. Variogram models of sand concentration. The variogram model for $15-30 \mathrm{~cm}$ is visually different than the other variogram models because stable model fit required established 'bin' sizes over which average semivariance values were calculated. Bin sizes were set to equal distances between the sampling levels as defined in the sample design.

Figure 4. Variogram models of clay concentration. The variogram model for $5-15 \mathrm{~cm}$ is visually different than the other variogram models because stable model fit required established 'bin' sizes over which average semivariance values were calculated. Bin sizes were set to equal distances between the sampling levels as defined in the sample design.10

Figure 5. Variogram model of soil depth....................................................................................................... 11
Figure 6. Sand concentration (\%) predictions and prediction uncertainty. Left hand figures are the predictions of sand concentration at $0-5 \mathrm{~cm}(\mathrm{~A}), 5-15 \mathrm{~cm}(\mathrm{C}), 15-30 \mathrm{~cm}(\mathrm{E})$, and $30-60 \mathrm{~cm}(\mathrm{G})$. Right hand figures are prediction uncertainty $(\mathrm{SD}=$ standard deviation of kriging variance) at 0-5 $\mathrm{cm}(B), 5-15 \mathrm{~cm}(\mathrm{D}), 15-30 \mathrm{~cm}(\mathrm{~F})$, and $30-60 \mathrm{~cm}(H)$. Filled circles on right hand figures are the sampling locations.

Figure 7. Clay concentration (\%) predictions and prediction uncertainty. Left hand figures are the predictions of sand concentration at $0-5 \mathrm{~cm}(\mathrm{~A}), 5-15 \mathrm{~cm}(\mathrm{C}), 15-30 \mathrm{~cm}(\mathrm{E})$, and $30-60 \mathrm{~cm}(\mathrm{G})$. Right hand figures are prediction uncertainty $(\mathrm{SD}=$ standard deviation of kriging variance) at 0-5 $\mathrm{cm}(B), 5-15 \mathrm{~cm}(\mathrm{D}), 15-30 \mathrm{~cm}(\mathrm{~F})$, and $30-60 \mathrm{~cm}(H)$. Filled circles on right hand figures are the sampling locations.

Figure 8. Soil depth predictions and prediction uncertainty. Top figure (A) is prediction. Bottom figure $(B)$ is prediction uncertainty $(\mathrm{SD}=$ standard deviation of the kriging variance). Filled circles on bottom figure are the sampling locations.

## LIST OF TABLES

Table 1. Terrain variables, unit, and interpretation of the variables.
Table 2. Summary statistics of sand and clay concentrations and soil depth for both measured values from field samplling and estimated values from soil survey. All values are in percent except for soil depth which is given in $\mathrm{cm} . \mathrm{SD}=$ standard deviation. $\mathrm{n}=$ number of observations.

Table 3. Terrain variables with statistically significant linear relationships between sand and clay at standard soil depths and soil depth. Multiple linear regression used if multiple variables significant.
Table 4. Variogram model parameters for geostatistical modeling of sand and clay concentrations and soil depth. .. 7


#### Abstract

Spatial patterns in soil properties such as particle size and soil depth significantly affect hydrological and ecological processes. Finely spatially resolved information about the spatial distribution of soil properties is needed for hydrological and ecohydrological modeling. This information is not currently provided by existing small-scale soil maps. This research uses geostatistical methods to interpolate soil depth as well as sand and clay concentrations at four harmonized depth increments ( $0-5,5-15,15-30$, and $30-60 \mathrm{~cm}$ ) within a single alluvial landform surrounding a small, heavily instrumented watershed at the Jornada Experimental Range in southern New Mexico. Soil depth and sand and clay concentration observations were obtained from two sampling campaigns. Each variable was analyzed for anisotropy and statistically significant relationships with nine terrain variables to account for non-stationarity. Spherical, circular, and exponential variogram models were fitted to all sand and clay concentrations and soil depth and compared using root-mean-square-error (RMSE) derived from leave-one-out cross validation. RMSE ranged between 4.8 and $5.9 \%$ for sand and between 1.3 and $1.9 \%$ for clay. RMSE for soil depth was 37.7 cm . In general, sand had a shorter range of spatial autocorrelation and a smaller nugget than did clay at all depths. The range of spatial autocorrelation for sand was between 150 and 225 m , while clay had a much more variable range of values between 90 and 3206 m . In general, nugget values were relatively low because of the sampling design that had a minimum distance of 3 m , which appears to have captured most of the small-scale variability. Spatial prediction was done using Kriging with External Drift. Uncertainty in sand and clay concentration predictions were low while the uncertainty of soil depth predictions was greater. Interpolated variables and the associated prediction uncertainty will be used to improve the parameterization of future ecohydrological modeling applications.


Keywords: soils, ecohydrological models, soil maps, Jornada Experimental Range, spatial patterns


## INTRODUCTION

Spatial patterns in soil properties such as soil particle size and soil depth significantly affect ecohydrological patterns and processes such as soil moisture, runoff generation, subsurface and groundwater flow (Freer et al., 2002; Stieglitz et al., 2003; Gribb et al., 2009) as well as vegetation community composition (English et al., 2005; Gremer et al., 2015). Finely spatially resolved information regarding the spatial distribution of soil properties is needed for improving ecohydrological models (Tesfa et al., 2009; Wood et al., 2011). Within the United States, the Natural Resource Conservation Service (NRCS) national soils database (SSURGO) has been the main source for soil property information used for ecohydrological modeling (Anderson et al., 2006). This information is provided in mapping units delineated with sharp boundaries. In arid western USA rangelands, these mapping units are often composed of multiple soil components, which often are not spatially represented. This representation of soils is discrete, highly generalized, and is often unsuited to work with other landscape data (Tesfa et al., 2009). Although a soil survey is an excellent tool to optimize land use and management, it was designed for county-level land management and does not provide detailed information required for environmental modeling or site-specific management (Moore et al., 1993; Duffera et al., 2007).

Spatially explicit soil information, specifically particle size and depth, at spatial resolutions finer than that provided by SSURGO, is needed to refine and constrain the parametrization of distributed ecohydrological models (Méndez-Barroso et al., 2016). The purpose of this study was to model soil particle size and soil depth for a small, heavily instrumented watershed with the ultimate goal of incorporating the resulting information into spatially distributed ecohydrological models. This study utilized geostatistics and Kriging with External Drift to produce this information.

## METHODS

## Site Information

The study area was co-located with the Tromble Weir Watershed (TWW, $32^{\circ} 35^{\prime} 4.62^{\prime \prime} \mathrm{N}$, $106^{\circ} 36^{\prime} 8.815 \mathrm{~W}$ ) in the northern part of the Chihuahuan Desert, 37 km northeast of Las Cruces, New Mexico at the Jornada Experimental Range. The TWW is a small experimental watershed ( 4.7 ha ) on the bajada of the San Andres Mountains (Templeton et al., 2014). Vegetation in the TWW is a mixed shrubland that has undergone historical changes in plant dominance throughout time. Throughout the TWW, hydrological instruments including an eddy covariance tower, flumes, and multiple soil moisture sensors were installed (Anderson and Vivoni, 2016). To capture soil spatial variability while avoiding excessive soil disturbance, soil sampling was performed across the entire ballena surrounding the TWW (Figure 1). A ballena ( $s p$. whale) is a remnant of fan alluvium that is distinctly round-topped and occurs along mountain fronts as groups of semiparallel ridges that reflect the incision of parallel drainageways (Peterson, 1981). This ballena was identified by selecting the map unit delineation from an existing NRCS soil survey, which adequately captured the entire landform.

Climate in this area is typical of the northern Chihuahuan Desert, with a mean annual precipitation of 247 $\mathrm{mm}, 53 \%$ of which occurs between July 1 and September 30 (Gibbens and Beck, 1987; Wainwright, 2006). Summer precipitation is mostly from short-duration high-intensity convective storms over small areas, while winter precipitation is mostly associated with low-intensity frontal storms over broad areas (Wainwright, 2006). Vegetative composition at the TWW includes four dominant plant communities being black grama grassland (Bouteloua eriopoda), creosotebush (Larrea tridentate), honey mesquite (Prosopis glandulosa), and tarbush (Flourensia cernua) (Anderson and Vivoni, 2016).


Figure 1. Study area. The solid white line outlines the landform (ballena) used to define the study area. The Tromble Weir Watershed (TWW) boundary is the black line inside the larger study area. The circles are sampling locations. The star on the inset map shows the location of the study area in southern NM.

Typical soils within the Jornada basin consist of Entisols and Aridisols. Within the study area, Aridisols were the dominant soil order. The soil map unit delineation used to define the study area boundary was a Doña Ana-Chutman Complex, with 1 to $10 \%$ slopes (Soil Survey Staff, 2017). This complex is comprised of the Doña Ana (fine-loamy, mixed, superactive, thermic Typic Calciargids) and the Chutman (fineloamy, mixed, superactive, thermic Typic Haplocalcids) series (Soil Survey Staff, 2017). Doña Ana soils ( $65 \%$ of the map unit) occur on fan piedmonts with alluvium parent material. Typical textures are sandy loam and sandy clay loam. Chutman soil ( $35 \%$ of the map unit) occur in drainageways and toeslopes of fan piedmonts and also contains alluvial parent material (Soil Survey Staff, 2017). Textures range from silt loam to clay loam. Diagnostic features for the soil series include an ochric epipedon (A and Bw horizons), cambic (Bw horizon), and calcic (Bk1 and Bk2 horizons) horizons (Soil Survey Staff, 2017).

## Data Collection and Laboratory Analysis

Sampling locations were generated using a modified balanced multi-stage sampling design (Webster et al., 2006). The concept behind this method is to hierarchically subdivide sampling distances using multiple stages to capture accurately enough observations to compute a semi-variogram with modest effort (Webster et al., 2006) This approach chooses several starting nodes and chooses subsequent nodes at set decreasing intervals in random directions (Figure 2). Soil sampling is then performed at the location of each sampling stage.

Starting nodes were generated by extracting centroids of seven spatially compact clusters (Walvoort et al., 2010) Seven levels of subsequent sampling locations were then chosen by decreasing distances by a factor of three from an initial sampling distance of 800 m . This resulted in sampling locations separated by the following distances: $800 \mathrm{~m}, 267 \mathrm{~m}, 89 \mathrm{~m}, 30 \mathrm{~m}, 10 \mathrm{~m}$, and 3 m . An initial distance of 800 m was chosen as it was approximately one-half the length of the longest axis of the study area. This resulted in 49 sampling locations (seven levels with seven samples at each level). Implementation of this balanced hierarchical sampling algorithm was done using a custom script written in R (R Core Team, 2018), which is included in Appendix A.


Figure 2. A schematic of the modified balanced nested sampling design used to identify sampling locations. Numbers indicate the physical distance between hierarchical stages. Adapted from Webster et al. (2006).

All 49 sampling locations were visited and sampled in June 2017. Sampling locations were navigated to by GPS, and physical soil sampling was located within a 3-m radius of the generated point according to estimated GPS accuracy. All sampling was done in intershrub areas. At each sampling location, general site information including slope, aspect, surface ground cover, and slope shape were collected. Soil profiles were then excavated ( $30-50 \mathrm{~cm}$ wide) to a depth of either 150 cm or to a root restrictive petrocalcic horizon. If a root restrictive horizon was not reached after approximately 100 cm , an auger was used to excavate from $100-150 \mathrm{~cm}$.

After each sampling location was excavated, $100-200 \mathrm{~g}$ soil samples were collected by genetic horizons ( $\sim 2-4$ horizons per soil sampling location) and soil profiles were described according to Schoeneberger et al. (2012). Soil profile descriptions included horizon depth and designation, rock fragments (percent, type, size), structure (grade, size, type), carbonate development stage, hand texture (textural class and clay percentage), and ped and void surface features (percent, distinction, continuity, kind, location). Field data descriptions are included in Appendix B.
After samples from each soil horizon were collected, air dried, and sieved to $<2 \mathrm{~mm}$, soil particle size distribution (i.e., sand, silt, and clay concentration) was measured by the hydrometer method. Briefly, 100 g of air-dry soil was mixed in a blender cup with $10 \mathrm{ml} 5 \%$ Sodium Hexametaphosphate and deionized water for five minutes. The mixture was quantitatively transferred to a graduated cylinder and the cylinder was then filled to 1000 ml . A stirring plunger was used to mix the sample for $\sim 30$ strokes, the hydrometer was inserted, and readings taken at 40 seconds, and again after six hours. With each set of measurements, the temperature of the hydrometer samples was recorded, and a blank was used to adjust for any differences found between actual readings and the blank.
In addition to the soil information collected at the 49 sampling locations, sand, silt, and clay concentrations from an exisiting dataset of 20 locations within the TWW were also included (Anderson, 2013). These samples were collected in June 2013 during the installation of soil moisture and temperature probes from the depth ranges of 0-7, 7-17, and $17-27 \mathrm{~cm}$ using a split-tube corer (AMS, 2"x12" Signature Split Soil Core Sampler) except for the depth range of $17-27 \mathrm{~cm}$ at one location, where excessively rocky soil prevented deep sampling (Anderson, 2013). Sand, silt, and clay was determined for each depth increment using the hydrometer method (Anderson, 2013). Because of the relatively shallow sampling depth, these additional 20 samples were used only for sand and clay predictions from the top three depth increments and were not included in the analysis of the $30-60 \mathrm{~cm}$ increment or for predicting soil depth. All numerical data used for analysis are included in Appendix C.

## Analysis

Measurements of sand, silt, and clay concentrations at each sampling location were standardized to the following depth increments by depth weighted median to facilitate interpolation: $0-5 \mathrm{~cm}, 5-15 \mathrm{~cm}$, $15-30 \mathrm{~cm}, 30-60 \mathrm{~cm}$ (Beaudette et al., 2013; Science Committee, 2015). Observations at depths below 60 cm were not included in the analysis because there were too few for robust analysis, and because soil moisture does not often infiltrate past this depth in this system (Schreiner-Mcgraw and Vivoni, 2018). Soil depth was defined as the distance to a root restricting horizon (e.g., petrocalcic or bedrock). If a root restricting horizon was not encountered before reaching the excavation depth of 150 cm , the soil depth was recorded as 150 cm .

Measured values at standardized depth increments were compared against estimated values of sand, silt, clay, and soil depth from the soil survey. Estimated values for the soil map unit delineation used to define the study area boundary were obtained from the Soil Survey Geographic (SSURGO) database for "White Sands Missile Range, New Mexico, Parts of Doña Ana, Lincoln, Otero, Sierra and Socorro Counties." (Soil Survey Staff, 2017). This data is created during soil survey by estimating low, representative, and high values of texture values (and other physical and chemical properties) for each component. Multiple components often exist in a single map unit. Component horizon values were standardized by weighted median to the same depth intervals as the measured data and used to calculate a weighted average for each depth interval using the proportion of the components in the map unit ( $65 \%$ Doña Ana, $30 \%$ Chutum).

Measured sand and clay were analyzed separately by depth interval. Silt was not analyzed as it could be calculated from the sum of sand + clay. Spatial non-stationarity, a key assumption of geostatistics, was evaluated by linear regression between soil texture fractions and soil depth and nine terrain variables. Terrain variables were derived from a $5-\mathrm{m}$ digital elevation model using SAGA-GIS and are listed in Table 1 (Conrad et al., 2015). Each variable was regressed against sand, clay, and depth one at a time. Significant variables ( $p<0.01$ ) were included in the kriging equation to remove any trend. If multiple variables were significant, the variables were used in a multiple-linear regression. Any non-significant variables ( $\mathrm{p}<0.01$ ) in the multilple-linear regression were removed and the process repeated until all variables were significant. Multiple-liner regression was only used in the analysis of sand at the $30-60 \mathrm{~cm}$. Anisotropy was found to exist and was included in each model at 120 degrees, which was approximately the longitudinal direction of the landform.
Spherical, circular, and exponential variogram models were fit to sand and clay concentrations and soil depth and compared using root-mean-square-error (RMSE) derived from leave-one-out cross validation. The model that returned the lowest RMSE was selected for each variable. If variogram models did not converge, 'bin' sizes equal to the distances used in the sampling design were used over which average semivariance was calculated.

Spatial prediction was done using Kriging with External Drift, which specifically accounts for correlations with auxiliary variables (i.e., terrain variables) (Hengl, 2007). Interpolated variables and the associated standard deviation (a measure of prediction uncertainty) were produced and are the digital soil mapping outputs that are intended for inclusion in future ecohydrological applications. Standard deviation was calculated as: $\sqrt{\text { kriging }}$ variance.

All analysis was performed using RStudio (RStudio Team, 2016) and the following packages: aqp (Beaudette et al., 2013); car (Fox and Weisberg, 2011); dplyr (Wickham et al., 2018); e1071 (Meyer et al., 2019); ggplot2 (Wickham, 2016); gstat (Pebesma, 2004); openxlsx (Walker, 2018); plyr (Wickham, 2011); raster (Hijmans, 2014); RColorBrewer (Neuwirth, 2014); and rgdal and sp (Bivand et al., 2018). R code used for the geostatistical analysis of sand and clay concentrations are included in Appendix D. R code use for the geostatistical analysis of depth are included in Appendix E.

Table 1. Terrain variables, unit, and interpretation of the variables.

| Terrain Variable | units | Interpretation |
| :--- | :--- | :--- |
| Northness* | degrees | Direction from north that the slope is facing |
| Convergence Index | unitless | Flow convergence and divergence |
| Cross-sectional Curvature | unitless | Flow convergence and divergence |
| Elevation | meters | Vertical distance above mean sea level |
| Flow Accumulation | $\mathrm{m}^{2}$ | Size of upslope area |
| Longitudinal Curvature | unitless | Flow convergence and divergence |
| Slope | degrees | Slope steepness |
| Topographic Wetness Index | unitless | Potential wetness |
| Valley Depth | meters | Elevation below the nearest ridge |
| ${ }^{\text {}}$ Northness calculated as: cosine(aspect) |  |  |

## RESULTS AND DISCUSSION

Summary statistics of measured sand and clay concentrations by standardized depth increment and total soil depth are presented in Table 2. Sand concentrations ranged between $35 \%$ and $84 \%$. Clay concentrations ranged between $3 \%$ and $19 \%$ (Table 3 ). Clay concentrations were about one-half as variable as were sand concentrations as quantified by the standard deviations of each harmonized horizon, but the variability was $\leq 8 \%$ for both sand and clay. Average depth to restrictive horizon was 77 cm , but this is calculated including ten observations that stopped at 150 cm because of limitations in the depth of excavation, which results in biased summary statistics. Summary statistics of estimated sand and clay concentrations by standardized depth increment and total soil depth from soil survey are also presented in Table 2. Although soil survey underestimated sand and overestimated clay at all most all depth increments when compared with the measured values, both measured and estimated values are similar. The maximum absolute difference between measured and estimated values is $19 \%$ clay at the $30-60 \mathrm{~cm}$ depth increment. The minimum absolute difference between measured and estimated values is $2 \%$ sand at the $5-15 \mathrm{~cm}$. However, total depth is poorly approximated by soil survey likely because of the spatial variability of total soil depth.

Table 2. Summary statistics of sand and clay concentrations and soil depth for both measured values from field samplling and estimated values from soil survey. All values are in percent except for soil depth which is given in cm. $S D=$ standard deviation. $n=$ number of observations.

|  | Depth | Measured values from field sampling |  |  |  |  |  | Estimated values from soil survey |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $n$ | Min. | Median | Mean | Max. | SD | Low | RV | High |
| Sand | 0-5 | 67 | 35 | 64 | 64 | 80 | 6 | 39 | 58 | 75 |
|  | 5-15 | 67 | 46 | 64 | 63 | 77 | 5 | 35 | 58 | 75 |
|  | 15-30 | 67 | 46 | 63 | 64 | 84 | 6 | 36 | 58 | 75 |
|  | 30-60 | 41 | 39 | 61 | 61 | 78 | 8 | 34 | 56 | 75 |
| Clay | 0-5 | 67 | 3 | 8 | 8 | 15 | 3 | 8 | 17 | 22 |
|  | 5-15 | 67 | 4 | 8 | 8 | 16 | 3 | 8 | 19 | 23 |
|  | 15-30 | 67 | 3 | 8 | 8 | 16 | 3 | 9 | 20 | 24 |
|  | 30-60 | 41 | 3 | 9 | 9 | 19 | 3 | 21 | 28 | 35 |
| Depth | - | 47 | 22 | 58 | 77 | 150 | 44 | - | 150+ | - |

[^0]Statistically significant linear relationships between sand and clay concentrations by harmonized depth increment and total soil depth are presented in Table 3. Elevation, topographic wetness index, valley depth, and cross-sectional curvature were linearly related with sand and clay concentrations and soil depth. The relationships are generally weak ( $\mathrm{R}^{2}<0.3$ for most variables), which is expected given the relatively small size of the study area and the general uniformity of the soil in this single landform. The relatively weak linear relationships may also be a result of the $5-\mathrm{m}$ resolution of the digital elevation model used to derive terrain parameters. It is possible that stronger relationships may have been found had terrain derivatives been calculated using a DEM with a finer resolution (e.g., $<1 \mathrm{~m}$ ) because such resolution could potentially capture variability in micro-relief between shrubs and inter-plant spaces that likely govern soil redistribution. The resolution of the DEM may also explain the generally increasing strength of the linear relationships between soil texture fractions with increasing depth. We assume that soil texture fractions become less related to surface features that redistribute soil particles, as the depth increases and internal pedological processes become more dominant. However, any interpretation of these relationships must be treated with caution as the range in sand and clay fractions was relatively narrow and the amount of variance explained by each variable was generally low.

Table 3. Terrain variables with statistically significant linear relationships between sand and clay at standard soil depths and soil depth. Multiple linear regression used if multiple variables significant.

|  | Depth Interval | Terrain Variable | $p$-value | ${ }^{*}$ Multiple-R ${ }^{2}$ |
| :--- | :--- | :--- | :--- | :--- |
| Sand | $0-5$ | Elevation | 0.005 | 0.113 |
|  | $5-15$ | Elevation | 0.006 | 0.109 |
|  | $15-30$ | Topographic Wetness Index | 0.001 | 0.169 |
|  | $30-60$ | Elevation | 0.000 | 0.555 |
| Clay | $0-5$ | Longitudinal Curvature | 0.002 |  |
|  | $5-15$ | Valley Depth | 0.001 | 0.165 |
|  | Valley Depth | 0.000 | 0.279 |  |
|  | $35-30$ | Valley Depth | 0.000 | 0.272 |
| Depth | - | Cross-sectional Curvature | 0.011 | 0.184 |
| ${ }^{*} R^{2}$ | Elevation | 0.008 | 0.145 |  |

${ }^{*} R^{2}$ is the coefficient of determination and indicates the variance explained
Variogram parameters are reported in Table 4. The RMSE is a measure of model performance, with lower values indicating a better model fit. In general, the RMSE values for sand are larger than RMSE values for clay, which is likely because observed clay concentrations were less variable than sand concentrations (Table 2). However, the RMSE for both sand and clay was relatively low indicating a good model fit and was approximately within the estimated accuracy of the hydrometer method used to measure the soil texture fractions. The RMSE for both sand and clay was similar to the range of measured values for each horizon (compare tables 2 and 4).

Table 4. Variogram model parameters for geostatistical modeling of sand and clay concentrations and soil depth.

|  | $\begin{gathered} \text { Depth } \\ \mathrm{cm} \end{gathered}$ | Model | $\begin{gathered} \text { RMSE } \\ \% \\ \hline \end{gathered}$ | $\begin{gathered} \text { Range } \\ \mathrm{m} \end{gathered}$ | Nugget $\mathrm{C}_{0}$ | $\begin{gathered} \text { Partial Sill } \\ \text { C } \\ \hline \end{gathered}$ | $\begin{gathered} \text { Sill } \\ \mathrm{C}_{0}+\mathrm{C} \\ \hline \end{gathered}$ | Nugget-to-Sill ratio $\mathrm{C}_{0} /\left(\mathrm{C}_{0}+\mathrm{C}\right)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sand | $0-5 \mathrm{~cm}$ | Cir | 5.8 | 185.1 | 10.7 | 27.5 | 38.2 | 0.3 |
|  | $5-15 \mathrm{~cm}$ | Cir | 4.8 | 152.0 | 18.7 | 7.7 | 26.4 | 0.7 |
|  | $15-30 \mathrm{~cm}$ | Cir | 5.9 | 35.0 | 5.7 | 30.4 | 36.1 | 0.2 |
|  | $30-60 \mathrm{~cm}$ | Sph | 5.1 | 225.6 | 15.8 | 8.9 | 24.6 | 0.6 |
| Clay | $0-5 \mathrm{~cm}$ | Cir | 1.9 | 3206.5 | 1.3 | 33.5 | 34.9 | 0.0 |
|  | $5-15 \mathrm{~cm}$ | Sph | 1.3 | 521.7 | 0.6 | 4.0 | 4.6 | 0.1 |
|  | $15-30 \mathrm{~cm}$ | Cir | 1.4 | 500.3 | 0.9 | 5.0 | 6.0 | 0.2 |
|  | $30-60 \mathrm{~cm}$ | Cir | 1.5 | 90.4 | 0.3 | 1.7 | 2.0 | 0.1 |
| Depth | - | Sph | 37.7* | 50.9 | 610.6 | 1067.8 | 1678.4 | 0.4 |

*Soil depth RMSE, nuggest and sill reported in cm
The RMSE for soil depth was 37.7 cm (Table 3). This RMSE value is very similar to the values reported by Tesfa et al. (2009) who modeled soil depth in a semi-arid environment using machine learning and Liu et al. (2013), who modeled soil depth in a humid area using an analytical terrain evolution model. Based on these results it may be that $\sim 35 \mathrm{~cm}$ is the average error that can be expected in soil depth predictions. This suggests that soil depth is rather difficult to accurately model. This is most likely because soil depth is controlled by processes such as deposition and weathering that are currently not approximated with terrain derivatives. However, the difficulty in dealing with observations where the soil is deeper than the excavation depth (e.g., $>150 \mathrm{~cm}$ in this study) is a problem that needs to be resolved. One possible approach may be to use maximum likelihood regression combined with kriging (Knotters et al., 1995). However, if soil depth estimates are required with greater than about 30 cm precision, geophysical methods such as ground penetrating radar may be more suited to estimating soil depth (Sucre et al., 2011).
In general, sand had a shorter range of spatial autocorrelation and a smaller nugget than did clay at all depths (Table 4). Semi-variogram ranges are interpreted as the range of spatial correlation. Samples separated by distances shorter than the range are spatially correlated and contribute to kriging predictions (Cambardella et al., 1994). Excluding the variogram models for sand $15-30 \mathrm{~cm}$ and clay $5-15 \mathrm{~cm}$, which required separate bin sizes for stable model fit, the range of spatial autocorrelation for sand was between 150 and 225 m , while clay had a much more variable range of values between 90 and 3206 m . The discrepancy in variogram ranges between sand and clay is a bit surprising, particularly the range of clay $0-5 \mathrm{~cm}$. We are unsure of the exact mechanism that would cause such differences, but it is likely related to the general paucity of clay in this landform and the sparseness of the sampling design. The variogram range of soil depth was much less than that of sand or clay and should be used to set the maximum distance between nodes in any subsequent grid sampling of this area.

The nugget value is the semivariance at separation distance equal to zero and can be interpreted as variability that is undetectable at the resolution of mapping (Cambardella et al., 1994). In general, nugget values were relatively low because of the sampling design that had a minimum distance of 3 m , which appears to have captured most of the small-scale variability. The nugget-to-sill ratio is an indicator of the strength of spatial dependency (Cambardella et al., 1994). Smaller ratios indicate stronger spatial dependency while a ratio of one would indicate no spatial correlation. Following the spatial dependence structure of Cambardella et al. (1994), all soil properties had moderate to high spatial dependence.

Figures 3, 4, and 5 plot each variogram model. The wide dispersion of points around the lines in each figure is a result of the relatively few observations used to build the variograms. In general, 150-200 observations are recommended for a robust variogram model, which is considerably more than were available in this study; and we acknowledge that if more observations were available for fitting each model, variograms would be more robust (Webster and Oliver, 2007). However, because this study occurred in an area that is part of ongoing hydrologic investigations, options for more intensive sampling may require the use of geophysical instruments that minimize sampling disturbance to produce enough observations for robust variogram modeling.

Figures 6,7 , and 8 show the kriging predictions and prediction uncertainty. The spatial patterns in Figs. 6E, 6 G , and 7 G (sand $15-30 \mathrm{~cm}$ and $30-60 \mathrm{~cm}$ and clay $30-60 \mathrm{~cm}$ predictions) are a result of the correlation with the terrain variables. The linear patterns in these predictions generally show a decrease in sand and an increase in clay concentrations. These patterns can be explained by the presense of shallow gullies in these locations where erosion has removed the overlying coarser textured soils and lowered the land surface closer to the siltier formation that underlies this area (the whitebottom surface; Gile et al., 1981). The gradient of soil depth (Fig. 8A, shallower in the west and gradually deepening to the east) is a result of the relationship between elevation and soil depth. Soils are generally shallower in the west where decreasing elevation exposes the relatively planar petrocalcic horizons that run throughout the landform. The spotty nature of soil depth predictions and uncertainty (Fig. 8B in particular) is a result of the range of autocorrelation ( $\sim 50 \mathrm{~m}$ ). Although not as visually obvious, all predictions (Figs. 6, 7, and 8) show the effect of including elevation as a variable.
Uncertainty in sand and clay concentration predictions were low, while the uncertainty of soil depth predictions was fairly large (compare areas of low vs. high uncertainty in Figures 6, 7, and 8). A sampling grid with nodes $\sim 50 \mathrm{~m}$ apart (less than the range of the soil depth variogram) would be required to reduce the uncertainty in soil depth predictions. These outputs are in a GIS-ready format and could be used as input to future distributed ecohydrological modeling efforts on the Tromble Weir Watershed.


Figure 3. Variogram models of sand concentration. The variogram model for $15-30 \mathrm{~cm}$ is visually different than the other variogram models because stable model fit required established 'bin' sizes over which average semivariance values were calculated. Bin sizes were set to equal distances between the sampling levels as defined in the sample design.


Figure 4. Variogram models of clay concentration. The variogram model for $5-15 \mathrm{~cm}$ is visually different than the other variogram models because stable model fit required established 'bin' sizes over which average semivariance values were calculated. Bin sizes were set to equal distances between the sampling levels as defined in the sample design.


Figure 5. Variogram model of soil depth.


Figure 6. Sand concentration (\%) predictions and prediction uncertainty. Left hand figures are the predictions of sand concentration at $0-5 \mathrm{~cm}(A), 5-15 \mathrm{~cm}(C), 15-30 \mathrm{~cm}(E)$, and $30-60 \mathrm{~cm}(G)$. Right hand figures are prediction uncertainty ( $S D=$ standard deviation of kriging variance) at 0-5 cm (B), 5-15 $\mathrm{cm}(D), 15-30 \mathrm{~cm}(F)$, and $30-60 \mathrm{~cm}(H)$. Filled circles on right hand figures are the sampling locations.


Figure 7. Clay concentration (\%) predictions and prediction uncertainty. Left hand figures are the predictions of sand concentration at $0-5 \mathrm{~cm}(A), 5-15 \mathrm{~cm}(C), 15-30 \mathrm{~cm}(E)$, and $30-60 \mathrm{~cm}(\mathrm{G})$. Right hand figures are prediction uncertainty ( $S D=$ standard deviation of kriging variance) at 0-5 $\mathrm{cm}(B), 5-15$ $\mathrm{cm}(D), 15-30 \mathrm{~cm}(F)$, and $30-60 \mathrm{~cm}(H)$. Filled circles on right hand figures are the sampling locations.


Figure 8. Soil depth predictions and prediction uncertainty. Top figure (A) is prediction. Bottom figure (B) is prediction uncertainty ( $S D=$ standard deviation of the kriging variance). Filled circles on bottom figure are the sampling locations.

## CONCLUSIONS

Sand and clay at four standardized depth intervals and soil depth were measured in the alluvial landform surrounding the Tromble Weir Watershed in southern New Mexico. Measured values were compared to estimated values from soil survey. Sand and clay were similar between measured and estimated values, while soil depth was overestimated by soil survey and much more variable in measured values. Geostatistical models were fit to observed data. In general, the accuracy of sand and clay concentration models were within the measurement accuracy and predictions are reliable. Soil depth models were less accurate than sand and clay models and had greater uncertainty. Denser observations of soil depth from a grid sampling effort or from geophysical methods are needed to reduce the uncertainty in soil depth predictions. Spatial predictions of sand, clay, and soil depth, and their accompanying uncertainty, may be used to test the effect of more finely resolved soil property values on distributed ecohydrological models and to explore patterns in vegetation density, structure, and distribution.

## REFERENCES

Anderson, C.A. 2013. Assessing Land-Atmosphere Interactions through Distributed Footprint Sampling at Two Eddy Covariance Towers in Semiarid Ecosystems of the Southwestern U.S. Masters's thesis, Arizona State University.
Anderson, R.M., V.I. Koren, and S.M. Reed. 2006. Using SSURGO data to improve Sacramento Model a priori parameter estimates. J. Hydrol. 320(1-2): 103-116.
Anderson, C.A., and E.R. Vivoni. 2016. Impact of land surface states within the flux footprint on daytime land-atmosphere coupling in two semiarid ecosystems of the Southwestern U.S. Water Resour. Res. 52(6): 4785-4800.
Beaudette, D.E., P. Roudier, and A.T. O’Geen. 2013. Algorithms for quantitative pedology: A toolkit for soil scientists. Comput. Geosci. 52: 258-268.
Bivand, R., T. Keitt, and B. Rowlingson. 2018. rgdal: Bindings for the "Geospatial" Data Abstraction Library. R Package Version 1.3-6. https://cran.r-project.org/package=rgdal.
Cambardella, C.A., T.B. Moorman, T.B. Parkin, D.L. Karlen, J.M. Novak, R.F. Turco, and A.E. Konopka. 1994. Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Sci. Soc. Am. J. 58(5): 1501-1511.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8(7): 1991-2007.
Duffera, M., J.G. White, and R. Weisz. 2007. Spatial variability of Southeastern U. S. Coastal Plain soil physical properties : Implications for site-specific management. Geoderma 137: 327-339.
English, N.B., J.F. Weltzin, A. Fravolini, L. Thomas, and D.G. Williams. 2005. The influence of soil texture and vegetation on soil moisture under rainout shelters in a semi-desert grassland. J. Arid Environ. 63(1): 324-343.
Fox, J., and S. Weisberg. 2011. An $\{$ R $\}$ Companion to Applied Regression. Second. Sage, Thousand Oaks $\{\mathrm{CA}\}$.
Freer, J., J.J. McDonnell, K.J. Beven, N.E. Peters, D.A. Burns, R.P. Hooper, B. Aulenbach, and C. Kendall. 2002. The role of bedrock topography on subsurface storm flow. Water Resour. Res. 38(12): 5-1-5-16.
Gibbens, R.P., and R.F. Beck. 1987. Changes in grass basal area and forb densities over a 64-year period on grassland types of the Jornada Experimental Range. J. Range Manag. 41(May): 186-192.
Gile, L.H., J.W. Hawley, and R.B. Grossman. 1981. Soils and geomorphology in the Basin and Range area of southern New Mexico: Guidebook to the Desert Project. Memoir 39—New Mexico Bureau of Mines and Mineral Resources, Soccoro, NM.
Gremer, J.R., J.B. Bradford, S.M. Munson, and M.C. Duniway. 2015. Desert grassland responses to climate and soil moisture suggest divergent vulnerabilities across the southwestern United States. Glob. Chang. Biol. 21(11): 4049-4062.
Gribb, M.M., I. Forkutsa, A. Hansen, D.G. Chandler, and J.P. McNamara. 2009. The Effect of Various Soil Hydraulic Property Estimates on Soil Moisture Simulations. Vadose Zo. J. 8(2): 321.
Hengl, T. 2007. A practical guide to geostatistical mapping of environmental variables. Lulu.com.
Hijmans, R.J. 2014. raster: Geographic data analysis and modeling. R package version 2.2-31. http://CRAN.R-project.org/package=raster.
Knotters, M., D.J. Brus, and J.H.O. Voshaar. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67: 227-246.
Liu, J., X. Chen, H. Lin, H. Liu, and H. Song. 2013. A simple geomorphic-based analytical model for predicting the spatial distribution of soil thickness in headwater hillslopes and catchments. Water Resour. Res. 49(11): 7733-7746.

Méndez-Barroso, L.A., E.R. Vivoni, and G. Mascaro. 2016. Impact of spatially-variable soil thickness and texture on simulated hydrologic conditions in a semiarid watershed in northwest Mexico. Rev. Mex. CIENCIAS GEOLÓGICAS 33(3): 365-377.
Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2019. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.
Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Sci. Soc. Am. J. 57(2): 443-452.
Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes.
Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30: 683-691.
Peterson, F.F. 1981. Landforms of the Basin and Range Province. Nevada Agricultural Experiment Station, University of Nevada Reno, Reno.
R Core Team. 2018. R: A Language and Environment for Statistical Computing.
RStudio Team. 2016. RStudio: Integrated Development Environment for R.
Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field Book for Describing and Sampling Soils, Version 3.0. Natural Resources Conservation Service. National Soil Survey Center, Lincoln, NE.
Schreiner-Mcgraw, A.P., and E. Vivoni. 2018. On the Sensitivity of Hillslope Runoff and Channel Transmission Losses in Arid Piedmont Slopes. Water Resour. Res. Res. 54(May): 1-21.
Science Committee, G.S.M. 2015. Specifications Tiered GlobalSoilMap products.
Soil Survey Staff. 2017. Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for White Sands Missile Range, New Mexico. Available online.
Stieglitz, M., J. Shaman, J. McNamara, V. Engel, J. Shanley, and G.W. Kling. 2003. An approach to understanding hydrologic connectivity on the hillslope and the implications for nutrient transport. Global Biogeochem. Cycles 17(4): 16-1-16-15.
Sucre, E.B., J.W. Tuttle, and T.R. Fox. 2011. The Use of Ground-Penetrating Radar to Accurately Estimate Soil Depth in Rocky Forest Soils. For. Sci. 57(1): 59-66.
Templeton, R.C., E.R. Vivoni, L.A. Méndez-Barroso, N.A. Pierini, C.A. Anderson, A. Rango, A.S. Laliberte, and R.L. Scott. 2014. High-resolution characterization of a semiarid watershed: Implications on evapotranspiration estimates. J. Hydrol. 509: 306-319.
Tesfa, T.K., D.G. Tarboton, D.G. Chandler, and J.P. McNamara. 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45(10).
Wainwright, J. 2006. Climate and climatological variations in the Jornada Basin. p. 44-80. In Havstad, K., Huenneke, L.F., Schlesinger, W.H. (eds.), Structure and Function of a Chihuahuan Desert Ecosystem. Oxford University Press, NY.
Walker, A. 2018. openxlsx: Read, Write and Edit XLSX Files. R package version 4.1.0. https://CRAN.Rproject.org/package=openxlsx.
Walvoort, D.J.J., D.J. Brus, and J.J. de Gruijter. 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36: 1261-1267.
Webster, R., and M.A. Oliver. 2007. Geostatistics for Environmental Scientists. 2 edition. Wiley, Chichester.
Webster, R., S.J. Welham, J.M. Potts, and M.A. Oliver. 2006. Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood. Comput. Geosci. 32(9): 1320-1333.
Wickham, H. 2011. The Split-Apply-Combine Strategy for Data Analysis. J. Stat. Softw. 40(1): 1-29.
Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wickham, H., R. François, L. Henry, and K. Müller. 2018. dplyr: A Grammar of Data Manipulation. R package version 0.7.7. https://CRAN.R-project.org/package=dplyr.

Wood, E.F., J.K. Roundy, T.J. Troy, L.P.H. van Beek, M.F.P. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek, J. Famiglietti, D. Gochis, N. van de Giesen, P. Houser, P.R. Jaffé, S. Kollet, B. Lehner, D.P. Lettenmaier, C. Peters-Lidard, M. Sivapalan, J. Sheffield, A. Wade, and P. Whitehead. 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res. 47(5): W05301.

## APPENDIX A CODE FOR NESTED SPATIAL SAMPLING

The following can be copied and pasted into an R script
\# Nested Spatial Sampling
\# This code implements a modified version of nested sampling in Webster et al., 2006. This code is implemented as follows: first, a polygon representing the study area is loaded, second the user creates a vector of the desired decreasing distances between sample points, thirdly the nestsamp function generates a series of initial sample points (the first level of hierarchy) by extracting centroids of compact clusters. Compact geographic clusters are created using the spcosa packages. Subsequent hierarchical levels are then created from each centroid while being restricted to remain inside the study area boundary.
\# Required arguments for the nestsamp function:
\# poly = polygon to sample in
\# $\mathrm{n}=$ number of samples at each level
\# dists = distances between each hierarchical level. Define before running function.
\# cellSize $=$ cellSize of grid used in spcosa. Start with 50 or greater (i.e., 50 meters) to quickly run, then set smaller to get a grid with higher fidelity to the original polygon.
\# hlevels = number of hierarchical levels to be run. Should match the number of desired hierarchical levels. e.g., if you want 7 levels then this should be 7 .
\#The number of resulting points will be n *hlevels.
\# Note, this code does not exactly follow the Webster et al. paper. Instead of choosing a random vector of length $\mathrm{h}-1$, this code simply samples the point and chooses the next point.

```
# Code written by:
# Colby Brungard, PhD
# Assistant Professor of Pedology
# Department of Plant and Environmental Sciences
# New Mexico State University
# Las Cruces, NM }8800
# cbrung@nmsu.edu
```

\# +1.575.646.1907

## \#Citation

\#' @article\{WEBSTER20061320,

```
#' title = "Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis
of variance and residual maximum likelihood",
#' journal = "Computers & Geosciences",
#' volume = "32",
#' number = "9",
#' pages = "1320-1333",
#' year = "2006",
#' note = "",
#' issn = "0098-3004",
#' doi = "http://dx.doi.org/10.1016/j.cageo.2005.12.002",
#' url = "http://www.sciencedirect.com/science/article/pii/S0098300405002761",
#' author = "R. Webster and S.J. Welham and J.M. Potts and M.A. Oliver",
#' keywords = "Nested sampling",
#' keywords = "Analysis of variance",
#' keywords = "Variance components",
#' keywords = "Variogram",
#' keywords = "Balance",
#' keywords = ""
#' }
#Begin
# Load necessary packages
library(sp)
library(rgdal)
library(spcosa)
library(plyr)
# Set working directory
setwd(".")
```

\# 1. Read in polygon. It is easiest if this is in a projection with meters, e.g., UTM
poly <- readOGR(dsn = ".", layer = "SoilMU26")
\# 2. Geographic distances between each subsequent hierarchy level. This should match the number of hierarchical levels you want minus one. e.g., if you want 7 levels then you should have six distances, because the first level is created using spcosa centroids. All subsequent hierarchical levels will be based off of this first level.
\# Inelegant way to get distances by decreasing factor of 3 . One could also set these manually if a nonexponential decrease was desired.
\# I chose 800 m because it seemed like a good idea and because it was approximatley $1 / 2$ the length of the longest axis of the soil map unit that I was interested in.

```
dists <- vector()
dists[1] <- 800
dists[2] <- dists[1]/3
dists[3] <- dists[2]/3
dists[4] <- dists[3]/3
dists[5] <- dists[4]/3
dists[6] <- dists[5]/3
```

\# To do an imbalanced sample I could just run the following balanced sampling for the number of desired balanced levels, then re-run for the following levels with $1 / 2$ of the sample points selected randomly.
\# 3. Function to apply modified version of fully balanced nested spatial sampling based on Webster et al. 2006

```
nestsamp <- function(poly, n, dists, cellSize, hlevels) {
# poly = polygon to sample in
# n = number of samples at each level
# dists = distances between each hierarchical level. Define before running function.
# cellSize = cellSize of grid used in spcosa. Start with 50 (i.e., }50\mathrm{ meters) to quickly run, then set smaller
to get a grid with higher fidelity to the original polygon.
# hlevels = number of hierarchical levels to be run. Should match the number of desired hierarchical levels.
e.g., if you want }7\mathrm{ levels then this should be 7.
```

\#The number of resulting points will be n *hlevels.

```
spreading the initial sample points across the area by compact clusters.
poly2 <- SpatialPolygons(poly@polygons)
strat <- stratify(poly2, nStrata = n, nTry = 5, cellSize = cellSize)
# Centroids in dataframe format
samp <- as(spsample(strat), "data.frame")
names(samp) <- c('X1', 'X2')
# Identify sampling lcoations for all hierarchical levels past the first level.
hsamps <- list(samp)
#-1 since the first hierarchical level is already done
for(k in 1:(hlevels-1)){
# Generate samples with in each hierarchical level
newSampX <- vector()
newSampY <- vector()
samps2 <- data.frame(matrix(ncol = 2, nrow = nrow(samp)))
for (i in 1:nrow(samp)){
# Generation of random direction
dir <- runif(1, min = 0, max = 360)
# Generation of new point
dx <- dists[k] * sin(dir)
dy <- dists[k] * cos(dir)
newSampX <- hsamps[[k]][i,1] + (dists[k] * sin(dir))
```

\# Define initial sample points (first level of hierarchy) by extracting centroids of compact clusters using
spcosa. One could also use spsample to generate random points in the polygon, but I like this idea of

```
newSampY <- hsamps[[k]][i,2] \(+(\operatorname{dists}[k] * \cos (\operatorname{dir}))\)
\# Convert new points to spatialpointsdataframe and assign projection to use the over function
newSamp <- data.frame(cbind(newSampX, newSampY))
coordinates(newSamp) <- ~ newSampX + newSampY
\(\operatorname{proj} 4\) string \((\) newSamp \()=\operatorname{proj} 4\) string \((\) poly \()\)
\# Is the new point in the boundaries of the polygon? If not, choose another point that is inside the boundaries.
inPoly <- !is.na(over(newSamp, poly))[1,1]
while(inPoly != TRUE) \{
    \# Generation of random direction
    dir \(<-\operatorname{runif}(1, \min =0, \max =360)\)
    \# Generation of new point
    dx <- dists[k] * \(\sin (\) dir \()\)
    dy <- dists[k] * \(\cos (\) dir \()\)
    newSampX <- hsamps[[k]][i,1] + (dists[k] * \(\sin (\) dir \())\)
    newSampY <- hsamps[[k]][i,2] + (dists[k] * \(\cos (\) dir \()\) )
    \# Convert new points to spatialpointsdataframe and assign projection to use the over function
    newSamp <- data.frame(cbind(newSampX, newSampY))
    coordinates(newSamp) <-~ newSampX + newSampY
    \(\operatorname{proj} 4\) string \((\) newSamp \()=\operatorname{proj} 4\) string \((\) poly \()\)
    \# Is the new point in the boundaries of the polygon?
    inPoly <- !is.na(over(newSamp, poly))[1,1]
    \} \# end while
```

```
    samps2[i,] <- data.frame(newSamp)
}# end inner for loop
# Join all samples into a list
hsamps[[k+1]] <- samps2
} # end outer for loop
return(hsamps)
}
# 4. Run nested sampling
try1 <- nestsamp(poly = poly, n = 7, dists = dists, cellSize = 5, hlevels = 7)
plot(poly)
points(try1[[1]], col = 'red', pch = 19)
points(tryl[[2]], col = 'blue', pch = 19)
points(try1[[3]], col = 'black', pch = 19)
points(try1[[4]], col = 'green', pch = 19)
points(tryl[[5]], col = 'orange', pch = 19)
points(try1[[6]], col = 'purple', pch = 19)
points(try1[[7]], col = 'grey', pch = 19)
```

\#Name each plot
$\operatorname{try} 1[[1]] \$ 1 e v e l<-\operatorname{rep}(' L e v e l 1 '$, nrow(try 1[[1]]))
$\operatorname{try} 1[[2]] \$$ level <- rep('Level2', nrow(try1[[2]]))
$\operatorname{try} 1[[3]] \$ l e v e l<-\operatorname{rep}(' L e v e l 3 '$, nrow(try1[[3]]))

```
try1[[4]]$level <- rep('Level4', nrow(try1[[4]]))
try1[[5]]$level <- rep('Level5', nrow(try1[[5]]))
try1[[6]]$level <- rep('Level6', nrow(try1[[6]]))
try1[[7]]$level <- rep('Level7', nrow(try1[[7]]))
```

\# Collapse to dataframe, add unique identifier, and write to csv.
dat <- ldply(try1, data.frame)
dat\$id <- paste0(0,seq(01,nrow(dat)))
write.csv(dat, "./SamplingPoints.csv", row.names = FALSE)
\# Convert to other file formats as needed in qgis as it is easier.
\#I imported the .csv file, assigned the right projection (same as SOILMU26.shp - WGS84 UTM 13N), then saved as .gpx and .kml in WGS84 lat/long geographic coordinates. I also saved these in WGS 84 UTM13N projection as a shapefile
\#End

## APPENDIX B SOIL PROFILE DESCRIPTIONS

The following paired images are the field data sheets collected by Mikalya Allan during her field sampling campaign. Each image pair consists of the front and back of one field sheet.




6

|  |  |  |  | -- Mmpunit sumbia |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mostot | (in) ${ }^{\text {Deith }}$ (em) | Horion | Prid |  | Texture | Rock Frats |  |  | consister | teree | Eiten |
|  | 0-6 | A. |  |  |  | $3 \% \mathrm{~m} \times \mathrm{R}$ \% |  | ion. | Ms ${ }^{\text {mis }}$ | Skx meper | \% S3: cn iso Mut sp too |
|  |  |  | W |  | VFSL | $2 \%$ MxR 6 | SG |  |  |  |  |
|  | 6-53 | BkI |  |  | VESL | 2\% MxRFG ${ }_{\text {a }}$ | FSBK |  |  |  |  |
|  | 53-82 | BK2 |  |  | L | 3\% MXEFG | FSBR |  |  |  |  |
| 5 | 82-105 | BKK | I |  | L |  | SG |  |  |  |  |
| 5 | 105t | BkKM |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |


|  |  |  |  |  |  | $\text { PH Effer Clay } \mathrm{CCE}$ | Notes |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 2 - | VFF RF CAF |  |  | - 8 |  |
| 2 |  | $22 C A N$ | $C D R F C A F$ |  |  | 6 | 土 |
|  |  | 42 CAN | MDRF CAF |  |  | 10 | \# |
|  |  | 73 CAN |  |  |  | 12 | II |
|  |  | . |  |  |  |  | IV |
|  |  |  |  |  |  |  |  |
| \% |  |  |  |  |  |  |  |
| ${ }^{8}$ |  |  |  |  |  |  |  |
| ${ }^{9}$ |  |  |  |  |  |  |  |
| , |  |  |  |  |  |  |  |
| USPA-NRCS ${ }^{\text {a }}$ 2.76 ${ }^{\text {September 2002 }}$ |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |






























































































)



## APPENDIX C DATA USED FOR GEOSTATISTICAL ANALYSIS

Soil observations used for analysis. Table C. 1 is sand, silt, and clay concentrations (measured by hydrometer) as sampled from 49 sample locations by genetic horizon. Table C. 2 is the location of each of the 49 pedons in Table C.1. Table C. 3 is the sand, silt, and clay concentrations (measured by hydrometer) as sampled at 20 locations by standardized depth increment. Table C. 4 is the location of each of the 20 observations in Table C.3. Data in tables C. 3 and C. 4 extracted from Cody Anderson's thesis.

Table C.1. Data used for analysis. PedonID is a unique identifier for each distinct sampling location. Designation is the horizon master and suffix designations used to describe each horizon. A combination of PedonID and Designation can be used as a unique identifier for each horizon. HZ top is the top of the horizon in cm. HZ Bottom is the bottom of the horizon in cm. Sand, Silt, and Clay are concentrations by genetic horizon in percent.

| Pedon <br> ID | Designation | HZ Top | HZ <br> Bottom | Sand | Silt | Clay |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | AB | 0 | 6 | 76 | 17 | 7 |
| 1 | Bk | 6 | 32 | 73 | 20 | 6 |
| 1 | Bkk | 32 | 52 | 78 | 17 | 6 |
| 2 | A | 0 | 6 | 77 | 15 | 9 |
| 2 | Bk1 | 6 | 53 | 77 | 13 | 10 |
| 2 | Bk2 | 53 | 82 | 76 | 14 | 10 |
| 2 | Bkk | 82 | 105 | 74 | 14 | 13 |
| 3 | AB | 0 | 7 | 82 | 9 | 9 |
| 3 | Bk | 7 | 22 | 78 | 15 | 7 |
| 3 | Bkk | 22 | 47 | 81 | 11 | 9 |
| 4 | AB | 0 | 6 | 75 | 15 | 10 |
| 4 | Bk | 6 | 22 | 70 | 22 | 9 |
| 4 | Bkk | 22 | 38 | 63 | 28 | 9 |
| 5 | AB | 0 | 7 | 65 | 28 | 8 |
| 5 | Bk | 7 | 27 | 65 | 25 | 10 |
| 5 | Bkkm | 27 | 36 | 74 | 15 | 11 |
| 6 | AB | 0 | 5 | 68 | 21 | 11 |
| 6 | Bkk | 5 | 29 | 68 | 24 | 8 |
| 6 | Bkkm | 29 | 65 | 67 | 26 | 7 |
| 7 | AB | 0 | 4 | 66 | 25 | 9 |
| 7 | Bkk | 4 | 24 | 65 | 27 | 8 |
| 7 | Bkkm | 24 | 52 | 65 | 26 | 9 |
| 8 | AB | 0 | 7 | 72 | 21 | 7 |
| 8 | Bk | 7 | 17 | 71 | 20 | 9 |
| 8 | Bkk | 17 | 32 | 70 | 21 | 9 |
| 9 | AB | 0 | 8 | 68 | 22 | 10 |
| 9 | Bkk1 | 8 | 29 | 65 | 26 | 9 |
| 9 | Bkk2 | 29 | 47 | 63 | 29 | 8 |
| 10 | AB | 0 | 4 | 66 | 24 | 10 |
| 10 | Bk | 4 | 28 | 62 | 30 | 8 |


| 10 | Bkk 1 | 28 | 62 | 64 | 26 | 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 10 | Bkk2 | 62 | 89 | 59 | 31 | 10 |
| 11 | A | 0 | 8 | 62 | 28 | 10 |
| 11 | Bk1 | 8 | 30 | 69 | 23 | 8 |
| 11 | Bk2 | 30 | 48 | 72 | 19 | 9 |
| 11 | Btk | 48 | 150 | 61 | 28 | 11 |
| 12 | A | 0 | 7 | 61 | 30 | 9 |
| 12 | Btkk | 7 | 32 | 60 | 33 | 7 |
| 12 | Bkkm | 32 | 50 | 60 | 31 | 9 |
| 13 | AB | 0 | 9 | 59 | 31 | 10 |
| 13 | Bk | 9 | 32 | 61 | 32 | 7 |
| 13 | Bkk | 32 | 75 | 65 | 32 | 3 |
| 14 | AB | 0 | 5 | 64 | 22 | 14 |
| 14 | Bkk 1 | 5 | 26 | 66 | 24 | 10 |
| 14 | Bkk2 | 26 | 70 | 59 | 22 | 19 |
| 14 | Bkkm | 70 | 98 | 56 | 31 | 13 |
| 15 | AB | 0 | 7 | 67 | 22 | 12 |
| 15 | Bk | 7 | 32 | 66 | 23 | 11 |
| 15 | Bwk1 | 32 | 60 | 49 | 47 | 4 |
| 15 | Bwk2 | 60 | 150 | 39 | 58 | 3 |
| 16 | AB | 0 | 9 | 64 | 24 | 12 |
| 16 | Bk | 9 | 50 | 46 | 50 | 4 |
| 16 | Btk | 50 | 150 | 61 | 28 | 11 |
| 17 | AB | 0 | 6 | 61 | 25 | 14 |
| 17 | Btk | 6 | 44 | 63 | 25 | 12 |
| 17 | Bt | 44 | 150 | 39 | 58 | 3 |
| 18 | A | 0 | 6 | 60 | 26 | 14 |
| 18 | Btkk | 6 | 36 | 58 | 26 | 16 |
| 18 | Bt1 | 36 | 74 | 48 | 49 | 3 |
| 18 | Bt2 | 74 | 150 | 39 | 58 | 3 |
| 19 | AB | 0 | 5 | 60 | 25 | 15 |
| 19 | Btk | 5 | 27 | 57 | 28 | 15 |
| 19 | Bt1 | 27 | 64 | 52 | 45 | 3 |
| 19 | Bt2 | 64 | 150 | 39 | 57 | 4 |
| 20 | AB | 0 | 9 | 80 | 13 | 7 |
| 20 | Bw1 | 9 | 52 | 61 | 30 | 9 |
| 20 | Bw2 | 52 | 90 | 62 | 26 | 12 |
| 20 | Bwk | 90 | 150 | 51 | 31 | 18 |
| 21 | AB | 0 | 8 | 64 | 28 | 8 |
| 21 | Bk1 | 8 | 25 | 64 | 29 | 7 |
| 21 | Bkk | 25 | 64 | 62 | 29 | 9 |
| 21 | Bk2 | 64 | 117 | 64 | 30 | 7 |


| 22 | AB | 0 | 6 | 63 | 27 | 11 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 22 | Bk | 6 | 48 | 64 | 29 | 8 |
| 22 | Bkk | 48 | 117 | 63 | 33 | 5 |
| 23 | AB | 0 | 5 | 65 | 28 | 8 |
| 23 | Bk | 5 | 27 | 65 | 29 | 6 |
| 23 | Bkk | 27 | 96 | 61 | 33 | 7 |
| 24 | AB | 0 | 7 | 64 | 28 | 9 |
| 24 | Bkk | 7 | 30 | 62 | 30 | 8 |
| 24 | Bkkm | 30 | 48 | 57 | 32 | 11 |
| 25 | AB | 0 | 7 | 70 | 22 | 8 |
| 25 | Bkk | 7 | 27 | 57 | 32 | 11 |
| 25 | Bkkm | 37 | 58 | 56 | 34 | 10 |
| 26 | AB | 0 | 7 | 64 | 28 | 8 |
| 26 | Bkk | 7 | 28 | 58 | 33 | 9 |
| 26 | Bkkm | 28 | 50 | 55 | 34 | 11 |
| 27 | AB | 0 | 6 | 65 | 26 | 9 |
| 27 | Bk | 6 | 30 | 61 | 31 | 8 |
| 27 | Btkk | 30 | 69 | 57 | 29 | 14 |
| 28 | AB | 0 | 9 | 61 | 28 | 11 |
| 28 | Bkk | 9 | 28 | 59 | 32 | 9 |
| 29 | AB | 0 | 6 | 61 | 30 | 9 |
| 29 | Bk | 6 | 32 | 58 | 34 | 8 |
| 29 | Bkk | 32 | 57 | 57 | 33 | 10 |
| 30 | AB | 0 | 7 | 57 | 33 | 10 |
| 30 | Bk | 7 | 28 | 57 | 35 | 8 |
| 30 | Bkk | 28 | 46 | 56 | 35 | 9 |
| 31 | AB | 0 | 5 | 61 | 30 | 9 |
| 31 | Bk | 5 | 22 | 58 | 34 | 8 |
| 31 | Bkkm | 22 | 43 | 61 | 29 | 10 |
| 31 | Btkk | 43 | 70 | 56 | 32 | 12 |
| 32 | A | 0 | 6 | 64 | 25 | 11 |
| 32 | Btk | 6 | 26 | 63 | 29 | 8 |
| 32 | Bkk | 26 | 44 | 63 | 29 | 8 |
| 33 | A | 0 | 10 | 64 | 28 | 8 |
| 33 | Bk | 10 | 27 | 61 | 30 | 9 |
| 33 | Bkk | 27 | 37 | 61 | 31 | 8 |
| 34 | AB | 0 | 8 | 58 | 32 | 10 |
| 34 | Bk | 8 | 20 | 65 | 25 | 10 |
| 34 | Bkkm | 20 | 30 | 62 | 30 | 8 |
| 35 | AB | 0 | 9 | 61 | 30 | 9 |
| 35 | Bk | 9 | 28 | 62 | 27 | 11 |
| 35 | Bkkm | 28 | 47 | 60 | 32 | 8 |
|  |  |  |  |  |  |  |


| 36 | AB | 0 | 9 | 61 | 29 | 10 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 36 | Bkk | 9 | 22 | 62 | 28 | 10 |
| 37 | AB | 0 | 6 | 58 | 33 | 9 |
| 37 | Bt1 | 6 | 44 | 52 | 38 | 10 |
| 37 | Bt2 | 44 | 115 | 55 | 35 | 10 |
| 37 | Btk | 115 | 150 | 54 | 34 | 12 |
| 38 | AB | 0 | 17 | 53 | 28 | 19 |
| 38 | Bt1 | 17 | 44 | 35 | 51 | 14 |
| 38 | Bt2 | 44 | 73 | 39 | 37 | 24 |
| 38 | Btk | 73 | 150 | 56 | 28 | 16 |
| 39 | AB | 0 | 8 | 61 | 31 | 8 |
| 39 | Btk | 8 | 37 | 62 | 30 | 8 |
| 39 | Bkk1 | 37 | 72 | 60 | 33 | 7 |
| 39 | Bkk2 | 72 | 105 | 61 | 31 | 8 |
| 40 | AB | 0 | 6 | 53 | 43 | 4 |
| 40 | Bk | 6 | 34 | 65 | 26 | 9 |
| 40 | Bkk | 34 | 87 | 64 | 29 | 7 |
| 41 | AB | 0 | 7 | 62 | 26 | 12 |
| 41 | Bk | 7 | 20 | 62 | 25 | 13 |
| 41 | Bkk | 20 | 39 | 61 | 30 | 9 |
| 42 | AB | 0 | 6 | 57 | 33 | 10 |
| 42 | Bk | 6 | 24 | 63 | 28 | 9 |
| 43 | AB | 0 | 5 | 64 | 29 | 7 |
| 43 | Bk | 5 | 27 | 60 | 28 | 12 |
| 43 | Bkk | 27 | 75 | 62 | 29 | 9 |
| 43 | Bt | 75 | 150 | 52 | 44 | 4 |
| 44 | AB | 0 | 7 | 35 | 61 | 4 |
| 44 | Bwk | 7 | 28 | 59 | 27 | 14 |
| 45 | AB | 0 | 8 | 58 | 33 | 9 |
| 45 | Bkk | 8 | 62 | 59 | 31 | 10 |
| 45 | Bkkm | 62 | 90 | 60 | 30 | 10 |
| 46 | AB | 0 | 6 | 61 | 36 | 3 |
| 46 | Bk | 6 | 22 | 62 | 29 | 9 |
| 46 | Bkk | 22 | 42 | 59 | 33 | 8 |
| 47 | AB | 0 | 7 | 77 | 18 | 5 |
| 47 | Bw | 7 | 24 | 69 | 24 | 7 |
| 48 | AB | 0 | 7 | 63 | 26 | 11 |
| 48 | Bk | 7 | 26 | 64 | 29 | 7 |
| 48 | Bkk | 26 | 50 | 63 | 27 | 10 |
| 49 | AB | 0 | 5 | 64 | 28 | 8 |
| 49 | Bk1 | 5 | 47 | 61 | 30 | 9 |
| 49 | Bk2 | 47 | 84 | 52 | 35 | 13 |
|  |  |  |  |  |  |  |

Table C.2. PedonID is a unique identifier for each sampling location and can be used to link sampling locations with horizon level data contained in Table C.1. Latitude and Longitude are GPS coordinates for each location. Coordinates are in WGS84 decimal degrees.

| Pedon ID | Latitude | Longitude |
| :---: | :---: | :---: |
| 1 | 32.5870910 | -106.6096710 |
| 2 | 32.5845890 | -106.6078140 |
| 3 | 32.5884608 | -106.6079170 |
| 4 | 32.5845820 | -106.6079340 |
| 5 | 32.5850240 | -106.6057790 |
| 6 | 32.5848680 | -106.6060360 |
| 7 | 32.5847900 | -106.6059400 |
| 8 | 32.5844060 | -106.6051760 |
| 9 | 32.5848400 | -106.6060240 |
| 10 | 32.5864240 | -106.6069420 |
| 11 | 32.5856800 | -106.6068600 |
| 12 | 32.5845080 | -106.5993850 |
| 13 | 32.5844800 | -106.5993800 |
| 14 | 32.5825330 | -106.5950810 |
| 15 | 32.5827400 | -106.5960000 |
| 16 | 32.5828400 | -106.5964900 |
| 17 | 32.5829000 | -106.5962230 |
| 18 | 32.5829000 | -106.5961900 |
| 19 | 32.5828300 | -106.5962900 |
| 20 | 32.5840000 | -106.6060200 |
| 21 | 32.5859300 | -106.6070100 |
| 22 | 32.5859200 | -106.6069800 |
| 23 | 32.5858400 | -106.6069200 |
| 24 | 32.5833000 | -106.6044900 |
| 25 | 32.5833900 | -106.6045200 |
| 26 | 32.5833800 | -106.6045500 |
| 27 | 32.5829700 | -106.6050100 |
| 28 | 32.5848400 | -106.6077200 |
| 29 | 32.5846700 | -106.5994100 |
| 30 | 32.5844100 | -106.5994400 |
| 31 | 32.5855800 | -106.6013400 |
| 32 | 32.5853400 | -106.6107900 |
| 33 | 32.5869900 | -106.6051600 |
| 34 | 32.5856400 | -106.6067100 |
| 35 | 32.5859700 | -106.6073300 |
| 36 | 32.5856200 | -106.6079600 |
| 37 | 32.5835400 | -106.6043500 |


| 38 | 32.5809500 | -106.6021600 |
| :--- | :--- | :--- |
| 39 | 32.5823200 | -106.6022700 |
| 40 | 32.5825800 | -106.6029100 |
| 41 | 32.5826200 | -106.6031500 |
| 42 | 32.5819600 | -106.5978400 |
| 43 | 32.5817800 | -106.5992600 |
| 44 | 32.5829300 | -106.5996000 |
| 45 | 32.5840900 | -106.6000600 |
| 46 | 32.5828000 | -106.6039800 |
| 47 | 32.5828100 | -106.6032300 |
| 48 | 32.5828400 | -106.6032400 |
| 49 | 32.5828400 | -106.6031600 |

Table C.3. Sand, Silt, Clay concentrations in percent by equal depth sampling interval. Data extracted from Cody Anderson's thesis available at: https://repository.asu.edu/items/21017. Additional details can be found in the thesis.

| Profile | Depth [cm] | \% coarse gravel $>19 \mathrm{~mm}$ | \% medium gravel 4.7519 mm | \% fine gravel 2- <br> 4.75 mm | \% very coarse sand 1-2mm | \% <br> coarse <br> sand <br> 0.5- <br> 1 mm | \% medium sand 250$\mathbf{5 0 0 \mu m}$ | $\begin{aligned} & \% \text { fine } \\ & \text { sand } \\ & \mathbf{1 0 6 -} \\ & \mathbf{2 5 0 \mu m} \end{aligned}$ | \% very <br> fine <br> sand <br> 53- <br> $106 \mu \mathrm{~m}$ | $\begin{aligned} & \hline \% \text { silt } \\ & 2-53 \mu \mathrm{~m} \end{aligned}$ | $\begin{aligned} & \% \text { clay } \\ & <2 \mu \mathrm{~m} \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| JER1 | 0-7 | 6.40 | 15.28 | 3.74 | 0.89 | 0.87 | 1.49 | 12.07 | 15.67 | 36.71 | 6.88 |
| JER1 | 7-17 | 0.00 | 14.91 | 4.99 | 1.17 | 0.82 | 1.74 | 12.85 | 17.17 | 38.41 | 7.94 |
| JER1 | 17-27 | 3.64 | 21.37 | 8.50 | 1.30 | 1.19 | 1.88 | 9.90 | 14.09 | 33.39 | 4.73 |
| JER2 | 0-7 | 0.00 | 16.61 | 6.54 | 2.26 | 1.73 | 2.44 | 14.53 | 17.27 | 32.63 | 5.98 |
| JER2 | 7-17 | 9.81 | 22.18 | 7.53 | 2.02 | 1.09 | 1.58 | 9.28 | 12.39 | 30.17 | 3.95 |
| JER2 | 17-27 | 18.60 | 11.93 | 7.77 | 2.43 | 1.55 | 1.95 | 9.99 | 13.63 | 28.08 | 4.06 |
| JER3 | 0-7 | 0.00 | 20.55 | 6.72 | 2.29 | 2.72 | 1.99 | 12.71 | 17.72 | 30.43 | 4.86 |
| JER3 | 7-17 | 9.90 | 17.23 | 7.98 | 1.80 | 1.31 | 1.87 | 9.03 | 14.71 | 32.57 | 3.60 |
| JER3 | 17-27 | 0.00 | 20.52 | 12.02 | 3.10 | 1.82 | 2.23 | 8.78 | 13.51 | 33.71 | 4.32 |
| JER4 | 0-7 | 4.94 | 14.39 | 7.75 | 2.58 | 2.01 | 3.25 | 18.44 | 18.84 | 24.24 | 3.56 |
| JER4 | 7-17 | 0.00 | 16.79 | 8.30 | 2.38 | 2.09 | 3.38 | 16.63 | 16.81 | 29.24 | 4.39 |
| JER4 | 17-27 | 9.15 | 17.51 | 11.57 | 3.62 | 3.62 | 3.40 | 11.25 | 11.01 | 23.74 | 5.13 |
| JER5 | 0-7 | 0.00 | 18.67 | 7.71 | 2.55 | 2.18 | 3.59 | 19.37 | 17.21 | 22.84 | 5.89 |
| JER5 | 7-17 | 0.00 | 21.71 | 7.95 | 2.35 | 1.78 | 2.63 | 13.74 | 14.37 | 29.61 | 5.85 |
| JER5 | 17-27 | 0.00 | 11.80 | 9.19 | 2.65 | 2.97 | 3.37 | 14.63 | 15.48 | 33.54 | 6.37 |
| JER6 | 0-7 | 0.00 | 18.83 | 7.47 | 2.21 | 1.52 | 1.96 | 12.67 | 18.65 | 31.62 | 5.07 |
| JER6 | 7-17 | 3.42 | 19.08 | 11.43 | 1.77 | 1.09 | 1.50 | 7.95 | 12.97 | 35.31 | 5.48 |
| JER6 | 17-27 | 25.46 | 27.83 | 9.40 | 3.65 | 2.14 | 1.78 | 4.36 | 5.75 | 16.38 | 3.26 |
| JER7 | 0-7 | 7.29 | 12.46 | 7.46 | 2.99 | 2.16 | 2.59 | 15.50 | 19.44 | 26.46 | 3.66 |
| JER7 | 7-17 | 15.40 | 10.10 | 6.04 | 1.97 | 1.52 | 2.32 | 13.12 | 16.56 | 28.52 | 4.45 |
| JER7 | 17-27 | 0.00 | 24.20 | 6.80 | 2.13 | 1.76 | 2.24 | 11.87 | 15.72 | 31.34 | 3.94 |
| JER8 | 0-7 | 0.00 | 6.94 | 6.43 | 2.78 | 2.00 | 3.02 | 19.28 | 24.70 | 31.08 | 3.77 |
| JER8 | 7-17 | 23.15 | 12.33 | 5.47 | 2.08 | 1.22 | 1.62 | 10.32 | 14.54 | 25.41 | 3.88 |
| JER8 | 17-27 | 4.58 | 21.18 | 10.26 | 2.58 | 1.66 | 1.81 | 9.77 | 14.49 | 29.22 | 4.44 |
| JER9 | 0-7 | 6.15 | 19.90 | 6.47 | 2.13 | 1.97 | 3.11 | 15.53 | 15.77 | 24.05 | 4.93 |
| JER9 | 7-17 | 34.89 | 12.66 | 8.69 | 2.06 | 0.99 | 1.31 | 6.58 | 8.57 | 19.80 | 4.46 |


|  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{aligned} & \infty \stackrel{\infty}{\underset{\sim}{N}} \underset{\sim}{\infty} \\ & \underset{\sim}{\lambda} \underset{\sim}{c} \end{aligned}$ |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | $\left\|\begin{array}{cccc} 0 & n & n & \stackrel{N}{\infty} \\ \underset{y}{i} & \underset{\sim}{*} & \infty & \underset{y}{c} \\ \infty \end{array}\right\|$ |  |  |  |  |  |
| $\underset{-}{\top} \underset{\sim}{i} \underset{i}{i}$ |  | $\begin{array}{lllll} \underset{N}{*} & n & n & n \\ & \cdots & n & n & n \\ n \end{array}$ |  |  |  | $\left\|\begin{array}{cccc} \hat{\infty} & \overrightarrow{0} & \cdots & \bar{n} \\ - & n & \cdots & n \end{array}\right\|$ |
| $\begin{aligned} & \exists \stackrel{n}{n} \stackrel{n}{i} \\ & \cdots \cdots \cdots \end{aligned}$ | $\left\lvert\,\right.$ |  | $\stackrel{r}{n} \underset{\sim}{0} 0_{0}^{\infty} 0$ |  |  | $\underset{\sim}{\sim}$ $\underset{\sim}{i} \text { in }$ |
| $\stackrel{\infty}{n} \stackrel{\infty}{\uparrow} \underset{\sim}{\infty}$ $\stackrel{n}{i} \stackrel{\sim}{i} \stackrel{\infty}{i}$ |  |  |  | $$ | $$ |  |
|  | $\left\lvert\, \begin{array}{llll} \infty & 0 & n & n \\ \infty & 0 & \infty \\ 0 & \ddots & 0 & n \\ 0 & n & \infty \\ \hline \end{array}\right.$ |  | $$ | $\begin{array}{lllll} 0 & n & 0 & n \\ 0 & 0 & n & n \\ 0 & 0 & n & n \end{array}$ |  |  |
| $\begin{aligned} & n \\ & \cdots \cdots \\ & \cdots \cdots \\ & m \end{aligned}$ |  | নへへ へ か $\dot{\sim} \dot{\sim} \dot{J} \dot{J}$ |  |  |  |  |
| $$ |  |  |  |  |  | $\left\lvert\, \begin{array}{llll} 0 & 0 & 0 \\ & -\infty & 8 & 8 \\ -1 & + & 0 & 0 \end{array}\right.$ |
|  |  |  |  | $\underset{\substack{r}}{\substack{1}}$ |  |  |
|  |  |  |  |  |  |  |

Table C.4. Location information for the texture fraction measurments in table C.3. Northing and Easting are given in UTM Zone 13N NAD83. ProfileID is a unique identifier that can be used to link tables C. 3 and C.4. Data extracted from Cody Anderson's thesi available at: https://repository.asu.edulitems/21017.

| JER | UTM Zone 13 <br> Location   |  | Vegetation cover |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{array}{\|l\|} \hline \text { Profile } \\ \hline \end{array}$ ID | Northing | Easting | Primary cover | Secondary cover | Coverage class |
| JER1 | 3606481.0 | 349470.1 | BA |  | BA |
| JER2 | 3606454.7 | 349467.3 | BM | BA | GR |
| JER3 | 3606426.7 | 349467.8 | BA |  | BA |
| JER4 | 3606396.8 | 349452.7 | BA | CR | BA |
| JER5 | 3606372.7 | 349448.5 | TB | BA | OS |
| JER6 | 3606478.3 | 349502.3 | CR/BM | BA | CR |
| JER7 | 3606451.0 | 349498.6 | CR | BA | CR |
| JER8 | 3606426.8 | 349492.7 | MQ | BA,BM/CR | MQ |
| JER9 | 3606392.4 | 349484.1 | BM/CR | BA | GR |
| JER10 | 3606366.8 | 349474.9 | BA |  | BA |
| JER11 | 3606472.7 | 349530.2 | BM/TB | BA | GR |
| JER12 | 3606447.3 | 349522.7 | BA | TB,MQ,CR | BA |
| JER13 | 3606420.2 | 349517.9 | CR | BA | CR |
| JER14 | 3606389.4 | 349508.8 | TB/BM | BA,MQ | OS |
| JER15 | 3606362.6 | 349506.1 | MQ/BM | BA | MQ |
| JER16 | 3606469.5 | 349558.7 | MA/BM/CR | BA | OS |
| JER17 | 3606444.6 | 349550.9 | MQ/TB | BM,BA | MQ |
| JER18 | 3606418.4 | 349543.2 | BA |  | BA |
| JER19 | 3606383.5 | 349538.8 | CR/PP | BA | CR |
| JER20 | 3606357.9 | 349535.0 | MA | TB,BA | OS |

```
APPENDIX D CODE USED FOR GEOSTATISTICAL MODELING OF SAND AND CLAY
R code used for geostatistical modeling of texture fraction (i.e., sand and clay).
#Geostatistical modeling of sand and clay by global maps standardized depth in the Tromble Weir
Watershed
# Colby Brungard, PhD
# Load libraries
library(aqp)
library(sp)
library(rgdal)
library(raster)
library(gstat)
library(ggplot2)
library(openxlsx)
library(plyr)
library(dplyr)
library(reshape2)
library(RColorBrewer)
library(e1071)
library(ggpubr)
# set working directory
setwd("D:/Tromble Weir")
#1. Data preprocessing
#1.1 Mikayla's sampling
mdat <- read.csv("./Mikayla Data/R_Pit_Data_cwb.csv")
mloc <- read.csv("./Mikayla Data/R_Pit_Site_Data.csv")
# Reproject coordinates to UTM Zone 13N
coordinates(mloc) <- ~ Longitude + Latitude
```

```
proj4string(mloc) <- '+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs +towgs84=0,0,0'
mloc <- spTransform(mloc, CRS('+proj=utm +zone=13 +ellps=GRS80 +datum=NAD83 +units=m
+ no_defs'))
mloc \(<-\) as.data.frame(mloc)
names(mloc)[2:3] <- c('Easting', 'Northing')
```

\# Points 3 and 38 were located outside of the study area in a different parent material. Remove them. Because they are far away they are unlikely to have much impact. Points $20 \& 37$ may also be in a different landform, but are likely not in different enough parent material to affect matters. Interesingly though, when I tested removing these two points I did get slighly better X-validation fits.
mdat <- mdat[mdat\$Pedon.ID != 3 \& mdat\$Pedon.ID != 38,]
mloc <- mloc[mloc\$Pedon.ID != 3 \& mloc\$Pedon.ID != 38,]

## \#1.2 Sampling by Cody Anderson

\# Tables from his thesis
adat <- readWorkbook("./Enriques Data/Anderson_asu_Thesis_TableB2.xlsx")
\# Get horizon top and bottom depths
adat\$HZ.Top <- sapply(strsplit(adat\$'Depth.[cm]`,"-"), ‘[’, 1) adat\$HZ.Bottom <- sapply(strsplit(adat\$’Depth.[cm]`,"-"), `[’, 2) \# Combine all gravel and sand percentages into only sand. Unfortunately it appears that Cody somehow divided gravel and sand values instead of dividing the coarse fraction from the fine earth fraction. I have decided to combine the gravel and sand values into a single 'Sand' column based on the following observations. 1) (Gravel+Sand)+Clay exactly equals Silt values. Since silt is calculated as 100 -(sand + clay) I'm reasonably confident that these values are correct. 2) The summary statistics of Gravel+Sand almost exactly match the summary statistics of sand values for the rest of the observation made by Mikayla. 3) The clay values appear correct. This assumption negates the use of the course fraction in any subsequent analysis. I have contacted Enrique about this data but have not gotten a response. adat\$Sand <- apply(adat[,c(3:10)], 1, sum) adat\$Silt <- adat\$ \% .silt.2-53 \(\mu \mathrm{m}\) ` \# Give silt a better name
adat\$Clay <- adat\$ $\%$.clay. $<2 \mu \mathrm{~m}$ ` \#Give clay a better name
\# Create psuedo horizon names (needed for AQP) and make factor
adat1 <- ddply(adat, .(Profile), mutate, Designation = seq_along(HZ.Top))
adat1\$Designation <- as.factor(as.character(adat1\$Designation))
\# Subset for relevant variables and name columns to match Mikayla's data

```
adat2 <- adat1[,c(1,18,13:17)]
names(adat2)[1] <- 'Pedon.ID'
# Location information
aloc <- readWorkbook("./Enriques Data/Anderson_asu_Thesis-Table2.xlsx", rows=c(25:45), cols = c(1:3))
names(aloc)[1] <- 'Pedon.ID'
aloc2<- aloc[,c(1,3,2)]
# 1.3 Join both datasets and format as needed
# Pedon data
dat <- rbind(mdat, adat2)
dat$HZ.Bottom <- as.numeric(dat$HZ.Bottom)
dat$HZ.Top <- as.numeric(dat$HZ.Top)
# Site data
sdat <- rbind(mloc, aloc2)
```

\#2. Convert to SPC and convert to standard depth intervals.
depths(dat) <- Pedon.ID $\sim$ HZ.Top + HZ.Bottom
\# Global soil map standard depth intervals: $0-5 \mathrm{~cm}, 5-15 \mathrm{~cm}, 15-30 \mathrm{~cm}, 30-60 \mathrm{~cm}, 60-100$, \& 100-200 cm. The following code modified from: https://ncss-tech.github.io/AQP/aqp/aqp-intro.html. However, many of these soils are $<100 \mathrm{~cm}$ deep. Because I have so few samples I am only going to map soil texture to 60 cm .
\# Use the slice and slab functions in AQP to average over these depths
s1 <- aqp::slice(dat, fm=0:100 ~Sand + Silt + Clay)
\# Subset to GSM depths and calculate weighted mean values (I'm pretty sure that this does a weighted mean).
gsm.depths $<-\mathrm{c}(0,5,15,30,60,100)$
d.gsm <- slab(s1, fm=Pedon.ID $\sim$ Sand + Silt + Clay, slab.structure $=$ gsm.depths, slab.fun $=$ median, na.rm=TRUE)

```
# reshape to wide format, convert to SPC, and make new hz names
gsmpedons <- dcast(d.gsm, Pedon.ID + top + bottom ~ variable, value.var = 'value')
depths(gsmpedons) <- Pedon.ID ~ top + bottom
gsmpedons$hzname <- profileApply(gsmpedons, function(i) {paste0('GSM-', 1:nrow(i))})
# Note: Use new gsmpedons with caution. It is very likely that values > 60 cm were calculated with very
few observtions.
# 2 Prepare for exploratory data analysis
# Convert to SpatialPointsDataframe, and reproject
site(gsmpedons) <- sdat #This automatically joins by id (cool!)
coordinates(gsmpedons) <- ~ Easting + Northing
proj4string(gsmpedons) <- '+proj=utm +zone=13 +ellps=GRS80 +datum=NAD83 +units=m +no_defs'
# Subset by GSM depth interval,
d1 <- gsmpedons[, 1]
d2 <- gsmpedons[, 2]
d3 <- gsmpedons[, 3]
d4 <- gsmpedons[, 4]
# Remove missing values (no data) from lower horizons
d4 <- as.data.frame(d4)
d4 <- d4[complete.cases(d4),]
coordinates(d4) <- ~ Easting + Northing
proj4string(d4) <- '+proj=utm +zone=13 +ellps=GRS80 +datum=NAD83 +units=m +no_defs'
```

\# Load all rasters. Rasters created from 5m ifsar DEM using geoprocess_by_area.bat. See readme file in tw folder.
brk <- do.call(brick, lapply(list.files(path = "./Terrain_derivatives/TD_5m", pattern = ".*tif", full.names = TRUE), raster))
\# Reproject rasters to points (if needed)

```
brk2<- projectRaster(brk, crs="+proj=utm +zone=13 +ellps=GRS80 + datum=NAD83 +units=m +no_defs
```

+towgs84=0,0,0")
\# Mask to study area, then crop extent (significantly reduces processing time).
studyarea <- readOGR("./NestedSampllingExample", "SoilMU26")
brk3 <- mask(brk2, mask $=$ studyarea $)$
brk4 <- crop(brk3, studyarea)
\# Extract covariate values
$\mathrm{ec}<-\operatorname{raster}:: \operatorname{extract}($ brk4, $\mathrm{y}=\mathrm{d} 1)$
ec4<- raster::extract(brk4, $y=d 4$ )
\# Join covariate values to soil depth data
de1 $<-\operatorname{cbind}(d 1$, ec)
de2 $<-\operatorname{cbind}(d 2$, ec)
de $3<-\operatorname{cbind}(d 3$, ec)
de4 $<-\operatorname{cbind}(d 4$, ec4)
\# Kriging and gaussian simulation requires a very fine underlying grid on which to predict.
\# Use rasters to create prediction grid
sgdf $<-$ as(brk4, 'SpatialGridDataFrame')
\# 3 Sand.
\# 3.1 $0-5 \mathrm{~cm}$.
\# Summary stats. Webster and Oliver suggest the transformation be applied if the skewness is $>0.5$.
summary(d1\$Sand) \# Median is close to mean so appears normally distributed.
skewness(d1\$Sand) \#-0.87
\# Histograms. This appears quite 'normal'
hist(d1\$Sand, col = "lightblue", border = "red")
rug(d1 \$Sand)

```
# Check for obvious spatial patterns
spplot(d1, zcol = 'Sand', col.regions=brewer.pal(5, "Set1"))
# Look for spatial outliers
# s1.sel = plot(variogram(Sand ~ 1, d1, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(s1.sel, d1)
# Fit linear models between Sand and covariates
# Significance only shows that the relationship is not-zero.
# Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary( }\operatorname{lm}(\mathrm{ Sand ~ Aspect , de1))
summary( }1\textrm{m}\mathrm{ (Sand }~\mathrm{ ConvergenceIndex , de1))
summary(lm(Sand ~ CrossSectionalCurvature, de1))
summary(lm(Sand ~ DEM_5_utm , de1)) #** Adj. R2:0.099
summary(lm(Sand ~ FlowAccumulation , de1))
summary(lm(Sand ~ LongitudinalCurvature, de1))
summary(lm(Sand ~ LSfactor , de1))
summary(}\operatorname{lm}(\mathrm{ Sand ~ Slope , de1))
summary(lm(Sand ~ TopographicWetnessIndex, de1))
summary(lm(Sand ~ ValleyDepth , de1))
# Only elevation is significant
plot(Sand ~ DEM_5_utm, data = de1)
#There isn't a very strong relationship, but it do need to account for this trend.
```

\# Check for Anisotropy. 120 seems best
$\operatorname{plot}(\operatorname{variogram}(S a n d \sim 1, \mathrm{~d} 1$, alpha $=c(0,15,30,45,60,75,90,105,120,135,155,180)))$
\# h-scatterplots. These are plots of $\mathrm{z}(\mathrm{x})$ against $\mathrm{z}(\mathrm{x}+\mathrm{h})$ for each lag interval and show the distribution of pairs of points for that interval. The closer the points lie to the diagional line, the stronger the correlation
and the smaller the semivariance. These distances were chosen because these are the distances that I used for the nested sampling.
hscat(Sand $\sim 1$, data $=$ de1, c(3, 9, 29, 88, 266, 800), variogram.alpha=120)
\# Most correlated (has the lowest semivariance) below $\sim 30 \mathrm{~m}$
\# Empirical (experimental or sample) variogram. It makes sense to use the distances over which I designed the sampling [boundaries $=c(3,9,29,88,266,800)$ ], but when I do this I have a great deal of trouble fiting a variogram model (most of the time I get a singular model or no convergence), so I decided not to implement this.
$\operatorname{svg} 1<-$ variogram(Sand $\sim$ DEM_5_utm, de1, alpha=120)
plot(svg1, plot.nu $=$ FALSE $)$
$\operatorname{svg} 1$
\# Variogram modeling
\# The experimental variogram is basically just two columns of numbers: distance and semivariance. To use this for predictions, we need to fit a model (like a regression line) to the variogram. Because the variogram modeling is a numerical optimization we need to provide starting values. psill is the partial sill which is the sill-nugget.

```
svgm1.s <- fit.variogram(object=svg1, model = vgm(nugget = 10, psill = 20, range= 300, model = 'Sph'))
svgm1.c <- fit.variogram(object=svg1,model = vgm(nugget = 10, psill = 20, range=300, model = 'Cir'))
svgm1.e <- fit.variogram(object=svg1, model = vgm(nugget = 10, psill = 20, range= 300, model = 'Exp'))
```

svgm1.s
svgm1.c
svgm1.e
plot(svg1, svgm1.s, pch = 19)
plot(svg1, svgm1.c, pch = 19)
plot(svg1, svgm1.e, pch = 19)
\# Leave-one-out cross validation
$\operatorname{scv} 1 . s=$ krige.cv $($ Sand $\sim$ DEM_5_utm, de1, model $=$ svgm1.s)
scv1.c $=$ krige.cv $($ Sand $\sim$ DEM_5_utm, de1, model $=\operatorname{svgm1} . c)$
scv1.e $=$ krige.cv(Sand ~DEM_5_utm, de1, model = svgm1.e)
\# MPE. Mean prediction error (predicted-observed) = bias. Positive $=$ under prediction, negative $=$ over prediction
mean(scv1.s\$residual) \#-0.10
mean(scv1.c\$residual) \#-0.19
mean(scv1.e\$residual) \# -0.07
\# MSE. Mean squared error measures on average how different predictions are from observations.
\# The MSE will be small if the predicted responses are very close to the true responses, and will be large if for some of the observations, the predicted and true responses differ substantially (ISL sixth printing).
mean(scv1.s\$residual^2) \# 34.0
mean(scv1.c\$residual^2) \# 33.2
mean(scv1.e\$residual^2) \# 34.1
\# RMSE (take the square root to get units in original units)
sqrt(mean(scv1.s\$residual^2)) \# 5.8 \%
sqrt(mean(scv1.c\$residual^2)) \# 5.8
sqrt(mean(scv1.e\$residual^2)) \# 5.8
\# What is the spatial distribution of the residuals?
bubble(scv1.s, "residual", main = "Sand 0-5 cm Spherical")
bubble(scv1.c, "residual", main = "Sand 0-5 cm Circular")
bubble(scv1.e, "residual", main = "Sand 0-5 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
sk1.s <- krige(Sand ~ DEM_5_utm, de1, model = svgm1.s, newdata = sgdf)
sk1.c <- krige(Sand ~ DEM_5_utm, de1, model = svgm1.c, newdata = sgdf)
sk1.e <- krige(Sand ~ DEM_5_utm, de1, model = svgm1.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
sk1.s $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,8)) + ggtitle('Spherical') + theme_bw()
sk1.c %>% as.data.frame %>%
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,8)) + ggtitle('Circular') + theme_bw()
```

sk1.e $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,8)) + ggtitle('Exponential') + theme_bw()
```

\# None of these models seemed to have much different prediction patterns or much different uncertainty than another, so I choose Circular because it had slightly lower MSE. Publication quality plotting and writing to raster are done below.

```
# 3.2 5-15 cm
# Summary stats. Appear fairly normal, no need to transform based on skewness
summary(d2$Sand)
skewness(d2$Sand) # -0.115
# Histograms. Very normally distributed
hist(d2$Sand, col = "lightblue", border = "red")
rug(d2$Sand)
# Plots to check for obvious spatial patterns
spplot(d2, zcol = 'Sand', col.regions=brewer.pal(5, "Set1"))
# # Look for spatial outliers. Nothing obvious
# s2.sel = plot(variogram(Sand ~ 1, d2, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(s2.sel, d1)
```

\# Fit linear models
summary $(\operatorname{lm}($ Sand $\sim$ Aspect $\quad$, de2 $)$ )
summary $(\operatorname{lm}$ (Sand $\sim$ ConvergenceIndex , de2) ) \#* Adj. R2 0.06
summary $(\operatorname{lm}($ Sand $\sim$ CrossSectionalCurvature, de2) ) \#* Adj. R2 0.08
summary (lm(Sand~DEM_5_utm , de2)) \#** Adj. R2 0.09
summary $(\operatorname{lm}($ Sand $\sim$ FlowAccumulation , de2) ) \#* Adj. R2 0.05
summary $(\operatorname{lm}$ (Sand $\sim$ LongitudinalCurvature , de2) ) \#** Adj. R2 0.004
summary $(\operatorname{lm}($ Sand $\sim$ LSfactor , de2) $)$
summary $(\operatorname{lm}$ (Sand $\sim$ Slope , de2) $)$
summary $(\operatorname{lm}$ (Sand $\sim$ TopographicWetnessIndex, de2) $)$ \#** Adj. R2 0.14
summary $(\operatorname{lm}($ Sand $\sim$ ValleyDepth , de2) $)$
plot(Sand ~ TopographicWetnessIndex, de2)
plot(Sand ~ DEM_5_utm, de2)
plot(Sand $\sim$ LongitudinalCurvature, data $=\operatorname{de} 2$ )
\# Hmmm, only TWI and elevation seem to have a strong relationship with Sand. I suspect that the strength (if it can be considered strong) of the relationship between sand and TWI is due to the few points located in areas with higher TWI values and that the relationship may not be as 'strong' if these points were removed. I'm still going to go with elevation as it seems less spurious.
\# Check for Anisotropy. 120 seems best and agreed with the direction of the landform.
$\operatorname{plot}(\operatorname{variogram}(\operatorname{Sand} \sim 1, \mathrm{~d} 2$, alpha $=c(0,15,30,45,60,75,90,105,120,135,155,180)))$
\# h-scatterplots. Strongest correlation at $<30 \mathrm{~m}$.
hscat(Sand $\sim 1$, data $=\operatorname{de} 2, c(3,9,29,88,266,800)$, variogram.alpha=120)
\# Empirical (experimental or sample) variogram. When I include boundaries $=c(3,9,29,88,266,800) \mathrm{I}$ am able to still get a model to fit, but it strongly reduces the range thus the prediction uncertainity is only concentrated around the sample locations and RMSE slightly increased, so I am not taking this approach.
$\operatorname{svg} 2<-$ variogram(Sand $\sim$ DEM_5_utm, de2, alpha=120)
plot(svg2, plot.nu = FALSE)
$\operatorname{svg} 2$

```
# Variogram modeling
svgm2.s <- fit.variogram(object=svg2, model = vgm(nugget = 20, psill = 1, range= 100, model = 'Sph'))
svgm2.c <- fit.variogram(object=svg2, model = vgm(nugget = 20, psill = 1, range= 100, model = 'Cir'))
svgm2.e <- fit.variogram(object=svg2, model = vgm(nugget = 20, psill = 1, range= 100, model = 'Exp'))
```

svgm2.s
svgm2.c
svgm2.e
$\operatorname{plot}(\operatorname{svg} 2, \operatorname{svgm} 2 . s, \operatorname{pch}=19)$
plot(svg2, svgm2.c, pch = 19)
plot(svg2, svgm2.e, pch = 19)
\# Leave-one-out cross validation
$\operatorname{scv} 2 . \mathrm{s}=$ krige.cv(Sand $\sim$ DEM_5_utm, de2, model $=$ svgm2.s)
scv2.c $=$ krige.cv $($ Sand $\sim$ DEM_5_utm, de2, model $=\operatorname{svgm} 2 . c)$
$\operatorname{scv} 2 . \mathrm{e}=$ krige.cv $($ Sand $\sim$ DEM_5_utm, de2, model $=\operatorname{svgm} 2 . e)$
\# MPE. Mean prediction error (predicted-observed) = bias. Positive $=$ under prediction, negative $=$ over prediction
mean(scv2.s\$residual) \#-0.02
mean(scv2.c\$residual) \#-0.02
mean(scv2.e\$residual) \#-0.02

```
# MSE. Mean squared error
mean(scv2.s$residual^2) # 23.0
mean(scv2.c$residual^2) # 23.2
mean(scv2.e$residual^2) # 23.5
```

\# RMSE (take the square root to get units in original units)
$\operatorname{sqrt}\left(\right.$ mean(scv2.s\$residual^$\left.\left.{ }^{\wedge} 2\right)\right)$ \# 4.79
sqrt(mean(scv2.c\$residual^2)) \# 4.82
sqrt(mean(scv2.e\$residual^2)) \# 4.84
\# What is the spatial distribution of the residuals?
bubble(scv2.s, "residual", main = "Sand 5-15 cm Spherical")
bubble(scv2.c, "residual", main = "Sand 5-15 cm Circular")
bubble(scv2.e, "residual", main = "Sand 5-15 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate, then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
sk2.s <- krige(Sand ~ DEM_5_utm, de2, model = svgm2.s, newdata = sgdf)
sk2.c <- krige(Sand ~ DEM_5_utm, de2, model = svgm2.c, newdata = sgdf)
sk2.e <- krige(Sand ~ DEM_5_utm, de2, model = svgm2.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
sk2.s $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
```

scale_fill_gradient(low = "yellow", high="red", limits = c(4,6)) + ggtitle('Spherical') + theme_bw()
sk2.c $\%>\%$ as.data.frame $\%>\%$

$$
\begin{aligned}
& \operatorname{ggplot}(\text { aes }(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\text { geom_tile(aes(fill=sqrt(var1.var))) }+ \text { coord_equal }()+ \\
& \text { scale_fill_gradient(low = "yellow", high="red", limits = c(4,6)) + ggtitle('Circular') + theme_bw() }
\end{aligned}
$$

sk2.e $\%>\%$ as.data.frame $\%>\%$

$$
\begin{aligned}
& \operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\text { geom_tile(aes(fill=sqrt(var1.var) }))+ \text { coord_equal }()+ \\
& \text { scale_fill_gradient(low }=\text { "yellow", high="red", limits }=c(4,6))+ \text { ggtitle('Exponential') }+ \text { theme_bw( })
\end{aligned}
$$

\# Little difference between the models. Chose to use a circular model to be consistent with the $0-5 \mathrm{~cm}$ layer and had slightly larger areas of lower uncertainty

```
# 3.3 15-30 cm
# Summary stats. Not much variability. Maybe I could just assume a mean value for this depth.
summary(d3$Sand)
skewness(d3$Sand) # 0.763
# Log transform makes < 0.5; skewness(log(d3$Sand)); but based on my attempts at back transform this
doesn't make much difference and only complicates analysis.
# Histograms. Not quite as 'normal' as the first two depths, but still pretty close.
hist(log(d3$Sand), col = "lightblue", border = "red")
rug(log(d3$Sand))
# Plots to check for obvious spatial patterns
spplot(d3, zcol = 'Sand', col.regions=brewer.pal(5, "Set1"))
# Look for spatial outliers. Nothing obvious
# s3.sel = plot(variogram(Sand ~ 1, d3, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# Fit linear models
summary(lm(Sand ~ Aspect , de3))
summary(lm(Sand ~ ConvergenceIndex , de3)) #* Adj. R2 0.05
summary(lm(Sand ~ CrossSectionalCurvature, de3)) #** Adj. R2 0.09
summary(lm(Sand ~ DEM_5_utm , de3))
summary( }\operatorname{lm}(\mathrm{ Sand ~FlowAccumulation , de3)) #* Adj. R2 0.05
summary(lm(Sand ~ LongitudinalCurvature , de3)) #** Adj. R2 0.14
summary(lm(Sand ~ LSfactor , de3))
summary(lm(Sand ~ Slope , de3)) #* Adj. R2 0.05
summary(lm(Sand ~ TopographicWetnessIndex, de3)) #*** Adj. R2 0.16
summary(lm(Sand ~ ValleyDepth , de3))
plot(Sand ~ LongitudinalCurvature, de3)
plot(Sand ~ TopographicWetnessIndex, de3)
```

\# \#Hmmm, I suspect that the 'strength' of these relationships is due to the few points located in areas with higher LongCurv and TWI values and that the relationship may not be as 'strong' if these points were removed. I tried removing what I thought were these points (Points $20 \& 37$ see data cleaning notes in section 1), but this didn't fully remove these points or change the relationships.
summary $(\operatorname{lm}$ (Sand $\sim$ LongitudinalCurvature+TopographicWetnessIndex, de3)) \# only TWI significant when run together. I'm going to use topographic wetness as the 'trend'

```
# Check for Anisotropy. Again 120 seems appropriate.
plot(variogram(Sand ~ 1, d3, alpha =c(0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 155, 180)))
```

\# h-scatterplots. Not much correlation beyond $\sim 30 \mathrm{~m}$.
hscat(Sand $\sim 1$, data $=$ de3, c(3, 9, 29, 88, 266, 800), variogram.alpha=120)
\# Empirical (experimental or sample) variogram.
\# For this depth interval I included the distances over which I designed the sampling [boundaries $=\mathrm{c}(3,9$,
$29,88,266,800)$ ], because I got singular variogram models if I didn't.
svg3 <- variogram(Sand $\sim$ TopographicWetnessIndex, de3, boundaries $=c(3,9,29,88,266,800)$,
alpha=120) \#
plot(svg3, plot.nu $=$ FALSE $)$
svg3
\# Variogram modeling
svgm3.s $<-$ fit.variogram(object=svg3, model $=$ vgm(nugget $=5$, psill $=30$, range $=50$, model $=$ 'Sph')
$\operatorname{svgm} 3 . c<-$ fit.variogram $($ object $=\operatorname{svg} 3$, model $=\operatorname{vgm}($ nugget $=5$, psill $=30$, range $=50$, model $=$ 'Cir' $)$ )
svgm3.e <- fit.variogram(object=svg3, model $=$ vgm(nugget $=5$, psill $=30$, range $=50$, model $=$ 'Exp' $)$ )
svgm3.s
svgm3.c
svgm3.e
$\operatorname{plot}(\operatorname{svg} 3, \operatorname{svgm} 3 . s, \operatorname{pch}=19)$
plot(svg3, svgm3.c, pch = 19)

```
plot(svg3, svgm3.e, pch = 19)
# Leave-one-out cross validation
scv3.s = krige.cv(Sand ~ TopographicWetnessIndex, de3, model = svgm3.s)
scv3.c = krige.cv(Sand ~ TopographicWetnessIndex, de3, model = svgm3.c)
scv3.e = krige.cv(Sand ~ TopographicWetnessIndex, de3, model = svgm3.e)
```

\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over
prediction
mean(scv3.s\$residual) \#-0.07
mean(scv3.c\$residual) \#-0.05
mean(scv3.e\$residual) \#-0.11
\# MSE. Mean squared error
mean(scv3.s\$residual^2) \# 34.1
mean(scv3.c\$residual^2) \# 34.7
mean(scv3.e\$residual^2) \# 32.2
\# RMSE. Removing topographicwetnessindex as a covariate results in an $\sim 0.5 \% \mathrm{RMSE}$ increase.
sqrt(mean(scv3.s\$residual^2)) \# 5.84
sqrt(mean(scv3.c\$residual^2)) \# 5.90
sqrt(mean(scv3.e\$residual^2)) \# 5.67
\# What is the spatial distribution of the residuals?
bubble(scv3.s, "residual", main = "Sand $15-30 \mathrm{~cm}$ Spherical")
bubble(scv3.c, "residual", main = "Sand $15-30 \mathrm{~cm}$ Circular")
bubble(scv3.e, "residual", main = "Sand $15-30 \mathrm{~cm}$ Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).
sk3.s <- krige(Sand $\sim$ TopographicWetnessIndex, de3, model $=$ svgm3.s, newdata $=$ sgdf) sk3.c $<-$ krige $($ Sand $\sim$ TopographicWetnessIndex, de3, model $=$ svgm3.c, newdata $=\operatorname{sgdf})$
sk3.e $<-$ krige $($ Sand $\sim$ TopographicWetnessIndex, de3, model $=$ svgm3.e, newdata $=\operatorname{sgdf}$ )
\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
sk3.s $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,12)) + ggtitle('Spherical') + theme_bw()
```

sk3.c $\%>\%$ as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+$ geom_tile(aes(fill=sqrt(var1.var)) $)+$ coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits $=c(2,12))+$ ggtitle('Circular') + theme_bw()

```
sk3.e %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,12)) + ggtitle('Exponential') + theme_bw()
```

\# I choose a circular model because it was the only model that returned a non-zero nugget.

```
# 3.4 30-60 cm
# Summary stats. Appears normally distributed and no need to transform.
summary(d4$Sand)
skewness(d4$Sand) # -0.14
# Histograms
hist(d4$Sand, col = "lightblue", border = "red")
rug(d4$Sand)
\# Plots to check for obvious spatial patterns
spplot(d4, zcol = 'Sand', col.regions=brewer.pal(5, "Set1"))
# It appears that the values around the Tromble Weir itself are a bit sandier
```

\# Fit Linear models

$$
\begin{aligned}
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { Aspect } \quad \text { de4 })) \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { ConvergenceIndex } \quad, \text { de4 })) \#^{*} \text { Adj. R2 }=0.12 \\
& \text { summary }\left(\operatorname{lm}(\text { Sand } \sim \text { CrossSectionalCurvature, de4) }) \#^{* *} \text { Adj. R2 }=0.15\right. \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { DEM_5_utm } \quad, \text { de4 })) \#^{* * *} \text { Adj. R2 }=0.42 \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { FlowAccumulation , de4) }) \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { LongitudinalCurvature }, \operatorname{de} 4)) \# * *^{*} \text { Adj. } \mathrm{R} 2=0.22 \\
& \operatorname{summary}(\operatorname{lm}(\operatorname{Sand} \sim \text { LSfactor } \quad, \operatorname{de} 4)) \#^{* *} \text { Adj. R2 }=0.14 \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { Slope } \quad \text {, de4 })) \#^{* * *} \text { Adj. R2 }=0.25 \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { TopographicWetnessIndex, de4 })) \#^{*} \text { Adj. R2 }=0.13 \\
& \text { summary }(\operatorname{lm}(\text { Sand } \sim \text { ValleyDepth } \quad \text {, de4 }))
\end{aligned}
$$

\# This is rather interesting. Perhaps the significance with more variables as depth increases suggests that the surface is affected by other variables that control erosion and deposition, and that these covariates don't become important until below the surface. The surface horizon of most pedons was $\sim 6 \mathrm{~cm}$. In any case, I believe that this shows a trend in the data that I will need to account for. However, I am uncertain of the physical significance of these since I am using weighted average values.
summary ( $\operatorname{lm}$ (Sand
DEM_5_utm+Slope+LongitudinalCurvature+CrossSectionalCurvature+LSfactor+TopographicWetnessIn dex, de4)) \# This reveals that only elevation (DEM_5_utm) is significant $\left({ }^{* * *}\right)$ when taken together.
\# Only the three variables with largest Adj. R2 values
summary $(\operatorname{lm}($ Sand $\sim$ DEM_5_utm+Slope+LongitudinalCurvature, de4)) \# Only elevation and longCurvature is significant
\# Elevation and slope
summary $(\operatorname{lm}($ Sand $\sim$ DEM_5_utm+Slope, de4 $))$ \# Both significant Adj. R2 $=0.46$
\# Elevation and LongCurvature
summary $(\operatorname{lm}($ Sand $\sim$ DEM_5_utm+LongitudinalCurvature, de4 $))$ \# Both significant Adj. R2 $=0.53$
\# I am going to use elevation and longitudinal curvature.
\# Check for Anisotropy. 120 is probably best, but 135 could also work
$\operatorname{plot}($ variogram $(\operatorname{Sand} \sim 1, d 4$, alpha $=c(0,15,30,45,60,75,90,105,120,135,155,180)))$
\# h-scatterplots. These show equivalent correlations between 120 and 135. So I chose 120 to be consistent. hscat(Sand $\sim 1$, data $=$ de4, c(3, 9, 29, 88, 266, 800), variogram.alpha=120) hscat(Sand $\sim 1$, data $=$ de4, c(3, 9, 29, 88, 266, 800), variogram.alpha=135)
\# Empirical (experimental or sample) variogram. Leaving out longitudinalCurvature increases RMSE by ~ 0.4 so I left it in.
$\operatorname{svg} 4<-$ variogram(Sand $\sim$ DEM_5_utm+LongitudinalCurvature, de4, alpha=120)
plot(svg4, plot.nu $=$ FALSE $)$
svg4
\# Variogram modeling. The exponential model doesn't converge, but the values are realistic and stable (even with different values) so I will include the model.
svgm4.s <- fit.variogram(object=svg4, model = vgm(nugget $=10$, psill $=20$, range $=300$, model $=$ 'Sph') $)$
$\operatorname{svgm} 4 . c<-$ fit.variogram(object=svg4, model $=\operatorname{vgm}($ nugget $=10$, psill $=20$, range $=300$, model $=$ 'Cir') $)$
svgm4.e $<-$ fit.variogram(object=svg4, model $=$ vgm(nugget $=10$, psill $=20$, range $=300$, model $=$ 'Exp' $)$ )
svgm4.s
svgm4.c
svgm4.e
$\operatorname{plot}(\operatorname{svg} 4, \operatorname{svgm} 4 . s, \operatorname{pch}=19)$
plot(svg4, svgm4.c, pch = 19)
plot(svg4, svgm4.e, pch = 19)
\# Leave-one-out cross validation
scv4.s $=$ krige.cv(Sand $\sim$ DEM_5_utm+LongitudinalCurvature, de4, model $=$ svgm4.s)
scv4.c $=$ krige.cv $($ Sand $\sim$ DEM_5_utm+LongitudinalCurvature, de4, model $=$ svgm4.c $)$
$\operatorname{scv} 4 . \mathrm{e}=$ krige.cv(Sand $\sim$ DEM_5_utm+LongitudinalCurvature, de4, model $=$ svgm4.e $)$
\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over prediction
mean(scv4.s\$residual) \# 0.19
mean(scv4.c\$residual) \# 0.19
mean(scv4.eSresidual) \# 0.19

```
\# MSE. Mean squared error
mean(scv4.s\$residual^2) \# 26.25
mean(scv4.c\$residual^2) \# 26.42
mean(scv4.e\$residual^2) \# 27.70
```

\# RMSE (take the square root to get units in original units)
sqrt(mean(scv4.s\$residual^2)) \# 5.12
sqrt(mean(scv4.c\$residual^2)) \# 5.14
$\operatorname{sqrt}($ mean(scv4.e\$residual^2) $)$ \# 5.26
\# What is the spatial distribution of the residuals?
bubble(scv4.s, "residual", main = "Sand 30-60 cm Spherical")
bubble(scv4.c, "residual", main = "Sand 30-60 cm Circular")
bubble(scv4.e, "residual", main = "Sand 30-60 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
sk4.s <- krige(Sand ~ DEM_5_utm+LongitudinalCurvature, de4, model = svgm4.s, newdata = sgdf)
sk4.c <- krige(Sand ~ DEM_5_utm+LongitudinalCurvature, de4, model = svgm4.c, newdata = sgdf)
sk4.e <- krige(Sand ~ DEM_5_utm+LongitudinalCurvature, de4, model = svgm4.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.

```
sk4.s %>% as.data.frame %>%
```

    \(\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\) geom_tile(aes(fill=sqrt(var1.var) \())+\) coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,12)) + ggtitle('Spherical') + theme_bw()
    sk4.c $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,12)) + ggtitle('Circular') + theme_bw()
sk4.e %>% as.data.frame %>%
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(2,12)) + ggtitle('Exponential') + theme_bw()
```

\# Based on RMSE and review of spatial predictions I choose the spherical model
\#4. Clay
\# Remove PedonID's 14:19 and 41. (possibly 48)
d1s <- d1[!(d1\$Pedon.ID \%in\% c(14:19,41)), ]
de1s $<-$ de1[!(de1 \$Pedon.ID \%in\% c(14:19,41)),]

## \# $4.10-5 \mathrm{~cm}$.

\# Summary stats. Webster and Oliver suggest the transformation be applied if the skewness is $>0.5$.
summary(d1s\$Clay) \# Median is close to mean so appears normally distributed.
skewness(d1s\$Clay) \# -0.19
\# Histograms. This appears quite 'normal'
hist(d1s\$Clay, col = "lightblue", border = "red")
rug(d1s\$Clay)
\# Check for obvious spatial patterns. Cody's values appear a bit low compared to Mikayla's sampling.

```
spplot(d1s, zcol = 'Clay', col.regions=brewer.pal(5, "Set1"))
# Look for spatial outliers. Maybe a few outliers.
# c1.sel = plot(variogram(Clay ~ 1, de1s, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(c1.sel, d1s)
# Fit linear models between Clay and covariates
# Significance only shows that the relationship is not-zero.
# Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary( }\operatorname{lm}(\mathrm{ Clay ~ Aspect , de1s))
summary(lm(Clay ~ ConvergenceIndex , de1s))
summary(lm(Clay ~ CrossSectionalCurvature, de1s))
summary(lm(Clay ~ DEM_5_utm , de1s)) #* Adj. R2 0.07
summary( }\operatorname{lm}(\mathrm{ Clay ~ FlowAccumulation , de1s))
summary(lm(Clay ~ LongitudinalCurvature , de1s))
summary( }\operatorname{lm}(\mathrm{ Clay ~ LSfactor , de1s))
summary( }\operatorname{lm}(Clay ~ Slope , de1s))
summary(lm(Clay ~ TopographicWetnessIndex, de1s))
summary(lm(Clay ~ ValleyDepth , de1s)) #** Adj. R2 0.15
# Only Convergence index is significant
plot(Clay ~ ValleyDepth, data = de1s)
abline(lm(Clay ~ ValleyDepth, de1s))
# Check for Anisotropy. 120 seems best
plot(variogram(Clay ~ ValleyDepth, de1s, alpha =c(105, 120, 135, 155, 180)))
# h-scatterplots. Correlation out to ~ 270 m.
hscat(Clay ~ ValleyDepth, data = de1, c(3, 9, 29, 88, 266, 800), variogram.alpha=120)
```

\# Empirical (experimental or sample) variogram. Including boundaries does not help with model fitting.

```
cvg1<- variogram(Clay ~ ValleyDepth, de1s, alpha=155)
plot(cvg1)
cvg1
```

\# Variogram modeling (also tried Pentaspherical and Matern, but they didn't fit either didn't work)
cvgm1.s <- fit.variogram(object=cvg1, model $=$ vgm(nugget $=1$, psill $=5$, range $=300$, model $=$ 'Sph') $)$
cvgm1.c $<-$ fit.variogram(object=cvg1, model $=\operatorname{vgm}($ nugget $=1$, psill $=5$, range $=300$, model $=$ 'Cir' $)$ )
cvgm1.e $<-$ fit.variogram(object $=$ cvg 1, model $=$ vgm(nugget $=1$, psill $=5$, range $=300$, model $=$ 'Exp' $)$ )
$\operatorname{plot}(\operatorname{cvg} 1, \operatorname{cvgm} 1 . s, \operatorname{pch}=19)$
plot(cvg1, cvgm1.c, pch = 19)
plot(cvg1, cvgm1.e, pch = 19)
cvgm1.s
cvgm1.c
cvgm1.e
\# Leave-one-out cross validation
$\operatorname{ccv} 1 . \mathrm{s}=$ krige.cv(Clay $\sim$ ValleyDepth, de1s, model $=\operatorname{cvgm} 1 . s)$
ccv1.c $=$ krige.cv $($ Clay $\sim$ ValleyDepth, de1s, model $=$ cvgm1.c $)$
ccv1.e $=$ krige.cv(Clay $\sim$ ValleyDepth, de1s, model $=$ cvgm1.e $)$
\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over prediction
mean(ccv1.s\$residual) \#-0.03
mean(ccv1.c\$residual) \# -0.03
mean(ccv1.e\$residual) \#-0.03
\# MSE. Mean squared error measures on average how different predictions are from observations.
\# The MSE will be small if the predicted responses are very close to the true responses, and will be large if for some of the observations, the predicted and true responses differ substantially (ISL sixth printing).
mean(ccv1.s\$residual^2) \# 3.45
mean(ccv1.c\$residual^2) \# 3.45
mean(ccv1.e\$residual^${ }^{\wedge}$ ) \# 3.44
\# RMSE (take the square root to get units in original units)
$\operatorname{sqrt}\left(\operatorname{mean}\left(\operatorname{ccv} 1 . s \$ r e s i d u a l^{\wedge} 2\right)\right)$ \# 1.86
$\operatorname{sqrt}\left(\operatorname{mean}\left(\operatorname{ccv} 1 . c \$ r e s i d u a{ }^{\wedge}{ }^{\wedge} 2\right)\right)$ \# 1.87
$\operatorname{sqrt}\left(\operatorname{mean}\left(\operatorname{ccv} 1 . e \$ r^{2} \operatorname{lidual}^{\wedge} 2\right)\right)$ \# 1.85
\# What is the spatial distribution of the residuals?
bubble(ccv1.s, "residual", main = "Clay 0-5 cm Spherical")
bubble(ccv1.c, "residual", main = "Clay $0-5 \mathrm{~cm}$ Circular")
bubble(ccv1.e, "residual", main = "Clay 0-5 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).
ck1.s $<-$ krige $($ Clay $\sim$ ValleyDepth, de1s, model $=$ cvgm1.s, newdata $=$ sgdf)
ck1.c $<-$ krige(Clay $\sim$ ValleyDepth, de1s, model $=$ cvgm1.c, newdata $=$ sgdf)
ck1.e $<-$ krige(Clay $\sim$ ValleyDepth, de1s, model $=$ cvgm1.e, newdata $=$ sgdf)
\# Plotting. sqrt(varl.var) returns the standard deviation rather than the variance.
ck1.s $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(0,4)) + ggtitle('Spherical') + theme_bw()
```

ck1.c $\%>\%$ as.data.frame $\%>\%$

$$
\begin{aligned}
& \operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\text { geom_tile(aes(fill=sqrt(var1.var) }))+ \text { coord_equal() }+ \\
& \text { scale_fill_gradient(low }=\text { "yellow", high="red", limits }=\mathrm{c}(1,3))+ \text { ggtitle('Circular') }+ \text { theme_bw( })
\end{aligned}
$$

ck1.e $\%>\%$ as.data.frame $\%>\%$

$$
\begin{aligned}
& \operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\text { geom_tile(aes(fill=sqrt(var1.var) }))+ \text { coord_equal() }+ \\
& \text { scale_fill_gradient(low }=\text { "yellow", high="red", limits }=\mathrm{c}(1,3))+ \text { ggtitle('Exponential') }+ \text { theme_bw( })
\end{aligned}
$$

\# I choose circular model because the model appeared to fit the data slightly better than the other models. Still, I'm not very happy with this data.

```
# 3.2 5-15 cm
# Remove PedonID's 14:19. Also remove 41 as it is an outlier and keeping it results in models that do not
converge.
d2s <- d2[!(d2$Pedon.ID %in% c(14:19,41)),]
de2s <- de2[!(de2$Pedon.ID %in% c(14:19,41)),]
# Summary stats. Appear fairly normal, skewness suggests some need to transform, but not a lot
summary(d2s$Clay)
skewness(d2s$Clay) # 0.21
# Histograms. Very normally distributed
hist(d2s$Clay, col = "lightblue", border = "red")
rug(d2s$Clay)
# Plots to check for obvious spatial patterns
spplot(d2s, zcol = 'Clay', col.regions=brewer.pal(5, "Set1"))
# Look for spatial outliers. Nothing obvious.
# c2.sel = plot(variogram(Clay ~ 1, d2s, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(c2.sel, d2)
# Fit linear models
summary(lm(Clay ~ Aspect , de2s))
summary(lm(Clay ~ ConvergenceIndex , de2s)) #* Adj. R2 0.08
summary(lm(Clay ~ CrossSectionalCurvature, de2s)) #* Adj. R2 0.07
summary(lm(Clay ~ DEM_5_utm , de2s))
summary(lm(Clay ~ FlowAccumulation , de2s)) #* Adj. R2 0.06
```

```
summary \((\operatorname{lm}\) (Clay \(\sim\) LongitudinalCurvature , de2s))
summary \((\operatorname{lm}(\) Clay \(\sim\) LSfactor , de2s) \()\)
summary \((\operatorname{lm}(\) Clay \(\sim\) Slope \(\quad\), de2s \()\) )
summary \((\operatorname{lm}(\) Clay \(\sim\) TopographicWetnessIndex, de2s))
summary \((\operatorname{lm}(\) Clay \(\sim\) ValleyDepth , de2s \())\) \#*** Adj. R2 0.21
summary \((\operatorname{lm}(\) Clay \(\sim\) CrossSectionalCurvature+ConvergenceIndex+ValleyDepth, de2s)) \#ValleyDepth
significant
summary( \(\operatorname{lm}(\) Clay \(\sim\) ConvergenceIndex+ValleyDepth, de2s)) \# ValleyDepth significant
summary \((\operatorname{lm}\) (Clay ~CrossSectionalCurvature+ValleyDepth, de2s)) \# ValleyDepty significant
plot(Clay ~ ConvergenceIndex, de2s)
abline( \(\operatorname{lm}\) (Clay \(\sim\) ConvergenceIndex, de2s)
plot(Clay ~ CrossSectionalCurvature, de2s)
abline ( \(\operatorname{lm}(\) Clay \(\sim\) CrossSectionalCurvature, de2s) \()\)
plot(Clay ~ ValleyDepth, de2s)
abline( \(\operatorname{lm}(\) Clay \(\sim\) ValleyDepth, de2s) \()\)
```

\# I'm choosing valley depth because it has the strongest correlation and because it makes sense to me.
\# Check for Anisotropy. 135 seems best and agreed with the direction of the landform.
$\operatorname{plot}($ variogram(Clay $\sim$ ValleyDepth, de2s, alpha $=c(105,120,135,155,180)))$
\# h-scatterplots. Not much correlation beyone 90 m .
hscat(Clay $\sim$ ValleyDepth, data $=\operatorname{de} 2, c(3,9,29,88,266,800)$, variogram.alpha=120)
\# Empirical (experimental or sample) variogram. When I include boundaries $=c(3,9,29,88,266,800) I$ am able to still get a model to fit, but it strongly reduces the range thus the prediction uncertainity is only concentrated around the sample locations and RMSE slightly increased, so I am not taking this approach. ConvergenceIndex+ValleyDepth
$\operatorname{cvg} 2<-$ variogram(Clay $\sim$ ValleyDepth, boundaries $=c(29,88,266,800)$, de2s, alpha $=135)$
plot(cvg2)

```
cvg2
```

\# Variogram modeling. Gaussian, Power, Log, Matern, none fit.
cvgm2.s $<-$ fit.variogram(object $=$ cvg2, model $=$ vgm(nugget $=1$, psill $=5$, range $=300$, model $=$ 'Sph') $)$
$\operatorname{cvgm} 2 . \mathrm{c}<-$ fit.variogram $($ object $=\mathrm{cvg} 2$, model $=\operatorname{vgm}($ nugget $=1$, psill $=5$, range $=300$, model $=$ 'Cir') )
cvgm2.e $<-$ fit.variogram $($ object $=\operatorname{cvg} 2$, model $=$ vgm(nugget $=1$, psill $=5$, range $=300$, model $=$ 'Exp' $)$ )
$\operatorname{plot}(\operatorname{cvg} 2, \operatorname{cvgm} 2 . s, \operatorname{pch}=19)$
plot(cvg2, cvgm2.c, pch $=19)$
$\operatorname{plot}(\operatorname{cvg} 2, \operatorname{cvgm} 2 . e, \operatorname{pch}=19)$
cvgm2.s
cvgm2.c
cvgm2.e
\# Leave-one-out cross validation
ccv2.s $=$ krige.cv $($ Clay $\sim$ ValleyDepth, de2s, model $=$ cvgm2.s $)$
$\operatorname{ccv} 2 . c=$ krige.cv(Clay $\sim$ ValleyDepth, de2s, model $=$ cvgm2.c)
$\operatorname{ccv} 2 . \mathrm{e}=$ krige.cv $($ Clay $\sim$ ValleyDepth, de2s, model $=$ cvgm2.e $)$
\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over prediction
mean(ccv2.s\$residual) \# 0.02
mean(ccv2.c\$residual) \# 0.03
mean(ccv2.e\$residual) \# 0.03
\# MSE. Mean squared error
mean(ccv2.s\$residual^2) \# 1.81
mean(ccv2.c\$residual^2) \# 1.88
mean(ccv2.e\$residual^2) \# 1.88
\# RMSE (take the square root to get units in original units)
sqrt(mean(ccv2.s\$residual^2)) \# 1.35
sqrt(mean(ccv2.c\$residual^2)) \# 1.37
sqrt(mean(ccv2.e\$residual^^2)) \# 1.37
\# What is the spatial distribution of the residuals?
bubble(ccv2.s, "residual", main = "Clay 5-15 cm Spherical")
bubble(ccv2.c, "residual", main = "Clay 5-15 cm Circular")
bubble(ccv2.e, "residual", main = "Clay 5-15 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
ck2.s <- krige(Clay ~ ValleyDepth, de2s, model = cvgm2.s, newdata = sgdf)
ck2.c <- krige(Clay ~ ValleyDepth, de2s, model = cvgm2.c, newdata = sgdf)
ck2.e <- krige(Clay ~ ValleyDepth, de2s, model = cvgm2.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
ck2.s \%>\% as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(0,3)) + ggtitle('Spherical') + theme_bw()
```

ck2.c $\%>\%$ as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+$ geom_tile(aes(fill=sqrt(var1.var)) $)+$ coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = $c(0,3))+$ ggtitle('Circular') + theme_bw()

```
ck2.e %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,3)) + ggtitle('Exponential') + theme_bw()
```

\# Little difference between the models. Chose to use a spherical model as it had slighly lower RMSE

```
# 3.3 15-30 cm
# Remove pedons 13:19. See explanation for 0-5 cm}
d3s <- d3[!(d3$Pedon.ID %in% c(14:19)),]
de3s <- de3[!(de3$Pedon.ID %in% c(14:19)),]
# Summary stats. Not much variability. No need to transform.
summary(d3s$Clay)
skewness(d3s$Clay) # -0.0007
# Histograms. Not quite as 'normal' as the first two depths, but still pretty close.
hist(d3s$Clay, col = "lightblue", border = "red")
rug(d3s$Clay)
# Plots to check for obvious spatial patterns
spplot(d3s, zcol = 'Clay', col.regions=brewer.pal(5, "Set1"))
# Look for spatial outliers. No outliers
# c3.sel = plot(variogram(Clay ~ 1, d3, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(c3.sel, d3)
# Fit linear models
summary(lm(Clay ~ Aspect , de3s))
summary(lm(Clay ~ ConvergenceIndex , de3s)) #* Adj. R2 0.06
summary(lm(Clay ~ CrossSectionalCurvature, de3s)) #* Adj. R2 0.07
summary(lm(Clay ~ DEM_5_utm , de3s)) #* Adj. R2 0.07
summary(lm(Clay ~ FlowAccumulation , de3s)) #* Adj. R2 0.06
summary(lm(Clay ~ LongitudinalCurvature , de3s)) #* Adj. R2 0.08
summary(lm(Clay ~ LSfactor , de3s))
summary(lm(Clay ~ Slope , de3s))
summary(lm(Clay ~ TopographicWetnessIndex, de3s))
```

```
summary( }\operatorname{lm}(Clay ~ ValleyDepth , de3s)) #*** Adj. R2 0.26
plot(Clay ~ ConvergenceIndex, de3s)
abline(lm(Clay ~ ConvergenceIndex, de3s))
plot(Clay ~ CrossSectionalCurvature, de3s)
abline( }\operatorname{lm}(Clay ~ CrossSectionalCurvature, de3s))
plot(Clay ~ LongitudinalCurvature, de3s)
abline(lm(Clay ~ LongitudinalCurvature, de3s))
plot(Clay ~ ValleyDepth, de3s)
abline(lm(Clay ~ ValleyDepth, de3s))
# Only valley depth is significant
summary(lm(Clay ~ ValleyDepth+ConvergenceIndex+CrossSectionalCurvature+LongitudinalCurvature,
de3s))
summary( }\operatorname{lm}(Clay ~ ValleyDepth+ConvergenceIndex +CrossSectionalCurvature, de3s))
summary(lm(Clay ~ ValleyDepth+ConvergenceIndex, de3s))
# Check for Anisotropy. }135\mathrm{ seems best as it has the most consistent variance
plot(variogram(Clay ~ ValleyDepth, de3s, alpha =c(105, 120, 135, 155, 180)))
# h-scatterplots. Not much correlation beyond ~ 88 m.
hscat(Clay ~ ValleyDepth, data = de3s, c(3, 9, 29, 88, 266, 800), variogram.alpha=135)
# Empirical (experimental or sample) variogram.
cvg3 <- variogram(Clay ~ ValleyDepth, de3s, alpha=135)
plot(cvg3, plot.nu = FALSE)
cvg3
# Variogram modeling
cvgm3.s <- fit.variogram(object=cvg3, model = vgm(nugget = 1, psill = 4, range = 350, model = 'Sph'))
cvgm3.c <- fit.variogram(object=cvg3, model = vgm(nugget = 1, psill = 4, range = 350, model = 'Cir'))
```

cvgm3.e $<-$ fit.variogram $($ object $=\operatorname{cvg} 3$, model $=$ vgm $($ nugget $=1$, psill $=4$, range $=350$, model $=$ 'Exp' $)$ )

```
plot(cvg3, cvgm3.s, pch = 19)
plot(cvg3, cvgm3.c, pch = 19)
plot(cvg3, cvgm3.e, pch = 19)
```

cvgm3.s
cvgm3.c
cvgm3.e
\# Leave-one-out cross validation
ccv3.s $=$ krige.cv $($ Clay $\sim$ ValleyDepth, de3s, model $=$ cvgm3.s $)$
ccv3.c $=$ krige.cv $($ Clay $\sim$ ValleyDepth, de3s, model $=$ cvgm3.c $)$
ccv3.e $=$ krige.cv $($ Clay $\sim$ ValleyDepth, de3s, model $=\operatorname{cvgm} 3 . e)$
\# MPE. Mean prediction error (predicted-observed) = bias. Positive $=$ under prediction, negative $=$ over
prediction
mean(ccv3.s\$residual) \# 0.036
mean(ccv3.c\$residual) \# 0.039
mean(ccv3.e\$residual) \# 0.033
\# MSE. Mean squared error
mean(ccv3.s\$residual^2) \# 2.22
mean(ccv3.c\$residual^2) \# 2.10
mean(ccv3.e\$residual^2) \# 2.25
\# RMSE (take the square root to get units in original units)
sqrt(mean(ccv3.s\$residual^2)) \# 1.50
sqrt(mean(ccv3.c\$residual^2)) \# 1.45
$\operatorname{sqrt}\left(\right.$ mean(ccv3.e\$residual^$\left.\left.{ }^{\wedge}\right)\right)$ ) 1.50
\# What is the spatial distribution of the residuals?
bubble(ccv3.s, "residual", main = "Clay 15-30 cm Spherical")
bubble(ccv3.c, "residual", main = "Clay $15-30 \mathrm{~cm}$ Circular")
bubble(ccv3.e, "residual", main = "Clay 15-30 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
ck3.s <- krige(Clay ~ ValleyDepth, de3s, model = cvgm3.s, newdata = sgdf)
ck3.c <- krige(Clay ~ ValleyDepth, de3s, model = cvgm3.c, newdata = sgdf)
ck3.e <- krige(Clay ~ ValleyDepth, de3s, model = cvgm3.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance. ck3.s \%>\% as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+\operatorname{geom}$ _tile(aes(fill=sqrt(var1.var) $))+$ coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(1,3)) + ggtitle('Spherical') + theme_bw()
ck3.c $\%>\%$ as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+$ geom_tile(aes(fill=sqrt(var1.var) $))+$ coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(1,3)) + ggtitle('Circular') + theme_bw()
ck3.e $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(1,3)) + ggtitle('Exponential') + theme_bw()
```

\# I'm going with circular as it has the lowest RMES, a reasonable and low partial sill, and a reasonable range.

## \# $3.430-60 \mathrm{~cm}$

\# Try 13-19. Wow, removing these totally changes which variables are significantly. It also makes variograms 'fit' the data better so that I got the models to converge.... it is suspicious to me that removing
this contigious 'batch' of pedon ids make the models fit. Also remove Pedon.ID 13. Including 13 (identified as an outlier) makes the models not converge and increases RMSE by $0.3 \%$.

```
d4s <- subset(d4, Pedon.ID <13 | Pedon.ID > 19)
de4s <- subset(de4, Pedon.ID <13 | Pedon.ID > 19)
# Summary stats.
summary(d4s$Clay)
skewness(d4s$Clay) # -0.33
# Histograms
hist(d4s$Clay, col = "lightblue", border = "red")
rug(d4s$Clay)
```

\# Plots to check for obvious spatial patterns
spplot(d4s, zcol = 'Clay', col.regions=brewer.pal(5, "Set1"))
\# Look for spatial outliers. 16 is probably an outlier.

```
# c4.sel = plot(variogram(Clay ~ 1, d4s, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(c4.sel, d4s)
```

\# Fit Linear models
summary $(\operatorname{lm}($ Clay $\sim$ Aspect , de4s) $)$ \# Adj. R2 0.09
summary $(\operatorname{lm}($ Clay $\sim$ ConvergenceIndex $\quad$, de4s $)$ ) \# Adj. R2 0.11
summary $(\operatorname{lm}($ Clay $\sim$ CrossSectionalCurvature, de4s) ) \# Adj. R2 0.16
summary $(\operatorname{lm}($ Clay ~ DEM_5_utm , de4s) $)$
summary $(\operatorname{lm}($ Clay $\sim$ FlowAccumulation , de4s $)$ )
summary $(\operatorname{lm}$ (Clay $\sim$ LongitudinalCurvature , de4s))
summary $(\operatorname{lm}($ Clay $\sim$ LSfactor , de4s $)$ )
summary $(\operatorname{lm}($ Clay $\sim$ Slope $\quad$, de4s $))$
summary $(\operatorname{lm}$ (Clay $\sim$ TopographicWetnessIndex, de4s))
summary $(\operatorname{lm}($ Clay $\sim$ ValleyDepth $\quad$, de4s $)$ )
plot(Clay ~ CrossSectionalCurvature, de4s)
abline ( $\operatorname{lm}($ Clay $\sim$ CrossSectionalCurvature, de4s) $)$
plot(Clay ~ Aspect, de4s)
abline( $\operatorname{lm}($ Clay $\sim$ Aspect, de4s $)$ ) \# land only faces west and north. I'm not sure that I can explain this so I'm not going to include it.

```
plot(Clay ~ ConvergenceIndex, de4s)
abline(lm(Clay ~ ConvergenceIndex, de4s))
```

\# Together neither are significant. I'm going to use CrossSectionalCurvature as it has highest R2 value. summary $(\operatorname{lm}$ (Clay $\sim$ CrossSectionalCurvature + ConvergenceIndex, de4s $)$ )
\# Check for Anisotropy. 120 seems best
$\operatorname{plot}($ variogram $($ Clay $\sim$ CrossSectionalCurvature, de4s, $\operatorname{alpha}=c(105,120,135,155,180)))$
\# h-scatterplots. Strongly correlated to $\sim 90 \mathrm{~m}$.
hscat(Clay $\sim$ CrossSectionalCurvature, data $=$ de4s, c(3, 9, 29, 88, 266, 800), variogram.alpha=120)

```
# Empirical (experimental or sample) variogram. Leaving out longitudinalCurvature increases RMSE by ~
0.4 so I left it in.
cvg4 <- variogram(Clay ~ CrossSectionalCurvature, de4s, alpha=135)
plot(cvg4, plot.nu = FALSE)
cvg4
```

\# Variogram modeling. The exponential model doesn't converge, but the values are realistic and stable (even with different values) so I will include the model.

```
cvgm4.s <- fit.variogram(object=cvg4, model = vgm(nugget = 0.5, psill = 1, range= 100, model = 'Sph'))
```

cvgm4.c $<-$ fit.variogram $($ object $=\operatorname{cvg} 4$, model $=\operatorname{vgm}($ nugget $=0.5$, psill $=1$, range $=100$, model $=$ 'Cir') $)$
cvgm4.e $<-$ fit.variogram $($ object $=c v g 4$, model $=\operatorname{vgm}($ nugget $=0.5$, psill $=1$, range $=100$, model $=$ 'Exp' $)$ )
plot $(\operatorname{cvg} 4, \operatorname{cvgm} 4 . s, \operatorname{pch}=19)$
plot $(\mathrm{cvg} 4, \operatorname{cvgm} 4 . \mathrm{c}, \mathrm{pch}=19)$
$\operatorname{plot}(\operatorname{cvg} 4, \operatorname{cvgm} 4 . e, \operatorname{pch}=19)$
cvgm4.s
cvgm4.c
cvgm4.e
\# Leave-one-out cross validation
ccv4.s $=$ krige.cv(Clay $\sim$ CrossSectionalCurvature, de4s, model $=$ cvgm4.s)
ccv4.c $=$ krige.cv(Clay $\sim$ CrossSectionalCurvature, de4s, model $=$ cvgm4.c $)$
ccv4. $=$ krige.cv(Clay $\sim$ CrossSectionalCurvature, de4s, model $=$ cvgm4.e $)$
\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over prediction
mean(ccv4.s\$residual) \# 0.008
mean(ccv4.c\$residual) \# 0.009
mean(ccv4.e\$residual) \# -0.004
\# MSE. Mean squared error
mean(ccv4.s\$residual^2) \# 2.36
mean(ccv4.c\$residual^2) \# 2.33
mean(ccv4.e\$residual^2) \# 2.37
\# RMSE (take the square root to get units in original units)
sqrt(mean(ccv4.s\$residual^2)) \# 1.54
sqrt(mean(ccv4.c\$residual^2)) \# 1.53
sqrt(mean(ccv4.e\$residual^2)) \# 1.54
\# What is the spatial distribution of the residuals?
bubble(ccv4.s, "residual", main = "Clay 30-60 cm Spherical")
bubble(ccv4.c, "residual", main = "Clay 30-60 cm Circular")
bubble(ccv4.e, "residual", main = "Clay 30-60 cm Exponential")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).

```
ck4.s <- krige(Clay ~ CrossSectionalCurvature, de4s, model = cvgm4.s, newdata = sgdf)
ck4.c <- krige(Clay ~ CrossSectionalCurvature, de4s, model = cvgm4.c, newdata = sgdf)
ck4.e <- krige(Clay ~ CrossSectionalCurvature, de4s, model = cvgm4.e, newdata = sgdf)
```

\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
ck4.s $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
```

scale_fill_gradient(low = "yellow", high="red", limits $=c(0,3))+$ ggtitle('Spherical') + theme_bw()
ck4.c $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(0,3)) + ggtitle('Circular') + theme_bw()
```

ck4.e $\%>\%$ as.data.frame $\%>\%$

```
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=sqrt(var1.var))) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,3)) + ggtitle('Exponential') + theme_bw()
```

\# I'm choosing circular as this has the lowest RMSE and largest range.

\# Notes:
\# On singular model fits: If your variogram turns out to be a flat, horizontal or sloping line, then fitting a three-parameter model such as the exponential or spherical with nugget is a bit heavy: there's an infinite number of possible combinations of sill and range (both very large) to fit to a sloping line. In this case, the returned, singular model may still be useful: just try and plot it. Gstat converges when the parameter values stabilize, and this may not be the case. Another case of singular model fit happens when a model that reaches the sill (such as the spherical) is fit with a nugget, and the range parameter starts, or converges to a value smaller than the distance of the second sample variogram estimate. In this case, again, an infinite number of possibilities occur essentially for fitting a line through a single (first sample variogram) point.

In both cases, fixing one or more of the variogram model parameters may help you out (from fit.variogram notes: https://cran.r-project.org/web/packages/gstat/gstat.pdf)
\# 5 Plotting kriging maps.
\#Code modified from https://rpubs.com/nabilabd/118172
\#Load libraries here so they don't mess with other packages
library(ggplot2)
library(dplyr)
\#Sand
\# Convert variance into standard deviation
sk1.c\$SD <- sqrt(sk1.c\$var1.var)
sk2.c\$SD <- sqrt(sk2.c\$var1.var)
sk3.c\$SD <- sqrt(sk3.c\$var1.var)
sk4.s\$SD <- sqrt(sk4.s\$var1.var)
\# Give better names
names(sk 1.c)[1] <- 'Sand'
names(sk2.c)[1] <- 'Sand'
names(sk3.c)[1] <- 'Sand'
names(sk4.s)[1] <- 'Sand'
\# Reproject and rename kriging SpatialGridDataFrame for better plotting library(plotKML)

```
sk1_1l <- reproject(sk1.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs84=0,0,0'))
sk2_ll <- reproject(sk2.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs84=0,0,0'))
sk3_1l <- reproject(sk3.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs84=0,0,0'))
sk4_1l <- reproject(sk4.s, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs84=0,0,0'))
```

\# Create dataframe to plot points on figures
\# How do I put popints on plot

| $\begin{aligned} & \text { d1_ll <- spTransform(d1, } \\ & \text { +towgs } \left.\left.84=0,0,0^{\prime}\right)\right) \end{aligned}$ | CRS('+proj=longlat | +ellps=WGS84 | +datum=WGS84 | +no_defs |
| :---: | :---: | :---: | :---: | :---: |
| d1_11_df <- as.data.frame(d1_ll) |  |  |  |  |
| d1s_ll<- spTransform(d1s, + towgs $84=0,0,0^{\prime}$ )) | CRS('+proj=longlat | +ellps=WGS84 | +datum=WGS84 | +no_defs |
| d1s_ll_df <- as.data.frame(d1s_11) |  |  |  |  |
| $\begin{aligned} & \text { d2s_ll<- } \quad \text { spTransform(d2s, } \\ & \text { +towgs } \left.\left.84=0,0,0^{\prime}\right)\right) \end{aligned}$ | CRS('+proj=longlat | +ellps=WGS84 | +datum=WGS84 | +no_defs |
| d2s_ll_df <- as.data.frame(d2s_11) |  |  |  |  |
| d3s_11<- spTransform(d3s, + towgs $84=0,0,0^{\prime}$ )) | CRS('+proj=longlat | +ellps=WGS84 | +datum=WGS84 | +no_defs |
| d3s_ll_df <- as.data.frame(d3s_11) |  |  |  |  |
| d4s_ll<- spTransform(d4s, + towgs $84=0,0,0^{\prime}$ )) | CRS('+proj=longlat | +ellps=WGS84 | +datum=WGS84 | +no_defs |
| d4s_ll_df <- as.data.frame(d4s |  |  |  |  |

\# Mean prediction
sk1_11 \%>\% as.data.frame $\%>\%$

```
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Sand)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(35,85)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "A") +
    ggtitle('0-5 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_0_5_mean.png", width=6, height=3, unit='in')
```

sk2_11 \%>\% as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+$ geom_tile(aes(fill=Sand) $)+$ coord_equal() +

```
scale_fill_gradient(low = "yellow", high="red", limits = c(35,85)) +
ylab("Latitude") + xlab("Longitude") +
scale_x_continuous() + scale_y_continuous() +
annotate("text", x = -106.611, y = 32.589, label = "C") +
ggtitle('5-15 cm') +
theme bw()
ggsave("./GeostatisticalModeling/Figures/Sand_5_15_mean.png", width=6, height=3, unit='in')
sk3_ll %>% as.data.frame %>%
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Sand)) + coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c(35,85)) +
ylab("Latitude") + xlab("Longitude") +
scale_x_continuous() + scale_y_continuous() +
annotate("text", x = -106.611, y = 32.589, label = "E") +
ggtitle('15-30 cm') +
theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_15_30_mean.png", width=6, height=3, unit='in')
```

sk4_11 \% $>\%$ as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(x=s 1, y=s 2))+$ geom_tile $(\operatorname{aes}(f i l l=$ Sand $))+$ coord_equal ()$+$
scale_fill_gradient(low = "yellow", high="red", limits $=c(35,85))+$
ylab("Latitude") + xlab("Longitude") +
scale_x_continuous ()$+$ scale_y_continuous ()$+$
annotate("text", x = -106.611, $y=32.589$, label = "G") +
ggtitle('30-60 cm') +
theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_30_60_mean.png", width=6, height=3, unit='in')
\# Standard deviation
sk1_11 $\%>\%$ as.data.frame $\%>\%$

```
ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
```

```
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,10)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "B") +
    geom_point(data = d1_11_df, aes(x=Easting, y = Northing)) +
    ggtitle('0-5 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_0_5_sd.png", width=6, height=3, unit='in')
sk2_1l %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,10)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "D") +
    geom_point(data = d1_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('5-15 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand 5_15_sd.png", width=6, height=3, unit='in')
sk3_ll %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,10)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "F") +
    geom_point(data = dl_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('15-30 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_15_30_sd.png", width=6, height=3, unit='in')
```

sk4_11 \% $>\%$ as.data.frame $\%>\%$

```
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,10)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "H") +
    geom_point(data = d4s_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('30-60 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Sand_30_60_sd.png", width=6, height=3, unit='in')
# Clay
ck1.c$SD <- sqrt(ck1.c$var1.var)
ck2.s$SD <- sqrt(ck2.s$var1.var)
ck3.c$SD <- sqrt(ck3.c$var1.var)
ck4.c$SD <- sqrt(ck4.c$var1.var)
```

\# Give better names
names(ck1.c)[1]<- 'Clay'
names(ck2.s)[1]<-'Clay'
names(ck3.c)[1] <- 'Clay'
names(ck4.c)[1] <- 'Clay'
\# Reproject and rename kriging SpatialGridDataFrame for better plotting
ck1_ll <- reproject(ck1.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs

+ towgs $\left.84=0,0,0^{\prime}\right)$ )
ck2_ll <- reproject(ck2.s, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+ towgs $\left.84=0,0,0^{\prime}\right)$ )
ck3_ll <- reproject(ck3.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs84=0,0,0'))
ck4_ll <- reproject(ck4.c, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
+towgs $\left.84=0,0,0^{\prime}\right)$ )

```
# Mean prediction
ck1_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Clay)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,16)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "A") +
    ggtitle('0-5 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_0_5_mean.png", width=6, height=3, unit='in')
ck2_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Clay)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,16)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "C") +
    ggtitle('5-15 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_5_15_mean.png", width=6, height=3, unit='in')
ck3_ll %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Clay)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,16)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "E") +
    ggtitle('15-30 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_15_30_mean.png", width=6, height=3, unit='in')
```

```
ck4_ll %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Clay)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(2,16)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "G") +
    ggtitle('30-60 cm') +
    theme bw()
ggsave("./GeostatisticalModeling/Figures/Clay_30_60_mean.png", width=6, height=3, unit='in')
# Standard deviation
ck1_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,5)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "B") +
    geom_point(data = d1s_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('0-5 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_0_5_sd.png", width=6, height=3, unit='in')
ck2_ll %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,5)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "D") +
    geom_point(data = d2s_ll_df, aes(x=Easting, y = Northing)) +
```

```
ggtitle('5-15 cm') +
theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_5_15_sd.png", width=6, height=3, unit='in')
ck3_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,5)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "F") +
    geom_point(data = d3s_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('15-30 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_15_30_sd.png", width=6, height=3, unit='in')
ck4_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(0,5)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "H") +
    geom_point(data = d4s_ll_df, aes(x=Easting, y = Northing)) +
    ggtitle('30-60 cm') +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Clay_30_60_sd.png", width=6, height=3, unit='in')
```

\# 6 Tables
\#Soil Survey Data (obtained from the White Sands Soil Survey by downloading the survey from WebSoilSurvey, opening the .mdb file and adding the tabular data, opening the component table and finding the chutum/dona ana complex component keys, then opening the horizon table and finding the component key. I then copied and pasted this data into excel (I had to do a bit of re-aranging to get the horizons right by depth).
ssd <- readWorkbook("./SoilData/physicalprop_chorion.xlsx", rows=c(1:10), cols = c(2,4:16))
\#Convert to SPC and convert to standard depth intervals.
depths(ssd) $<$ - Component.Key $\sim$ top + bottom
\# Global soil map standard depth intervals: $0-5 \mathrm{~cm}, 5-15 \mathrm{~cm}, 15-30 \mathrm{~cm}, 30-60 \mathrm{~cm}, 60-100, \& 100-200 \mathrm{~cm}$. The following code modified from: https://ncss-tech.github.io/AQP/aqp/aqp-intro.html. However, many of these soils are $<100 \mathrm{~cm}$ deep. Because I have so few samples, I am only going to map soil texture to 60 cm .
\# Use the slice and slab functions in AQP to average over these depths
ssd.pc <- aqp::slice(ssd, fm=0:100 ~total.sand.low +

```
total.sand.rv +
    total.sand.high +
    total.silt.low +
    total.silt.rv +
    total.silt.high +
    total.clay.low +
    total.clay.rv +
    total.clay.high)
```

\# Subset to GSM depths and calculate weighted mean values (I'm pretty sure that this does a weighted mean).
gsm.depths $<-\mathrm{c}(0,5,15,30,60,100)$
ssd.gsm <- slab(ssd.pc, fm=Component.Key $\sim$ total.sand.low +
total.sand.rv +
total.sand.high +
total.silt.low +
total.silt.rv +
total.silt.high +
total.clay.low +
total.clay.rv +
total.clay.high,slab.structure $=$ gsm.depths, slab.fun $=$ median, na.rm=TRUE)

```
# Reshape to wide format, convert to SPC, and make new hz names
ssd.d2 <- dcast(ssd.gsm, Component.Key + top + bottom ~ variable, value.var = 'value')
depths(ssd.d2) <- Component.Key ~ top + bottom
ssd.d2$hzname <- profileApply(ssd.d2, function(i) {paste0('GSM-', 1:nrow(i))})
```

\# Copy and paste the following into an excel spreadsheet, calculate weighted average from the proportions of components in the map unit, and reformat to make publication quality.
ssd.d2@horizons
\# Table 2
tab2a <- data.frame(cbind(rep('Sand', 4), rbind('0-5', '5-15', '15-30', '30-60'),
rbind(length(d1),
length(d2),
length(d3),
length(d4)),
rbind(summary(d1\$Sand),
summary (d2\$Sand),
summary (d3\$Sand),
summary(d4\$Sand)),
rbind(sd(d1\$Sand),
sd(d2\$Sand),
sd(d3\$Sand),
sd(d4\$Sand))))

```
names(tab2a)[10] <- 'SD'
```

tab2b <- data.frame(cbind(rep('Clay', 4),
rbind('0-5',
'5-15',
'15-30',
'30-60'),
rbind(length(d1),
length(d2),
length(d3),
length(d4)),
rbind(summary(d1\$Clay),
summary(d2\$Clay),
summary(d3\$Clay),
summary(d4\$Clay)),
rbind(sd(d1\$Clay),
sd(d2\$Clay),
sd(d3\$Clay),
sd(d4\$Clay))))
names(tab2b)[10] <- 'SD'
\# Table 3
tab3a <- data.frame(cbind(
rbind((as.character(svgm1.c\$model)[2]),
(as.character(svgm2.c\$model)[2]),
(as.character(svgm3.c\$model)[2]),
(as.character(svgm4.s\$model)[2])),
rbind(sqrt(mean(scv1.c\$residual^2)),

```
        sqrt(mean(scv2.c$residual^2)),
        sqrt(mean(scv3.c$residual^2)),
        sqrt(mean(scv4.s$residual^2))),
    rbind(svgm1.c$range[2],
    svgm2.c$range[2],
    svgm3.c$range[2],
    svgm4.s$range[2]),
    rbind(svgm1.c$psill,
        svgm2.c$psill,
        svgm3.c$psill,
        svgm4.s$psill)))
names(tab3a) <- c('model', 'rmse', 'range','nugget', 'sill')
tab3b <- data.frame(cbind(
    rbind((as.character(cvgm1.c$model)[2]),
        (as.character(cvgm2.s$model)[2]),
        (as.character(cvgm3.c$model)[2]),
    (as.character(cvgm4.c$model)[2])),
    rbind(sqrt(mean(ccv1.c$residual^2)),
        sqrt(mean(ccv2.s$residual^2)),
        sqrt(mean(ccv3.c$residual^2)),
        sqrt(mean(ccv4.c$residual^2))),
    rbind(cvgm1.c$range[2],
        cvgm2.s$range[2],
        cvgm3.c$range[2],
        cvgm4.c$range[2]),
    rbind(cvgm1.c$psill,
        cvgm2.s$psill,
        cvgm3.c$psill,
        cvgm4.c$psill)))
```

```
# 7. Variogram plotting
# Make nice variogram lines for plotting
# sand
slline = variogramLine(svgm1.c, maxdist = max(svg1$dist))
s2line = variogramLine(svgm2.c, maxdist = max(svg2$dist))
s3line = variogramLine(svgm3.c, maxdist = max(svg3$dist))
s4line = variogramLine(svgm4.s, maxdist = max(svg4$dist))
# clay
c1line = variogramLine(cvgm1.c, maxdist = max(cvg1$dist))
c2line = variogramLine(cvgm2.s, maxdist = max(cvg2$dist))
c3line = variogramLine(cvgm3.c, maxdist = max(cvg3$dist))
c4line = variogramLine(cvgm4.c, maxdist = max (cvg4$dist))
# Sand
splot1<-
ggplot(svg1, aes(x = dist, y = gamma)) +
    geom_point() +
    geom_line(data = slline) +
    ylim(c(0,80)) +
    annotate("text", x = 75, y = 75, label = "Sand 0-5 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
splot2 <-
```

```
ggplot(svg2, aes(x = dist, y = gamma ) ) +
geom_point() +
geom_line(data = s2line ) +
ylim(c(0,80)) +
annotate("text", x = 75, y = 75, label = "Sand 5-15 cm") +
theme_bw() +
theme(axis.title.x=element_blank(),
    axis.title.y=element_blank(),
    axis.text=element_text(size=11))
splot3 <-
ggplot(svg3, aes(x = dist, y = gamma ) ) +
    geom_point() +
    geom_line(data = s3line) +
    ylim(c(0,80)) +
    annotate("text", x = 75, y = 75, label = "Sand 15-30 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
splot4<-
ggplot(svg4, aes(x = dist, y = gamma ) ) +
    geom_point() +
    geom_line(data = s4line ) +
    ylim}(c(0,80))
    annotate("text", x = 75, y = 75, label = "Sand 30-60 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
```

```
# Clay
cplot1 <-
    ggplot(cvg1, aes(x = dist, y = gamma)) +
    geom_point() +
    geom_line(data = clline) +
    ylim(c(0,10)) +
    annotate("text", x = 75, y = 9.2, label = "Clay 0-5 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
cplot2 <-
    ggplot(cvg2, aes(x = dist, y = gamma)) +
    geom_point() +
    geom_line(data = c2line) +
    ylim(c(0,10)) +
    annotate("text", x = 75, y = 9.2, label = "Clay 5-15 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
cplot3 <-
    ggplot(cvg3, aes(x = dist, y = gamma) ) +
    geom_point() +
    geom_line(data = c3line) +
```

```
ylim(c(0,10)) +
annotate("text", x = 75, y = 9.2, label = "Clay 15-30 cm") +
theme bw() +
theme(axis.title.x=element_blank(),
axis.title.y=element_blank(),
axis.text=element_text(size=11))
cplot4<-
    ggplot}(\operatorname{cvg}4,\operatorname{aes}(\textrm{x}=\mathrm{ dist, }\textrm{y}=\operatorname{gamma}))
    geom_point() +
    geom_line(data = c4line ) +
    ylim(c(0,10)) +
    annotate("text", x = 75, y = 9.2, label = "Clay 30-60 cm") +
    theme_bw() +
    theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text=element_text(size=11))
```

\# Arrange into one plot
\# sand
sfig $<-$ ggarrange $($ splot 1, splot2, splot3, splot4, ncol $=2$, nrow $=2$ )
sfig <- annotate_figure(sfig,
bottom = text $\_$grob("Distance (m)"),
left $=$ text $\_$grob $($"Semivariance", rot $=90)$ )
ggsave(sfig, filename="Fig3.png")
\# clay
cfig $<-$ ggarrange $(\operatorname{cplot} 1$, cplot 2, cplot $3, \operatorname{cplot} 4$, ncol $=2$, nrow $=2)$

```
cfig <- annotate_figure(cfig,
    bottom = text_grob("Distance (m)"),
    left = text_grob("Semivariance", rot = 90))
```

ggsave(cfig, filename="Fig4.png")
\# \# X. Stochastic Simulations, i.e., equiprobable realizations of the variable that replicate the spatial characteristics found in the sample data. When all the simulated surfaces are assembled, they provide a distribution of values for each location in the study area. Models that well fit the data will have little variability between realizations. The nmax parameter results in local kriging, but without it, the command seems to go into an infinite loop.
\#
\# \#Kriging is a deterministic method whose function has a unique solution and does not attempt to represent the actual variability of the studied attribute. The smoothing property of any interpolation algorithm replaces local detail with a good average value; however, the geologist and reservoir engineer are more interested in finer-scaled details of reservoir heterogeneity than in a map of local estimates of the mean value. Like the traditional deterministic approach, stochastic methods preserve hard data where known and soft data where informative. Unlike the deterministic approach, though, it provides geoscientists and reservoir engineers with many realizations. The kriged solution is the average of numerous realizations, and the variability in the different outcomes is a measure of uncertainty at any location. Thus, the standard deviation of all values simulated at each grid node is the quantification of uncertainty.[2] [3] http://petrowiki.org/Geostatistical_conditional_simulation
\# I decided against stochastic simulation as I was more interested in getting a good prediction than in assessing local variability because kriging is twice as good at estimation as is stochastic simulation (Webster and Oliver, 2007, Geostats for Env. Sci, pg. 271) and because the standard deviation of the simulation was often $>100$.

```
#0-5 cm
```

set.seed(4801)
sSS1 $<-\operatorname{krige}($ Sand $\sim$ DEM_5_utm, de1, model $=\operatorname{svgm} 1 . c$, newdata $=\operatorname{sgdf}, \operatorname{nsim}=100, \operatorname{nmax}=67$ )
\#5-15 cm
set.seed(4801)
sSS2 $<-$ krige $(S a n d \sim$ DEM_5_utm, de2, model $=\operatorname{svgm} 2 . c$, newdata $=\operatorname{sgdf}, \operatorname{nsim}=100$, nmax $=67$ )
set.seed(4801)
sSS3 <- krige(Sand $\sim$ TopographicWetnessIndex, de3, model $=$ svgm3.e, newdata $=$ sgdf, nsim=100, nmax =67)

## \#30-60 cm

set.seed(4801)
sSS4 <- krige(Sand ~ DEM_5_utm+LongitudinalCurvature, de4, model $=$ svgm4.s, newdata $=$ sgdf, $n \operatorname{sim}=100$, $n \max =67$ )
\# Convert simulations to raster brick
sSS $1<-\operatorname{brick}(\mathrm{sSS} 1)$
sSS2 <- brick(sSS2)
sSS3 <- brick(sSS3)
sSS4 <- brick(sSS4)
\# Calculate mean and standard deviation of soil depth
s1.m<-calc(sSS1, mean)
s2.m <- calc(sSS2, mean)
s3.m $<-$ calc(sSS3, mean)
s4.m <- calc(sSS4, mean)
s1.sd $<-$ calc(sSS1, sd)
s2.sd $<-$ calc(sSS2, sd)
s3.sd <- calc(sSS3, sd)
s4.sd $<-$ calc $(s S S 4, ~ s d)$
$\operatorname{par}(\mathrm{mfrow}=\mathrm{c}(2,2))$
plot(s1.m)
plot(s2.m)
plot(s3.m)
plot(s4.m)

```
plot(s1.sd)
plot(s2.sd)
plot(s3.sd)
plot(s4.sd)
# Clay
set.seed(4801)
cSS1.c <- krige(Clay ~ ValleyDepth, de1s, model = cvgm1.s, newdata = sgdf, nsim=100, nmax =60)
cSS1.s <- krige(Clay ~ x, de1s, model = x, newdata = sgdf, nsim=100, nmax =60)
cSS1.c <- krige(Clay ~ x, de1s, model = x, newdata = sgdf, nsim=100, nmax =60)
```



```
# Convert simulations to raster brick
cSS1.r.s<- brick(cSS1.s)
# Calculate mean and standard deviation of soil depth
c1.s.m <- calc(cSS1.r.s, mean)
cl.s.sd <- calc(cSS1.r.s, sd)
par(mfrow = c(1,2))
plot(c1.s.m)
plot(c1.s.sd)
```

```
APPENDIX E CODE USED FOR GEOSTATISTICAL MODELING OF SOIL DEPTH
R code used for geostatistical modeling of soil depth
# Geostatistical modeling of soil depth in the the Tromble Weir Watershed
# Colby Brungard, PhD
# Load libraries
library(aqp)
library(sp)
library(rgdal)
library(raster)
library(gstat)
library(dplyr)
library(ggplot2)
# Set working directory
setwd("D:/Tromble Weir")
#1. Data preprocessing
# read in and check data
dat <- read.csv("./Mikayla Data/R_Pit_Data_cwb.csv")
sdat <- read.csv("./Mikayla Data/R_Pit_Site_Data.csv")
head(dat)
# Convert to SPC
depths(dat) <- Pedon.ID ~ HZ.Top + HZ.Bottom
site(dat) <- sdat
# Create depth variable
dat$depth <- profileApply(dat, FUN = max)
# Convert site data to spatialpointsdataframe for further analysis
```

```
dsp1<- dat@site
coordinates(dsp1) <- ~ Longitude + Latitude
proj4string(dsp1) <- '+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs'
# Reproject
dsp <- spTransform(dsp1, CRS('+proj=utm +zone=13 +ellps=GRS80 +datum=NAD83 +units=m
+no_defs'))
# Write to file for visulization in gis
# writeOGR(dsp, "./Mikayla Data", "SoilDepthObservations", driver = "ESRI Shapefile")
# Points 3 and 37 were located outside of the study area. Remove them
dsp <- dsp[dsp$Pedon.ID != 3 & dsp$Pedon.ID != 38,]
# Remove dsp1 so I'm not confused
rm(dsp1)
# Read in the study boundary and plot points over it
sarea2 <- readOGR(dsn = "./SoilData/spatial", layer = "SoilMU26")
sarea <- spTransform(sarea2, projection(dsp))
plot(sarea)
points(dsp, pch = 19, col = 'blue')
```

\#2 Exploratory data analysis
summary(dsp\$depth)
sd(dsp\$depth) \# 1919
\# Histogram

```
hist(dsp\$depth, col = "lightblue", border = "red", main = "Depth")
```

rug(dsp\$depth)
\# This appears to be bi-modal distribution with soils $<120$ and $>150 \mathrm{~cm}$.
\# I could separate these by depths, but it doesn't make a lot of sense to separate by depth

```
# Does a log transform help? Somewhat, I think, so I'll try it, but bimodal distribution is largest problem.
dsp$ldepth <- log(dsp$depth)
hist(dsp$ldepth, col = "grey", border = "red", main = 'Log (depth)')
rug(dsp$ldepth)
```

\# No spatial patterns readily apparent
bubble(obj = dsp, z = "depth", pch=1)
bubble (obj = dsp, $\mathrm{z}=$ "ldepth", $\mathrm{pch}=1$ )
\# Look for outliers
$\#$ sel $=\operatorname{plot}($ variogram $($ depth $\sim 1$, dsp, cloud $=$ TRUE $)$, col $=$ 'blue', pch $=19$, digitize $=$ TRUE $)$
\#plot(sel, dsp) \# No outliers readily apparent
\#2.1 Exploratory relationships with terrain variables
\#Load all rasters. Rasters created from 5m ifsar DEM using geoprocess_by_area.bat. See readme file in tw folder.
brk <- do.call(brick, lapply(list.files(path = "./Terrain_derivatives/TD_5m", pattern = ".*tif", full.names = TRUE), raster))
\# Reproject rasters to points (if needed)
brk2 $<$ - projectRaster(brk, crs="+proj=utm+zone= $=13+$ ellps=GRS80 + datum=NAD $83+$ units=m+no_defs + towgs $84=0,0,0$ ")
\# Mask to study area, then crop extent (significantly reduces processing time).
studyarea <- readOGR("./NestedSampllingExample", "SoilMU26")
brk3 <- mask(brk2, mask = studyarea)
brk4 <- crop(brk3, studyarea)

```
# Extract covariate values
ec <- raster::extract(brk4, y = dsp)
# Kriging and gaussian simulation requires a very fine underlying grid on which to predict.
# Use rasters to create prediction grid
sgdf<- as(brk4, 'SpatialGridDataFrame')
# Join covariate values to soil depth data
dsp2<- cbind(dsp, ec)
# Plotting
and topographic position index. It also shows a lot of co-linearity between covariates.
scatterplotMatrix(as.data.frame(dsp2[,-1]))
\# Fit linear models between Sand and covariates
\# Significance only shows that the relationship is not-zero.
\#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' \(0.1^{\prime \prime}\) ' 1
summary \((\operatorname{lm}(\) depth \(\sim\) Aspect , dsp2) \()\)
summary \((\operatorname{lm}(\) depth \(\sim\) ConvergenceIndex , dsp2) \()\)
summary ( \(\operatorname{lm}\) (depth ~ CrossSectionalCurvature, dsp2))
summary \((\operatorname{lm}(\) depth ~ DEM_5_utm , dsp2) \()\) \#** 0.1454
summary \((\operatorname{lm}\) (depth \(\sim\) FlowAccumulation , dsp2) \()\)
summary \((\operatorname{lm}\) (depth \(\sim\) LongitudinalCurvature , dsp2) )
summary \((\operatorname{lm}(\) depth \(\sim\) LSfactor \(\quad, \operatorname{dsp} 2))\)
summary \((\operatorname{lm}(\) depth \(\sim\) Slope \(\quad\), dsp2 \()\) )
summary \((\operatorname{lm}(\) depth \(\sim\) TopographicWetnessIndex, dsp2) \()\)
summary \((\operatorname{lm}(\) depth \(\sim\) ValleyDepth , dsp2) \()\)
```

\# This plot shows a somewhat linear relationship between depth and all variables except upslope curvature
\# Only elevation is significant. Not a very strong relationship, but I do need to account for this relationship.

```
\# Check for Anisotropy
\(\operatorname{plot}(\operatorname{variogram}(\operatorname{depth} \sim 1, \operatorname{dsp}\), alpha \(=c(0,15,30,45,60,75,90,105,120,135,140,155,180)))\)
\# Looks like a variogram at 120 degrees would be best so I'm going with this.
```

\# h-scatterplots
hscat(depth $\sim 1$, data $=\operatorname{dsp} 2, \mathrm{c}(3,9,29,88,266,800)$, variogram.alpha=120) \# Most correlated at $<30 \mathrm{~m}$, maybe 88 m .
\# Look for spatial outliers. One posible outlier, but I do not interpret this as a 'real' outlier as this is likely in the area next to the small drainage.

```
# s1.sel = plot(variogram(depth ~ 1, dsp2, cloud = TRUE), col = 'blue', pch = 19, digitize = TRUE)
# plot(s1.sel, dsp2)
```

\# Empirical (experimental or sample) variogram.
dvg <- variogram(depth $\sim$ DEM_5_utm, dsp2, alpha=120)
plot(dvg)
dvg
\#Variogram modeling
\#The experimental variogram is basically just two columns of numbers: distance and semivariance. To use this for predictions, we need to fit a model (like a regression line) to the variogram. Because the variogram modeling is a numerical optimization we need to provide starting values. psill is the partial sill, which is the sill-nugget.
dvgm1.s $<-$ fit.variogram $($ object $=\operatorname{dvg}$, model $=\operatorname{vgm}($ nugget $=400$, psill $=1600$, range $=300$, model $=' \operatorname{Sph} '))$
dvgm1.c $<-$ fit.variogram $\left(\right.$ object $=$ dvg, model $=\operatorname{vgm}\left(\right.$ nugget $=400$, psill $=1600$, range $=300$, model $={ }^{\prime}$ Cir' $\left.\left.^{\prime}\right)\right)$
dvgm1.e $<-$ fit.variogram(object=dvg, model $=\operatorname{vgm}($ nugget $=400$, psill $=1600$, range $=300$, model $=$ 'Exp' $)$ )
dvgm1.s
dvgm1.c
dvgm1.e

```
plot(dvg, dvgm1.s, pch = 19)
\(\operatorname{plot}(\operatorname{dvg}, \operatorname{dvgm} 1 . c, \operatorname{pch}=19)\)
plot(dvg, dvgm1.e, pch = 19)
\# Leave-one-out cross validation
dcv1.s \(=\) krige.cv(depth \(\sim\) DEM_5_utm, dsp2, model \(=\operatorname{dvgm} 1 . s)\)
dcv1.c \(=\) krige.cv(depth \(\sim\) DEM_5_utm, dsp2, model \(=\operatorname{dvgm1.c})\)
dcv1.e \(=\) krige.cv(depth \(\sim\) DEM_5_utm, dsp2, model \(=\operatorname{dvgm} 1 . e)\)
```

\# MPE. Mean prediction error (predicted-observed) $=$ bias. Positive $=$ under prediction, negative $=$ over
prediction
mean(dcv1.s\$residual) \# -0.351
mean(dcv1.c\$residual) \# -0.165
mean(dcv1.e\$residual) \# 0.205
\# MSE. Mean squared error measures on average how different predictions are from observations.
\# The MSE will be small if the predicted responses are very close to the true responses, and will be large if
for some of the observations, the predicted and true responses differ substantially (ISL sixth printing).
mean(dcv1.s\$residual^2) \# 1418
mean(dcv1.c\$residual^2) \# 1452
mean(dcv1.e\$residual^2) \# 1753
\# RMSE (take the square root to get units in original units)
sqrt(mean(dcv1.s\$residual^2)) \# 37.7 cm
sqrt(mean(dcv1.c\$residual^2)) \# 38.1
sqrt(mean(dcv1.e\$residual^2)) \# 41.9
\# The spherical model has the lowest RMSE, largest range, and lowest nugget so I choose spherical. Still it his hard to model soil depth.
\# What is the spatial distribution of the residuals?
bubble(dcv1.s, "residual", main = "Sand 0-5 cm Spherical")
\# Kriging + uncertainty. Because I use elevation as a covariate then this is Kriging with an External Drift, rather than universal kriging (which is only if I use the coordinates as variables).
$\mathrm{dk}<-$ krige (depth $\sim$ DEM_5_utm, dsp2, model $=$ dvgm1.s, newdata $=$ sgdf)
\# Plotting. sqrt(var1.var) returns the standard deviation rather than the variance.
$\mathrm{dk} \%>\%$ as.data.frame $\%>\%$
$\operatorname{ggplot}(\operatorname{aes}(\mathrm{x}=\mathrm{s} 1, \mathrm{y}=\mathrm{s} 2))+$ geom_tile(aes(fill=sqrt(var1.var) $))+$ coord_equal() +
scale_fill_gradient(low = "yellow", high="red", limits = c( 2,8$)$ ) + ggtitle('Spherical') + theme_bw()
\# 5 Plotting kriging maps for publication.
\# Convert variance into standard deviation
dk\$SD <- sqrt(dk\$var1.var)
\# Give better names
names(dk)[1] <- 'Depth'
\# Reproject and rename kriging SpatialGridDataFrame for better plotting
library (plotKML)
dk_1l <- reproject(dk, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs +towgs84=0,0,0'))
\# Create dataframe to plot points on figures
\# How do I put points on plot?
dsp2_11 <- spTransform(dsp2, CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs + towgs $\left.84=0,0,0^{\prime}\right)$ )
dsp2_11_df <- as.data.frame(dsp2_11)

```
# Mean prediction
dk_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=Depth)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(30,150)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "A") +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Depth_mean.png", width=6, height=3, unit='in')
# Standard deviation
dk_11 %>% as.data.frame %>%
    ggplot(aes(x=s1, y=s2)) + geom_tile(aes(fill=SD)) + coord_equal() +
    scale_fill_gradient(low = "yellow", high="red", limits = c(30,45)) +
    ylab("Latitude") + xlab("Longitude") +
    scale_x_continuous() + scale_y_continuous() +
    annotate("text", x = -106.611, y = 32.589, label = "B") +
    geom_point(data = dsp2_ll_df, aes(x=Longitude, y = Latitude)) +
    theme_bw()
ggsave("./GeostatisticalModeling/Figures/Depth_sd.png", width=6, height=3, unit='in')
# Variogram model plotting
# Make nice variogram lines for plotting
dline = variogramLine(dvgm1.s, maxdist = max(dvg$dist))
dplot <-
    ggplot(dvg, aes(x = dist, y = gamma)) +
    geom_point() +
```

```
geom_line(data = dline) +
ylim(c(0,2500)) +
ylab('Semivariance') +
xlab('Distance (m)') +
annotate("text", x = 75, y = 2400, label = "Depth") +
theme_bw() +
theme(axis.text=element_text(size=13))
```

ggsave(dplot, filename="Fig5.png")
\# CONVERT spatialgriddataframe to raster and write to file. These are the predictions that could be used for ecohydrological modeling.

```
dPred <- raster(dk)
writeGDAL(dk, "test2.tif", band=1)
writeRaster(d.sd, "./Predictions/SoilDepth_sd.tif")
```

\# 4. Build a sampling grid for stage II sampling.
\# The key result from this geostatistical analysis is the range of the ordinary variogram.
d.vgm \#range $=46 \mathrm{~m}$. So I need to sample at distances closer than this for better modeling.

TWW <- readOGR(dsn = "./Enriques Data", layer = "Watershed2_Dissolve")
\#Buffer out a few meters to be able to sample surrounding areas. I chose the buffer distance iteratively so that I felt that I had enough points outside of the actual study area to make good predictions.

TWWb <- buffer(TWW, width = 30)
\# Make a sampling grid and select only the points inside the study area
grid $<-$ makegrid $($ TWWb, cellsize $=45)$ \#Cell size is 45 m , $<$ range of variogram.
grid $<-$ SpatialPointsDataFrame $($ coords=grid[,c(1,2)], data=grid, proj4string $=\operatorname{CRS}(\operatorname{proj} 4$ string $(T W W b))$ )
sampGrid <- grid[TWWb, ]
\# Plot
plot(TWWb)
points(sampGrid, pch = 19)
\# Write to file
\# writeOGR(sampGrid, ".", "SoilDepthSamplePoints", driver = "ESRI Shapefile")


[^0]:    Sand and clay values are in percent. Depth values are in cm

